Moving Forward from Predictive Regressions: Boosting Asset Allocation Decisions *

Lauri Nevasalmi † Henri Nyberg ‡

January 29, 2021

Abstract

We introduce a flexible utility-based empirical approach to directly determine asset allocation decisions between risky and risk-free assets. This is in contrast to the commonly used two-step approach where least squares optimal statistical equity premium predictions are first constructed to form portfolio weights before economic criteria are used to evaluate resulting portfolio performance. Our single-step customized gradient boosting method is specifically designed to find optimal portfolio weights in a direct utility maximization. Empirical results of the monthly U.S. data show the superiority of boosted portfolio weights over several benchmarks, generating interpretable results and profitable asset allocation decisions.

Keywords: Utility maximization, return predictability, machine learning, gradient boosting

JEL classification: C22, C53, C58, G11, G17

* We thank Charlotte Christiansen, Mika Hannula, Heikki Kauppi, Matthijs Lof, Seppo Pynnönen and Mika Vaihekoski and seminar participants at the University of Turku for useful comments. We gratefully acknowledge the financial support from the Emil Aaltonen Foundation and (Nyberg) the Academy of Finland (grant 321968) and the Finnish Foundation for Advancement of Securities Markets.

† Department of Mathematics and Statistics, University of Turku. FI-20014 University of Turku, Finland. E-mail: lauri.nevasalmi@utu.fi. Corresponding author.

‡ Department of Mathematics and Statistics, University of Turku. FI-20014 University of Turku, Finland. E-mail: henri.nyberg@utu.fi

1 Introduction

Asset allocation decisions have always been at the heart of finance and asset pricing research. The focus in the existing empirical and econometric implementations has been to predict stock returns to generate profitable portfolio decisions (see, e.g., the survey of Brandt, 2010). These have fundamentally been two-step approaches where especially linear predictive regressions, optimized via ordinary least squares, are first used to predict stock returns and resulting predictions are subsequently utilized to generate portfolio weights and portfolio returns evaluated with economic goodness-of-fit criteria. However, Leitch and Tanner (1991), Kandel and Stambaugh (1996), Xu (2004), Guidolin and Timmermann (2007) and Cenesizoglu and Timmermann (2012), among others, have shown from various perspectives that statistically optimal return predictions do not automatically imply economic gains and that even weak return predictability can have important economic value in terms of utility- and profit-based metrics. These views challenge the foundations of conventional two-step modelling as eventually investors are interested in the received economic gains on their investments, not arguably convenient statistical criteria such as least squares.

In this study, we consider a classic and simple asset allocation problem where an investor trades between risky market return and risk-free rate. Our contribution is to introduce a flexible single-step empirical approach built upon direct utility maximization in finding portfolio weights. That is, we set our objective function in accordance with theoretical utility maximization premises also in empirical implementation instead of statistical criteria such as least squares used in linear predictive regressions. We show that this advanced direct portfolio weight determination can interestingly be combined and implemented with modern machine learning, specifically a customized gradient boosting algorithm, enabling various advantages over the past approaches. Our results can help investors with different risk aversion preferences to determine which predictive variables and their combinations are useful in optimizing their asset allocation decisions.

This study follows the footsteps of the seminal works by Welch and Goyal (2008) and Campbell and Thompson (2008), but importantly this is not the next attempt in already extensive literature modifying (linear) predictive regressions to find statistical (out-of-sample) predictability in stock returns. Instead, our single-step approach sets portfolio weights directly by maximizing the underlying empirical utility function. In addition to the fundamental difference in the underlying objective function, we incorporate typically pre-determined lower and upper bounds of the portfolio weights as a part of the method, whereas the weights obtained with predictive regressions are typically truncated subsequently to lie between these bounds (see, e.g., Campbell and Thompson, 2008; Rapach and Zhou, 2013; Neely et al., 2014). This 'post-truncation' step has been found successful in out-of-sample stock return forecasting, but it does not give much economic intuition why certain state variables are important predictors to portfolio weight decisions. Our single-step approach circumvents these complications and produces easily interpretable results.

Our contribution is strongly connected to a more general issue in (financial) econometrics on what is the appropriate objective (loss) function in econometric inference: See general discussion and arguments for much more detailed examination in this respect in Elliott and Timmermann (2016, Chapter 2). Moreover, as argued by Brandt (2010) in his literature review (see also Brandt, 1999; Aït-Sahalia and Brandt, 2001; Brandt and Santa-Clara, 2006; Brandt, Santa-Clara and Valkanov, 2009), besides the obvious intuitive fact that asset allocation is the ultimate object of interest, there are several benefits when focusing portfolio weights directly with available predictive information. Ultimately, in contrast to direct utility maximization, the usual two-step statistical approach requires first to learn the notoriously complicated data generating process of stock returns, with various misspecification possibilities, before obtaining necessary input predictions to be plugged in (ad-hoc) trading rules. To evaluate the resulting asset allocation decisions with several economic criteria does not remove the fact that the underlying inference is not directed to optimize economic performance. To address this challenge, we develop a customized gradient boosting algorithm to empirically optimize asset allocation decisions in a flexible and single-step manner.

In economics and finance, machine learning-based methods have so far been (rightly) somewhat criticized about lack of intuition and interpretability. Due to its flexibility to allow custom objective functions, gradient boosting is an excellent workhorse for our purposes instead of 'textbook' machine learning algorithms often binded with their specific objectives. The usual 'statistical' gradient boosting algorithm with regression trees has generally been one of the most successful machine learning algorithms so far in different applications: See Rossi and Timmermann (2015) and Rossi (2018) as recent examples in finance to predict stock returns.

Our customized version diverges from the usual gradient boosting by utilizing the (negative) utility function as the objective function. We are hence able to establish an innovative synthesis between financial economics and machine learning practices instead of following everlasting least squares (mean square error) and hypothesis testing-based inference. That is, we get interpretable results on which predictive variables are truly important for asset allocation decisions when optimizing portfolio performance. Intuitively, our customized gradient boosting approach is iteratively learning from training sample asset allocation mistakes, measured by the gradient of the objective function (the selected utility function), before finally resulting in superior portfolio weight forecasts for the next period. This is intuitively in line with investors' attempts in practice to continuously update their trading strategy before taking future positions.

In empirical analysis, following the past (out-of-sample) return predictability studies for a comparison sake, we use a general form of the utility function that resembles to connect quadratic preferences to investors' decision making. This is the evaluation context typically considered in the past studies with an emphasis on out-of-sample forecasting performance, providing thus a natural building block for our approach. With the large updated dataset of macroeconomic and technical indicator predictive variables, originally compiled by Welch and Goyal (2008) and Neely et al. (2014), we are able to examine which predictors are truly important to asset allocation decisions rather than concentrating on the statistical predictability of stock returns.

Our empirical results on the monthly U.S. market returns and predictors show that substantial and quantitatively meaningful economic value can be obtained with our utility boosting method. This is the case even despite the fact that monthly stock returns are from the statistical perspective at most only weakly predictable. Technical indicators yield as a group the largest benefits in out-of-sample forecasting experiments. This is generally in line with the conclusions of Neely et al. (2014) and now confirmed with very different methodology. In the full sample estimation and model selection results, the gains obtained with the utility boosting are broad, also containing some specific macroeconomic variables. Interestingly, theoretically well-motivated inflation and partly also the dividend-price ratio stand out both in- and out-of-sample predictions for which the past empirical findings on their usefulness, obtained with statistical predictive regressions, have been quite inconclusive.

The rest of the paper is organized as follows. In Section 2, we frame and present the starting point of our contribution, determined by the related past out-of-sample predictive regression studies, before setting up our utility boosting approach. Empirical results are reported in Section 3, before discussion on the main general findings and final conclusions in Sections 4–5. Various additional and robustness analyses are compiled into the attached Internet Appendix.

2 Methodology

2.1 Starting point and two-step statistical approach

Consider a classic and commonly examined simple asset allocation decision problem for an investor with a single-period horizon aiming to optimally compose portfolio value between risky asset $r_{m,t}$ (market return) and risk-free asset return r_{ft}.[1] For a given level of (initial) wealth, the classic asset pricing perspective to the investor's optimization problem is to find portfolio weights maximizing the underlying utility function. Let w_t denote the proportion of the portfolio value allocated to the risky asset at time t, which

[1] An extension to multiple asset case requires a separate treatment to extend our methods to multiple-equation case (see details in Discussion in Section 4). This is yet out of scope of our advancement and left for the future research.

is based on the predictive information available at time $t-1$. The resulting portfolio return (at time t) is hence

$$r_{p,t} = r_{f,t} + w_t r_{e,t}, \tag{1}$$

where $r_{e,t}$ is the excess return on the (broad) stock market index in excess of the risk-free rate $r_{f,t}$ from the period $t-1$ to t.

Our aim is to determine (empirical) portfolio weights w_t in (1) directly using one or multiple state variables (predictors) contained in $\boldsymbol{x}_{t-1}$ at time $t-1$ (i.e. $w_t \equiv w(\boldsymbol{x}_{t-1})$). By focusing directly on the weights w_t, we aim to capture predictable and possibly time-varying patterns in weights as opposed to just relying predictive regressions on expected returns. Despite this objective and the fact that we are not explicitly interested in stock return predictions, we connect and frame our approach to (out-of-sample) return predictability examination originating from the seminal contributions by Welch and Goyal (2008) and Campbell and Thompson (2008).

Throughout this study and as a necessary selection to concretely setting up our approach, we consider a general utility function which resembles to attach quadratic (mean-variance) preferences to investor decision making. Following (1) and the formulation of Marquering and Verbeek (2004) and Fleming, Kirby, and Ostdiek (2001), among others, we consider to maximize the ex-ante (quadratic) expected utility of form

$$\max_{w_t} \left\{ E_{t-1}(r_{p,t}) - \frac{1}{2}\gamma \operatorname{Var}_{t-1}(r_{p,t}) \right\} = \max_{w_t} \left\{ E_{t-1}(r_{p,t}) - \frac{1}{2}\gamma\, w_t^2 \operatorname{Var}_{t-1}(r_{e,t}) \right\}, \tag{2}$$

where $\gamma > 0$ is the investor's risk aversion coefficient, representing the degree of risk aversion, and $E_{t-1}(\cdot)$ and $\operatorname{Var}_{t-1}(\cdot)$ denote the conditional expectation and conditional variance, given the information set at time $t-1$. The aim is to maximize (2) by determining w_t, using the information included in $\boldsymbol{x}_{t-1}$, resulting to portfolio returns $r_{p,t}$.

The utility scheme (2) is directly built upon the large majority of past (out-of-sample) return predictability studies and portfolio performance evaluation therein (see, e.g., Campbell and Thompson (2008), Rapach et al. (2010), Neely et al. (2014), Rossi (2018), among others, including the survey of Rapach and Zhou (2013)). They, likewise we, are not explicitly claiming that quadratic (mean-variance) utility function is necessarily exactly correct utility configuration in their two-step approaches. Our goal is simply to use an empirical counterpart of (2) as the selected underlying objective function in a single-step approach, to be developed in Sections 2.2–2.3, to determine asset allocation decisions.

Solving the maximization problem (2) leads to the solution of the optimal weights

$$w_t^* = \frac{E_{t-1}(r_{e,t})}{\gamma\,\mathrm{Var}_{t-1}(r_{e,t})} = \frac{E_{t-1}(r_{m,t}) - r_{f,t}}{\gamma\,\mathrm{Var}_{t-1}(r_{e,t})}, \tag{3}$$

written here in terms of the expected excess stock return $E_{t-1}(r_{e,t})$. If the expected return on the risky asset increases (ceteris paribus), an investor increases his/her weight on the risky asset, whereas increasing risk (conditional variance) involved is negatively related to the optimal weights. However and importantly, the optimal theoretical solution (3) does not yet tell how to obtain weights empirically.

As the weights w_t are functions of predictors $\boldsymbol{x}_{t-1}$, our goal is to determine empirical weights and asset allocation decisions in direct utility maximization using the 'training' (estimation) data $\{\boldsymbol{x}_{t-1}, r_{e,t}\}_{t=1}^{T}$ where T denotes the sample size. Our utility boosting approach in Sections 2.2–2.3 follows the steps opened by Brandt (1999), Aït-Sahalia and Brandt (2001) and Brandt and Santa-Clara (2006) also looking beyond predicting conditional moments of excess stock returns when setting portfolio weights. Our utility boosting, however, emphasizes different final goals (i.e. forecasting) than their contributions and, specifically, integrate recent advancement in machine learning to obtain profitable asset allocation decisions in genuine forecasting situations.

We develop our empirical approach as a complement to the dominant two-step empirical practice followed in the past (in and out-of-sample) return predictability research. It is built upon a simple linear predictive regression

$$r_{e,t} = \boldsymbol{z}_{t-1}'\boldsymbol{\beta} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{4}$$

where $\boldsymbol{z}_t = [1 \quad \boldsymbol{x}_t']'$, $E(\boldsymbol{z}_{t-1}\varepsilon_t) = \mathbf{0}$ and $\boldsymbol{\beta}$ is the vector of unknown parameters, containing also a constant term. The parameters $\boldsymbol{\beta}$ will be estimated by the method of ordinary least squares (OLS) where the underlying objective function is statistical minimizing the least squares criterion, given the estimation data $\{\boldsymbol{x}_{t-1}, r_{e,t}\}_{t=1}^{T}$ (including the (known) initial value $\boldsymbol{x}_0$):

$$\widehat{\boldsymbol{\beta}}_{OLS} = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \left(r_{e,t} - \boldsymbol{z}_{t-1}'\boldsymbol{\beta}\right)^2. \tag{5}$$

The resulting expected (predicted) excess returns

$$\widehat{r}_{e,t} = \boldsymbol{z}_{t-1}'\widehat{\boldsymbol{\beta}}_{OLS} \tag{6}$$

are then used as the empirical proxy for $E_{t-1}(r_{e,t})$ needed in (3).

In a similar fashion, an empirical proxy for the conditional variance of excess returns (hereafter denoted by $\sigma_t^2 \equiv \mathrm{Var}_{t-1}(r_{e,t})$) is required given the information available at time $t-1$. To allow for well-documented volatility clustering, we follow Campbell and Thompson (2008) and the subsequent studies (see, e.g., Rapach et al., 2010; Neely et al., 2014) by using a five-year rolling window variance of historical excess returns. This simplification can be relaxed by using, for example, the GARCH model or another realized volatility-based proxy (see the Internet Appendix A.4). Therefore, given the (pre-determined) risk aversion coefficient γ and the result (3), an investor allocates the following share of the portfolio value to risky equity for the period t:

$$\widehat{w}_t = \frac{\widehat{r}_{e,t}}{\gamma\,\widehat{\sigma}_t^2} = \frac{\boldsymbol{z}_{t-1}'\widehat{\boldsymbol{\beta}}_{OLS}}{\gamma\,\widehat{\sigma}_t^2}. \tag{7}$$

Expression (7) summarizes how the portfolio weights can empirically be obtained as a function of predictors $\boldsymbol{x}_{t-1}$ in two steps relying on the commonly considered linear predictive regressions (4) (i.e. first constructing $\widehat{r}_{e,t}$ (and $\widehat{\sigma}_t^2$) before plugging them into (7) in the second stage).[2] This approach has, however, several critical issues and

[2] Several (statistical) nonlinear predictive regressions and systems have been considered to predict stock returns with subsequent asset allocation decision objectives. See, e.g., the regime switching models surveyed by Guidolin (2011), including Guidolin and Timmermann (2007) who find that asset allocations guided by the forecasts of stock and bond returns from the Markov switching models yield utility gains relative to constant expected excess return predictions as described in equations (4)–(6).

complications from the final and arguable the most important portfolio performance perspective that we aim to tackle in this study:

(i) The resulting estimated weights $\widehat{w}_t$ in (7) are not necessarily nowhere near between typically pre-determined (assumed) bounds

$$w_t \in [w^{\min}, w^{\max}]. \tag{8}$$

The common practice (see the above-mentioned return predictability studies) has been to set the lower bound to zero ($w^{\min} = 0$), i.e. no short selling is allowed, and to consider mainly two different selections for the upper bound ($w^{\max}$): In Campbell and Thompson (2008), Neely et al. (2014) and Zhu (2015), w_t lies between 0 and 1.5 (i.e. $w^{\max} = 1.5$). They argue that these impose realistic portfolio constraints by precluding short sales and preventing more than 50% leverage. On the contrary, Aït-Sahalia and Brandt (2001), Marquering and Verbeek (2004) and Rossi (2018), among others, set the upper bound to $w^{\max} = 1$ (with $w^{\min} = 0$). This selection is also essentially the same as $w^{\max} = 0.99$ in, e.g., Kandel and Stambaugh (1996) and Cenesizoglu and Timmermann (2012). The empirical weights constructed as in (7) by no means guarantee the bounds (8) without strict additional and complicated restrictions on the predictive regression (4).

(ii) The empirical findings of Campbell and Thompson (2008), Rapach et al. (2010), Neely et al. (2014) and Pettenuzzo et al. (2014) emphasize the importance of restrictions (8) on the weights (7) to improve both statistical and economic out-of-sample predictive performance. To impose these subsequent constraints (8) turn out to modify the initial weights (7) substantially. It is important to realize that this post-truncation is not part of the predictive regression (4) by any means, and hence all the usual conventional statistical interpretations, such as t-test statistics, goodness-of-fit measures and model selection conclusions on useful predictors $\boldsymbol{x}_t$ are lost for further interpretations. In other words, even if a predictive variable is deemed statistically useful (i.e. 'statistically significant'), this does not necessarily mean that it really has important predictive information on asset allocation decisions when respecting the bounds (8) in the end. This logic works also other way round so that statistically poor predictors can be useful in direct portfolio weight determination.

(iii) Consider a commonly used one predictor ($x_{i,t}$) special case of (4)

$$r_{e,t} = \beta_{i,0} + x_{i,t-1}\beta_i + \varepsilon_{i,t}. \tag{9}$$

In the vast (in-sample) return predictability research, this is the common specification where testing the null hypothesis $\beta_i = 0$ of no (conditional) mean predictability in stock returns with a t-test statistic has been of particular interest. This is important because if (correctly) rejecting the null $\beta_i = 0$, the resulting statistical predictability is widely interpreted to imply also useful predictive power for systematic asset allocation decisions via equations (3)–(7). This 'significance testing' setting is, however, commonly reported to suffer 'Stambaugh bias', i.e. substantial size distortions when $x_{i,t}$ is highly persistent and correlated with return innovations $\varepsilon_{i,t}$: See Stambaugh (1999), Rapach and Zhou (2013, Section 3.1) and the recent contributions of Demetrescu et al. (2020) as representative references of voluminous in-sample return predictability results and techniques aiming to improve statistical inference in model (9).

(iv) Finally, and importantly as also argued by Brandt (2010) and Elliott and Timmermann (2016, Section 4.2) in their surveys, the two-step predictive regressions-based approach to set the weights (7) belongs to the 'plug-in' methods: The statistical least squares criterion (5) in econometric modelling does not (most likely) line up with investors' true preferences and the final objective which is in asset allocation decisions. Even though the two-step approach is arguably a simple way to find required predictions to be plugged in (7), it generally leads to a discrepancy between the original goal (utility maximization) and the objective function in econometric inference.[3]

2.2 Objective function

To acknowledge all the difficulties (i)–(iv) reviewed in Section 2.1 and connected to the two-step approach around (linear) predictive regressions, we introduce a flexible non-

[3] Sentana (2005) considers formal conditions and assumptions under which least squares-based predictions and mean-variance analyses are connected in associated market timing strategies.

parametric and nonlinear approach which is strictly building upon utility maximization also in empirical implementation to obtain weights w_t. We are thus integrating asset allocation decision making and machine learning via a customized gradient boosting algorithm with specific and important modifications over the mechanical use of existing machine learning algorithms. Section 2.3 presents details of the 'utility boosting' algorithm, with various favourable properties.

Our utility boosting approach follows the general arguments made by Brandt (1999), Aït-Sahalia and Brandt (2001), Brandt and Santa-Clara (2006) and Brandt et al. (2009) in the early literature, arguing the importance of direct portfolio decisions instead of the two-step 'plug-in' statistical approach. Along their studies, we are not explicitly relying on specific assumptions on the excess stock return data generating mechanism, such as the arbitrage pricing theory (APT) or the capital asset pricing model (CAPM). Instead, our (empirical) view is that the portfolio weight w_t is a direct, potentially highly nonlinear, function of state variables $\boldsymbol{x}_{t-1}$ maximizing the investor's utility. This linkage incorporates all the predictive information contained in $\boldsymbol{x}_{t-1}$ to determine the weights, irrespective of the conclusions on statistical mean return predictability as discussed in (ii) and (iii) in Section 2.1, including also the possible impact of higher conditional moments than just mean and variance.

On these past utility-based approaches, the closest to our approach seems Brandt and Santa-Clara (2006) where they parametrize the portfolio weight as a linear function of predictors (state variables) $\boldsymbol{x}_{t-1}$ and solve the optimal values of the present parameters maximizing expected quadratic utility function similar to (2). Their approach is empirically, however, much more restrictive than ours (see details in the Appendix A) and designed more closely on portfolio choice problems with a genuine cross-sectional dimension (i.e. multiple risky assets). Moreover, their resulting portfolio weights can be interpreted as being proportional to the standard OLS regression of a vector of ones on the excess returns and, importantly, additional subsequent constraints to maintain the bounds (8) are required to address the point (i) in Section 2.1.[4]

[4] In addition to the OLS interpretation of the portfolio weights as obtained in Brandt and Santa-Clara (2006), Brandt (1999) and Brandt et al. (2009) build upon on the (statistical) method of moments and

To set a tractable empirical counterpart of (2), we first convert the maximization problem to a minimization problem. We will train our boosting algorithm with the same training (estimation) data $\{\boldsymbol{x}_{t-1}, r_{e,t}\}_{t=1}^{T}$ (including also the initial values for the volatility proxy) as in the least squares-optimal statistical two-step approach in equations (4)–(7) where $r_{e,t}$ implicitly contains required information on both $r_{m,t}$ and $r_{f,t}$. Our utility-based empirical objective function (cf. (2)) is

$$\arg\min_{w_t} \left\{ \frac{1}{T} \sum_{t=1}^{T} -u_t \right\} = \arg\min_{w_t} \left\{ \frac{1}{T} \sum_{t=1}^{T} - \left(r_{p,t} - \frac{1}{2} \gamma\, w_t^2\, \sigma_t^2 \right) \right\}, \tag{10}$$

where $r_{p,t}$ denotes the resulting portfolio returns (see (1)). The (negative) utility contribution of the tth observation, $-u_t$, is crucially dependent on the weights w_t constructed with the information contained in $\boldsymbol{x}_{t-1}$. The selected volatility proxy σ_t^2 is already introduced in connection to (7): It is the same as in (10) and in the two-step approaches throughout this study for comparison reasons (cf. a different formulation in this respect in Brandt and Santa-Clara (2006)). The form (10) is the one examined in various return predictability studies as an evaluation diagnostic tool measuring portfolio performance (see Marquering and Verbeek, 2004; Campbell and Thompson, 2008; Rapach et al., 2010; Rossi, 2018) and hence it acts as a natural choice for our advancement.

For simplicity and in accordance with past closely related studies, throughout this study, we set the lower bound in (8) as $w^{\min} = 0$ and hence excluding short selling.[5] Moreover, to respect the pre-determined maximum weight $w^{\max}$ explicitly as a part of our procedure (cf. the subsequent weight truncation needed in (7)), we set

$$w_t = w^{\max} \lambda_t, \tag{11}$$

'maximum utility estimator', respectively. At the end, they also emphasize statistical hypothesis testing with the aim to obtain 'statistically significant' results in the evaluation stage, whereas our objective is different and specifically in prediction (forecasting) performance for asset allocation decisions after direct utility-based modelling.

[5] An extension allowing also for short selling (i.e. negative weights) is possible but means a different and slightly more complicated parametrization than the one in (11)–(12).

where the portfolio weight is essentially specified by the logistic function

$$\lambda_t = \frac{1}{1 + \exp\left(-\frac{1}{\gamma\,\sigma_t^2} F(\boldsymbol{x}_{t-1})\right)}. \tag{12}$$

Combined with (11), the logistic growth curve form (12) guarantees that the weights (11) are all the time inside the interval $w_t \in [0, w^{\max}]$. In (12), the essential ingredient to determine portfolio weights (cf. (4)) is the component $F(\boldsymbol{x}_{t-1})$, which is possibly a complex function of the predictive information $\boldsymbol{x}_{t-1}$ that we aim to teach with our training data. To this end, in Section 2.3, we specifically develop a customized version of the gradient boosting, which is in principle only one but seemingly highly relevant empirical algorithm (implementation strategy) over alternatives to determine $F(\boldsymbol{x}_{t-1})$ when aiming to maximize the empirical utility.

In specification (12), the impact of $F(\boldsymbol{x}_{t-1})$ is adjusted by the risk proxy σ_t^2 and the risk aversion coefficient γ along the expression (3).[6] We can also strengthen the linkage to (7) by the following 'payoff to stake' representation

$$\log\left(\frac{\lambda_t}{1-\lambda_t}\right) = \log\left(\frac{w^{\max}\lambda_t}{w^{\max}(1-\lambda_t)}\right) = \frac{F(\boldsymbol{x}_{t-1})}{\gamma\,\sigma_t^2}.$$

This is the log odds ratio of λ_t where higher λ_t reflects the likelihood of high portfolio weight should take place. In addition to this representation, the inclusion of the volatility proxy σ_t^2 in (12) is strongly motivated by the early empirical evidence of Fleming et al. (2001) and Marquering and Verbeek (2004) on the importance of an explicit volatility component to determine portfolio weights.

In practice, the utility boosting algorithm in Section 2.3 provides the practical method to determine the weights. It can intuitively be interpreted so that an investor, with given risk aversion preferences and constraints (8), aim to continuously optimize and update his or her asset allocation decision mechanism with the available past predictive information targeting to find optimal portfolio weight for the next period. This emphasizes

[6] The identity $E_{t-1}(r_{e,t}) = w_t^*\,\gamma\,\sigma_t^2$ in (3) shows the linkage between the estimated weights and implied expected excess returns. Even though it is not of our main interest, the extracted weights (11), obtained with (12), can thus also be interpreted to imply a proxy for expected excess stock returns.

the training step of our algorithm before predicting the weight for the next period out of sample. This thinking again diverges from the goals of Brandt and Santa-Clara (2006), and a few closely related studies, concentrating on full sample portfolio policies.

2.3 Customized gradient boosting

In utility maximization (i.e. to minimize (10)), we determine the optimal weights directly by using a customized gradient boosting algorithm, acknowledging all the complications (i)–(iv) of the current two-step statistical approach presented in Section 2.1. Specifically, this means to specify the ingredient $F(\boldsymbol{x}_{t-1})$ in (12), leading to the 'boosted' portfolio weights maximizing the empirical utility as the objective function instead of aiming to predict excess stock returns with predictive regressions.

Boosting is a powerful technique originating from the machine learning community. The general idea is that simple weak models, also known as 'base learners', are combined in a stagewise manner to form a 'boosting ensemble' with strong predictive performance, leading to superior portfolio weights in our context. Friedman et al. (2000) introduced the statistical framework for boosting which enables additional theoretical insights into the success of boosting and has led to a variety of new boosting algorithms.

Provided with sufficient amount of data and flexible base learners, such as regression trees or smoothing splines, boosting can basically approximate any kind of functional form to determine weights w_t (cf. the linearity assumption in (4) before the weight truncation (8)). Boosting also performs model selection simultaneously with estimation as each new optimal base learner function is found by conducting an extensive search involving all the predictor variables. This makes boosting a viable algorithm for (relatively) large predictor sets, such as the one of interest in Section 3. Another major advantage is the interpretability of the final model: Unlike the mechanical use of existing machine learning algorithms, as we are now building the method explicitly on the financial economics and asset pricing bases, the final outcome provides important insights on the most relevant predictors (state variables) specifically for portfolio weight determination.

Boosting has shown considerable success in different fields including robotics, medical statistics and economics. Past financial applications range from stock return predictions (Rossi and Timmermann, 2015; Rossi, 2018), volatility forecasting (Mittnik, Robinzonov and Spindler, 2015), yield curve modelling (Audrino and Trojani, 2007) and failures in banking sector (Carmona, Climent and Momparler, 2019). Our context is, however, different than allowed by the basic setup of gradient boosting. That is, here the boosting algorithm contains the customized objective function, the (negative) utility function (10), while the usual boosting algorithms are fully statistical containing mean square error (MSE) and likelihood-based ingredients. This shows that instead of the mechanical use of gradient boosting, our goal is to specifically integrate a flexible modern machine learning algorithm with the asset allocation decision objective.

Following Bühlmann and Hothorn (2007) and the bulk of the boosting estimators, our customized algorithm can be described as follows:

Input: The training data $\{(\boldsymbol{x}_{t-1}, r_{e,t})\}_{t=1}^{T}$ is the same as in linear predictive regressions (4)–(7). The essential part is to determine $F(\boldsymbol{x}_{t-1})$ in (12) using a differentiable objective function (10), given the fixed risk aversion coefficient γ, selected volatility proxy σ_t^2 and the number of iterations M (see below).

Algorithm[7]:

1. Initialize the algorithm by a constant value: We select, for simplicity, $F_0(\boldsymbol{x}_{t-1}) = 0$.

2. For $m = 1, \ldots, M$:

   (a) Compute 'pseudo-residuals':

$$\widetilde{y}_{t,m} = -\left[\frac{\partial(-u_t)}{\partial F(\boldsymbol{x}_{t-1})}\right]_{F(\boldsymbol{x}_{t-1})=F_{m-1}(\boldsymbol{x}_{t-1})} \quad \text{for } t = 1, \ldots, T.$$

   The gradient is obtained by the chain rule. Using the expression (12), given

[7] A slightly modified version will be considered in the Internet Appendix A.3. to obtain evidence for the robustness of our main findings.

the selected volatility proxy σ_t^2 and noticing the negative sign in (10), we get:

$$\frac{\partial(-u_t)}{\partial F(\boldsymbol{x}_{t-1})} = \frac{\partial(-u_t)}{\partial w_t}\frac{\partial w_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial F(\boldsymbol{x}_{t-1})} = -\Big(r_{e,t} - \gamma\, w_t\, \sigma_t^2\Big)\frac{w^{\max}\,\lambda_t(1-\lambda_t)}{\gamma\,\sigma_t^2}. \quad (13)$$

(b) Fit a base learner $h_m(\boldsymbol{x}_{t-1})$ to pseudo-residuals, i.e. train it using the training set $\{(\boldsymbol{x}_{t-1}, \widetilde{y}_{t,m})\}_{t=1}^{T}$.

(c) Update the prediction: $F_m(\boldsymbol{x}_{t-1}) = F_{m-1}(\boldsymbol{x}_{t-1}) + \upsilon h_m(\boldsymbol{x}_{t-1})$.

3. Output $F_M(\boldsymbol{x}_{t-1})$, leading to the solution of (10), given (11) and (12).

As a whole, mathematically the above algorithm can be understood to minimize (10) by iterative steepest descent in function space. Bühlman and Hothorn (2007) call it the functional gradient descent algorithm. After initializing the algorithm in step 1 with a simple scalar value the pseudo-residuals in step 2.(a) can be constructed. These are the negative gradients of the objective function evaluated with the boosting ensemble learnt so far with $F_{m-1}(\boldsymbol{x}_{t-1})$. The base learner model that best fits the negative gradient is then selected and added to the ensemble. The impact of each update is shrunk towards zero using a step-length factor $\upsilon \in \{0, 1\}$, which we set to 0.001 according to the past boosting studies. The whole process (step 2) is repeated 500 times (i.e. $M = 500$). For the performance of the boosting estimator, the base procedure and the stopping rule in step 2 are the most important ones. These and other tuning parts of the estimator will be presented and discussed in more detail in the Appendix B.[8]

With the above gradient boosting, we can, in theory, reach the ultimate equilibrium with excessive boosting iterations M where the algorithm is overfitted to the training (in-sample) data. However, it should be emphasized that especially in out-of-sample predictions, we only use the data that we have at hand.[9] We will control the boosting iterations and the underlying speed of learning in estimation (learning) stage to terminate

[8] Following various utility configurations (see, e.g., by Aït-Sahalia and Brandt (2001)), also examined within the numerical 'full-scale optimization' search algorithms (see Adler and Kritzman, 2007), our method can also, in principle, be extended to various other (differentiable) utility functions than the one in (2) and (10). The quadratic form (2) is expected to provide at least a good approximation for most variations of power utility (see also, e.g., Campbell and Thompson, 2008, footnote 11) and is, in particular, in line with the past return predictability studies that we are mainly linking our approach and empirical analysis in Section 3.

[9] To clarify our notation, as our estimation (training) data contains the observations $\{(\boldsymbol{x}_{t-1}, r_{e,t})\}_{t=1}^{T}$,

the iterative fitting (steps 1–3) at the right time, to get portfolio weights based on their genuinely predictable patterns of interest rather than ending up to overfitted weights. This means that the resulting weights and asset allocation decisions are obtained directly with the aim to maximize (2) but at the same time circumventing overfitting of the training data. These selections enable also meaningful full sample analyses, as generally examined in empirical finance research so far, on the importance of different predictors (state variables) in portfolio weight determination.

3 Empirical results

In empirical analysis below, we compare our utility boosting method to the existing return predictability studies, with out-of-sample forecasting emphasis, using the same dataset (Section 3.1) and evaluation criteria (Section 3.2) as in Neely et al. (2014), Welch and Goyal (2008), Campbell and Thompson (2008), Rapach et al. (2010) and Rapach and Zhou (2013). In Sections 3.3–3.4, we report the full sample (in-sample) estimation results, with the special interest on comparing the best predictors for direct portfolio weight determination. Section 3.5 reports the out-of-sample asset allocation results. In Sections 3.3–3.5, we concentrate on reporting the results, whereas in Section 4 we summarize and discuss the obtained empirical results for more general conclusions. Various additional results and robustness checks for our main findings are compiled into the attached Internet Appendix.

3.1 Dataset

Following the closely related return predictability studies, we consider an updated monthly dataset compiled by Welch and Goyal (2008) and extended by Neely et al. (2014), containing various predictive variables to determine portfolio weights directly in our single-step and indirectly in the conventional statistical two-step plug-in methods.

in forecasting situation portfolio weight forecasts are made for the time $T+1$, i.e. w_{T+1}, using the information up to time T. This explains the difference in 'forecasting' and more general 'prediction' goals including full sample (in-sample) results.

As in Neely et al. (2014), our sample period starts from December 1950 and is now updated until December 2018.[10]

Excess stock returns (i.e. the equity risk premium) is obtained as the difference between the return on the S&P 500 index (including dividends) and the risk-free interest rate. The set of macroeconomic predictive variables contains 14 variables as originally in Welch and Goyal (2008). Detailed descriptions of all predictors are presented in Table 1. Table 2 reports the summary statistics of monthly excess stock returns (Panel A) and macroeconomic variables (Panel B). The average monthly equity risk premium is 0.60% with monthly standard deviation of 4.14%. This produces a monthly Sharpe ratio of 0.144. As the Panel B shows, most of the macroeconomic variables are rather persistent (cf. the challenge (iii) in Section 2.1) including, as expected, the valuation ratios, nominal interest rates and partly also interest rate spreads.

Together with the macroeconomic variables, Table 1 also lists 14 binary-valued technical indicators as predictors. These are based on the same three representative trend-following technical indicators as in Neely et al. (2014). As they summarize it (see also, e.g., Rapach and Zhou, 2013), instead of macroeconomic fundamentals, technical indicators rely on past price and volume patterns with the idea that they will identify future price trends. The theoretical models explaining the predictive power of technical indicators are based on how investors can be heterogeneous regarding the availability of new information, their response to this information and how they view the overall investor sentiment (see Neely et al., 2014, and the references therein).

The first indicators are moving average (MA) buy and sell signal rules

$$\mathrm{MA}(s,l) = I(\mathrm{MA}_{s,t} \geq \mathrm{MA}_{l,t}), \tag{14}$$

[10] The Welch-Goyal dataset, containing macroeconomic variables and the S&P 500 index (returns), is from Amit Goyal's website at http:/www.hec.unil.ch/agoyal. Technical indicators are obtained with the S&P 500 data. For a general and extensive summary of the past return prediction findings obtained with different predictors and existing methodological approaches (see Section 2.1), see the survey of Rapach and Zhou (2013).

where $I(\cdot)$ is an indicator function and

$$\mathrm{MA}_{j,t} = \frac{1}{j}\sum_{i=0}^{j-1} P_{t-i}, \quad j = \{s, l\}, \quad s = \{1, 2, 3\}, \quad l = \{9, 12\},$$

where P_t is the level of the S&P 500 index. These MA rules reflect the short- and long-run trends in stock price movements. Moreover, the second set of technical indicators are based on the momentum signals

$$\mathrm{MOM}(m) = I(P_t \geq P_{t-m}), \quad m = \{9, 12\}, \tag{15}$$

where a positive (negative) momentum signal means that the current value of the index is higher (smaller) than m periods ago. Finally, the third set of technical indicators incorporates also the volume data to identify market trends. Define

$$\mathrm{OBV}_t = \sum_{k=1}^{t} \mathrm{VOL}_k D_k, \qquad D_k = 2\, I(P_k - P_{k-1}) - 1,$$

where VOL_k is the trading volume during the period k. The trading signal is then

$$\mathrm{VOL}(s, l) = I(\mathrm{MA}_{s,t}^{\mathrm{OBV}} \geq \mathrm{MA}_{l,t}^{\mathrm{OBV}}), \tag{16}$$

where the volume data of the S&P 500 index is obtained from Finance Yahoo and

$$\mathrm{MA}_{j,t}^{\mathrm{OBV}} = \frac{1}{j}\sum_{i=0}^{j-1} \mathrm{OBV}_{t-i}, \quad j = \{s, l\}, \quad s = \{1, 2, 3\}, \quad l = \{9, 12\}.$$

Intuitively, say, relatively high recent trading volume combined with recent price increases indicates a strong positive market trend. Like with macroeconomic variables, the predictive information of the technical indicators (14)–(16) at time $t-1$ is used in $\boldsymbol{x}_{t-1}$ and in the weight determination for the period t. Due to the binary nature of (14)–(16), we do not report the descriptive statistics of technical indicators in Table 2.

3.2 Evaluation and benchmarks

Before reporting our in- and out-of-sample prediction results in Sections 3.3–3.5, we introduce our main evaluation criteria and benchmark approaches. In the past research, economic evaluation criteria, such as resulting utilities and portfolio returns after fitting the linear predictive regression (4) as in (5)–(6), have typically been treated as secondary over the statistical ones and return predictability considerations. As briefly reviewed in Section 2.1, model (4), and (9) as a typically considered special case, has potentially difficult econometric issues related to the evaluation of statistical significance of persistent predictors in $\boldsymbol{x}_t$. Importantly from our point of view, we are stressing that the usefulness of a certain model, with one or more predictors, is not determined by the conventional statistical significance of individual coefficients. Instead, it is directly its usefulness in asset allocation decisions what matters.[11]

The first economic and profit-based evaluation criteria is naturally the resulting average utility (cf. the objective function (10), now percentages per month)

$$\bar{u}(\widehat{\boldsymbol{w}}) \quad = \quad 100 \times \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t = \frac{100}{T}\sum_{t=1}^{T}\left\{\widehat{r}_{p,t} - \frac{1}{2}\gamma\, \widehat{w}_t^2\, \widehat{\sigma}_t^2\right\}, \tag{17}$$

where $\widehat{\boldsymbol{w}} = (\widehat{w}_1, \ldots, \widehat{w}_T)$ is the vector of estimated weights, $\widehat{u}_t$ is the utility contribution of tth observation and $\widehat{r}_{p,t}$ is the resulting portfolio return (i.e. expression (1) evaluated with $\widehat{\boldsymbol{w}}$). To enable explicit comparison between different methods, throughout this study, we utilize the same conditional variance estimate $\widehat{\sigma}_t^2$ (i.e. the 60-month rolling window volatility) constructed using the information available at time $t-1$. Robustness checks with the GARCH model and realized variance (RVOL, see Table 1) based volatility proxies, reported in the Internet Appendix A.4, lead to largely the same main empirical conclusions. The risk aversion coefficient γ is fixed to the typical and commonly used value $\gamma = 5$ (see, e.g., Neely et al., 2014; Cenesizoglu and Timmermann, 2012; Brandt and Santa-Clara, 2006) or $\gamma = 3$ (see, e.g., Campbell and Thompson, 2008; Rapach et al.

[11] As in the footnote 9, it is important to point out that below the notation $t = 1, \ldots, T$ refers to the in-sample (training) performance. However, the same criteria will be used to evaluate out-of-sample portfolio performance with necessary changes in notation. In other words, portfolio weight forecasts for the period $T+1$ are constructed using the information available at time T.

2010; Zhu, 2015).[12]

An important and commonly used benchmark to our utility boosting method, and also for the predictive regressions with the subsequent restrictions (8), is that predictors $\boldsymbol{x}_{t-1}$ do not have useful predictive power to determine portfolio weights. For the linear predictive regressions, this implies that the expected excess return (6) is constant over time and (4) contains only the constant term. In the two-step approach, this leads to the use of historical average ('HA') return, $\bar{r}_{e,t}$, and hence the weights

$$\bar{w}_t = \frac{\bar{r}_{e,t}}{\gamma\,\widehat{\sigma}_t^2}. \tag{18}$$

This is the simple benchmark commonly employed in the past (out-of-sample) return predictability literature and subsequent portfolio performance evaluation. As argued by Welch and Goyal (2008), and various subsequent studies, it is a highly adequate statistical description for excess stock returns. Notice that despite the constant expected return, the weights (18) are time-varying and governed by the estimated volatility proxy $\widehat{\sigma}_t^2$. Building upon the thinking of (18), we consider a restricted utility boosting approach where $\boldsymbol{x}_{t-1}$ contains only the constant (i.e. $\boldsymbol{x}_{t-1} = 1$) for all t. That is 'Const' approach (cf. (12)) hereafter, where $F(\boldsymbol{x}_{t-1})$ in the customized gradient boosting algorithm does not include any predictive information over the conditional volatility σ_t^2, which is in line with the idea of volatility timing (see the general arguments favouring such approach in Fleming et al. (2001)).

A closely related evaluation criterion to the average realized utility (17) is the certainty equivalent return gain over the historical average (HA) weights (18):

$$\text{CER gain} = 1200 \times (\text{CER} - \text{CER}_{\text{HA}}), \tag{19}$$

[12] Selections $\gamma = 4$ and $\gamma = 6$ have also commonly been used (see, e.g., Marquering and Verbeek, 2004; Rossi, 2018) and hence $\gamma = 5$ turns up a suitable compromise between them. Moreover, following the reasoning of Rossi (2018), moving from $\gamma = 3$ to $\gamma = 5$ provides guard and a reasonable check against the impact of estimation uncertainty by increasing the value of γ. Additional arguments favouring the importance of incorporating estimation uncertainty in asset allocation decisions can be found, e.g., in Kandel and Stambaugh (1996), Barberis (2000) and Kan and Zhou (2007).

where

$$\text{CER} = \bar{r}_{p,t} - \frac{\gamma}{2}\,\text{Var}(\widehat{r}_{p,t}), \qquad \bar{r}_{p,t} = \frac{1}{T}\sum_{t=1}^{T}\widehat{r}_{p,t}, \tag{20}$$

and CER_{HA} is obtained when replacing $\widehat{w}_t$ by $\bar{w}_t$ in $\widehat{r}_{p,t}$ construction. The difference to (17) is that in (20) the volatility estimate is the resulting portfolio variance. This change does not turn out to change our main empirical results at all, providing important robustness for our findings. The CER gain (19), likewise differences in the estimated average utilities (17), can be interpreted as the received additional economic value of the utility boosting method over the historical average benchmark (18). We report the CER gain (19) multiplied by 1200 so that it can be interpreted as the annual percentage portfolio management fee that an investor would be willing to pay to have an access to the utility boosting-based asset allocations instead of (18).

In addition to the above criteria, we also consider the traditional (monthly) Sharpe ratio

$$\text{Sharpe} = \frac{\frac{1}{T}\sum_{t=1}^{T}\widehat{r}_{ep,t}}{\sqrt{\text{Var}(\widehat{r}_{ep,t})}}, \tag{21}$$

where $\widehat{r}_{ep,t} = \widehat{r}_{p,t} - r_{f,t}$ denotes the resulting portfolio return in excess of the risk-free rate. The Sharpe ratio (21) is hence the mean portfolio return in excess of the risk-free rate divided by the standard deviation of excess portfolio returns. As a reward-to-variability ratio, the Sharpe ratio measures the additional amount of excess return that an investor receives per unit of increase in risk. Asset allocation decisions (strategies) with a high Sharpe ratio are preferable to those with a low Sharpe ratio.

3.3 In-sample (full sample) results

Following the common practice in empirical finance and specifically in stock return predictability studies, before examining out-of-sample forecasting results in Section 3.5, we consider the full sample results for the sample period 1951:1–2018:12 (816 observations). The main interest in this and Section 3.4 is to examine differences between the introduced utility boosting method and various benchmarks to explore which predictive variables are the most useful ones for portfolio weight determination when the maximum amount of information (the longest data availability) is employed in the analysis.

As explained in Section 2.3, it should be emphasized that our utility boosting approach contains shrinkage type of elements and hence provides guard against potential overfitting concerns, even for in-sample analysis. This means, together with the fact that the utility boosting and the two-stage predictive regression-based weighting are not generally nested approaches (i.e. either one is not obtained as a special case of the other one), that utilities resulting from the former are not automatically higher than in the latter in and out of sample. This should happen especially if the predictive power of a certain variable $x_{i,t}$ for portfolio weight determination is indeed non-existent or negligible. Moreover, in line with the arguments of Inoue and Kilian (2005), it should be kept in mind that there are also potential complications in out-of-sample forecasting evaluation, such as the selection of the evaluation period and other potentially random (outlier-type) events, which may obscure forecasting results as well.

Table 3 presents the in-sample (full sample) results for the single-predictor models, and the benchmarks described in Section 3.2, for the investor with the risk aversion coefficient of $\gamma = 5$. We report the average utilities (17) and Sharpe ratios (21) for the single-step utility boosting and the two-step (statistical, 'linear') approaches. The resulting portfolio weights fulfil the bounds (8), which are $w^{\min} = 0$ and $w^{\max} = 1$ in these results. For illustrative reasons, we present both the average utilities (17) (util%) and CER gains (19) for the boosting method, leading to essentially the same conclusions here and also in other analysis settings. Moreover, even though the statistical significance of the individual predictors is indeed secondary in this study, to establish comparisons to the past linear predictive regressions studies, we also report the typical heteroskedasticity-autocorrelation consistent (HAC) t-statistics[13] and adjusted-R^2s

[13] Throughout this study, we systematically report the HAC t-statistics by means of the Newey-West (1987) estimator with lags determined by the rule (integer) $\text{floor}(4 \times (T/100)^{(2/9)}$. In other words, we do not take any specific standpoint on the statistical significance of the estimated β_i coefficients in (9), or the appropriateness of the HAC standard errors for all the predictors, due to the potential Stambaugh bias as discussed in the point (iii) in Section 2.1.

('adj-R^2') to measure the degree of statistical return predictability. As pointed out in Section 2.1, these statistics provide only partial evidence when the final objective is in asset allocation decisions and economic value of predictions.

Starting with the conventional statistical criteria, given the inherently substantial unpredictable component in monthly stock returns, Campbell and Thompson (2008) and Neely et al. (2014) concluded that a monthly adjusted-R^2 near 0.5% might represent economically significant degree of equity risk premium predictability. Together with using, e.g., the t-value close to 2 in absolute value as an indicative threshold for statistically significant predictors (at the 5% significance level), Table 3 shows as a whole that technical indicators seem to express useful but small statistical predictive power. There are also a few macroeconomic predictors, mainly TBL, LTR and RVOL, with statistically significant predictive content. When moving to realized utilities (util% and CER gain), utility boosting yields consistently higher average realized utility levels over the two-step 'linear' approach and benchmarks ('Const' and 'HA').[14] It is not surprising that in the full sample results our flexible non-parametric approach can find higher economic predictive power in the best (mostly real-valued) macroeconomic variables with potentially larger amount of information than the binary-valued technical indicators. This fact already shows, together with the analysis of Neely et al. (2014), that it is reasonable to treat these two categories of predictors partly separately at least in this section.

The best macroeconomic predictors in terms of average utility for the investor (when $\gamma = 5$ and $w^{\max} = 1$) are DY, LTR and RVOL. In contrast to other best predictions in terms of statistical criteria, DFR (default return spread) and inflation (INFL) have in relative sense much more useful information content for the direct portfolio weight determination than obtained with the statistical predictive regressions and statistical goodness-of-fit measures. Moving to the results of the technical indicators, the shortest

[14] There is one exception, MA(2,12), with a very marginal difference. As pointed out above, this is possible due to the elements controlling overfitting of the boosting algorithm. Similarly, for DE and DFY the utility boosting does not find substantial additional value in this setting ($\gamma = 5$ and $w^{\max} = 1$). These cases importantly show that the flexible utility boosting do not automatically outperform the linear two-step approach in sample and hence some overfitting concerns can already be eased.

MA rules (MA(1,9)–MA(3,9)), together with VOL(3,12), have the highest predictive power. These are largely the best ones also in terms of statistical criteria, even though some minor differences occur.

In Table 4, we provide the first of many alternative specifications to the one considered in Table 3. The risk aversion parameter is now lower ($\gamma = 3$) but still a common selection in the past related studies (see Section 3.2), implying less risk aversive investor profile. The main results are, however, largely the same as in Table 3 where $\gamma = 5$. The biggest differences are the rise of TMS and TBL in terms of received utilities whereas specifically LTR performs well in predictive regressions but does not stand out as a strong predictor in the utility boosting relative to other best predictors. As in Table 3, DFR and INFL are again examples of the best performing state variables in which the utility boosting finds much more useful predictive content than in the past two-step statistical approach. As the level of risk aversion, and hence the impact of conditional variance in (17), decreases due to the lower value of γ, the realized utility levels are throughout somewhat higher than in Table 3.

One of the main empirical results of this study is already evident in Tables 3–4: In terms of the Sharpe ratio (21), the utility boosting method strongly outperforms the conventional two-step statistical approach. That is the received risk-adjusted portfolio returns are substantially and throughout higher in the boosting method. The exact additional value varies between the predictors, but the Sharpe ratios of the boosted weights and resulting portfolio returns are mostly about 10–40% higher than in the two-step approach, but still within the realistic values in sample. This superior risk-return compromise is coming largely from the smaller portfolio return variance: The mitigated volatility in the utility boosting applies to portfolio weights over the sample period ($t = 1, \ldots, T$) and hence the average utility.[15]

Another interesting predictor-specific empirical finding, together with the rise of inflation (INFL) and default return spread (DFR), is related to the dividend-price ratio (DP

[15] The resulting excess portfolio returns are also higher in the utility boosting approach and the best performing models are basically exactly the same as obtained with presented criteria. Moreover, the returns are in line with the conventional levels (about between 5-8% in annualized terms).

and the dividend yield (DY)). This is the variable which has especially been examined a lot in the past return predictability research due to its tight linkage to asset pricing theory and the relationship with the present value model. It is also probably the most commonly considered predictive variable in connection to the potential Stambaugh bias (point (iii) in Section 2.1), with ambiguous empirical conclusions. Goyal and Welch (2003, 2008) specifically argue that despite of their wide attempts, they could not find robust statistical predictive ability in the dividend-price ratio (dividend yield). The utility boosting approach clearly supports the usefulness of DY and DP much more when the objective is in the asset allocation decisions instead of the statistical performance.

As an illustrative example, in Figure 1 we depict the estimated portfolio weights $\widehat{w}_t$ in the utility boosting and the linear predictive regression-based two-step approaches using the dividend-price ratio (DP) as a predictor. There are several periods with clearly different weights between the methods. The deviation in the estimated weights is especially large during the 1960s and 1970s and from the year 2000 onwards. The impact of weight truncation to fulfil bounds (8) is also evident in the two-step approach, while the utility boosting keeps the weights automatically inside the selected interval. Weight truncation in the predictive regression-based approach is especially needed for the upper limit ($w^{\max} = 1$) and can be seen as multiple flat sections in Figure 1.

Similarly as in various return predictability studies, instead of single-predictor analyses, next we utilize multiple predictive variables in $\boldsymbol{x}_{t-1}$ simultaneously. As in Neely et al. (2014), we first consider the macro variables (MACRO) and technical indicators (TECH) separately and then finally jointly (ALL, i.e. combining MACRO and TECH) to determine asset allocation decisions. One of the advantages of the utility boosting method is that no pre-selection is required as the algorithm performs model selection internally. In contrast, following Neely et al. (2014), among others, in the two-step predictive regression approach we first extract the principal components of the candidate set of predictors (MACRO, TECH and ALL). Notice that the exact multicollinearity between some of the macro variables (see Table 1) implies that the direct use of OLS, as in (5), is not even possible without these additional steps.

Table 5 reports the results of multivariate predictor models in otherwise the same setting as above in predictor-specific analyses in Tables 3–4. In the Panels A and B, the risk aversion coefficients are $\gamma = 5$ and $\gamma = 3$, respectively. In the linear predictive regressions, we utilize the same principal component approach as in Neely et al. (2014).[16] This comparison strengthens the superior performance of the utility boosting: All the combinations outperform the benchmark cases reported in Tables 3–4 as well as the two-step 'linear' alternatives.

Neely et al. (2014) conclude that the macroeconomic variables and technical indicators provide almost completely complementary predictor sets to predict equity risk premium. When moving to direct asset allocation decisions, we can confirm this when evaluating linear predictive regressions with economic goodness-of-fit measures while in the utility boosting macroeconomic variables dominate technical indicators in sample. That is, in Table 5, the case ALL is driven by fluctuations in the macroeconomic variables with only minor role for technical indicators. This is not surprising at all: As discussed above, this is largely due to the nature of macroeconomic variables almost necessarily containing more information in sample than binary-valued technical indicators. Tree-based models are often reported to favor continuous-type predictors (see, e.g., Loh and Shih, 1997). It is hence also important to consider out-of-sample forecasting results before further conclusions.

In accordance with the above views, Table 6 presents the top-10 predictors chosen by the internal model selection capability of the customized gradient boosting. The relative influence criterion gives the normalized empirical improvement as a result of including a particular predictor in the final model. The most contributing macroeconomic and technical analysis predictors are in line with the univariate results in Tables 3–4. As in Table 5, it is noteworthy how the final model is essentially based on different macroeconomic predictors.

[16] This means that the maximum amount of principal components is set to 3 for predictor group MACRO, 1 for TECH and 4 for ALL, and the final selection is made with the adjusted-R^2 (adj-R^2).

3.4 In-sample extensions

As described in Section 2, our utility boosting method builds upon empirical utility maximization with the explicit linkage to the investor's preferences implemented via the sample objective function (10) and bounds (8). This leads to an important general point already validated in Section 3.3: Utility boosting is more relevant over the past two-step approach when the final interest is in asset allocation decisions. Empirically, this all of course indicate that the exact numbers, as already seen in Section 3.3 over the selections $\gamma = 5$ and $\gamma = 3$, might naturally be somewhat dependent on the risk aversion level, the maximum weight bound $w^{\max}$ and the volatility proxy σ_t^2. Therefore, in this section we still briefly present various additional and robustness analyses with detailed results compiled to the Internet Appendix.

In Section 3.3, we considered the setting where taking a leveraged position $w^{\max} > 1$ was not possible. Therefore, it is meaningful to also consider empirical results when the maximum weight is $w^{\max} = 1.5$. This has also been a rather common selection in the past predictive regression studies (see the point (ii) in Section 2.1). It turns out that the main findings on the best single predictors and differences between the utility boosting and the two-step approaches are essentially intact (see the Internet Appendix A.1). Mainly DY, LTR and TMS seem to perform in relative sense even somewhat better when an investor has an access to 50% leverage (vs. the case $w^{\max} = 1$).

One important extension to the analysis presented in Section 3.3 is to impose transaction costs as a part of the analysis. It is, however, largely an open issue how much emphasis we should put on this view. In addition to the fact that transaction costs are seemingly in practice becoming smaller and smaller all the time, from the methodological point of view imposing transaction costs means that otherwise optimal portfolio weight determination might be severely disrupted by effectively unnecessary (continuous) portfolio rebalancing activity. As an example, is it optimal to rebalance the portfolio at all when moving from, say, $w_t = 0.75$ to $w_{t+1} = 0.7$, given the loss of portfolio value due to transaction costs? Following this lead, in the optimal case transaction costs should be part of the econometric procedure and utility maximization. This is not, however, clear cut to implement and requires additional complicated steps extending the utility boosting introduced in Sections 2.2–2.3. Therefore, we content ourselves to the same view as in the past return prediction studies: We evaluate the received portfolio returns when transaction costs are imposed on the evaluation stage, after constructing first the same 'optimal' asset allocation decisions as above.

Following Marquering and Verbeek (2004), in the Internet Appendix A.2 we present the corresponding results as in Tables 3–4 but now also incorporating low and high transaction costs scenarios given as percentage points (low (0.1%) and high (0.5%)) of the value traded (see also Rossi, 2018, and the references therein). As in the various previous studies, utility gains naturally decrease due to transaction costs, but especially in the low transaction costs scenario the empirical results are still favourable for the utility boosting over the benchmarks. All in all, the results in Tables 3–4 can be interpreted as upper bound estimates for the received economic gains.

To get additional robustness for our findings, we also consider alternative volatility proxies to the well-established 5-year rolling window estimate. In the Appendix A.4., we consider the natural alternatives where the conditional mean of excess stock returns is constant and the GARCH(1,1) model equation is assumed for the conditional variance providing the volatility proxy σ_t^2. Another check is performed with the realized volatility series RVOL as described in Table 1 and Mele (2007). Again the exact numbers in the economic goodness-of-fit measures naturally differ, and certain variables (mainly the term spread (TMS) and earnings-price ratio (EP)) perform somewhat better than in our main analysis, but importantly the main conclusions are still intact.

Finally, the results in Tables 3–4, and more generally the coming Section 3.5, show that less persistent macroeconomic predictors perform relatively well in the utility boosting method. This suggests still to consider whether taking the first differences of the (highly) persistent predictors (i.e. all the variables except DFR, LTR and INFL) changes the big picture. It turns out that (see the Internet Appendix A.5), when concentrating on out-of-sample forecasting results as in the next section, especially the lagged changes in the dividend price ratio contain even higher predictive power than the level of the

DP, emphasizing the conclusion that this theoretically well-motivated state variable contains useful information in direct portfolio weight determination.

3.5 Out-of-sample forecasting results

In this section, we report out-of-sample asset allocation results based on portfolio weights obtained with the (single-step) utility boosting and (two-step) linear predictive regression-based methods using the same macroeconomic and technical indicator predictors as in Sections 3.3–3.4. The portfolio weight for the month t is thus constructed using the information at time $t-1$ and contained in $\boldsymbol{x}_{t-1}$ and the volatility proxy σ_t^2.

Due to additional flexibility allowed by our utility boosting method over the simple linear predictive regressions, we believe that the initial and expanding estimation sample in out-of-sample forecasting should be slightly longer than in various past studies. Therefore, we use January 1951 to December 1989 as the first initial estimation training sample window to generate portfolio weights for January 1990. When moving the forecast origin ahead the initial estimation sample is expanding by one new observation when constructing portfolio weights one month ahead. The forecasting evaluation period hence contains the predicted weights from January 1990 to December 2018 (348 observations).

We analyze portfolio (asset allocation) performance in terms of received average utility, CER gain and the Sharpe ratio (see equations (17), (19) and (21)), all now defined for out-of-sample forecasting evaluation. These are the economic evaluation criteria of interest, enabling a comparison between different approaches to determine portfolio weights. In line with the past out-of-sample return predictability studies, we do not claim that a representative investor would have ended up exactly to the best performing model at each step. Instead and completely in line with the past studies, we are interested in general findings on the usefulness of our new method over the past ones in a relatively large set of predictors.

The main interest and views to be considered in this section are at least the following ones, together with a comparison to the full sample results:

(a) To which extent our utility boosting method can outperform the benchmarks ('Const' and 'HA'), essentially claiming that there is no useful predictive information in the state variable (variables) $\boldsymbol{x}_{t-1}$? Specifically the historical average ('HA'), i.e. the constant expected equity premium forecast (see (18)), has been a popular and very stringent out-of-sample benchmark (see, e.g., Welch and Goyal, 2003, 2008; Campbell and Thompson, 2008; Neely et al., 2014). Linear predictive regressions containing individual (macroeconomic) variables typically fail to outperform this historical average statistically. Moreover, there is also a quite unanimous stylized fact in empirical finance that the time period since 1990 (as examined here) is the most challenging among the past decades in terms of out-of-sample predictability of stock returns (see, e.g., Campbell and Thompson, 2008; Lettau and Van Nieuwerburgh, 2008), presumably making it difficult to any new method to find meaningful additional predictive value.

(b) The post-truncation of the weights obtained with linear predictive regressions (4) to fulfil the bounds (8) has been found very useful in the past out-of-sample forecasting studies (see (ii) in Section 2.1). Therefore, this past 'truncated linear' approach is expected to be a tough and well-trained competitor to beat when the out-of-sample economic forecasting performance is of interest.

(c) Following the main findings of Neely et al. (2014), obtained with the past two-step method, technical indicators turned out to outperform macroeconomic predictors in terms of both statistical and economic out-of-sample evaluation criteria. Whether this is the case in our approach is also of particular interest.

Tables 7–8 report the out-of-sample forecasting results for the benchmarks and single-predictor models. In these results, we have assumed that the relative risk aversion coefficient γ is fixed to either $\gamma = 5$ (Table 7) or $\gamma = 3$ (Table 8) with the upper bound weight constraint set as $w^{\max} = 1$. As expected, the utility levels and Sharpe ratios are generally somewhat lower than in the in-sample results and that the historical average (HA) and the two-step linear approaches (with weight truncation to fulfil (8)) are indeed and in accordance with the past studies performing well and tough competitors to the utility boosting. There are also several candidate predictors that either the utility boosting or the two-step approach cannot find useful out-of-sample predictive power over the historical average weights (18).

As a whole, the results confirm that when a certain predictor has (some) reasonable out-of-sample asset allocation predictive power in this sample period, the Sharpe ratios are almost uniformly higher (i.e., higher risk-adjusted returns) in the boosting approach over the conventional statistical approaches. The results are also largely intact to the change of risk aversion level ($\gamma = 5$ and $\gamma = 3$).

In accordance with Neely et al. (2014), technical indicators perform as a group better than macroeconomic variables in out-of-sample forecasting purposes and the best (single) indicators are the same ones (i.e. MA(1,9), MA(2,12), VOL(1,12) and VOL(3,12)) as in the in-sample analysis, showing their robustness as predictors. In these cases, economic gains over the benchmarks are clear. For the MA(2,12) and VOL(3,12), the linear approach performs equally well as the utility boosting, which again points out the successfulness of the past truncated two-step approach. The simple binary nature of technical indicators is likely an important contributing factor for seemingly better out-of-sample performance over macroeconomic variables.[17] The latter ones are much more vulnerable to possible, and also arguably happened, time-varying instability issues in this period of interest, including, among others, the IT-bubble around the millennium and low interest rates at the end of the evaluation sample period.

Figure 2 illustrates the estimated out-of-sample portfolio weights for both methods using the well-performing technical indicator MA(1,9) (see Table 1). Like in the in-sample results (cf. Figure 1), there are periods where the weight truncation is indeed strongly taking place in the two-step statistical approach. It is noteworthy how these truncation periods can last even for several years.

As mentioned above and in line with past return predictability findings, there are only a few macroeconomic variables which seem to have reasonable out-of-sample predictive power in this sample period. From the utility boosting perspective, in line with

[17] Due to the binary nature of the variables, the superiority of the utility boosting is coming solely from the different objective function as non-linearities behind the construction of $F(\boldsymbol{x}_{t-1})$ in the boosting estimator play no role for binary-valued predictors.

the full sample results, inflation (INFL) and default return spread (DFR) are the best ones and perform relatively much better than in the 'linear' statistical approach. On the other hand, supposedly the well-documented erratic behaviour of the dividend-price ratio (DP) around the end of the 1990s and the beginning 2000s costs us in its performance when compared with the simple benchmarks. However, the utility boosting clearly outperforms the linear approach also with these predictors, as concluded already in the in-sample results. Moreover and importantly, when considering the monthly changes of the dividend-price ratio (ΔDP), instead of the levels, the utility boosting is able to find substantial out-of-sample predictive power (see details in the Internet Appendix A.5).

Tables 7–8 present and concentrate on univariate (single-predictor) out-of-sample forecasting results. As mentioned, we do not claim that the best-performing predictors were known in advance. Therefore, multivariate out-of-sample analysis provides important information on the internal model selection capability of the utility boosting to find useful predictors at a time and without any prior knowledge. Table 9 reports the multivariate results with both cases of $\gamma = 5$ and $\gamma = 3$. The 'linear' multivariate benchmark is built upon the principal component analysis (PCA) as in Neely et al. (2014). The multivariate out-of-sample results are very much in line with the single-predictor as well as the in-sample results: Each of the utility boosting-based predictions from different groups of predictors outperform the benchmarks. Models based on macroeconomic variables (MACRO) yield only slightly higher utilities compared to the benchmarks, whereas technical indicators (TECH) is the best performing group of predictors also in this multivariate setting.

When comparing the multivariate utility boosting to the two-step predictive regressions (combined with the PCA), we can see that the former outperforms the latter for both MACRO and ALL (all the predictors combined). For the technical indicators (TECH), the performances of the two methods are essentially the same, which is also evident in Figure 3 where the estimated portfolio weights using TECH are graphically illustrated for both methods (given $\gamma = 5$). Despite the slightly higher utility levels in the two-step approach, the risk-adjusted portfolio returns (and the Sharpe ratios) are higher in the utility boosting.

4 Discussion

The current stance in empirical finance appears that stock returns are time-varying and at least at times somewhat predictable. Whether these predictable statistical patterns translate to useful asset allocation decisions is arguably of the main interest for investors in practice instead of commonly used statistical criteria. However, the past empirical findings have been quite ambiguous in this respect.

Most academic interest and professional investment advice is so far directed to two-step plug-in approaches to find first useful (macroeconomic) predictive variables such as the dividend-price ratio and information contained in the term structure of interest rates, to predict excess stock returns with linear predictive regressions before constructing portfolio weights. In contrast to these premises, the utility boosting approach developed in this study establishes a direct relationship between the predictive variables and portfolio weights. This linkage is econometrically feasible when combined with a specifically designed customized gradient boosting algorithm to this objective instead of existing machine learning algorithms, resulting in superior and less noisy portfolio performance than obtained with the conventional two-step statistical approach relying on attempts to predict weakly predictable stock returns. Albeit only experience on even more extensive empirical examinations will determine the extent to which our utility boosting approach actually improve investment decisions, we believe it offers a valuable prospect for dealing with a number of complicated aspects in asset allocation decisions in an integrated and single-step manner.

Brandt (1999), Aït-Sahalia and Brandt (2001), Brandt and Santa-Clara (2006), and Brandt et al. (2009) have already emphasized the advantages of focusing portfolio weights directly. Our findings, obtained with a very different econometric method and emphasizing prediction (forecasting) aspects, point out the advantages of this type of general thinking. The customized gradient boosting algorithm learns, likewise investors in their practical decisions, from portfolio weighting mistakes in the training stage before selecting the portfolio weight for the next period. This importantly bypasses the intermediate construction of expected stock returns, which is the fundamental part of the two-step approach, involving hence greater risk for misspecification from the final asset allocation objective due to likely difficulties related to statistical determination of expected returns at most weakly predictable monthly excess stock returns of interest in this study.

The promising utility boosting approach also internally respects the pre-determined lower and upper bounds of the portfolio weights. That is continuum of the past out-of-sample return predictability studies pointing out the usefulness of theoretically motivated subsequent restrictions on the linear predictive regressions (see Campbell and Thompson, 2008; Rapach et al., 2010; Pettenuzzo et al., 2014, and the survey of Rapach and Zhou 2013). Imposing such post restrictions turn out to modify the equity premium and portfolio weight predictions substantially: See Figures 1–3 on both in- and out-of-sample evidence in this respect. This all implies that, even successful from the prediction perspective, the conventional statistical interpretations of predictive regressions are lost, while the utility boosting produces interpretable results on the importance of different predictors and their combinations. It is also noteworthy that our out-of-sample forecasting evaluation sample, starting from the beginning of 1990s, is often found very challenging from the prediction perspective (see, e.g., Campbell and Thompson, 2008; Lettau and Van Nieuwerburgh, 2008). This makes it particularly notable that the utility boosting produces substantial additional value over the past methods and benchmarks.

Our empirical results generally suggest that there is some but nowhere near one-to-one connection between conventional statistical criteria, such as t-values or adjusted-R^2, and the realized utility levels and economic goodness-of-fit criteria. This is in line with the arguments of Elliott and Timmermann (2016, chapter 2, and the references therein) and their call for a closer look to strengthen the linkage of the appropriate objective function and the employed econometric method. Our general empirical findings coincide with the conclusions of Leitch and Tanner (1991), Kandel and Stambaugh (1996), Xu (2004) and Cenesizoglu and Timmermann (2012), among others, that some poor return prediction models, producing even worse statistical (out-of-sample) forecasts than the simple benchmarks, may add economic value when used to guide portfolio decisions. In accordance with this view, our findings generally emphasize that statistical mean predictability of equity premium over time is not necessary that some advanced econometric methods can be useful for investors' decision making. Similar argumentation is made, for example, when predicting the direction (sign) of stock returns (see, e.g., Pesaran and Timmermann, 1995; Christoffersen and Diebold, 2006) or in the recent findings of only episodic ('local') and time-varying return predictability ('pockets of predictability') in stock return (see Farmer, Schmidt and Timmermann, 2018; Demetrescu et al., 2020).

We obtain several robust empirical conclusions in different empirical settings specified by the investor risk aversion levels, portfolio weight upper bounds and volatility proxies. First, as in Neely et al. (2014), technical indicators perform the best in out-of-sample forecasting and provide more stable performance in different settings than macroeconomic variables. Second, utility boosting mitigates excessive volatility in portfolio weights and resulting portfolio performance, leading to generally superior return-risk compromise over the past two-step statistical approach. Third, utility boosting backs certain macroeconomic variables in their relative performance versus the evidence in the predictive regression models. In particular, theoretically motivated inflation and the dividend-price ratio (cf. Welch and Goyal (2008) and the subsequent studies, surveyed by Rapach and Zhou (2013)), benefit substantially from the direct portfolio weight determination. Notably in the latter case, we find that specifically the changes, rather than the level, in the dividend-price ratio provides important information for portfolio weights.

As emphasized by, e.g., Ban, El Karoui and Lim (2018), and the references therein, several modern academic portfolio optimization models, with large cross-sections of assets, are intractable when applied to real data due to difficulties in estimation although they present solid theoretical properties. Ban et al. (2018) address this by adapting regularization and cross-validation approaches for portfolio optimization. We don't concentrate on large cross-sections of assets in this study: We are instead integrating ongoing advancement in machine learning and financial economics practice to the customized gradient boosting approach that can be extended to handle multiple risky assets, after some modifications and requiring a separate attempt. This advancement is also partly dependent on the development of boosting-based methods to multiple-equation cases which are still largely non-existent in the machine learning literature.

Finally, to enable the utility boosting method, we need to assume some form for the underlying utility function. We rely on the quadratic form commonly used in forecast evaluation of relevant past (out-of-sample) return predictability studies. We are well aware of its limitations and we do not claim that the conditions behind the mean-variance analysis are strictly satisfied in practice. In the future research, not just to explore alternative utility schemes, another important connected point is to consider whether it is optimal to rebalance the portfolio at all in certain time points. This view is linked with the impact of transaction costs, larger cross-sections of assets and longer than one period investment horizon. These all extensions require additional steps to be taken with the utility boosting approach opened in this study.

5 Conclusions

In contrast to commonly used linear predictive regressions, we introduce a flexible utility-based empirical approach to directly determine asset allocation decisions in a simple setting between risky and risk-free assets. From our standpoint and diverging substantially from the usual financial economics perspective, whether stock returns are statistically predictable is not of main interest. Instead, we focus directly on the portfolio weights and their dependence on the predictive variables by maximizing a sample analogy of the utility function characterizing investors' preferences in their asset allocation decisions.

Our utility boosting approach arises from the synthesis between practices in financial economics and the recent advancements in machine learning. It builds upon a customized gradient boosting introduced in this study, selecting and combining predictive variables to form optimal portfolio weights designed specifically to that objective instead of using general textbook machine learning algorithms. Methodologically our approach contains built-in mechanisms circumventing overfitting, keeping the portfolio weights inside pre-specified bounds and not basing the method on the usual statistical significance testing framework to determine the usefulness of certain predictive variables in asset allocation decisions.

When applied to monthly U.S. excess stock returns, the utility boosting method generates superior and economically meaningful utility gains over the typical and commonly used benchmarks. These gains apply both full sample and out-of-sample predictions, providing systematically higher risk-adjusted portfolio returns than the past two-step statistical approach based on linear predictive regressions. We find that especially various technical indicators and some specific macroeconomic variables perform better in terms of portfolio performance than the benchmarks based on historical average stock returns when the objective is in the economic gains of asset allocations decisions.

Appendix A: Comparison to Brandt and Santa-Clara (2006)

In this Appendix, we briefly state the main points of Brandt and Santa-Clara (2006) relevant to the comparison with our method and analysis. Their study is seemingly the closest one to our approach in terms of the general idea of direct portfolio weight determination. However, their approach is still in various ways very different and aiming to answer different objectives than ours. First of all, their empirical context is directed to potentially large cross-sections of assets. That is at the centre of attention also in Brandt (1999), Aït-Sahalia and Brandt (2001) and Brandt et al. (2009), whereas here we are introducing a new empirical approach in a simple setting between one risky asset and the risk-free rate with mainly forecasting purposes in mind.

Methodologically, in contrast to our highly flexible, nonlinear and advanced machine learning-motivated approach, Brandt and Santa-Clara (2006) assume that the optimal portfolio weights are linear in parameters:

$$w_t = \boldsymbol{\theta} \boldsymbol{x}_{t-1}, \tag{A.1}$$

where $\boldsymbol{\theta}$ is the (row) vector of parameter coefficients. Likewise in (2), they consider the decision making problem of an investor who maximizes the conditional expected value of a quadratic utility function over the next period's wealth. When writing their objective function in terms of our notation, we get

$$\max_{\boldsymbol{\theta}} \Big\{ r_{f,t} + \boldsymbol{\theta} \boldsymbol{x}_{t-1} r_{e,t} - \frac{\gamma}{2} (\boldsymbol{\theta} \boldsymbol{x}_{t-1})^2 r_{e,t}^2 \Big\}. \tag{A.2}$$

In our single risky asset setting (market portfolio and risk-free rate), the solution of the above optimization problem leads to the estimate

$$\widehat{\boldsymbol{\theta}} = \frac{1}{\gamma} \Big(\sum_{t=1}^{T} (\boldsymbol{x}_{t-1} \boldsymbol{x}_{t-1}') r_{e,t}^2 \Big)^{-1} \sum_{t=1}^{T} \boldsymbol{x}_{t-1} r_{e,t}. \tag{A.3}$$

From this solution, the (empirical) weight invested in risky asset is obtained by adding the corresponding products of elements of $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{x}_{t-1}$. This solution depends only on the data and does not require any assumptions about the distribution of stock returns, apart from stationarity (as in our utility boosting approach) and that returns are assumed non-i.i.d. (identically and independently distributed).

The final solution (A.3) relies on the selected (simple) conditional variance specification in (A.2) (cf. equation (2)). This does not allow a straightforward use of rolling window (moving-average) based volatility proxy σ_t^2 (as in the utility boosting) which importantly acknowledges the well-documented volatility clustering in asset prices and found an important ingredient in the past two-step statistical approaches (see Section 2.1). Together with this point, another key point not allowing for a completely direct comparison to our empirical perspective, framed by the recent large branch of return predictability studies, is the fact that the solution (A.3) does not automatically respect the pre-determined bounds (8), which are guaranteed and an important internal ingredient in the utility boosting approach.

Table A.1 presents the (in-sample) results when combining the approach of Brandt and Santa-Clara (2006) and weight restrictions (8) for comparison reasons to our approach. The results show how utility boosting outperforms the approach of Brandt and Santa-Clara (2006) in terms of both the average utilities and Sharpe ratios (cf. Table 3).

Table A.1: In-sample results combining (A.1)–(A.3) and weight restrictions (8).

Variable	Util(%)	Sharpe
Panel A: Macroeconomic variables		
DP	0.596	0.152
DY	0.599	0.154
EP	0.576	0.149
DE	0.570	0.145
RVOL	0.592	0.149
BM	0.567	0.145
NTIS	0.564	0.145
TBL	0.617	0.167
LTY	0.596	0.156
LTR	0.621	0.161
TMS	0.605	0.163
DFY	0.562	0.146
DFR	0.580	0.156
INFL	0.577	0.149
Panel B: Technical indicators		
MA(1,9)	0.610	0.179
MA(1,12)	0.638	0.188
MA(2,9)	0.621	0.181
MA(2,12)	0.661	0.196
MA(3,9)	0.624	0.179
MA(3,12)	0.582	0.166
MOM(9)	0.586	0.168
MOM(12)	0.589	0.167
VOL(1,9)	0.593	0.170
VOL(1,12)	0.621	0.180
VOL(2,9)	0.597	0.171
VOL(2,12)	0.610	0.177
VOL(3,9)	0.577	0.164
VOL(3,12)	0.638	0.183

Notes: See also the notes to Table 3.

Overall, the results in Table A.1 are really close to the ones obtained with the linear two-stage approach. Utilities are slightly higher in the two-stage approach, whereas determining the portfolio weights directly as in Brandt and Santa-Clara (2006) leads to somewhat higher Sharpe ratios. The estimated weights in both of these approaches often lie outside the pre-determined bounds and need to be truncated for the usual interpretations and bounds as in equation (8) and discussion around it.

Appendix B: Tuning customized gradient boosting

As was shown in the proposed algorithm in Section 2.3, gradient boosting builds the final model in a forward stagewise manner by adding new base learner functions to the ensemble that best fit the negative gradient of the loss function. The final additive boosting ensemble can be written as a sum of the base learner functions

$$F_M(\boldsymbol{x}_{t-1}) = \sum_{m=1}^{M} \upsilon h_m(\boldsymbol{x}_{t-1}).$$

This equation also summarizes the tuning parameters related to gradient boosting, which are the base learner function $h_m(\boldsymbol{x}_{t-1})$ and the amount of iterations M. In order to make the optimization in functional space feasible, the base learners are assumed to belong to certain parameterized class of functions. Regression trees and smoothing splines are two common choices. Simple linear functions is another alternative, leading to the linear final model.

Regression trees split the predictor space into J disjoint rectangles and attach a simple constant as the functional estimate in each of these rectangles. This is the method that we incorporate to the practical algorithm in Section 2.3. Mathematically these J-terminal node regression trees can be written as

$$h_m(\boldsymbol{x}_{t-1}; \{c_{jm}, R_{jm}\}_{j=1}^{J}) = \sum_{j=1}^{J} c_{jm} I(\boldsymbol{x}_{t-1} \in R_{jm}),$$

where $c_{jm} \in \mathbb{R}$ is the functional estimate in region R_{jm} at the mth iteration. The complexity of the regression tree base learner can be controlled by the amount of terminal nodes J, and the amount of observations required at each terminal node. The simplest regression tree with two terminal nodes is sufficient in our predictor-specific analysis. Each of the terminal nodes must contain at least 10 observations, a value commonly used in previous literature.

The optimal amount of iterations M could be determined by setting aside part of the dataset (i.e. excluding this part from fitting the model) and then this separate test set is used to evaluate the generalization ability of the model. To this end K-fold cross-validation is a commonly used method where the dataset is randomly split into K non-overlapping folds. Each of the K folds (concretely $K = 5$ in this study) is used as a test set once while the model is trained using the remaining $K - 1$ folds. Using the utility function presented in (17), the validation utility is the average utility produced by the K independent folds

$$CV = \frac{1}{K} \sum_{k=1}^{K} \bar{u}_k^{(-k)},$$

where $\bar{u}_k$ is the average utility when the data points of fold k are used as an independent test set and not in fitting the model. The cross-validation estimate for the amount of iterations M is the one producing the maximum validation utility.

Friedman (2002) introduces an additional randomization step to the algorithm in Section 2.3 to further enhance the generalization ability of the algorithm. At each of the m repeats a random subsample is drawn without replacement from the entire dataset. The pseudo-residuals (see Section 2.3) and the new base learner is then constructed using this random sample instead of the entire dataset. A subsampling rate of one half is a commonly used alternative.

References

Adler, T., and M. Kritzman (2007). Mean-variance versus full-scale optimisation: In and out of sample. *Journal of Asset Management* 7, 302–311.

Aït-Sahalia, Y., and M.W. Brandt (2001). Variable selection for portfolio choice. *Journal of Finance* 56(4), 1297–1351.

Audrino, F., and F. Trojani (2007). Accurate short-term yield curve forecasting using functional gradient descent. *Journal of Financial Econometrics* 5, 591–623.

Ban, G-Y., El Karoui, N., and A.E.B. Lim (2018). Machine learning and portfolio optimization. *Management Science* 64(3), 1136–1154.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance* 55, 225–264.

Brandt, M.W. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *Journal of Finance* 54(5), 1609–1645.

Brandt, M.W. (2010). Portfolio Choice Problems. Handbook of Financial Econometrics, Yacine Aït-Sahalia and Lars Hansen (eds), Vol. 1, Chapter 5.

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

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

Bühlmann, P., and T. Hothorn (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science* 22, 477–505.

Campbell J.Y., and S.B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21, 1509–1531.

Carmona, P., Climent, F., and A. Momparler (2019). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance* 61, 304–323.

Cenesizoglu T., and A. Timmermann (2012). Do return prediction models add economic value? *Journal of Banking & Finance* 36, 2974–2987.

Christoffersen, P.F., and F.X. Diebold (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52, 1273–1287.

Demetrescu, M., Georgiev, I., Rodrigues, P.M.M., and A.M.R. Taylor (2020). Testing for episodic predictability in stock returns. *Journal of Econometrics*, in press.

Elliott, G., and A. Timmermann (2016). *Economic Forecasting*. Princeton University Press, Princeton and Oxford.

Farmer, L., Schmidt, L., and A. Timmermann (2018). Pockets of predictability. CEPR Discussion Paper 12885.

Fleming, J., Kirby, C., and B. Ostdiek (2001). The economic value of volatility timing. *Journal of Finance* 56, 329–352.

Friedman, J.H., Hastie, T., and R. Tibshirani (2000). Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *Annals of Statistics* 28(2), 337–407.

Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29(5), 1189–1232.

Friedman, J.H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38(4), 367–378.

Goyal, A., and I. Welch (2003). Predicting the equity premium with dividend ratios. *Management Science* 49(5), 639–654.

Guidolin, M. (2011). Markov switching in portfolio choice and asset pricing models: A survey. *Advances in Econometrics* 27B, 87–178.

Guidolin, M., and A. Timmermann (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control* 31, 3503–3544.

Inoue, A., and L. Kilian (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews* 23, 371–402.

Kan, R., and G. Zhou (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42, 621–656.

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

Leitch, G., and J. Tanner (1991). Economic forecast evaluation: Profits versus the conventional error measures. *American Economic Review* 81, 580–590.

Lettau, M., and S. Van Nieuwerburgh (2008). Reconciling the return predictability evidence. *Review of Financial Studies* 21(4), 1607–1652.

Loh, W-Y., and Y-S. Shih (1997). Split selection methods for classification trees. *Statistica Sinica* 7(4), 815–840.

Marquering, W., and M. Verbeek (2004). The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis* 39, 407–429.

Mele, A. (2007). Asymmetric stock market volatility and the cyclical behavior of expected returns. *Journal of Financial Economics* 86, 446–478.

Mittnik, S., Robinzonov, N., and M. Spindler (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of Banking & Finance* 58, 1–14.

Neely, C.J., Rapach, D.E., Tu, J., and G. Zhou (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science* 60, 1772–1791.

Pesaran, H., and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance* 50, 1201–1228.

Pettenuzzo D., Timmermann, A., and R. Valkanov (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics* 114, 517–553.

Rapach D., and G. Zhou (2013). Forecasting stock returns. In Handbook of Economic Forecasting, G. Elliott and A. Timmermann (eds), vol. 2A, North-Holland.

Rapach, D.E., Strauss, J.K., and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies* 23, 821–862.

Rossi, A. (2018). Predicting stock market returns with machine learning. Manuscript, University of Maryland (August 2018).

Rossi, A.G., and A. Timmermann (2015). Modeling covariance risk in Merton's ICAPM. *Review of Financial Studies* 28, 1429–1461.

Sentana, E. (2005). Least squares predictions and mean-variance analysis. *Journal of Financial Econometrics* 5, 56–78.

Stambaugh, R.F. (1999). Predictive regressions. *Journal of Financial Economics* 54, 375–421.

Welch, I., and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21, 1455–1508.

Xu, Y. (2004). Small levels of predictability and large economic gains. *Journal of Empirical Finance* 11, 247–275.

Zhu, X. (2015). Tug-of-war: Time-varying predictability of stock returns and dividend growth. *Review of Finance* 19, 2317–2358.

Tables and Figures

Table 1: Predictive variables.

	Panel A: Macroeconomic variables
DP	Log dividend-price ratio $\log(D/P)$
DY	Log dividend yield $\log(D/Y)$, where Y is the lagged P
EP	Log earnings-price ratio $\log(E/P)$
DE	Log dividend-payout ratio $\log(D/E)$
RVOL	Equity risk premium volatility, 12-month moving standard deviation (Mele, 2007)
BM	Book-to-market value ratio for the DJIA (Dow Jones Industrial Average)
NTIS	Net equity expansion
TBL	Treasury bill rate (three-month Treasury bill, secondary market)
LTY	Long-term government bond yield
LTR	Return on long-term government bonds
TMS	Term spread: LTY-TBL
DFY	Default yield spread
DFR	Default return spread
INFL	Inflation (CPI inflation), lagged by one period due to the delay in CPI releases.
	Panel B: Technical indicators
MA(1,9)	Moving average indicator (14) with $s = 1$ and $l = 9$
MA(1,12)	Moving average indicator (14) with $s = 1$ and $l = 12$
MA(2,9)	Moving average indicator (14) with $s = 2$ and $l = 9$
MA(2,12)	Moving average indicator (14) with $s = 2$ and $l = 12$
MA(3,9)	Moving average indicator (14) with $s = 3$ and $l = 9$
MA(3,12)	Moving average indicator (14) with $s = 3$ and $l = 12$
MOM(9)	Momentum indicator (15) with $m = 9$
MOM(12)	Momentum indicator (15) with $m = 12$
VOL(1,9)	Volume indicator (16) with $s = 1$ and $l = 9$
VOL(1,12)	Volume indicator (16) with $s = 1$ and $l = 12$
VOL(2,9)	Volume indicator (16) with $s = 2$ and $l = 9$
VOL(2,12)	Volume indicator (16) with $s = 2$ and $l = 12$
VOL(3,9)	Volume indicator (16) with $s = 3$ and $l = 9$
VOL(3,12).	Volume indicator (16) with $s = 3$ and $l = 12$

Notes: D and E refer to log of a 12-month moving sum of dividends paid (D) and earnings (E) on the S&P 500 index (P). Net equity expansion (NTIS) is a ratio of a 12-month moving sum of net equity issues by NYSE-listed stocks to the total end-of-year market capitalization of NYSE stocks. Default yield spread (DFY) is the difference between BAA and AAA-rated corporate bond yields, whereas the default spread (DFR) is defined as the difference between corporate bond return minus LTR. Following Welch and Goyal (2008), and subsequent follow-up studies, inflation is the Consumer Price Index (All Urban Consumers) and we consider its first lag as inflation information is released only in the following month. Binary-valued technical indicators are defined in (14)–(16) with different price window lengths.

Table 2: Descriptive statistics of excess stock returns ($r_{e,t}$) and macroeconomic predictive variables.

Variable	Mean	Std	Min	Max	$\widehat{\rho}_1$	$\widehat{\rho}_2$	$\widehat{\rho}_3$
Panel A: Excess stock returns							
$r_{e,t}$	0.60	4.14	-22.11	16.14	0.04	-0.03	0.04
Panel B: Macroeconomic variables							
DP	-3.53	0.42	-4.52	-2.60	0.99	0.98	0.97
DY	-3.53	0.42	-4.53	-2.59	0.99	0.98	0.97
EP	-2.80	0.42	-4.84	-1.90	0.99	0.97	0.94
DE	-0.73	0.29	-1.24	1.38	0.99	0.95	0.90
RVOL	0.002	0.001	0.0002	0.008	0.96	0.91	0.86
BM	0.51	0.25	0.12	1.21	0.99	0.99	0.98
NTIS	0.01	0.02	-0.06	0.05	0.98	0.95	0.93
TBL	4.25	3.08	0.01	16.30	0.99	0.98	0.96
LTY	5.95	2.76	1.75	14.82	0.99	0.99	0.98
LTR	0.52	2.75	-11.24	15.23	0.04	-0.07	-0.02
TMS	1.70	1.38	-3.65	4.55	0.96	0.91	0.86
DFY	0.96	0.44	0.32	3.38	0.97	0.92	0.88
DFR	0.02	1.40	-9.75	7.37	-0.08	-0.09	-0.02
INFL	0.29	0.36	-1.92	1.81	0.55	0.39	0.27

Notes: This table presents descriptive statistics for the excess stock returns (Panel A, i.e. simple equity risk premium) in percent (i.e. returns multiplied by 100) and 14 macroeconomic variables (Panel B) where LTR, DFR, and INFL (TBL, LTY, TMS, and DFY) are measured in percent (annual percent). The sample period is December 1950–December 2018. The first three sample autocorrelation coefficients of each variable are denoted by $\widehat{\rho}_1$, $\widehat{\rho}_2$ and $\widehat{\rho}_3$.

Table 3: In-sample results for different predictors with $\gamma = 5$ and $w^{\max} = 1$.

	Utility boosting			Linear, constrained weights			
Variable	Util(%)	CER(%)	Sharpe	Util(%)	Sharpe	t-val	adj-R^2
Panel A: Macroeconomic variables							
DP	0.684	1.706	0.185	0.581	0.137	1.76	0.28
DY	0.731	2.296	0.201	0.581	0.138	1.87	0.33
EP	0.656	1.573	0.181	0.569	0.135	0.77	0.04
DE	0.593	0.727	0.159	0.587	0.139	0.69	-0.02
RVOL	0.730	2.355	0.206	0.601	0.137	2.80	0.63
BM	0.718	2.101	0.202	0.568	0.132	0.58	-0.07
NTIS	0.661	1.546	0.180	0.583	0.138	0.27	-0.10
TBL	0.668	1.617	0.185	0.634	0.164	2.26	0.55
LTY	0.634	1.268	0.181	0.606	0.150	1.49	0.18
LTR	0.727	2.249	0.197	0.659	0.164	2.68	0.69
TMS	0.695	1.914	0.192	0.657	0.169	1.90	0.41
DFY	0.603	0.816	0.163	0.587	0.138	0.38	-0.08
DFR	0.715	2.455	0.208	0.602	0.152	0.96	0.08
INFL	0.687	1.945	0.192	0.578	0.139	0.35	-0.09
Panel B: Technical indicators							
MA(1,9)	0.634	1.557	0.181	0.613	0.166	1.63	0.29
MA(1,12)	0.656	1.895	0.190	0.653	0.184	1.97	0.55
MA(2,9)	0.642	1.631	0.184	0.628	0.172	1.83	0.39
MA(2,12)	0.676	2.149	0.197	0.677	0.193	2.34	0.76
MA(3,9)	0.651	1.745	0.185	0.642	0.176	1.78	0.41
MA(3,12)	0.613	1.217	0.170	0.594	0.153	1.02	0.08
MOM(9)	0.618	1.294	0.173	0.598	0.156	1.11	0.11
MOM(12)	0.618	1.273	0.171	0.596	0.156	1.10	0.12
VOL(1,9)	0.624	1.358	0.175	0.593	0.154	1.30	0.16
VOL(1,12)	0.640	1.593	0.181	0.623	0.170	1.73	0.39
VOL(2,9)	0.613	1.236	0.171	0.598	0.157	1.31	0.18
VOL(2,12)	0.619	1.331	0.174	0.610	0.164	1.49	0.29
VOL(3,9)	0.606	1.112	0.167	0.585	0.149	0.95	0.04
VOL(3,12)	0.651	1.692	0.183	0.643	0.175	1.93	0.52
Panel C: Benchmarks							
Const	0.570	0.264	0.141				
HA	0.570		0.135				

Notes: This table reports portfolio performance measures for an investor, with the utility function (2) and the relative risk-aversion coefficient of five ($\gamma = 5$), who allocates portfolio value between the risky market portfolio and risk-free rate in each month. The bounds for portfolio weights are $w^{\min} = 0$ and $w^{\max} = 1$. The reported predictions are from the customized gradient boosting (Section 2.3) with different single macroeconomic (Panel A) or technical indicators (Panel B) in $\boldsymbol{x}_{t-1}$ at a time and two benchmarks ('Const' refers to the case, where $\boldsymbol{x}_{t-1} = 1$ and 'HA' to the historical average weights (18)). On the right, we report the results of the linear predictive regressions ('Linear') with the subsequent weight constraints (8). 'Util(%)' refers to the realized utility (17), CER(%) is the CER gain (19). Finally, adj-R^2 is the adjusted-R^2 and 't-val' is the heteroskedasticity-autocorrelation consistent (robust) t-statistic for the null hypothesis $\beta_i = 0$ in the predictive regression (9).

Table 4: In-sample results for different predictors with $\gamma = 3$ and $w^{\max} = 1$.

	Utility boosting			Linear, constrained weights			
Variable	Util(%)	CER(%)	Sharpe	Util(%)	Sharpe	t-val	adj-R^2
Panel A: Macroeconomic variables							
DP	0.814	1.526	0.187	0.711	0.147	1.76	0.28
DY	0.858	2.084	0.202	0.705	0.146	1.87	0.33
EP	0.739	0.760	0.173	0.698	0.145	0.77	0.04
DE	0.701	0.209	0.159	0.687	0.139	0.69	-0.02
RVOL	0.829	1.801	0.201	0.698	0.139	2.80	0.63
BM	0.823	1.628	0.183	0.702	0.143	0.58	-0.07
NTIS	0.779	1.209	0.178	0.702	0.143	0.27	-0.10
TBL	0.897	2.694	0.212	0.728	0.160	2.26	0.55
LTY	0.789	1.340	0.192	0.698	0.148	1.49	0.18
LTR	0.870	2.290	0.205	0.771	0.167	2.68	0.69
TMS	0.921	2.958	0.223	0.777	0.171	1.90	0.41
DFY	0.706	0.255	0.163	0.705	0.144	0.38	-0.08
DFR	0.879	2.654	0.221	0.721	0.155	0.96	0.08
INFL	0.881	2.577	0.210	0.693	0.142	0.35	-0.09
Panel B: Technical indicators							
MA(1,9)	0.739	1.061	0.181	0.700	0.161	1.63	0.29
MA(1,12)	0.766	1.405	0.190	0.751	0.185	1.97	0.55
MA(2,9)	0.747	1.110	0.183	0.717	0.169	1.83	0.39
MA(2,12)	0.786	1.650	0.197	0.780	0.195	2.34	0.76
MA(3,9)	0.756	1.211	0.184	0.722	0.169	1.78	0.41
MA(3,12)	0.716	0.680	0.169	0.685	0.147	1.02	0.08
MOM(9)	0.721	0.795	0.172	0.683	0.148	1.11	0.11
MOM(12)	0.723	0.778	0.171	0.687	0.149	1.10	0.12
VOL(1,9)	0.724	0.788	0.174	0.679	0.148	1.30	0.16
VOL(1,12)	0.746	1.106	0.181	0.719	0.169	1.73	0.39
VOL(2,9)	0.712	0.590	0.168	0.699	0.154	1.31	0.18
VOL(2,12)	0.729	0.892	0.175	0.717	0.165	1.49	0.29
VOL(3,9)	0.707	0.581	0.167	0.684	0.145	0.95	0.04
VOL(3,12)	0.759	1.230	0.184	0.745	0.178	1.93	0.52
Panel C: Benchmarks							
Const	0.698	0.182	0.144				
HA	0.695		0.141				

Notes: See the notes to Table 3.

Table 5: In-sample results with different predictor groups.

	Utility boosting			Linear (PCA), constrained weights		
Variables	Util(%)	CER(%)	Sharpe	Util(%)	Sharpe	adj-R^2
Panel A: Selections $\gamma = 5$ and $w^{\max} = 1$						
MACRO	0.975	5.480	0.303	0.648	0.166	0.54
TECH	0.680	2.144	0.199	0.640	0.178	0.45
ALL	0.978	5.541	0.306	0.671	0.184	0.92
Panel B: Selections $\gamma = 3$ and $w^{\max} = 1$						
MACRO	1.066	4.882	0.293	0.742	0.160	0.54
TECH	0.835	2.215	0.211	0.768	0.187	0.45
ALL	1.070	4.895	0.293	0.782	0.187	0.92

Notes: In the utility boosting, all the macroeconomic variables (MACRO) and technical indicators (TECH) or both of them (ALL) are simultaneously included in $\boldsymbol{x}_{t-1}$ and the customized algorithm includes only the best of them in the final predictive models (see also Table 6). In the linear predictive regressions, the principal components of the predictor groups are first constructed (e.g., due to the perfect multicollinearity between the variables), combined with weight restrictions (8). The amount of principal components is chosen according to the adjusted R^2 as in Neely et al. (2014).

Table 6: Top-10 in-sample predictors in the multivariate utility boosting models with $\gamma = 5$ and $w^{\max} = 1$.

MACRO		TECH		ALL	
Variable	Rel.inf(%)	Variable	Rel.inf(%)	Variable	Rel.inf(%)
LTR	0.158	MA(2,12)	0.432	LTR	0.159
TMS	0.148	VOL(1,12)	0.120	TMS	0.134
NTIS	0.130	VOL(3,12)	0.108	NTIS	0.111
DFR	0.101	VOL(1,9)	0.051	DFR	0.093
RVOL	0.090	MA(2,9)	0.050	RVOL	0.089
EP	0.074	MA(3,12)	0.042	EP	0.068
DY	0.060	MA(1,9)	0.041	DY	0.065
BM	0.051	MA(3,9)	0.036	BM	0.054
TBL	0.046	VOL(3,9)	0.035	TBL	0.048
DE	0.034	VOL(2,12)	0.026	DE	0.031

Notes: For more information on the relative influence measure, see Friedman (2001).

Table 7: Out-of-sample forecasting results with $\gamma = 5$ and $w^{\max} = 1$.

	Utility boosting			Linear, constrained weights		
Variable	Util(%)	CER(%)	Sharpe	Util(%)	CER(%)	Sharpe
Panel A: Macroeconomic variables						
DP	0.427	0.058	0.133	0.245	-2.325	0.032
DY	0.420	-0.017	0.131	0.259	-2.032	0.041
EP	0.412	0.358	0.150	0.499	0.948	0.167
DE	0.371	-0.236	0.124	0.299	-1.600	0.085
RVOL	0.444	0.302	0.141	0.382	-1.123	0.104
BM	0.376	-0.644	0.110	0.412	-0.562	0.115
NTIS	0.451	0.656	0.152	0.372	-0.751	0.114
TBL	0.423	0.082	0.135	0.472	0.746	0.155
LTY	0.433	0.507	0.148	0.463	0.298	0.142
LTR	0.421	0.019	0.133	0.413	-0.453	0.122
TMS	0.420	0.193	0.138	0.430	0.268	0.144
DFY	0.405	-0.191	0.125	0.417	-0.399	0.122
DFR	0.472	1.095	0.172	0.462	0.553	0.148
INFL	0.450	0.445	0.145	0.391	-0.647	0.117
Panel B: Technical indicators						
MA(1,9)	0.511	1.805	0.193	0.474	1.056	0.164
MA(1,12)	0.496	1.769	0.187	0.528	2.065	0.197
MA(2,9)	0.494	1.745	0.189	0.478	1.349	0.174
MA(2,12)	0.542	2.437	0.209	0.565	2.580	0.214
MA(3,9)	0.470	1.319	0.171	0.446	0.869	0.158
MA(3,12)	0.470	1.184	0.172	0.453	0.765	0.155
MOM(9)	0.484	1.402	0.178	0.487	1.211	0.169
MOM(12)	0.485	1.392	0.178	0.486	1.130	0.166
VOL(1,9)	0.492	1.513	0.181	0.442	0.654	0.152
VOL(1,12)	0.527	2.004	0.195	0.484	1.372	0.175
VOL(2,9)	0.461	1.137	0.169	0.446	0.759	0.155
VOL(2,12)	0.495	1.507	0.181	0.504	1.417	0.175
VOL(3,9)	0.474	1.177	0.171	0.448	0.500	0.147
VOL(3,12)	0.535	2.056	0.198	0.547	2.043	0.195
Panel C: Benchmarks						
Const	0.435	0.222	0.139			
HA	0.445		0.133			

Notes: See the notes to Table 3.

Table 8: Out-of-sample forecasting results with $\gamma = 3$ and $w^{\max} = 1$.

	Utility boosting			Linear, constrained weights		
Variable	Util(%)	CER(%)	Sharpe	Util(%)	CER(%)	Sharpe
Panel A: Macroeconomic variables						
DP	0.525	-0.408	0.137	0.296	-3.324	0.051
DY	0.478	-0.973	0.122	0.289	-3.389	0.048
EP	0.526	-0.118	0.155	0.589	0.514	0.165
DE	0.426	-1.394	0.112	0.438	-1.423	0.109
RVOL	0.509	-0.641	0.129	0.523	-0.595	0.127
BM	0.507	-0.681	0.134	0.508	-0.704	0.127
NTIS	0.545	0.073	0.150	0.511	-0.428	0.131
TBL	0.561	0.222	0.147	0.590	0.689	0.153
LTY	0.519	-0.326	0.140	0.581	0.484	0.150
LTR	0.533	-0.157	0.140	0.505	-0.603	0.128
TMS	0.528	-0.108	0.141	0.608	0.869	0.158
DFY	0.456	-1.275	0.114	0.541	-0.122	0.138
DFR	0.594	0.847	0.179	0.562	0.441	0.154
INFL	0.597	0.640	0.163	0.527	-0.253	0.135
Panel B: Technical indicators						
MA(1,9)	0.609	1.190	0.186	0.557	0.548	0.159
MA(1,12)	0.609	1.352	0.192	0.621	1.649	0.196
MA(2,9)	0.595	1.132	0.184	0.551	0.625	0.163
MA(2,12)	0.654	1.959	0.212	0.660	2.151	0.212
MA(3,9)	0.575	0.809	0.174	0.508	0.100	0.149
MA(3,12)	0.555	0.439	0.163	0.522	0.094	0.147
MOM(9)	0.596	1.023	0.180	0.544	0.425	0.157
MOM(12)	0.601	1.039	0.178	0.559	0.516	0.158
VOL(1,9)	0.583	0.887	0.179	0.523	0.130	0.149
VOL(1,12)	0.616	1.349	0.191	0.566	0.839	0.171
VOL(2,9)	0.565	0.653	0.171	0.540	0.374	0.156
VOL(2,12)	0.603	1.093	0.179	0.594	1.011	0.171
VOL(3,9)	0.543	0.255	0.156	0.528	0.020	0.144
VOL(3,12)	0.650	1.663	0.200	0.638	1.650	0.192
Panel C: Benchmarks						
Const	0.563	0.320	0.146			
HA	0.551		0.141			

Notes: See the notes to Table 3.

Table 9: Out-of-sample forecasting results for predictor groups with $\gamma = 5$ or $\gamma = 3$ and $w^{\max} = 1$.

	Utility boosting			Linear (PCA), constrained weights		
Variables	Util(%)	CER(%)	Sharpe	Util(%)	CER(%)	Sharpe
Panel A: Selections $\gamma = 5$ and $w^{\max} = 1$						
MACRO	0.480	1.075	0.164	0.269	-1.658	0.075
TECH	0.513	1.840	0.191	0.522	1.885	0.191
ALL	0.487	1.273	0.171	0.348	-0.136	0.132
Panel B: Selections $\gamma = 3$ and $w^{\max} = 1$						
MACRO	0.566	0.322	0.159	0.377	-1.932	0.097
TECH	0.628	1.561	0.196	0.631	1.717	0.195
ALL	0.571	0.441	0.164	0.441	-0.729	0.148

Notes: See the notes to Table 5.

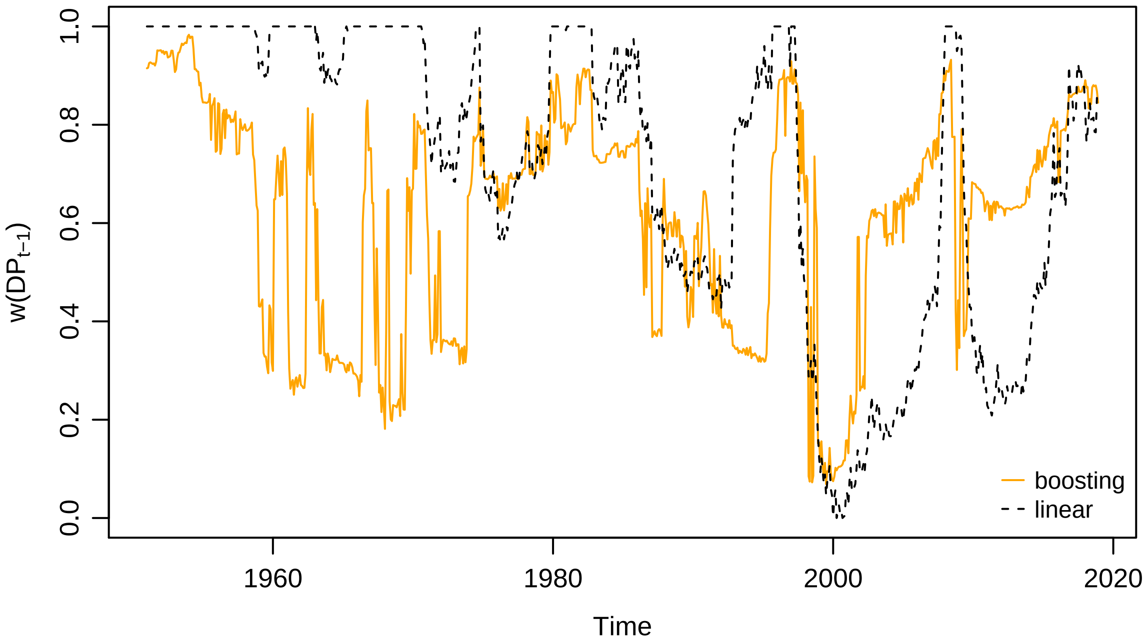

Figure 1: In-sample portfolio weights using the predictor DP (dividend-price ratio).

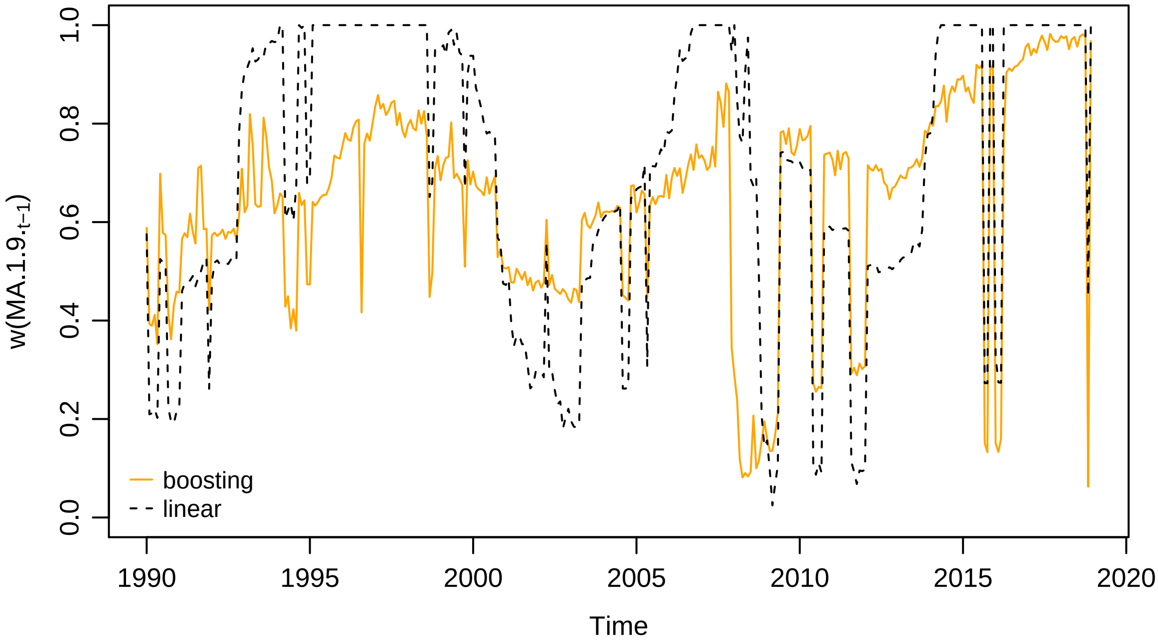

Figure 2: Out-of-sample portfolio weights using the predictor MA(1,9).

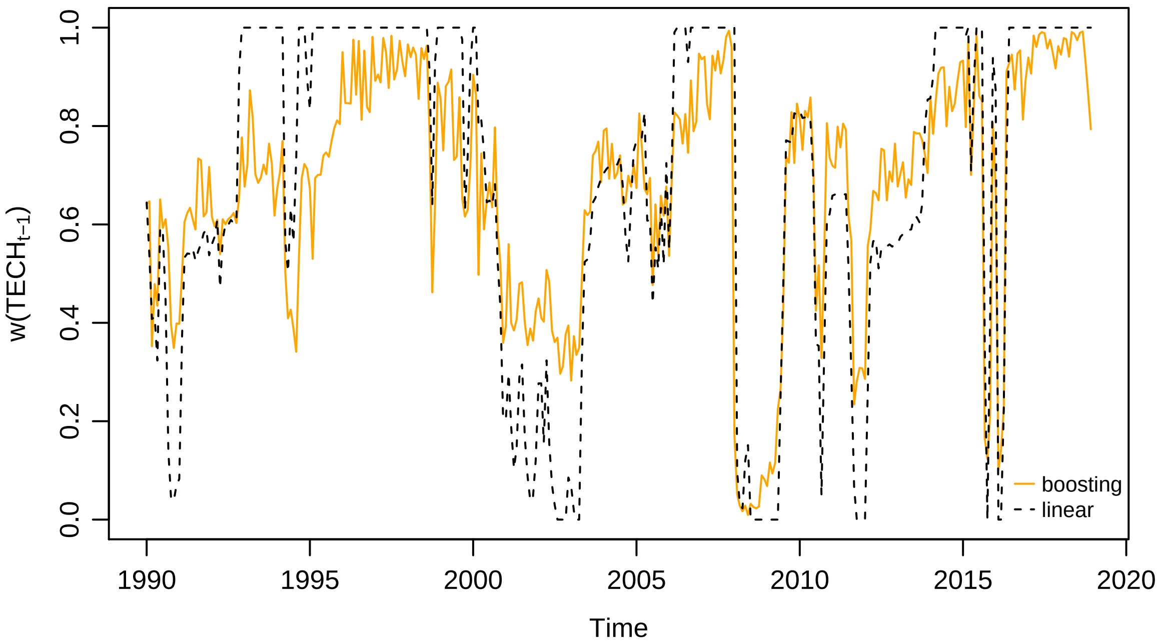

Figure 3: Out-of-sample portfolio weights using the predictor group TECH.

Internet Appendix:
Moving Forward from Predictive Regressions: Boosting Asset Allocation Decisions

January 29, 2021

This Internet Appendix presents several additional analyses and robustness checks for the main empirical results reported in Section 3 of the main paper. If not otherwise mentioned, we consider the same monthly sample period and constructed volatility proxies as in Section 3. In Section A.1, we explore the impact of allowing for leveraged positions while Section A.2 concentrates on the results when low and high transaction costs scenarios are also imposed on the evaluation stage. Section A.3 contains an alternative setting of the customized gradient boosting algorithm where the regression tree-based base learner function is replaced by a spline-based learner. Section A.4 examines the robustness of our findings on the selected volatility proxy by utilizing the GARCH model and a shorter-horizon realized variance as volatility proxies. Finally, Section A.5 considers the predictive power of monthly changes in the highly persistent macroeconomic predictor variables, such as the dividend-price ratio, instead of their levels.

A.1 Leveraged weights

In Tables A.1.1 and A.1.2, we report the results when the upper bound for the portfolio weights is set to $w^{\max} = 1.5$. This allows an investor to take leverage in his or her positions. When comparing to the results of Tables 3-4 and 7–8 in the main analysis, the essential conclusions are intact (with some minor predictor-specific differences).

Table A.1.1: In-sample results of predictor-specific models with $\gamma = 5$ and $\gamma = 3$ when $w^{\max} = 1.5$.

	$\gamma = 5$				$\gamma = 3$			
	Utility boosting		Linear, constr. weights		Utility boosting		Linear, constr. weights	
Variable	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe
Panel A: Macroeconomic variables								
DP	0.677	0.177	0.587	0.136	0.929	0.191	0.738	0.141
DY	0.831	0.223	0.595	0.141	0.999	0.208	0.734	0.140
EP	0.716	0.193	0.593	0.138	0.848	0.181	0.717	0.138
DE	0.636	0.170	0.573	0.127	0.806	0.172	0.740	0.141
RVOL	0.695	0.185	0.616	0.136	1.114	0.243	0.758	0.138
BM	0.719	0.186	0.566	0.128	0.890	0.183	0.712	0.135
NTIS	0.668	0.178	0.576	0.129	0.869	0.182	0.739	0.141
TBL	0.782	0.212	0.642	0.155	0.877	0.187	0.808	0.165
LTY	0.694	0.193	0.600	0.140	0.866	0.195	0.766	0.151
LTR	0.803	0.215	0.688	0.161	1.059	0.220	0.841	0.164
TMS	0.768	0.206	0.642	0.155	1.139	0.245	0.846	0.170
DFY	0.650	0.170	0.565	0.124	0.823	0.175	0.745	0.141
DFR	0.652	0.180	0.593	0.144	1.069	0.237	0.766	0.155
INFL	0.690	0.187	0.571	0.128	0.931	0.197	0.733	0.142
Panel B: Technical indicators								
MA(1,9)	0.643	0.182	0.633	0.164	0.810	0.181	0.770	0.164
MA(1,12)	0.667	0.190	0.672	0.179	0.845	0.190	0.837	0.184
MA(2,9)	0.651	0.183	0.649	0.171	0.824	0.184	0.797	0.172
MA(2,12)	0.691	0.198	0.703	0.190	0.876	0.197	0.874	0.194
MA(3,9)	0.662	0.185	0.664	0.174	0.836	0.185	0.816	0.175
MA(3,12)	0.616	0.171	0.605	0.150	0.777	0.170	0.741	0.152
MOM(9)	0.622	0.173	0.612	0.153	0.784	0.172	0.745	0.153
MOM(12)	0.622	0.172	0.612	0.153	0.786	0.171	0.745	0.154
VOL(1,9)	0.633	0.176	0.617	0.154	0.793	0.174	0.737	0.152
VOL(1,12)	0.648	0.181	0.643	0.167	0.820	0.181	0.788	0.169
VOL(2,9)	0.617	0.172	0.606	0.153	0.780	0.172	0.752	0.156
VOL(2,12)	0.622	0.174	0.611	0.156	0.792	0.176	0.773	0.164
VOL(3,9)	0.609	0.168	0.597	0.146	0.766	0.167	0.730	0.148
VOL(3,12)	0.659	0.182	0.654	0.169	0.838	0.183	0.824	0.176
Panel C: Benchmarks								
Const	0.564	0.144			0.715	0.142		
HA	0.564	0.127			0.714	0.137		

Notes: See the notes to Tables 3 and 4 in the body text.

Table A.1.2: Out-of-sample results of predictor-specific models with $\gamma = 5$ and $\gamma = 3$ when $w^{\max} = 1.5$.

	$\gamma = 5$				$\gamma = 3$			
	Utility boosting		Linear, constr. weights		Utility boosting		Linear, constr. weights	
Variable	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe
Panel A: Macroeconomic variables								
DP	0.403	0.125	0.245	0.032	0.574	0.137	0.255	0.031
DY	0.383	0.119	0.259	0.041	0.563	0.135	0.280	0.041
EP	0.404	0.145	0.510	0.162	0.547	0.150	0.665	0.167
DE	0.339	0.126	0.228	0.067	0.393	0.103	0.383	0.092
RVOL	0.414	0.135	0.381	0.097	0.578	0.139	0.502	0.110
BM	0.421	0.124	0.398	0.102	0.487	0.116	0.536	0.118
NTIS	0.441	0.152	0.364	0.109	0.588	0.148	0.489	0.119
TBL	0.363	0.131	0.412	0.134	0.620	0.152	0.627	0.155
LTY	0.436	0.156	0.441	0.126	0.520	0.135	0.615	0.144
LTR	0.427	0.139	0.439	0.122	0.547	0.134	0.525	0.122
TMS	0.425	0.146	0.345	0.125	0.566	0.143	0.590	0.148
DFY	0.432	0.136	0.351	0.092	0.436	0.103	0.556	0.128
DFR	0.471	0.167	0.460	0.136	0.659	0.179	0.618	0.152
INFL	0.417	0.137	0.368	0.105	0.622	0.153	0.514	0.122
Panel B: Technical indicators								
MA(1,9)	0.514	0.188	0.509	0.160	0.709	0.196	0.627	0.164
MA(1,12)	0.531	0.197	0.560	0.187	0.680	0.193	0.706	0.197
MA(2,9)	0.506	0.189	0.515	0.172	0.675	0.190	0.628	0.173
MA(2,12)	0.561	0.207	0.609	0.206	0.748	0.213	0.763	0.214
MA(3,9)	0.455	0.166	0.470	0.154	0.636	0.176	0.572	0.156
MA(3,12)	0.466	0.169	0.480	0.149	0.645	0.176	0.586	0.154
MOM(9)	0.482	0.175	0.516	0.161	0.685	0.186	0.635	0.167
MOM(12)	0.476	0.172	0.515	0.159	0.689	0.185	0.638	0.165
VOL(1,9)	0.487	0.177	0.475	0.150	0.658	0.181	0.572	0.151
VOL(1,12)	0.509	0.183	0.520	0.170	0.700	0.193	0.634	0.173
VOL(2,9)	0.449	0.164	0.460	0.147	0.632	0.175	0.582	0.155
VOL(2,12)	0.486	0.176	0.523	0.163	0.689	0.185	0.670	0.175
VOL(3,9)	0.463	0.165	0.463	0.139	0.635	0.171	0.583	0.147
VOL(3,12)	0.543	0.194	0.566	0.179	0.736	0.199	0.742	0.196
Panel C: Benchmarks								
Const	0.413	0.138			0.578	0.141		
HA	0.438	0.120			0.588	0.136		

Notes: See the notes to Tables 3 and 4 in the body text.

Both in- and out-of-sample results reflect the fact that the utility boosting approach takes into account the risk awareness in estimation. Naturally the utility boosting benefits less than the two-step 'linear' approach from the ability to take leverage as risky leveraged positions are less frequently taken when $\gamma = 5$ (in the latter the upper bound $w^{\max} = 1.5$, and even levels above that, are often reached without restrictions on the maximum portfolio weight). However, the Sharpe ratios are consistently higher in the utility boosting. With the higher tolerance to risk ($\gamma = 3$), riskier positions are allowed and hence resulting utilities are then naturally also somewhat higher.

A.2 Transaction costs involved

In the following Tables A.2.1 and A.2.2, we impose transaction costs along the lines of Neely et al. (2014) in the evaluation stage when computing the resulting portfolio returns and the values of different evaluation metrics. We consider two transaction costs scenarios (see Marquering and Verbeek, 2004; Rossi, 2018): Low transaction costs case means that 0.1% of the value traded is lost as a result of rebalancing portfolio weights whereas high transaction costs corresponds the case of 0.5%.

As discussed in Section 3.3, here we take the standpoint that asset allocation decisions are first determined at each step so that the potential impact of transaction costs is not taken into account. This corresponds the idea of continuous portfolio rebalancing. Integrating transaction costs to optimal portfolio weight determination requires several additional steps and detailed examination, which are hence left for the future research. First of all, to determine whether it is optimal to update the portfolio weight from period t to $t + 1$ at all needs to be addressed. As discussed in the main text, say that the portfolio weight for the risky asset is 0.7 at time t (i.e. $w_t = 0.7$) and then the optimal prediction (not acknowledging the impact of transaction costs) is to change it to 0.75 for the period $t + 1$. Is this change of 0.05 large enough that we should really rebalance the portfolio again or should we just continue with the weight 0.7? Seemingly, as also pointed out already by Brandt (1999), the impact of the investment horizon must also be addressed simultaneously when addressing this question, whereas throughout this study we concentrate on the one-month horizon.

The above points show that the potential impact of the transaction costs in various stages of our analysis is much more fundamental to the method than is possible to deal with in this paper yet, given all the other advancements. Even though these extensions are left for the future research, below we report the results when taking the same stand-

Table A.2.1: In-sample results with low transaction costs when $\gamma = 5$ and $\gamma = 3$ with the portfolio weight upper bound $w^{\max} = 1$.

	$\gamma = 5$				$\gamma = 3$			
	Utility boosting		Linear, constr. weights		Utility boosting		Linear, constr. weights	
Variable	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe
Panel A: Macroeconomic variables								
DP	0.680	0.183	0.579	0.136	0.808	0.186	0.710	0.147
DY	0.725	0.199	0.579	0.137	0.850	0.199	0.704	0.146
EP	0.653	0.180	0.567	0.134	0.736	0.172	0.697	0.144
DE	0.591	0.159	0.586	0.139	0.698	0.158	0.686	0.138
RVOL	0.716	0.201	0.597	0.136	0.815	0.196	0.695	0.139
BM	0.711	0.200	0.566	0.132	0.819	0.182	0.702	0.143
NTIS	0.657	0.179	0.581	0.137	0.775	0.177	0.701	0.143
TBL	0.666	0.184	0.632	0.164	0.891	0.210	0.726	0.160
LTY	0.632	0.180	0.604	0.150	0.784	0.191	0.697	0.148
LTR	0.702	0.188	0.632	0.155	0.845	0.198	0.748	0.161
TMS	0.690	0.190	0.652	0.168	0.909	0.219	0.773	0.170
DFY	0.600	0.162	0.585	0.137	0.702	0.162	0.705	0.143
DFR	0.695	0.200	0.586	0.147	0.857	0.214	0.711	0.152
INFL	0.672	0.186	0.574	0.138	0.863	0.204	0.690	0.141
Panel B: Technical indicators								
MA(1,9)	0.626	0.177	0.605	0.163	0.729	0.178	0.692	0.158
MA(1,12)	0.649	0.187	0.645	0.181	0.758	0.187	0.742	0.182
MA(2,9)	0.635	0.181	0.620	0.169	0.739	0.180	0.710	0.166
MA(2,12)	0.670	0.195	0.670	0.190	0.779	0.195	0.773	0.192
MA(3,9)	0.645	0.183	0.635	0.174	0.749	0.182	0.716	0.167
MA(3,12)	0.609	0.168	0.590	0.152	0.711	0.167	0.682	0.146
MOM(9)	0.611	0.170	0.592	0.154	0.714	0.170	0.679	0.147
MOM(12)	0.613	0.169	0.592	0.154	0.718	0.169	0.684	0.148
VOL(1,9)	0.614	0.170	0.583	0.150	0.711	0.169	0.670	0.145
VOL(1,12)	0.631	0.178	0.613	0.167	0.736	0.178	0.709	0.166
VOL(2,9)	0.606	0.168	0.591	0.155	0.706	0.166	0.694	0.152
VOL(2,12)	0.613	0.171	0.604	0.162	0.722	0.173	0.711	0.164
VOL(3,9)	0.600	0.165	0.580	0.147	0.701	0.164	0.681	0.144
VOL(3,12)	0.645	0.181	0.636	0.173	0.753	0.182	0.739	0.176
Panel C: Benchmarks								
Const	0.570	0.141			0.698	0.144		
HA	0.569	0.134			0.695	0.141		

Notes: In these results, the low transaction cost scenario (0.10% of the traded portfolio value) is taken into account when computing the values of economic goodness-of-fit measures.

Table A.2.2: In-sample results with high transaction costs when $\gamma = 5$ and $\gamma = 3$ with the portfolio weight upper bound $w^{\max} = 1$.

	$\gamma = 5$				$\gamma = 3$			
	Utility boosting		Linear, constr. weights		Utility boosting		Linear, constr. weights	
Variable	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe
Panel A: Macroeconomic variables								
DP	0.663	0.177	0.571	0.134	0.786	0.179	0.705	0.145
DY	0.701	0.190	0.571	0.135	0.818	0.189	0.699	0.144
EP	0.641	0.176	0.559	0.132	0.727	0.169	0.692	0.143
DE	0.584	0.155	0.580	0.137	0.689	0.155	0.683	0.138
RVOL	0.662	0.180	0.578	0.130	0.757	0.177	0.683	0.135
BM	0.686	0.190	0.559	0.130	0.801	0.177	0.699	0.142
NTIS	0.642	0.173	0.575	0.135	0.760	0.172	0.699	0.142
TBL	0.656	0.180	0.625	0.161	0.867	0.203	0.720	0.158
LTY	0.624	0.176	0.599	0.148	0.766	0.184	0.692	0.146
LTR	0.603	0.155	0.523	0.120	0.744	0.167	0.658	0.135
TMS	0.671	0.183	0.632	0.161	0.861	0.204	0.758	0.165
DFY	0.589	0.157	0.578	0.135	0.686	0.156	0.702	0.143
DFR	0.616	0.171	0.524	0.126	0.771	0.186	0.670	0.141
INFL	0.614	0.165	0.556	0.132	0.787	0.181	0.679	0.138
Panel B: Technical indicators								
MA(1,9)	0.592	0.163	0.571	0.149	0.690	0.163	0.662	0.148
MA(1,12)	0.621	0.175	0.614	0.168	0.729	0.176	0.710	0.170
MA(2,9)	0.608	0.170	0.591	0.158	0.710	0.170	0.682	0.157
MA(2,12)	0.646	0.185	0.642	0.179	0.754	0.186	0.744	0.182
MA(3,9)	0.619	0.173	0.609	0.164	0.721	0.172	0.690	0.158
MA(3,12)	0.590	0.161	0.574	0.146	0.692	0.161	0.670	0.143
MOM(9)	0.587	0.160	0.571	0.146	0.686	0.159	0.663	0.142
MOM(12)	0.594	0.162	0.573	0.147	0.696	0.161	0.668	0.143
VOL(1,9)	0.570	0.153	0.543	0.136	0.662	0.152	0.637	0.135
VOL(1,12)	0.593	0.163	0.572	0.151	0.695	0.163	0.669	0.152
VOL(2,9)	0.578	0.157	0.565	0.145	0.680	0.157	0.672	0.146
VOL(2,12)	0.589	0.162	0.579	0.152	0.697	0.164	0.689	0.157
VOL(3,9)	0.576	0.156	0.561	0.140	0.674	0.156	0.668	0.141
VOL(3,12)	0.621	0.172	0.611	0.164	0.728	0.173	0.714	0.168
Panel C: Benchmarks								
Const	0.567	0.140			0.697	0.144		
HA	0.564	0.133			0.693	0.141		

point as in the past two-step statistical approaches. That is, the impact of transaction costs is taken into account in the evaluation stage when evaluating the resulting portfolio returns and utility levels. This view is natural in the two-step statistical approach where the predictive regressions do not contain utility optimization, whereas transaction costs would be an important part of direct portfolio optimization problem.

The full sample results in Tables A.2.1 and A.2.2 show that when comparing to the main conclusions of Tables 3 and 4 the best predictors are still essentially the same. In other words, transaction costs naturally diminish the economic gains in different predictors quite symmetrically. The essential differences between the utility boosting and the current two-step approach are still valid. Less persistent variables (DFR and INFL) naturally suffer somewhat more on inevitably stronger impact of transaction costs (i.e. due to more active portfolio rebalancing), especially in the high cost scenario. On the other hand, as technical indicators are defined as binary variables, there is especially in full sample estimations no portfolio rebalancing in most months as the value of the predictor remains the same (i.e. due to persistent runs of either zeros and ones as defined in Section 3.1).

In Tables A.2.3 and A.2.4, we present the corresponding out-of-sample forecasting results as in Tables 7 and 8 in the main text but now incorporating also transaction costs in the evaluation stage. The obtained realized utility levels naturally drop due to transaction costs. The drop appears to be about the same magnitude in both the utility boosting and the ’linear’ two-step statistical approach. As in the in-sample results above, the smaller amount of trading activity in the case of technical indicators imply that the impact of transaction costs is substantially smaller versus the macroeconomic variables where partly quite erratic behaviour (such as the IT bubble during the millennium 2000) complicates their use as predictors.

Table A.2.3: Out of-sample results with low transaction costs when $\gamma = 5$ and $\gamma = 3$ with the portfolio weight upper bound $w^{\max} = 1$.

	$\gamma = 5$				$\gamma = 3$			
	Utility boosting		Linear, constr. weights		Utility boosting		Linear, constr. weights	
Variable	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe
Panel A: Macroeconomic variables								
DP	0.420	0.131	0.242	0.030	0.518	0.134	0.293	0.049
DY	0.413	0.128	0.256	0.039	0.470	0.119	0.285	0.046
EP	0.403	0.146	0.497	0.166	0.514	0.151	0.586	0.164
DE	0.366	0.121	0.296	0.084	0.418	0.109	0.435	0.108
RVOL	0.434	0.137	0.377	0.103	0.496	0.125	0.519	0.126
BM	0.368	0.107	0.411	0.115	0.499	0.131	0.506	0.126
NTIS	0.446	0.150	0.369	0.113	0.539	0.148	0.508	0.131
TBL	0.418	0.133	0.470	0.155	0.556	0.145	0.590	0.153
LTY	0.429	0.146	0.461	0.142	0.514	0.138	0.580	0.150
LTR	0.396	0.124	0.383	0.113	0.505	0.132	0.477	0.120
TMS	0.412	0.135	0.426	0.143	0.519	0.138	0.605	0.157
DFY	0.396	0.122	0.414	0.121	0.445	0.111	0.538	0.137
DFR	0.462	0.168	0.450	0.144	0.578	0.173	0.552	0.150
INFL	0.435	0.139	0.381	0.114	0.581	0.158	0.519	0.133
Panel B: Technical indicators								
MA(1,9)	0.506	0.191	0.469	0.162	0.602	0.184	0.553	0.157
MA(1,12)	0.488	0.184	0.522	0.194	0.601	0.189	0.614	0.193
MA(2,9)	0.487	0.186	0.472	0.171	0.588	0.181	0.545	0.161
MA(2,12)	0.536	0.207	0.559	0.212	0.647	0.210	0.654	0.210
MA(3,9)	0.462	0.169	0.439	0.155	0.567	0.172	0.501	0.147
MA(3,12)	0.465	0.169	0.449	0.153	0.549	0.161	0.519	0.146
MOM(9)	0.478	0.176	0.483	0.167	0.590	0.178	0.540	0.155
MOM(12)	0.480	0.176	0.482	0.164	0.595	0.176	0.556	0.157
VOL(1,9)	0.481	0.177	0.433	0.148	0.572	0.175	0.514	0.146
VOL(1,12)	0.516	0.191	0.473	0.170	0.604	0.187	0.555	0.167
VOL(2,9)	0.454	0.166	0.440	0.153	0.558	0.168	0.534	0.154
VOL(2,12)	0.490	0.179	0.499	0.173	0.597	0.177	0.590	0.170
VOL(3,9)	0.468	0.169	0.444	0.145	0.536	0.154	0.524	0.143
VOL(3,12)	0.529	0.195	0.541	0.193	0.643	0.197	0.633	0.190
Panel C: Benchmarks								
Const	0.435	0.139			0.563	0.146		
HA	0.443	0.133			0.551	0.141		

Notes: As in Tables A.2.1 and A.2.2, transaction costs are taken into account during the out-of-sample forecasting evaluation.

Table A.2.4: Out of-sample results with high transaction costs when $\gamma = 5$ and $\gamma = 3$ with the portfolio weight upper bound $w^{\max} = 1$.

	$\gamma = 5$				$\gamma = 3$			
	Utility boosting		Linear, constr. weights		Utility boosting		Linear, constr. weights	
Variable	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe
Panel A: Macroeconomic variables								
DP	0.394	0.121	0.233	0.025	0.491	0.126	0.281	0.044
DY	0.383	0.117	0.243	0.031	0.441	0.110	0.269	0.039
EP	0.369	0.129	0.487	0.161	0.466	0.132	0.578	0.161
DE	0.344	0.113	0.283	0.080	0.385	0.098	0.423	0.104
RVOL	0.397	0.123	0.358	0.097	0.445	0.110	0.503	0.122
BM	0.337	0.093	0.404	0.112	0.466	0.119	0.501	0.125
NTIS	0.427	0.143	0.357	0.109	0.517	0.141	0.499	0.128
TBL	0.401	0.128	0.465	0.153	0.534	0.139	0.588	0.153
LTY	0.414	0.140	0.455	0.140	0.494	0.131	0.577	0.149
LTR	0.296	0.090	0.262	0.075	0.393	0.098	0.366	0.089
TMS	0.381	0.125	0.412	0.139	0.481	0.127	0.595	0.155
DFY	0.364	0.109	0.404	0.117	0.399	0.096	0.530	0.135
DFR	0.421	0.149	0.406	0.128	0.511	0.149	0.512	0.138
INFL	0.376	0.118	0.343	0.102	0.517	0.137	0.487	0.124
Panel B: Technical indicators								
MA(1,9)	0.483	0.181	0.449	0.154	0.574	0.173	0.535	0.151
MA(1,12)	0.458	0.172	0.496	0.183	0.570	0.177	0.586	0.182
MA(2,9)	0.457	0.174	0.447	0.161	0.557	0.170	0.521	0.153
MA(2,12)	0.510	0.196	0.537	0.202	0.619	0.199	0.629	0.201
MA(3,9)	0.432	0.157	0.412	0.145	0.537	0.161	0.471	0.137
MA(3,12)	0.445	0.161	0.434	0.148	0.525	0.153	0.505	0.142
MOM(9)	0.456	0.166	0.465	0.160	0.562	0.168	0.525	0.150
MOM(12)	0.462	0.168	0.468	0.159	0.571	0.168	0.544	0.153
VOL(1,9)	0.439	0.159	0.395	0.133	0.528	0.158	0.478	0.134
VOL(1,12)	0.474	0.174	0.432	0.153	0.560	0.171	0.509	0.151
VOL(2.9)	0.426	0.154	0.413	0.142	0.527	0.157	0.508	0.145
VOL(2,12)	0.471	0.171	0.480	0.165	0.572	0.168	0.572	0.164
VOL(3,9)	0.446	0.159	0.426	0.139	0.507	0.144	0.509	0.138
VOL(3,12)	0.505	0.186	0.520	0.184	0.618	0.188	0.609	0.182
Panel C: Benchmarks								
Const	0.432	0.138			0.561	0.146		
HA	0.439	0.131			0.548	0.140		

A.3 Splines as baselearner functions

Regression trees, which are now used as the base learner functions in the customized gradient boosting in the main analysis, have several advantages. Trees are highly flexible, easily interpretable and computationally efficient. However, trees provide only one possible alternative in this respect. This section presents results when the base learner function is changed to a slightly more constrained spline-based alternative (i.e. using so called P-splines).

Table A.3.1: Spline-based in-sample predictive results with $\gamma = 5$ or $\gamma = 3$ and $w^{\max} = 1$.

	$\gamma = 5$			$\gamma = 3$		
Variable	Util(%)	CER(%)	Sharpe	Util(%)	CER(%)	Sharpe
Panel A: Macroeconomic variables						
DP	0.651	1.309	0.172	0.762	0.929	0.172
DY	0.647	1.270	0.170	0.763	0.953	0.171
EP	0.648	1.434	0.174	0.779	1.304	0.175
DE	0.588	0.556	0.150	0.706	0.211	0.148
RVOL	0.619	1.000	0.164	0.708	0.233	0.151
BM	0.566	0.341	0.145	0.741	0.705	0.160
NTIS	0.655	1.419	0.173	0.771	1.095	0.170
TBL	0.638	1.289	0.170	0.740	0.772	0.168
LTY	0.603	0.849	0.164	0.740	0.675	0.161
LTR	0.649	1.244	0.168	0.780	1.207	0.170
TMS	0.681	1.728	0.186	0.793	1.368	0.181
DFY	0.597	0.750	0.158	0.684	-0.035	0.150
DFR	0.605	0.909	0.162	0.720	0.514	0.160
INFL	0.643	1.321	0.173	0.766	1.114	0.177
Panel B: Technical indicators						
MA(1,9)	0.633	1.529	0.181	0.739	1.055	0.181
MA(1,12)	0.655	1.835	0.189	0.766	1.405	0.190
MA(2,9)	0.641	1.625	0.184	0.749	1.162	0.183
MA(2,12)	0.675	2.103	0.197	0.786	1.651	0.197
MA(3,9)	0.650	1.698	0.185	0.756	1.208	0.184
MA(3,12)	0.612	1.163	0.169	0.717	0.722	0.169
MOM(9)	0.616	1.255	0.172	0.721	0.788	0.172
MOM(12)	0.617	1.232	0.171	0.724	0.800	0.170
VOL(1,9)	0.623	1.296	0.174	0.724	0.785	0.173
VOL(1,12)	0.639	1.555	0.181	0.748	1.130	0.181
VOL(2,9)	0.612	1.183	0.171	0.719	0.754	0.171
VOL(2,12)	0.618	1.291	0.173	0.731	0.931	0.175
VOL(3,9)	0.604	1.049	0.167	0.708	0.583	0.166
VOL(3,12)	0.649	1.627	0.182	0.759	1.226	0.183

Table A.3.2: Spline-based out of-sample predictive results with $\gamma = 5$ or $\gamma = 3$ and $w^{\max} = 1$.

	$\gamma = 5$			$\gamma = 3$		
Variable	Util(%)	CER(%)	Sharpe	Util(%)	CER(%)	Sharpe
Panel A: Macroeconomic variables						
DP	0.431	0.084	0.135	0.545	-0.094	0.143
DY	0.395	-0.425	0.120	0.529	-0.311	0.137
EP	0.446	0.804	0.164	0.584	0.678	0.174
DE	0.348	-0.604	0.112	0.433	-1.352	0.111
RVOL	0.434	0.147	0.136	0.521	-0.522	0.131
BM	0.401	-0.440	0.118	0.484	-0.953	0.122
NTIS	0.489	1.152	0.165	0.615	0.917	0.167
TBL	0.392	-0.128	0.130	0.543	-0.041	0.141
LTY	0.424	0.228	0.138	0.505	-0.579	0.131
LTR	0.404	-0.237	0.127	0.535	-0.167	0.138
TMS	0.429	0.306	0.141	0.544	0.078	0.145
DFY	0.391	-0.457	0.117	0.462	-1.140	0.119
DFR	0.441	0.538	0.150	0.503	-0.517	0.132
INFL	0.460	0.525	0.147	0.576	0.320	0.153
Panel A: Technical indicators						
MA(1,9)	0.505	1.689	0.187	0.605	1.144	0.183
MA(1,12)	0.508	1.953	0.193	0.618	1.513	0.195
MA(2,9)	0.481	1.514	0.180	0.589	1.067	0.181
MA(2,12)	0.553	2.573	0.213	0.660	2.063	0.214
MA(3,9)	0.473	1.373	0.173	0.570	0.794	0.172
MA(3,12)	0.459	1.011	0.164	0.562	0.588	0.166
MOM(9)	0.475	1.244	0.172	0.581	0.843	0.174
MOM(12)	0.480	1.268	0.173	0.585	0.852	0.173
VOL(1,9)	0.478	1.328	0.174	0.590	0.979	0.181
VOL(1,12)	0.499	1.674	0.184	0.611	1.333	0.190
VOL(2,9)	0.456	1.028	0.165	0.559	0.583	0.167
VOL(2,12)	0.493	1.429	0.178	0.592	0.960	0.173
VOL(3,9)	0.470	1.055	0.166	0.556	0.401	0.160
VOL(3,12)	0.527	1.972	0.194	0.645	1.678	0.199

Table A.3.3: Spline-based in-sample predictive results for the predictor groups with $\gamma = 5$ or $\gamma = 3$ and $w^{\max} = 1$.

	$\gamma = 5$			$\gamma = 3$		
Variables	Util(%)	CER(%)	Sharpe	Util(%)	CER(%)	Sharpe
MACRO	0.840	3.819	0.252	0.959	3.504	0.253
TECH	0.686	2.227	0.200	0.822	2.040	0.207
ALL	0.851	4.016	0.261	0.993	3.975	0.270

Table A.3.4: Spline-based out of-sample predictive results for the predictor groups with $\gamma = 5$ or $\gamma = 3$ and $w^{\max} = 1$.

	$\gamma = 5$			$\gamma = 3$		
Variables	Util(%)	CER(%)	Sharpe	Util(%)	CER(%)	Sharpe
MACRO	0.501	1.139	0.166	0.582	0.411	0.158
TECH	0.511	1.907	0.194	0.588	1.106	0.183
ALL	0.541	1.747	0.187	0.613	0.901	0.173

The results in Tables A.3.1–A.3.4 are largely similar to the ones obtained with regression trees in Tables 3–4 and 7–9. This concerns single predictors and multivariate predictor groups. Here using all the predictors (ALL) even lead to somewhat superior performance than obtained with the regression tree case, further strengthening the usefulness of the utility boosting method and its internal model selection capability.

A.4 GARCH model and realized variance based volatility proxies

In accordance with various closely related return predictability studies (see e.g., Campbell and Thompson, 2008; Rapach et al., 2010; Rapach and Zhou, 2013; Neely et al., 2014), we have used a five-year rolling window (moving average) based volatility proxy computed using historical excess returns. In this section, we consider two alternative ways to extract the necessary volatility proxy σ_t^2: The Generalized Autoregressive Conditional Heteroskedastic (GARCH) model and an alternative realized variance-based approach.

In this section, we concentrate on the full sample results. Specifically here, as this is already a robustness check for the main analysis, detailed out-of-sample forecasting results, containing continuous parameter updating, depends crucially on finding the optimal GARCH model specification as well and its detailed forecasting performance evaluation. Hence, we believe relying on the maximum data availability provides the best additional value over the main results.

In the GARCH model specification for the excess market returns, the conditional variance h_t^2 is extracted with a GARCH error term, combined with the constant conditional mean in line with the evidence of Welch and Goyal (2008), among others (i.e. no mean return predictability). Formally, we set hence

$$r_{e,t} = \alpha_0 + h_t \, \nu_t,$$

where ν_t is an independent and identically distributed error term with zero mean and unit variance ($\nu_t \sim \mathrm{iid}(0,1)$). Following the large majority of past GARCH specifications, the normality assumption (i.e. $\nu_t \sim \mathrm{nid}(0,1)$) is assumed for maximum likelihood estimation purposes, and the conditional variance is assumed to follow the GARCH(1,1) process

$$h_t^2 = c + b_1 h_{t-1}^2 + a_1 u_{t-1}^2, \qquad c > 0, b_1 \geq 0, a_1 > 0,$$

where $u_t = r_{e,t} - \alpha_0$. There are, of course, various extensions available for the above specifications of the conditional mean and conditional variance, including, for example, asymmetric conditional variance equations (asymmetric GARCH models) and non-Gaussian innovations, but here we restrict ourselves to this commonly used specification. Finally, the extracted estimate of the conditional variance $\widehat{h}_t^2$ acts as our volatility proxy in the procedures described in Section 2.

Table A.4.1 presents the (full sample) results obtained with our benchmark selections $\gamma = 5$ and $w^{\max} = 1$ when the GARCH-based volatility proxy is employed. It turns out that within technical indicators the same indicators perform the best (MA(1,12), MA(2,12) and VOL(3,12)) as in the case of main volatility proxy. In the case of macroeconomic variables, some changes occur as now EP and TMS are among the best performing variables in this setting. All in all, all the same main conclusions on the superiority of the utility boosting over the traditional two-step 'linear' approach are intact (seen as almost uniformly higher Sharpe ratios and realized utility levels).

Table A.4.1: In-sample predictive results with $\gamma = 5$ and $w^{\max} = 1$ when using the extracted conditional variance from the GARCH(1,1) model as the underlying volatility proxy σ_t^2.

	Utility boosting			Linear, constrained weights			
Variable	Util(%)	CER(%)	Sharpe	Util(%)	Sharpe	t-val	adj-R^2
Panel A: Macroeconomic variables							
DP	0.664	1.692	0.186	0.562	0.142	1.76	0.28
DY	0.656	1.633	0.184	0.563	0.144	1.87	0.33
EP	0.706	2.375	0.204	0.554	0.140	0.77	0.04
DE	0.579	0.711	0.157	0.537	0.134	0.69	-0.02
RVOL	0.690	2.229	0.209	0.555	0.145	2.80	0.63
BM	0.666	1.755	0.186	0.536	0.133	0.58	-0.07
NTIS	0.637	1.451	0.177	0.532	0.131	0.27	-0.10
TBL	0.637	1.503	0.184	0.584	0.155	2.26	0.55
LTY	0.625	1.308	0.180	0.563	0.146	1.49	0.18
LTR	0.678	1.895	0.190	0.601	0.155	2.68	0.69
TMS	0.704	2.280	0.202	0.595	0.154	1.90	0.41
DFY	0.667	1.848	0.191	0.526	0.129	0.38	-0.08
DFR	0.603	1.078	0.169	0.575	0.154	0.96	0.08
INFL	0.622	1.302	0.174	0.540	0.135	0.35	-0.09
Panel B: Technical indicators							
MA(1,9)	0.621	1.541	0.178	0.584	0.159	1.63	0.29
MA(1,12)	0.662	2.039	0.192	0.628	0.175	1.97	0.55
MA(2,9)	0.642	1.751	0.184	0.606	0.166	1.83	0.39
MA(2,12)	0.683	2.293	0.199	0.654	0.184	2.34	0.76
MA(3,9)	0.645	1.739	0.183	0.606	0.164	1.78	0.41
MA(3,12)	0.614	1.316	0.171	0.569	0.148	1.02	0.08
MOM(9)	0.612	1.320	0.172	0.571	0.151	1.11	0.11
MOM(12)	0.617	1.350	0.171	0.571	0.150	1.10	0.12
VOL(1,9)	0.603	1.253	0.170	0.575	0.153	1.30	0.16
VOL(1,12)	0.627	1.583	0.179	0.596	0.162	1.73	0.39
VOL(2,9)	0.613	1.368	0.174	0.578	0.154	1.31	0.18
VOL(2,12)	0.620	1.502	0.177	0.587	0.158	1.49	0.29
VOL(3,9)	0.600	1.162	0.168	0.562	0.146	0.95	0.04
VOL(3,12)	0.642	1.693	0.181	0.609	0.164	1.93	0.52
Panel C: Benchmarks							
Const	0.556	0.425	0.143				
HA	0.529		0.131				

Table A.4.2: In-sample predictive results with $\gamma = 5$ and $w^{\max} = 1$ when RVOL is the volatility proxy.

	Utility boosting			Linear, constrained weights			
Variable	Util(%)	CER(%)	Sharpe	Util(%)	Sharpe	t-val	adj-R^2
Panel A: Macroeconomic variables							
DP	0.667	1.877	0.186	0.585	0.142	1.76	0.28
DY	0.639	1.556	0.174	0.588	0.144	1.87	0.33
EP	0.728	2.740	0.207	0.571	0.136	0.77	0.04
DE	0.584	0.972	0.157	0.547	0.128	0.69	-0.02
RVOL	0.587	0.982	0.160	0.563	0.144	2.80	0.63
BM	0.707	2.396	0.205	0.550	0.129	0.58	-0.07
NTIS	0.646	1.724	0.180	0.535	0.123	0.27	-0.10
TBL	0.629	1.599	0.180	0.587	0.148	2.26	0.55
LTY	0.726	2.740	0.218	0.570	0.140	1.49	0.18
LTR	0.681	2.068	0.190	0.596	0.146	2.68	0.69
TMS	0.735	2.897	0.215	0.596	0.147	1.90	0.41
DFY	0.590	1.036	0.159	0.537	0.124	0.38	-0.08
DFR	0.622	1.418	0.170	0.587	0.148	0.96	0.08
INFL	0.601	1.165	0.164	0.550	0.129	0.35	-0.09
Panel B: Technical indicators							
MA(1,9)	0.615	1.639	0.176	0.590	0.155	1.63	0.29
MA(1,12)	0.654	2.122	0.190	0.628	0.172	1.97	0.55
MA(2,9)	0.627	1.732	0.180	0.609	0.163	1.83	0.39
MA(2,12)	0.673	2.360	0.197	0.651	0.180	2.34	0.76
MA(3,9)	0.636	1.838	0.181	0.610	0.162	1.78	0.41
MA(3,12)	0.606	1.403	0.168	0.571	0.142	1.02	0.08
MOM(9)	0.601	1.373	0.168	0.577	0.145	1.11	0.11
MOM(12)	0.623	1.587	0.173	0.587	0.148	1.10	0.12
VOL(1,9)	0.598	1.359	0.168	0.587	0.149	1.30	0.16
VOL(1,12)	0.621	1.711	0.178	0.601	0.161	1.73	0.39
VOL(2,9)	0.601	1.435	0.171	0.581	0.149	1.31	0.18
VOL(2,12)	0.608	1.559	0.174	0.588	0.155	1.49	0.29
VOL(3,9)	0.589	1.207	0.163	0.562	0.138	0.95	0.04
VOL(3,12)	0.631	1.759	0.178	0.607	0.160	1.93	0.52
Panel C: Benchmarks							
Const	0.562	0.611	0.143				
HA	0.541		0.125				

As another realized volatility proxy and robustness check for the 5-year rolling window volatility, we utilize RVOL as defined in Table 1 (see Mele (2007), without annualizing RVOL). This means that the lagged RVOL (lagged by one month) then acts as our volatility proxy σ_t^2. As a 12-month moving average volatility estimator, RVOL is also rather persistent, as the 60-month (5-year) moving average in the main analysis. However, due to the shorter computation window the 12-month estimator naturally reacts somewhat sharper to volatility changes.

Table A.4.3: In-sample predictive results with $\gamma = 5$ and $w^{\max} = 1$ when using the extracted conditional variance from the GARCH(1,1) model and RVOL as the volatility proxy.

	Utility boosting			Linear (PCA), constrained weights		
Variables	Util(%)	CER(%)	Sharpe	Util(%)	Sharpe	adj-R^2
Panel A: GARCH-based volatility proxy						
MACRO	0.934	5.223	0.305	0.584	0.158	0.54
TECH	0.676	2.186	0.198	0.614	0.169	0.45
ALL	0.944	5.377	0.313	0.650	0.188	0.92
Panel B: RVOL-based volatility proxy						
MACRO	0.896	4.866	0.292	0.605	0.159	0.54
TECH	0.648	1.982	0.191	0.605	0.161	0.45
ALL	0.958	5.664	0.321	0.645	0.183	0.92

Table A.4.2 shows that when relying on the RVOL-based volatility proxy, basically the same empirical findings between the methods and predictor-specific performances arise as already pointed out in the GARCH case. With the technical indicators, the same indicators are again performing the best whereas with the macroeconomic variables there is more variation. Interestingly in both robustness checks (GARCH and RVOL), EP and TMS stand out much more clearer among the macroeconomic variables than in the case of slowly evolving 5-year moving average volatility proxy.

Instead of the predictor-specific considerations, let us still consider multivariate results with different volatility proxies. Table A.4.3 reports the same groups of predictors and analysis settings as in Table 5 in the main text, but now for the GARCH and RVOL-based volatility proxies. The main results are largely the same as in Table 5. Larger information sets of predictors stabilizes the performances over certain noise accompanied to predictor-specific results. Compared with Table 5, maybe the only slight change is that now TECH appears to contribute the combined ALL case marginally stronger both in the utility boosting and in the two-step linear PCA-based approaches.

A.5 First differences of macroeconomic predictors

The descriptive statistics in Table 2 shows that majority of macroeconomic predictive variables are highly persistent. The only exceptions are LTR, DFR and INFL. This section considers whether taking the first differences of the persistent macroeconomic predictors affect the out-of-sample forecasting results.

Table A.5.1: Out of-sample results in predictor-specific models with $\gamma = 5$ and $\gamma = 3$ ($w^{\max} = 1$) after taking the first differences of the persistent macroeconomic predictors.

	$\gamma = 5$				$\gamma = 3$			
	Utility boosting		Linear, constr. weights		Utility boosting		Linear, constr. weights	
Variable	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe	Util(%)	Sharpe
ΔDP	0.510	0.182	0.438	0.139	0.649	0.194	0.520	0.139
ΔDY	0.417	0.128	0.430	0.125	0.510	0.130	0.521	0.131
ΔEP	0.380	0.123	0.353	0.104	0.491	0.135	0.478	0.123
ΔDE	0.492	0.172	0.440	0.138	0.624	0.184	0.546	0.145
ΔRVOL	0.396	0.127	0.419	0.125	0.519	0.138	0.535	0.137
ΔBM	0.441	0.142	0.456	0.139	0.536	0.141	0.554	0.146
ΔNTIS	0.435	0.139	0.416	0.124	0.530	0.145	0.561	0.145
ΔTBL	0.417	0.136	0.452	0.135	0.527	0.136	0.541	0.138
ΔLTY	0.428	0.138	0.435	0.130	0.541	0.142	0.545	0.139
ΔTMS	0.418	0.138	0.406	0.121	0.457	0.120	0.542	0.139
ΔDFY	0.421	0.137	0.464	0.139	0.546	0.144	0.569	0.147

Notes: In this table, we report the out-of-sample forecasting results when taking the first differences of the persistent macroeconomic predictive variables (see Tables and 1 and 2). Compare the out-of-sample forecasting results (with the levels) in Tables 7 and 8.

Table A.5.2: Out of-sample predictive results of the predictor groups (MACRO and ALL) with $\gamma = 5$ and $w^{\max} = 1$ using different volatility proxies and taking the first differences of persistent macroeconomic predictors.

	Utility boosting			Linear (PCA), constrained weights		
Variables	Util(%)	CER(%)	Sharpe	Util(%)	CER(%)	Sharpe
Panel A: 5-year moving average volatility proxy						
MACRO	0.538	1.795	0.186	0.375	-0.985	0.108
ALL	0.553	2.035	0.195	0.371	-0.520	0.119
Panel B: GARCH-based volatility proxy						
MACRO	0.571	1.444	0.205	0.377	-1.356	0.118
ALL	0.574	1.552	0.209	0.357	-1.375	0.116
Panel C: RVOL-based volatility proxy						
MACRO	0.564	1.454	0.199	0.416	-1.003	0.123
ALL	0.594	1.893	0.215	0.375	-1.244	0.115

Forecasting results with the changes in the dividend-price ratio (ΔDP), and partly also the differenced payout ratio (ΔDE), pops out from the results. In other words, using the changes of the strongly asset pricing-motivated dividend-price ratio supports its use in portfolio weight determination even further. Together with the fact that less persistent inflation (INFL) and partly also default return spread (DFR) perform well in Tables 7 and 8, this all suggests that less persistent state variables seem generally more useful in the utility boosting than highly persistent ones.

Multivariate results in Table A.5.2 (cf. Table 9) again strongly support the usefulness of the utility boosting method. Using all the predictors, including thus also the technical indicators (TECH), show that the additional predictive value can be obtained over using just the MACRO variables (here the first differences of the persistent macro variables augmented with LTR, DFR and INFL). It is also noteworthy how the multivariate results for utility boosting in Panel A are slightly higher using each evaluation criteria when compared to the ones obtained without differencing in Table 9 of the main text.

References

Brandt, M.W. (1999). Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach. *The Journal of Finance* 54(5), 1609–1645.

Campbell J.Y., and S.B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21, 1509–1531.

Marquering, W., and M. Verbeek (2004). The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis* 39, 407–429.

Mele, A. (2007). Asymmetric stock market volatility and the cyclical behavior of expected returns. *Journal of Financial Economics* 86, 446–478.

Neely, C.J., D.E. Rapach, J. Tu, and G. Zhou (2014). Forecasting the Equity Risk Premium: The Role of Technical Indicators. *Management Science* 60, 1772–1791.

Rapach, D.E., J.K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies* 23, 821–862.

Rapach D., and G. Zhou (2013). Forecasting stock returns. In G. Elliott and A. Timmermann (eds.), Handbook of Economic Forecasting, vol. 2A, North-Holland.

Rossi, A. (2018). Predicting Stock Market Returns with Machine Learning. Manuscript, University of Maryland (August 2018).

Welch, I., and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21, 1455–1508.