Predicting Stock Market Returns with Machine Learning

Alberto G. Rossi[†]
University of Maryland

August 21, 2018

Abstract

We employ a semi-parametric method known as Boosted Regression Trees (BRT) to forecast stock returns and volatility at the monthly frequency. BRT is a statistical method that generates forecasts on the basis of large sets of conditioning information without imposing strong parametric assumptions such as linearity or monotonicity. It applies soft weighting functions to the predictor variables and performs a type of model averaging that increases the stability of the forecasts and therefore protects it against overfitting. Our results indicate that expanding the conditioning information set results in greater out-of-sample predictive accuracy compared to the standard models proposed in the literature and that the forecasts generate profitable portfolio allocations even when market frictions are considered. By working directly with the mean-variance investor's conditional Euler equation we also characterize semi-parametrically the relation between the various covariates constituting the conditioning information set and the investor's optimal portfolio weights. Our results suggest that the relation between predictor variables and the optimal portfolio allocation to risky assets is highly non-linear.

Keywords: Equity Premium Prediction, Volatility Forecasting, GARCH, MIDAS, Boosted Regression Trees, Mean-Variance Investor, Portfolio Allocation.

[†]Smith School of Business, University of Maryland, 4457 Van Munching Hall, College Park, MD 20742. Email: arossi@rhsmith.umd.edu.

1 Introduction

Information plays a central role in modern finance. Investors are exposed to an ever-increasing amount of new facts, data and statistics every minute of the day. Assessing the predictability of stock returns requires formulating equity premium forecasts on the basis of large sets of conditioning information, but conventional statistical methods fail in such circumstances. Non-parametric methods face the so-called "curse-of-dimensionality". Parametric methods are often unduly restrictive in terms of functional form specification and are subject to data overfitting concerns as the number of parameters estimated increases. The common practice is to use linear models and reduce the dimensionality of the forecasting problem by way of model selection and/or data reduction techniques. But these methods exclude large portions of the conditioning information set and therefore potentially reduce the accuracy of the forecasts. To overcome these limitations we employ a novel semi-parametric statistical method known as Boosted Regression Trees (BRT). BRT generates forecasts on the basis of large sets of conditioning variables without imposing strong parametric assumptions such as linearity or monotonicity. It does not overfit because it performs a type of model combination that features elements such as shrinkage and subsampling. Our forecasts outperform those generated by established benchmark models in terms of both mean squared error and directional accuracy. They also generate profitable portfolio allocations for mean-variance investors even when market frictions are accounted for. Our analysis also shows that the relation between the predictor variables constituting the conditioning information set and the investors' optimal portfolio allocation to risky assets is, in most cases, non-linear and non-monotonic.

Our paper contributes to the long-standing literature assessing the predictability of stock returns. Over the nineties and the beginning of the twenty-first century the combination of longer time-series and greater statistical sophistication have spurred a large number of attempts to add evidence for or against the predictability of asset returns and volatility. In-sample statistical tests show a high degree of predictability for a number of variables: Rozeff (1984), Fama and French (1988), Campbell and Shiller (1988a,b), Kothari and Shanken (1997) and Pontiff and Schall (1998) find that valuation ratios predict stock returns, particularly so at long horizons; Fama and Schwert (1977), Keim and Stambaugh (1986), Campbell (1987), Fama and French (1989), Hodrick (1992) show that short and long-term treasury and corporate bonds explain variations in stock returns; Lamont (1998), Baker and Wurgler (2000) show that variables related to aggregate corporate payout and financing activity are useful predictors as well. While these results are generally encouraging, there are a number of doubts regarding their accuracy as most of the regressors considered are very persistent, making statistical inference less than straightforward; see, for example, Nelson and Kim (1993), Stambaugh (1999), Campbell and

Yogo (2006) and Lewellen, Nagel, and Shanken (2010). Furthermore, data snooping may be a source of concern if researchers are testing for many different model specifications and report only the statistically significant ones; see, for example, Lo and MacKinlay (1990), Bossaerts and Hillion (1999) and Sullivan, Timmermann, and White (1999). While it is sometimes possible to correct for specific biases, no procedure can offer full resolution of the shortcomings that affect the in-sample estimates.

Due to the limitations associated with in-sample analyses, a growing body of literature has argued that out-of-sample tests should be employed instead; see, for example, Pesaran and Timmermann (1995, 2000), Bossaerts and Hillion (1999), Marquering and Verbeek (2005), Campbell and Thompson (2008), Goyal and Welch (2003) and Welch and Goyal (2008). There are at least two reasons why out-of-sample results may be preferable to in-sample ones. The first is that even though data snooping biases can be present in out-of-sample tests, they are much less severe than their in-sample counterparts. The second is that out-of-sample tests facilitate the assessment of whether return predictability could be exploited by investors in real time, therefore providing a natural setup to assess the economic value of predictability.

The results arising from the out-of-sample studies are mixed and depend heavily on the model specification and the conditioning variables employed.[1] In particular, many of the studies conducted so far are characterized by one or more of these limitations. First, the forecasts are generally formulated using simple linear regressions. The choice is dictated by simplicity and the implicit belief that common functional relations can be approximated reasonably well by linear ones.[2] Most asset pricing theories underlying the empirical tests, however, do not imply linear relationships between the equity premium and the predictor variables, raising the issue whether the mis-specification implied by linear regressions is economically large. Second, linear models overfit the training dataset and generalize poorly out-of-sample as the number of regressors increases, so parsimonious models need to be employed at the risk of discarding valuable conditioning information. Approaching the forecasting exercise by way of standard non-parametric or semi-parametric methods is generally not a viable option because these methods encounter "curse-of-dimensionality" problems rather quickly as the size of the conditioning information set increases. Third, the models tested are generally constant: different model specifications are proposed and their performance is assessed ex-post. Although interesting from an econometric perspective, these findings are of little help for an investor interested in exploiting the condition-

[1]The data frequency also affects the results. Stock returns are found to be more predictable at quarterly, annual or longer horizons, while returns at the monthly frequency are generally considered the most challenging to predict.

[2]Another reason underlying the use of linear frameworks is that those statistical techniques were known by investors since the beginning of the twentieth century. For this and other issues related to "real-time" forecasts, see Pesaran and Timmermann (2005).

ing information in real time as he would not know what model to choose ex-ante.[3] Finally, apart from some important exceptions, much of the literature on financial markets prediction focuses on formulating return forecasts and little attention is dedicated to analyzing quantitatively the economic value associated with them for a representative investor.[4]

While conditional returns are a key element needed by risk-averse investors to formulate asset allocations, the conditional second moments of the return distribution are crucial as well. In fact, they are the only two pieces of information required by a mean-variance investor to formulate optimal portfolio allocations. It is widely known that stock market volatility is predictable and a number of studies attempts to identify which macroeconomic and financial time-series can improve volatility forecasts at the monthly or longer horizons.[5] But it is still unclear whether that conditioning information could have been incorporated in real-time and how much an investor would have benefitted from it.

In this paper we consider a representative mean-variance investor that exploits publicly available information to formulate excess returns and volatility forecasts using Boosted Regression Trees (BRT). BRT finds its origin in the machine learning literature, it has been studied extensively in the statistical literature and has been employed in the field of financial economics by Rossi and Timmermann (2010) to study the relation between risk and return. The appeal of this method lies in its forecasting accuracy as well as its ability to handle high dimensional forecasting problems without overfitting. These features are particularly desirable in this context, because they allow us to condition our forecasts on all the major conditioning variables that have been considered so far in the literature, guaranteeing that our analysis is virtually free of data-snooping biases. BRT also provide a natural framework to assess the relative importance of the various predictors at forecasting excess returns and volatility. Finally, the method allows for semi-parametric estimates of the functional form linking predictor and predicted variables, giving important insights on the limitations of linear regression.

Our analysis answers three questions. The first is whether macroeconomic and financial variables contain information about expected stock returns and volatility that can be exploited in real time by a mean-variance investor. For stock returns we use the major conditioning variables proposed so far in the literature and summarized by Welch and Goyal (2008). We propose two models of volatility forecasts. The first models volatility as a function of monthly macroeconomic and financial time-series as well as past volatility. The second is inspired by the family of MIDAS models proposed by Ghysels, Santa-Clara, and Valkanov (2006) and models monthly volatility

[3] For exceptions, see Dangl and Halling (2008) and Johannes, Korteweg, and Polson (2009).

[4] For exceptions, see Campbell and Thompson (2008) and Marquering and Verbeek (2005).

[5] See, for example, Campbell (1988), Breen, Glosten, and Jagannathan (1989), Marquering and Verbeek (2005), Engle and Rangel (2005) and Engle, Ghysels, and Sohn (2006), Lettau and Ludvigson (2009) and Paye (2010).

as a function of lagged daily squared returns. We call this model "semi-parametric MIDAS" and show that its performance is superior to that of its parametric counterpart. Genuine out-of-sample forecasts require not only that the parameters are estimated recursively, but also that the conditioning information employed is selected in real-time. For this reason, every predictive framework under consideration starts from the large set of predictor variables employed by Welch and Goyal (2008) and selects recursively the model specification. Our estimates show that BRT forecasts outperform the established benchmarks and possess significant market timing in both returns and volatility.

A related question we address is whether the conditioning information contained in macro and financial time-series can be exploited to select the optimal portfolio weights directly, as proposed by Ait-Sahalia and Brandt (2001). Rather than forecasting stock returns and volatility separately and computing optimal portfolio allocations in two separate steps, we model directly the optimal portfolio allocation as a target variable. Our approach can be interpreted as the semi-parametric counterpart of Ait-Sahalia and Brandt (2001),[6] because instead of reducing the dimensionality of the problem faced by the investor using a single index model, we employ a semi-parametric method that avoids the so-called "curse of dimensionality". Our analysis gives rise to two findings. First, formal tests of portfolio allocation predictability show that optimal portfolio weights are time-varying and forecastable; second, we show that the relation between the predictor variables constituting the conditioning information set and the mean-variance investor's optimal portfolio allocation to risky assets is highly non-linear.

The third question we analyze is whether the generated forecasts are economically valuable in terms of the profitability of the portfolio allocations they imply. We assess this by computing excess returns, Sharpe ratios and Treynor-Mazuy market timing tests for the competing investment strategies. Our results highlight that BRT forecasts translate into profitable portfolio allocations. We also compute the realized utilities and the break-even monthly portfolio fees that a representative agent would be willing to pay to have his wealth invested through the strategies we propose, compared to the benchmark of placing 100% of his wealth in the market portfolio. We show that the break-even portfolio fees are sizable even when transaction costs as well as shortselling and borrowing constraints are considered. For example, a representative investor with a risk-aversion coefficient of 4 who faces short-selling and borrowing constraints as well as transaction costs would be willing to pay yearly fees equal to 4% of his wealth to have his capital invested in the investment strategy we propose rather than the market portfolio.

The rest of the paper is organized as follows. Section 2 introduces our empirical framework and describes how stock returns and volatility are predicted. In Section 3 we show how we employ

[6]It is important to clarify that our analysis applies only to the mean-variance investor, while Ait-Sahalia and Brandt (2001) work with power utility investors as well.

boosted regression trees to directly select optimal portfolio allocations. Section 4 presents results for the out-of-sample accuracy of the model, conducts formal tests of market timing in both returns and volatility and evaluates the performance of empirical trading strategies based on BRT forecasts. Section 5 concludes.

2 Empirical Framework and Full-Sample Results

Consider a representative agent that has access to a risk-free asset paying a return of $r_{f,t+1}$ and the market portfolio with a return r_{t+1} and volatility σ_{t+1}. The agent's utility function is affected only by the first and second moments of the returns distribution, i.e. his utility function takes the form

$$U_t(\cdot) = E_t\{r_{p,t+1}\} - \frac{1}{2}\gamma Var_t\{r_{p,t+1}\}, \tag{1}$$

where γ is the coefficient of risk-aversion, $r_{p,t+1} = w_{t+1|t}\ r_{t+1} + (1 - w_{t+1|t})\,(r_{f,t+1})$ and $w_{t+1|t}$ is the proportion of wealth allocated to the risky asset for period $t+1$ given the information available as of time t. Given the expected returns and volatility of the market portfolio, the investor chooses his asset allocation by solving the maximization problem

$$\max_{w_{t+1|t}} \left[E_t\left\{w_{t+1|t}\ r_{t+1} + (1 - w_{t+1|t})\,(r_{f,t+1})\right\} - \frac{1}{2}\gamma Var_t\left\{w_{t+1|t}\ r_{t+1} + (1 - w_{t+1|t})\,(r_{f,t+1})\right\}\right],$$

leading to the optimal portfolio weights

$$w^*_{t+1|t} = \frac{E_t\{r_{t+1}\} - r_{f,t+1}}{\gamma Var_t\{r_{t+1}\}}. \tag{2}$$

When we impose realistic short-selling and borrowing constraints, the optimal weights have to lie between 0 and 1, so they become

$$w^r_{t+1|t} = \begin{cases} 0 & if \quad w^*_{t+1|t} < 0, \\ w^*_{t+1|t} & if \quad 0 \le w^*_{t+1|t} \le 1 \\ 1 & if \quad w^*_{t+1|t} > 1. \end{cases}$$

The objects $E_t\{r_{t+1}\} = \mu_{t+1|t}$ and $Var_t\{r_{t+1}\} = \sigma^2_{t+1|t}$ in Eq. 2 represent conditional expectations of returns and variance on the basis of the investor's conditioning information at time t. In this paper we allow these conditional expectations to be non-linear functions of observable macroeconomic and financial time-series, the idea being that the linearity assumption generally adopted in financial economics may be costly in terms of forecasting accuracy and portfolio allocation profitability.

The conditioning information we use are the twelve predictor variables previously analyzed in Welch and Goyal (2008) and by many others subsequently. Stock returns are tracked by the S&P 500 index and include dividends. A short T-bill rate is subtracted to obtain excess returns. The predictor variables from the Goyal and Welch analysis are available during 1927-2005 and we extend their sample up to the end of 2008.[7] The predictor variables pertain to three large categories. The first goes under the heading of "risk and return" and contains lagged returns (exc), long-term bond returns (ltr) and volatility (vol). The second, called "fundamental to market value" includes the log dividend-price ratio (dp) and the log earnings-price ratio (ep). The third category comprises measures of interest rate term structure and default risk and includes the three-month T-bill rate (Rfree), the T-bill rate minus a three-month moving average (rrel), the yield on long term government bonds (lty), the term spread measured by the difference between the yield on long-term government bonds and the three-month T-bill rate (tms) and the yield spread between BAA and AAA rated corporate bonds (defspr). We also include inflation (infl) and the log dividend-earnings ratio (de). Additional details on data sources and the construction of these variables are provided by Welch and Goyal (2008). All predictor variables are appropriately lagged so they are known at time t for purposes of forecasting returns in period $t+1$.

For stock returns, conditional expectations are commonly generated according to the following linear model

$$\widehat{\mu}_{t+1|t} = \beta_{\mu}' x_t,$$

where x_t represents a set of publicly available predictor variables and β_μ is a vector of parameter estimates obtained via ordinary least squares. The linear specification is generally imposed for simplicity at the expense of being potentially misspecified. The sources of misspecification are at least two. The first relates to what information is incorporated in the formulation of the forecasts. Asset pricing models suggest a wide array of economic state variables for both returns and volatility, but linear frameworks are prone to over-fitting if the number of parameters to be estimated is large compared to the number of observations, forcing the agent to exclude a large portion of the conditioning information available. The second relates to how information is incorporated in the forecasts: theoretical frameworks rarely identify linear relations between the

[7] We are grateful to Amit Goyal and Ivo Welch for providing this data. A few variables were excluded from the analysis since they were not available up to 2008, including net equity expansion and the book-to-market ratio. We also excluded the CAY variable since this is only available quarterly since 1952.

variables at hand, so empirical estimates based on ordinary least squares may not be appropriate. Note however that, in our context, misspecification *per se* is not a source of concern as long as it does not translate into lower predictive accuracy, which is ultimately what matters for portfolio allocation.

To address this issue, we extend the basic linear regression model to a class of more flexible models known as Boosted Regression Trees. These have been developed in the machine learning literature and can be used to extract information about the relationship between the predictor variables x_t and r_{t+1} based only on their joint empirical distribution. To get intuition for how regression trees work and explain why we use them in our analysis, consider the situation with a continuous dependent variable Y (e.g., stock returns) and two predictor variables X_1 and X_2 (e.g., the volatility and the default spread). The functional form of the forecasting model mapping X_1 and X_2 into Y_t is unlikely to be known, so we simply partition the sample support of X_1 and X_2 into a set of regions or "states" and assume that the dependent variable is constant within each partition.

More specifically, by limiting ourselves to lines that are parallel to the axes tracking X_1 and X_2 and by using only recursive binary partitions, we carve out the state space spanned by the predictor variables. We first split the sample support into two states and model the response by the mean of Y in each state. We choose the state variable (X_1 or X_2) and the split point to achieve the best fit. Next, one or both of these states is split into two additional states. The process continues until some stopping criterion is reached. Boosted regression trees are additive expansions of regression trees, where each tree is fitted on the residuals of the previous tree. The number of trees used in the summation is also known as the number of boosting iterations.

This approach is illustrated in Figure 1, where we show boosted regression trees that use two state variables, namely the lagged values of the default spread and market volatility, to predict excess returns on the S&P500 portfolio. We use "tree stumps" (trees with only two terminal nodes), so every new boosting iteration generates two additional regions. The graph on the left uses only three boosting iterations, so the resulting model splits the space spanned by the two regressors in six regions with one split along the default spread axis and two splits along the volatility axis. Within each state the predicted value of stock returns is constant. The predicted value of excess returns is smallest for high values of volatility and low values of the default spread, and highest for medium values of volatility and high values of the default spread. So already at three boosting iterations BRT highlights non-linearities in the functional form relating volatility and stock returns. With only three boosting iterations the model is quite coarse, but the fit becomes more refined as the number of boosting iterations increase. To illustrate this we plot on right the fitted values for a BRT model with 5,000 boosting iterations. Now the plot is much more smooth, but clear similarities between the two graphs remain.

Figure 1 illustrates how boosted regression trees can be used to approximate the relation between the dependent and independent variables by means of a series of piece-wise constant functions. This approximation is good even in situations where, say, the true relation is linear, provided that sufficiently many boosting iterations are used. Next, we provide a more formal description of the methodology and how we implement it in our study.[8]

2.1 Regression Trees

Suppose we have P potential predictor ("state") variables and a single dependent variable over T observations, i.e. (x_t, y_{t+1}) for $t = 1, 2, ..., T$, with $x_t = (x_{t1}, x_{t2}, ..., x_{tp})$. As illustrated in Figure 1, fitting a regression tree requires deciding (i) which predictor variables to use to split the sample space and (ii) which split points to use. The regression trees we use employ recursive binary partitions, so the fit of a regression tree can be written as an additive model:

$$f(x) = \sum_{j=1}^{J} c_j I\{x \in S_j\}, \tag{3}$$

where $S_j,\ j = 1, ..., J$ are the regions we split the space spanned by the predictor variables into, $I\{\}$ is an indicator variable and c_j is the constant used to model the dependent variable in each region. If the L^2 norm criterion function is adopted, the optimal constant is $\hat{c}_j = mean(y_{t+1}|x_t \in S_j)$, while it is $\hat{c}_j = median(y_{t+1}|x_t \in S_j)$ for the L^1 norm instead.

The globally optimal splitting point is difficult to determine, particularly in cases where the number of state variables is large. Hence, a sequential greedy algorithm is employed. Using the full set of data, the algorithm considers a splitting variable p and a split point s so as to construct half-planes

$$S_1(p, s) = \{X|X_p \leq s\} \qquad and \qquad S_2(p, s) = \{X|X_p > s\}$$

that minimize the sum of squared residuals:

$$\min_{p,s} \left[\min_{c_1} \sum_{x_t \in S_1(p,s)} (y_{t+1} - c_1)^2 + \min_{c_2} \sum_{x_t \in S_2(p,s)} (y_{t+1} - c_2)^2 \right]. \tag{4}$$

[8] Our description draws on Hastie, Tibshirani, and Friedman (2009) and Rossi and Timmermann (2010) who provide a more in-depth coverage of the approach.

For a given choice of p and s the fitted values, $\hat{c}_1$ and $\hat{c}_2$, are

$$\begin{aligned}\widehat{c}_1 &= \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_1(p,s)\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_1(p,s)\}, \\ \widehat{c}_2 &= \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_2(p,s)\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_2(p,s)\}. \end{aligned} \tag{5}$$

The best splitting pair (p, s) in the first iteration can be determined by searching through each of the predictor variables, $p = 1, .., P$. Given the best partition from the first step, the data is then partitioned into two additional states and the splitting process is repeated for each of the subsequent partitions. Predictor variables that are never used to split the sample space do not influence the fit of the model, so the choice of splitting variable effectively performs variable selection.

Regression trees are generally employed in high-dimensional datasets where the relation between predictor and predicted variables is potentially non-linear. This becomes important when modeling stock returns because numerous predictor variables have been proposed so far in the literature. Furthermore, the theoretical frameworks rarely imply a linear or monotonic relation between predictor and predicted variable. On the other hand, the approach is sequential and successive splits are performed on fewer and fewer observations, increasing the risk of fitting idiosyncratic data patterns. Furthermore, there is no guarantee that the sequential splitting algorithm leads to the globally optimal solution. To deal with these problems, we next consider a method known as boosting.

2.2 Boosting

Boosting is based on the idea that combining a series of simple prediction models can lead to more accurate forecasts than those available from any individual model. Boosting algorithms iteratively re-weight data used in the initial fit by adding new trees in a way that increases the weight on observations modeled poorly by the existing collection of trees. From above, recall that a regression tree can be written as:

$$\mathcal{T}\left(x; \{S_j, c_j\}_{j=1}^{J}\right) = \sum_{j=1}^{J} c_j I\{x \in S_j\} \tag{6}$$

A boosted regression tree is simply the sum of regression trees:

$$f_B(x) = \sum_{b=1}^{B} \mathcal{T}_b\left(x; \{S_{b,j}, c_{b,j}\}_{j=1}^{J}\right), \tag{7}$$

where $\mathcal{T}_b\left(x;\{S_{b,j},c_{b,j}\}_{j=1}^{J}\right)$ is the regression tree used in the b-th boosting iteration and B is the number of boosting iterations. Given the model fitted up to the $(b-1)-th$ boosting iteration, $f_{b-1}(x)$, the subsequent boosting iteration seeks to find parameters $\{S_{j,b},c_{j,b}\}_{j=1}^{J}$ for the next tree to solve a problem of the form

$$\{\hat{S}_{j,b},\hat{c}_{j,b}\}_{j=1}^{J}=\min_{\{S_{j,b},c_{j,b}\}_{j=1}^{J}}\sum_{t=0}^{T-1}\left[y_{t+1}-\left(f_{b-1}(x_t)+\mathcal{T}_b\left(x_t;\{S_{j,b},c_{j,b}\}_{j=1}^{J}\right)\right)\right]^2. \tag{8}$$

For a given set of state definitions ("splits"), $S_{j,b}$, $j=1,..,J$, the optimal constants, $c_{j,b}$, in each state are derived iteratively from the solution to the problem

$$\begin{aligned}\hat{c}_{j,b} &= \min_{c_{j,b}}\sum_{x_t\in S_{j,b}}\left[y_{t+1}-(f_{b-1}(x_t)+c_{j,b})\right]^2\\ &= \min_{c_{j,b}}\sum_{x_t\in S_{j,b}}\left[e_{t+1,b-1}-c_{j,b}\right]^2,\end{aligned} \tag{9}$$

where $e_{t+1,b-1}=y_{t+1}-f_{b-1}(x_t)$ is the empirical error after $b-1$ boosting iterations. The solution to this is the regression tree that most reduces the average of the squared residuals $\sum_{t=1}^{T}e_{t+1,b-1}^2$ and $\hat{c}_{j,b}$ is the mean of the residuals in the jth state.

Forecasts are simple to generate from this approach. The boosted regression tree is first estimated using data from $t=1,...,t^*$. Then the forecast of y_{t^*+1} is based on the model estimates and the value of the predictor variable at time t^*, x_{t^*}. Boosting makes it more attractive to employ small trees (characterized by only two terminal nodes) at each boosting iteration, reducing the risk that the regression trees will overfit. Moreover, by summing over a sequence of trees, boosting performs a type of model averaging that increases the stability and accuracy of the forecasts.[9]

2.3 Implementation

Our estimations follow the stochastic gradient boosting approach of Friedman (2001) and Friedman (2002) with $J=2$ nodes. The baseline implementation employs 10,000 boosting iterations, but we conduct a number of robustness checks to show that the results are not very sensitive to this choice.

We adopt three refinements to the basic boosted regression tree methodology. The first is *shrinkage*. As with ridge regression and neural networks, shrinkage is a simple regularization technique that diminishes the risk of over-fitting by slowing the rate at which the empirical risk is minimized on the training sample. We use a shrinkage parameter, $0<\lambda<1$, which determines

[9]See Rapach, Strauss, and Zhou (2010) for similar results in the context of linear regression.

how much each boosting iteration contributes to the overall fit:

$$f_b(x) = f_{b-1}(x) + \lambda \sum_{j=1}^{J} c_{j,b} I\{x \in S_{j,b}\}. \tag{10}$$

Following common practice we set $\lambda = 0.001$ as it has been found (Friedman (2001)) that the best empirical strategy is to set λ very small and correspondingly increase the number of boosting iterations.

The second refinement is *subsampling* and is inspired by "bootstrap aggregation" (bagging), see Breiman (1996). Bagging is a technique that computes forecasts over bootstrap samples of the data and averages them in a second step, therefore reducing the variance of the final predictions. In our context, the procedure is adapted as follows: at each boosting iteration we sample without replacement one half of the training sample and fit the next tree on the sub-sample obtained.

Finally, our empirical analysis minimizes mean absolute errors, i.e. $T^{-1} \sum_{t=1}^{T} |y_{t+1} - f(x_t)|$. Under this criterion function, the optimal forecast is the conditional median of y_{t+1} rather than the conditional mean entailed by squared error loss. We do this in the light of a large literature which suggests that squared-error loss places too much weight on observations with large absolute residuals. This is a particularly important problem for fat-tailed distributions such as those observed for stock returns and volatility. By minimizing absolute errors, our regression model is likely to be more robust to outliers such as returns during October 1987, thus reducing the probability of overfitting.

2.4 Conditional Volatility Estimates

Conditional volatility estimates have taken various shapes and forms in the recent past. Paye (2010), Ludvigson and Ng (2007) and Marquering and Verbeek (2005) propose a linear specification, whereby estimates of monthly return volatility, σ_{t+1}, are obtained from high frequency data and are modeled according to the following linear specification:

$$\widehat{\sigma}^{lin}_{t+1|t} = \beta'_{\sigma} x_t$$

Even though these methods give interesting in-sample results, their out-of-sample performance was never shown to be convincing and this is probably why linear regression estimates are not so popular in this strand of the literature.

Volatility forecasts are generally based on GARCH-type frameworks. These assume a model for the variance of the residuals in a regression $r_t = \mu_t + \sigma_t \ \epsilon_t$, where $var(\epsilon_t) = 1$. For example, the GARCH (1,1) model for the variance is defined as follows:

$$\widehat{\sigma}_{t+1|t}^{Garch} = \sqrt{\omega + \alpha(r_t - \mu_t)^2 + \beta\sigma_t} = \sqrt{\omega + \alpha\sigma_t\epsilon_t^2 + \beta\sigma_t}$$

where ω, α and β are estimated via maximum likelihood.

More recently, Ghysels, Santa-Clara, and Valkanov (2005) proposed a new family of models under the heading MIDAS (MIxed DAta Sampling). Their model adopts the following estimator for the conditional variance of monthly returns:

$$\widehat{\sigma}_{t+1|t}^{MIDAS} = 22\sum_{d=0}^{D} w_d r_{t-d}^2, \tag{11}$$

where

$$w_d(\kappa_1, \kappa_2) = \frac{\left(\frac{d}{D}\right)^{\kappa_1 - 1}\left(1 - \frac{d}{D}\right)^{\kappa_2 - 1}}{\sum_{i=0}^{D}\left(\frac{i}{D}\right)^{\kappa_1 - 1}\left(1 - \frac{i}{D}\right)^{\kappa_2 - 1}} \tag{12}$$

for the "Beta" weights model that we adopt in this paper. "D" represents the maximum lag length which is set to 250 days following Ghysels, Santa-Clara, and Valkanov (2005).

We propose two models for conditional volatility. The first extends the linear framework presented above as we estimate

$$\widehat{\sigma}_{t+1|t}^{BRT} = f_\sigma(x_t|\hat{\theta}_\sigma), \tag{13}$$

where x_t represents the same set of publicly available predictor variables we employ to construct conditional expected returns and $\hat{\theta}_\sigma$ are estimates of the parameters obtained via boosted regression trees. The second specification is similar to that of Eq. 13, with the difference that x_t is now a vector consisting of 250 days of lagged squared returns. Because semi-parametric in nature and inspired by MIDAS, we call this model "Semi-Parametric MIDAS".

The exercise we undertake has three main objectives. The first is to understand what kind of information Semi-Parametric MIDAS exploits compared to the parametric MIDAS specification. Second, we are able to assess whether relaxing the MIDAS parametric assumptions results in more accurate out-of-sample forecasts. Third, we can compare the forecasts of "Semi-Parametric MIDAS" to those of Boosted Regression Trees that exploit macroeconomic and financial variables and assess what type of information is more valuable to forecast aggregate market volatility at the monthly frequency.

2.5 MIDAS and Semi-Parametric MIDAS

MIDAS originates from the idea that expanding the information set on which volatility forecasts are conditioned should increase their precision. While realized volatility frameworks generally use arbitrary volatility lags to forecast the future, MIDAS models fit volatility by weighting past observations optimally. For the case of monthly data, this translates into forecasting a given month's volatility using daily squared returns lagged up to a year: i.e. 250 days. In order to reduce the dimensionality of the problem specific weighting functions are assumed (the Beta weights reported in Eq. 12 is one possible example) and their parameters are obtained via Quasi-Maximum Likelihood. The optimal weights determine how past information is incorporated in the predictive model to generate volatility forecasts.

For boosted regression trees we can construct a comparable measure of how past information affects volatility forecasts. We consider the reduction in the empirical error every time one of the 250 lagged squared return, $x_{l,\cdot}$, is used to split the tree. Summing the reductions in empirical errors (or improvements in fit) across the nodes in the tree gives a measure of the variable's influence (Breiman (1984)):

$$I_l(\mathcal{T}) = \sum_{j=2}^{J} \Delta|e(j)| I(LSR(j) = l), \tag{14}$$

where $\Delta|e(j)| = T^{-1}\sum_{t=1}^{T}(|e_t(j-1)| - |e_t(j)|)$, is the reduction in the (absolute) empirical error at the j'th node and $LSR(j)$ is the lagged squared return chosen at this node, so $I(LSR(j) = l)$ equals one if lag l is chosen and zero otherwise. The sum is computed across all time periods, $t = 1, ..., T$ and over the $J - 1$ internal nodes of the tree.

The rationale for this measure is that at each node, one of the 250 lagged squared returns gets selected to partition the sample space into two sub-states. The particular lag chosen at node j achieves the greatest reduction in the empirical risk of the model fitted up to node $j - 1$. The importance of each lagged return, $x_{l,\cdot}$, is the sum of the reductions in the empirical errors computed over all internal nodes for which it was chosen as the splitting variable. If a lag never gets chosen to conduct the splits, its influence is zero. Conversely, the more frequently a lag is used for splitting and the bigger its effect on reducing the model's empirical risk, the larger its influence.

This measure of influence can be generalized by averaging over the number of boosting iterations, B, which generally provides a more reliable measure of influence:

$$\bar{I}_l = \frac{1}{B}\sum_{b=1}^{B} I_l(\mathcal{T}_b). \tag{15}$$

This is best interpreted as a measure of relative influence that can be compared across squared lagged returns. We therefore report the following measure of relative influence, $\overline{RI}_l$, which sums to one:

$$\overline{RI}_l = \bar{I}_l / \sum_{l=1}^{L} \bar{I}_l. \tag{16}$$

It should be clear that the relative influence measure we propose is the semi-parametric counterpart of the popular MIDAS weights. In Figure 2 we aggregate individual days' relative influence and MIDAS (Beta) weights in 1, 2, 3, and 4 - week periods and compare the relative importance that the two procedures give to past information. The weights obtained via the semi-parametric procedure are less smooth than the MIDAS (Beta) ones as we do not impose any parametric restriction on their shape, and yet it is comforting to see that the weighting functions are similar across the two procedures. There are however some important differences as semi-parametric MIDAS gives greater weight to more recent observations compared to MIDAS (Beta): the first six weeks of lagged squared returns are given more weight in our model at the expense of the following 10-15 weeks of returns that are underweighted compared to MIDAS (Beta). The most distant thirty weeks of data are given very little weight by both models.

The fact that the two procedures exploit differently the same set of conditioning information raises the question of which one exploits it in a more effective fashion. We answer this question in Section 4 that is dedicated to assessing the out-of-sample performance of the forecasting models at hand.

3 Predicting Optimal Portfolio Allocations

In the previous section we have focused our attention on forecasting returns and volatility separately and constructing optimal portfolio allocations in a second step. Given the set of conditioning information we rely on, a natural question is whether it is possible to forecast directly the optimal portfolio allocation and whether such approach leads to more or less profitable forecasts in a portfolio allocation setting. Furthermore, given that the optimal portfolio allocation for a mean-variance investor is the ratio of expected excess return and variance, scaled by the coefficient of risk-aversion,[10] conditional optimal portfolio weights represent a natural measure of the expected price of risk.

We construct optimal portfolio allocations based on market returns r_{t+1} and volatility σ_{t+1}

[10]We set it to 4 as a commonly chosen parameter in the literature

as well as the risk-aversion coefficient γ

$$w_{t+1}^{real} = \frac{r_{t+1} - r_{f,t+1}}{\gamma\ \sigma_{t+1}^2}, \tag{17}$$

and use it as our target variable. We then estimate optimal portfolio allocations based on time t information set

$$\widehat{w}_{t+1|t}^{real} = f_w(x_t|\hat{\theta}_w), \tag{18}$$

where x_t represents the same set of publicly available predictor variables we use for returns and volatility and $\hat{\theta}_w$ are estimates of the parameters obtained via Boosted Regression Trees. There are very few papers that have analyzed the direct relationship between economic activity and optimal portfolio weights, see Ait-Sahalia and Brandt (2001) as one example, and there are virtually no economic theories that model this relation. It then seems particularly appropriate to present our exploratory findings on the relation between conditional portfolio weights and the variables constituting the conditioning information set.

This task is more complicated than usual as regression trees do not impose any restrictions on the functional form of the relationship between the dependent variable $-$ the optimal portfolio weight$-$and the predictor variables, X. To address this point, we proceed as follows. Suppose we select a particular predictor variable, X_p, from the set of P predictor variables $X = (X_1, X_2, ..., X_P)$ and denote the remaining variables X_{-p}, i.e. $X_{-p} = X \backslash \{X_p\}$. We use the following measure of the average marginal effect of X_p on the dependent variable

$$f_p(X_p) = E_{X_{-p}} f(X_p, X_{-p}). \tag{19}$$

This is called the average partial dependence measure. It fixes the value of X_p and averages out the effect of all other variables. By, repeating this process for different values of X_p, we trace out the marginal effect this covariate has on the predicted variable.

An estimate of $f_p(X_p)$ can be computed by averaging over the sample observations

$$\bar{f}_p(X_p) = \frac{1}{T} \sum_{t=1}^{T} f(X_p, x_{t,-p}), \tag{20}$$

where $x_{t,-p} = \{x_{1,-p}, ..., x_{T,-p}\}$ are the values of X_{-p} occurring in the data.

Over the full sample, the predictors with the greatest relative influence are the log earnings price ratio (ep), inflation (infl), the log dividend earnings ratio (de), the yield on long term government bonds (lty), stock market volatility (vol) and the three-month T-bill rate (Rfree) that have a relative influence of 16.63%, 15.88%, 13.94%, 9.38%, 7.23% and 6.67%, respectively.

In Figure 3 we present their partial dependence plots. The relationship between the optimal weight in the risky asset and the log earnings price ratio is positive, it is strongest at low or high levels of this ratio and gets weaker at medium levels. The variation in the expected optimal portfolio weight as a function of the earnings-price ratio is economically large, spanning a range of 500% and entailing long and highly levered positions for log earnings price ratios greater than -2.2 and short positions for ratios smaller than -3.3. The partial dependence plot for inflation is particularly interesting. At negative levels of inflation the relationship between the rate of inflation and the optimal risky investment is either flat or rising. Thus in a state of deflation, rising consumer prices leads to a higher exposure to risky assets by the investor. Conversely, at positive levels of inflation, higher consumer prices become bad news for stocks. Again, the effect of inflation is quite strong in economic terms: an inflation rate equal to zero is associated with a long and levered position in the stock market, while inflation rates greater than 1% are associated with no exposure to the risky asset. The log dividend earnings ratio is inversely related to the investment in risky assets and its sensitivity is greatest for medium values of the ratio and weakest at low and high levels of the ratio. The variation is again quite large in economic terms as the difference between optimal portfolio weights at small and large values of the log dividend earnings ratio is 200%. The relation between optimal portfolio weights and long-term government bonds, volatility and the T-bill rate are all highly nonmonotonic. The optimal investment in risky asset is increasing for low values of the predictors and decreasing for high values of them.

Rossi and Timmermann (2010) perform a similar analysis for return and volatility prediction. Their results highlight that the three most important predictors for the equity premium are inflation, the log earnings price ratio and the de-trended T-bill rate; the ones for volatility are the lag volatility, the default spread and the lag return on the market portfolio. They also uncover strong non-linearities in the functional form relating the predictor and predicted variables. These findings imply that linear models are potentially misspecified. In order to assess whether correcting for such misspecification translates into greater predictive accuracy, we assess next the out-of-sample forecasting performance of boosted regression trees and other benchmark models commonly employed in the literature.

4 Out-of-Sample Results.

In this section we present the out-of-sample forecasting performance of boosted regression trees and other benchmark forecasting methods. We present the results in terms of Mean Squared Error as it is the most common loss function for forecasting problems with continuous dependent variables. We also present directional accuracy results as it is known since Leitch and Tanner (1991) that the latter forecast evaluation criterion is closely related to the profitability of financial forecasts. Finally, we implement a number of forecasting accuracy tests proposed in the literature. Throughout the empirical section we use 1969 as the starting value for our out-of-sample evaluation period. The choice is driven by the need to have a large training sample and allow for a 40-year test sample.

4.1 Returns

Panel A of Table 1 reports the results for return prediction in terms of mean-squared error (MSE) and directional accuracy. The results are reported for BRT as well as two competing models proposed in the literature: the prevailing mean proposed by Welch and Goyal (2008) and a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2. Over the sample 1969-2008 BRT outperforms both benchmarks in terms of L^2 norm and achieves an out-of-sample R^2 of 0.30%. The second best is the prevailing mean, with an R^2 of -0.79%. Finally, the performance of the multivariate linear model is disappointing, with an R^2 of -2.28%.[11]

Our results are consistent and extend the ones obtained by Welch and Goyal (2008). They are consistent in the sense that the linear framework is outperformed by the prevailing mean. They extend them as our specification simulates the real-time decision-making of the forecaster that chooses the conditioning information in real time by way of recursive model selection on the whole set of predictor variables. The results indicate that even though the linear framework and boosted regression trees condition their forecasts on the same information, only the latter is capable of exploiting the information effectively and generate valuable forecasts.[12]

As mentioned above, MSE may not be an appropriate loss function in financial settings as directional accuracy has, in some circumstances, proved to be more closely related to the profitability of financial forecasts. For this reason, we also analyze the directional accuracy of the competing models. Because there is no obvious benchmark of no-predictability in this case, we employ the recursively estimated prevailing direction as a benchmark. Furthermore, because the prevailing direction has been that of positive excess returns over the sample under consideration, our benchmark is equivalent to predicting positive excess returns at all times and has a predictive accuracy of 55.26%. BRT has a hit ratio of 57.35%, 2% greater than the prevailing direction.

[11]Our sample comprises the years 2006-2008 that are not included in the analysis of Welch and Goyal (2008). Our results for the test sample ending in 2005 are much stronger. BRT has an R^2 of 1.56% while the prevailing mean and the linear model have an R^2 equal to -0.58% and -3.64%, respectively.

[12]In the context of financial markets prediction, very low (but positive) R^2's generally translate into profitable investment strategies, see Campbell and Thompson (2008).

The linear model has a rather disappointing accuracy of 52.17% and so ranks last among the models considered.

In the top panel of Figure 4, we present the Goyal and Welch "CumSum analysis" for stock returns. It is defined as

$$CumSum(T) = \sum_{t=\tau}^{T-1} \left(\widehat{\mu}_{t+1|t}^{pm} - r_{t+1}\right)^2 - \left(\widehat{\mu}_{t+1|t}^{model} - r_{t+1}\right)^2, \tag{21}$$

where τ is the first month we start our calculations from, "*pm*" stands for prevailing mean and "*model*" is alternatively BRT or the multivariate linear model. This measure is useful at tracking the performance of the forecasting framework on each period because an upward sloping "CumSum" curve over a given month entails that the predictive model outperforms the prevailing mean, and vice versa. The plot shows that BRT outperforms the multi-variate linear model and the prevailing mean over the sample under consideration. BRT outperforms the prevailing mean from 1970 to 1975 and from 1980 to 2005. It underperforms the prevailing mean over the period 1975-1980 and the last few years characterized by the financial crisis. The multivariate linear model outperforms the prevailing mean over the period 1970-1975 and in 2008. It underperforms consistently the prevailing mean from 1975 until the end of 2007.

To parallel the "CumSum" analysis presented above, we present in the top panel of Figure 5 the cumulative out-of-sample directional accuracy difference between BRT, the linear model and the prevailing direction that we use as a benchmark. The vertical axis counts the differential performance between the prediction model and the benchmark in terms of correct directional forecasts. BRT performs very well over the years 1969-1988. Its performance is rather poor between 1989-1995 and picks up again towards the end of the sample. The performance for the linear model is always worse than that of BRT and is particularly disappointing over the period 1990-2005.

Our benchmark analysis uses 10,000 boosting iterations to estimate the regression trees. In unreported results we show that our findings are not sensitive to this choice. For example, BRT with 5,000 and 15,000 boosting iterations outperform the benchmarks in Table 1 Panel A. We also consider two alternative ways of selecting the number of boosting iteration that could be used in real time, a point emphasized by Bai and Ng (2009). The first chooses the best model, i.e. the optimal number of boosting iterations, recursively through time. Thus, at time t, the number of boosting iterations is only based on model performance up to time t. Second, we use forecast combinations as a way to lower the sensitivity of our results to the choice of B by using the simple average of the forecasts from regression trees with B =1, 2, ..., 10,000 boosting iterations. Both models beat the benchmarks with the combined average being particularly effective.

4.2 Volatility

Panel B of Table 1 presents forecasting accuracy results for volatility. BRT that uses both macroeconomic and financial variables as conditioning information performs best with an R^2 of 40.84%. The "Semi-Parametric MIDAS" specification we propose has an R^2 of 34.95%, which is comparable to the 35.69% of MIDAS (beta). Garch (1,1) has the worst performance as its R^2 is only 14.13%.[13]

In Panel B we also report the results for directional accuracy. Directional volatility is not a natural concept because volatility is always positive, so we follow Marquering and Verbeek (2005) and define as high/low volatility those periods characterized by a volatility greater/smaller than the full-sample median volatility. Relative to the other models, GARCH (1,1) has a rather poor directional forecasting accuracy with a hit ratio of only 62.31%. This is expected, given its rather slow mean reverting nature. MIDAS (Beta), Semi-Parametric MIDAS and BRT perform very well: the directional accuracy is above 75% for all of them. With a hit ratio of 77.02% MIDAS (beta) is the best model, followed by Semi-Parametric MIDAS and BRT that have a directional accuracy of 76.40% and 75.98%, respectively.

In the middle panel of Figure 4 we present the results for a modified "CumSum analysis" defined as

$$CumSum(T) = \sum_{t=\tau}^{T-1} \left(\widehat{\sigma}_{t+1|t}^{GARCH(1,1)} - \sigma_{t+1}\right)^2 - \left(\widehat{\sigma}_{t+1|t}^{model} - \sigma_{t+1}\right)^2, \tag{22}$$

where GARCH (1,1) now replaces the prevailing mean in the definition and acts as the benchmark.[14]The models compared are BRT, MIDAS (Beta) and semi-parametric MIDAS. A close examination of the plot reveals that semi-parametric MIDAS performs better than what transpires from the results reported in Panel B of Table 1. BRT and semi-parametric MIDAS perform equally well on most of the sample, except for the period associated with the recent financial crisis where BRT dominates. MIDAS (Beta), instead, performs rather poorly on most of the sample, but predicts very well volatility during the current financial crisis.

The middle panel of Figure 5 reports the out-of-sample directional accuracy difference between BRT, Semi-Parametric MIDAS, MIDAS (Beta) and GARCH (1,1) that we employ as a benchmark. The vertical axis counts the differential performance between the prediction model and the benchmark in terms of correct volatility directional forecasts. The plot shows that the BRT model outperforms its competitors on most of the sample only to lose its lead in the last

[13]Conducting the analysis until 2005 gives stronger results for the methods proposed here as the R^2's of BRT, Semi-Parametric MIDAS, MIDAS(beta) and Garch (1,1) are 35.51%, 32.07%, 23.90% and 7.69%, respectively.

[14]For the predictive accuracy of GARCH(1,1) compared to other ARCH-type models, see Hansen and Lunde (2005).

few years. MIDAS (Beta) and Semi-Parametric MIDAS display a virtually identical performance throughout the test period.

In unreported results we show that the performance of boosted regression trees is not sensitive to the choice of boosting iterations. BRT with 5,000 and 15,000 boosting iterations outperforms the benchmarks reported in Table 1 Panel B. Also, the combined average model and the best model selected recursively beat the benchmarks with the latter being particularly effective.

4.3 Weight Prediction

We report the results for BRT-based optimal weight predictions in Panel C of Table 1 and we compare it to the prevailing mean, as no other empirical framework has been proposed until now in such context. BRT has an R^2 of 0.51% and does significantly better than the prevailing mean, whose R^2 is -6.24%. In terms of directional accuracy, the performance of BRT is superior to that of the prevailing mean with a performance differential of 1.0% (56.11% compared to 55.26%).

The bottom panel of Figure 4 presents estimates of

$$CumSum(T) = \sum_{t=\tau}^{T-1} (\widehat{w}_{t+1|t}^{pm} - w_{t+1}^{real})^2 - (\widehat{w}_{t+1|t}^{BRT} - w_{t+1}^{real})^2. \tag{23}$$

The plot confirms that boosted regression trees consistently outperform the prevailing mean throughout the sample. Finally, the bottom panel of Figure 5 plots the cumulative out-of-sample directional accuracy difference between BRT and the prevailing direction that we use as a benchmark. The plot highlights that while a great deal of directional accuracy was present in the 1973-1976 period, over the rest of the sample boosted regression trees have performed as well as the prevailing mean in terms of directional accuracy.

Overall, the results reported in this section highlight the superiority of BRT compared to the established benchmarks at predicting the equity premium and volatility. The results hold whether mean squared error or directional accuracy is employed as an evaluation criterion. As far as the optimal portfolio allocation is concerned, our preliminary results suggest that BRT performs well in terms of mean squared error, but the evidence for directional accuracy is not convincing.

In order to draw more precise conclusions on the predictive accuracy of BRT in the context of financial markets prediction, we conduct next a number of formal tests of return and volatility timing that have been proposed in the literature.

4.4 Formal Tests of Predictive Accuracy

We employ three tests of return and volatility timing that have been widely used in the literature. The first is designed to test for directional accuracy only, while the second and third focus on magnitudes as well.

The first test we employ is the Henriksson and Merton (1981) (HM) test, which is motivated by the idea that a forecasting framework has directional accuracy if the sum of the conditional probabilities of forecasting correctly the target variable when it takes positive and negative values is greater than one: i.e. if $p_1 + p_2 = 1$, where $p_1 = Pr(\hat{y}_t > 0|y_t > 0)$, $p_2 = Pr(\hat{y}_t < 0|y_t < 0)$, $\hat{y}$ is the forecast and y is the target variable. It is easy to show that the non-parametric test proposed by Henriksson and Merton (1981) is asymptotically equivalent to a one-tailed test on the significance of the coefficient α_1 in the regression

$$I_{\{\hat{\mu}_{t+1|t}>0\}} = \alpha_0 + \alpha_1 I_{\{r_{t+1}>0\}} + \epsilon_{t+1}, \tag{24}$$

where r_{t+1} and $\hat{\mu}_{t+1|t}$ are the realized and predicted market excess returns at time $t+1$ and $I_{\{\cdot\}}$ represents an indicator function.

The second test we employ is the one developed by Cumby and Modest (1987)(CM). The CM test is constructed following the intuition that it is more important to correctly predict the direction of positive or negative returns whenever they are large in magnitude. The coefficient of interest is α_1 in the regression specification

$$r_{t+1} = \alpha_0 + \alpha_1 I_{\{\hat{\mu}_{t+1|t}>0\}} + \epsilon_{t+1}. \tag{25}$$

The null is that α_1 is zero and is tested against the alternative of it being greater than zero. Finally, following Bossaerts and Hillion (1999)(BH) that employ a Mincer-Zarnowitz-type test, we run the regression:

$$r_{t+1} = \alpha_0 + \alpha_1 \hat{\mu}_{t+1|t} + \epsilon_{t+1} \tag{26}$$

and test for the significance of the α_1 coefficient. The values and t-tests of the α_1 coefficients for the tests described above are reported in Table 2 for the sample 1969-2008. Panel A reports the results for stock returns. The HM test indicates that BRT is able to forecast market positive and negative excess returns at a statistically significant level, while the timing of the linear model is negative and not significant. The results for the CM and BH tests are similar. The null of no-predictive ability is rejected for BRT for both tests, while we fail to reject the null hypothesis for the linear model. The results are in line with the findings displayed in Table 1 as BRT has a rather strong forecasting accuracy in terms of directional accuracy and mean-squared error, while the opposite holds true for the linear model selected recursively using BIC.

We conduct the same tests for volatility, *mutatis mutandis*, in Panel B of Table 2. The results are more clear-cut in this context as all four models analyzed have a significant predictive ability across all tests and samples. This should come at no surprise as volatility is easier to predict than returns and volatility models have been used for many years because of their forecasting accuracy. Still, it is worth noting that the BRT model has higher and more significant α_1 coefficients compared to its competitors for the CM and BH tests, precisely those tests where not only the sign of realized volatility, but also its magnitude matters. In the BH test, BRT is followed by Semi-Parametric MIDAS, MIDAS (Beta) and GARCH (1,1),[15] while in the CM test the rank of the MIDAS (Beta) and Semi-Parametric MIDAS is inverted. Finally, the results of the HM test confirm the superiority of the MIDAS (Beta) model in terms of directional accuracy displayed in Table 1.

In Panel C of Table 2 we report the market timing tests for the optimal portfolio weights. In accordance with the rather weak directional accuracy results of Table 1, BRT does not display significant directional accuracy according to the HM test. In the tests where not only directional accuracy but also magnitudes matter (CM and BH tests) BRT displays instead significant market timing, possibly implying profitable portfolio allocations.

The results for return, volatility and optimal weight prediction highlight an overall better performance of BRT compared to the frameworks currently employed in the literature. Among the volatility models under consideration, the BRT model that exploits information from macroeconomic and financial time-series displays a better performance than the semi-parametric MIDAS model that conditions volatility forecasts on the basis of past returns only. In turn, semi-parametric MIDAS performs better than its parametric counterpart. We draw two important conclusions from these results. First, at the monthly horizon our semi-parametric specification leads to a more effective use of the conditioning information contained in lagged squared returns, compared to its parametric counterpart. Second, incorporating economic information leads to more precise volatility forecasts compared to the models that exploit only the time series of returns. We analyze next whether the predictive accuracy presented here translates into profitable portfolio allocations for the mean-variance investor.

[15] We also consider an alternative specification for the BH test that includes the lagged realized volatility in an effort to control for volatility's persistence, i.e. we estimate a linear model of the form: $\sigma_{t+1} = \alpha_0 + \alpha_1 \hat{\sigma}_{t+1|t} + \alpha_2 \sigma_t + \epsilon_{t+1}$. BRT, MIDAS(Beta) and Semi-Parametric MIDAS maintain their significance, while GARCH(1,1) does not.

4.5 Mean-Variance Investor and Portfolio Allocation

We construct portfolio weights for the mean-variance investor described in Section 2 who uses boosted regression trees to formulate return and volatility predictions.[16]

We have a free parameter γ that determines the investor's degree of risk-aversion. We set it to 4 as that corresponds to a moderately risk-averse investor. We first present the results for unconstrained portfolio weights and evaluate the performance of two models. In the first, market returns and volatility are forecasted separately and the optimal mean-variance portfolio allocation is computed in a second stage (two-step BRT model). In the second, the optimal portfolio allocation is forecasted directly following the procedure described in Section 3 (one-step BRT model).

We compare our results to three benchmark strategies. The first uses stock returns predictions from a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2 and volatility predictions from a GARCH (1,1) model. The second model replaces the GARCH (1,1) model with a MIDAS (Beta) model (Ghysels et al (2005)). The third benchmark is a passive investment strategy that allocates 100% of wealth in the market portfolio at all times.

In Panel A of Table 3 we report the mean and standard deviation of returns for each investment strategy. We also report the Sharpe Ratio, the Jensen's *Alpha* measure of abnormal returns, i.e. the coefficient and t-statistic for the α_0 coefficient in the regression

$$r_{p,t+1} = \alpha_0 + \alpha_1 r_{t+1} + \epsilon_{t+1}, \tag{27}$$

where $r_{p,t+1}$ is the return on the investment portfolio and r_{t+1} is the return on the market portfolio. Finally, we report the Treynor-Mazuy (TM) test statistic, i.e. the coefficient and t-statistic for the α_2 coefficient in the regression

$$r_{p,t+1} = \alpha_0 + \alpha_1 \ r_{t+1} + \alpha_2 \ r_{t+1}^2 + \epsilon_{t+1}. \tag{28}$$

The performance is evaluated over the sample 1969-2008. The best model is the BRT that forecasts the optimal portfolio allocation in two steps with a monthly Sharpe ratio of 15.81%. The second best is the one-step BRT model with a Sharpe ratio of 14.81%. The two-step BRT model has a higher and more significant Jensen's *Alpha* measure of market outperformance, but a lower and less significant Treynor-Mazuy (TM) measure of market timing compared to the one-step BRT model. The benchmark models perform considerably worse. The two active

[16] The performance of the portfolio allocations that use semi-parametric MIDAS volatility forecasts are not reported for brevity.

benchmark strategies have a negative and insignificant Jensen's *Alpha*: -0.072% for the model that uses a GARCH (1,1) and -0.069% for the model that uses a MIDAS (Beta) specification. The Treynor-Mazuy measure of market timing is positive but insignificant for both models. The Sharpe ratios, 2.70% and 3.18% respectively, are smaller than that of the passive strategy (Panel D) that invests 100% of wealth in the market portfolio (7.71%).

In Panel B of Table 3 we repeat the same exercise imposing short-selling and borrowing constraints. As expected, the constraints reduce the profitability of both BRT-based investment strategies. The Sharpe ratio of the two-step BRT model drops by approximately 1.8% to 14.01%, while the Sharpe ratio of the one-step BRT model drops by approximately 2.2% to 12.59%. The lower Sharpe ratios are due to both lower mean returns and volatility of the constrained portfolio allocations compared to the unconstrained ones. The magnitudes of both Jensen's *Alpha*'s and TM market timing measures are lower than their unconstrained counterparts, but they maintain a strong level of statistical significance. The opposite holds true for the two active benchmark strategies. The Sharpe ratio for the investment strategy that exploits GARCH (1,1) volatility forecasts increases by 2.8% to 5.52% and the one for the MIDAS (Beta) increases by 3.5% to 6.69%. There are a number of reasons for this, the most important being that there is a fair degree of estimation error in the estimated optimal portfolio allocations, particularly so for the linear regression, GARCH (1,1) and MIDAS (Beta) models that minimize L^2 criterion functions. We show in section 4.7 that accounting for estimation uncertainty improves the performance of the investment strategies.

Table 3 Panel C reports results for models that exploit only the predictability of stock returns: i.e. the investment strategy entails investing 100% of wealth in the risky asset if expected excess returns are greater than zero and 0% otherwise. For the one-step BRT model the decision is based on the sign of the expected Sharpe ratio rather than that of expected returns. For BRT, the portfolio allocations that do not exploit volatility predictions are less profitable than those that do. The Sharpe ratio of the two-step BRT model drops by 1.2% to 12.83% and the one for the one-step BRT model drops by 1.4% to 11.18%. The results for the BRT models indicate that the investment strategies that explicitly model volatility lead to more profitable portfolio allocations. The opposite holds true the two active benchmark strategies as the Sharpe ratio of the linear model that does not exploit volatility forecasts is 8.09%, higher than the 5.52% and 6.69% obtained by the models that use GARCH (1,1) and MIDAS (Beta) volatility forecasts, respectively.

Overall, this section highlights that BRT models outperform the established benchmarks in term of constrained and unconstrained portfolio allocation performance. They also show that, for BRT models, portfolio allocations based on both conditional returns and volatility estimates are more profitable than those that exploit return predictability only. Next, we focus our attention on BRT models with transaction costs and show that the outperformance of our framework is robust to the inclusion of such frictions.

4.6 Portfolio Allocation Performance with Transaction Costs

In Table 4 we report the same estimates of Table 3 for the boosted regression tree models, while accounting for the effect of transaction costs. Following Marquering and Verbeek (2005), we impose the assumption that transaction costs "τ" are equal to percentage points of the value traded. The wealth of the agent at time $t+1$ "W_{t+1}" is equal to

$$\begin{aligned} W_{t+1} &= W_t r_{p,t+1} - \tau W_t |\Delta w_{t+1}| \\ &= W_t \left(r_{p,t+1} - \tau |\Delta w_{t+1}| \right), \end{aligned}$$

where $|\Delta w_{t+1}|$ is the absolute change in the portfolio weights between t and $t+1$ and $r_{p,t+1}$ is the corresponding portfolio return. The returns net of transaction costs are therefore $r_{p,t+1} - \tau|\Delta w_{t+1}|$. We consider two scenarios: low (0.1%) and high (0.5%) transaction costs.

The investment strategies display positive and significant Jensen's *Alpha* and TM market timing when low transaction costs are considered. This holds for the two-step BRT model (Panel A) as well as the one-step BRT model (Panel B). The two-step models display a better performance with Sharpe ratios of 14.33%, 13.28% and 12.24% for the unconstrained, constrained and "return predictability only" investment strategies. The one-step BRT counterparts are 13.77%, 12.07% and 10.73%. The investment performance deteriorates slightly in the presence of high transactions costs: all market timing measures are positive and significant, the Jensen's *Alpha*'s are positive, but not significant, and the Sharpe ratios are greater than the market Sharpe ratio (7.71%). Our findings highlight that the investment strategies based on BRT forecasts remain profitable even in the presence of transaction costs. The results are generally strong if we consider that the transaction costs are not present in the maximization problem that determines the portfolio weights. This is particularly true for the unconstrained models, whose weights can be rather extreme and volatile and therefore vulnerable to the introduction of unexpected market frictions. We conduct next further analyses on the economic value generated by the trading strategies based on BRT forecasts.

4.7 Economic Value of the Investment Strategies

Even though very popular and widely used, the performance measures employed until now are affected by a number of shortcomings. As argued by Fleming, Kirby, and Ostdiek (2001) and

Marquering and Verbeek (2005), the Sharpe ratio is not a correct measure of risk-adjusted returns in the presence of time-varying volatility and Jensen's *Alpha* is a biased estimator of abnormal returns in the presence of time-varying portfolio weights. Also, the TM measure provides an indication of whether the portfolio exhibits market timing, but it does not quantify the economic value it generates .

Following Marquering and Verbeek (2005) and Fleming, Kirby, and Ostdiek (2001), we employ a utility-based performance measure based on the investor's (ex-ante) utility function

$$U_t(\cdot) = E_t\{r_{p,t+1}\} - \frac{1}{2}\gamma Var_t\{r_{p,t+1}\}. \tag{29}$$

By using the definition of $r_{p,t+1}$ we can re-write the expression above as:

$$U_t(\cdot) = r_{f,t+1} + w_{p,t+1}E_t\{r_{t+1} - r_{f,t+1}\} - \frac{1}{2}\gamma w_{p,t+1}^2 Var_t\{r_{t+1} - r_{f,t+1}\}. \tag{30}$$

This allows the construction of an (ex-post) average utility level for each investment strategy using the following formula

$$U_p(\cdot) = \frac{1}{T}\sum_{t=0}^{T-1}\left[r_{f,t+1} + w_{p,t+1}r_{t+1} - \frac{1}{2}\gamma w_{p,t+1}^2\sigma_{t+1}^2\right], \tag{31}$$

where r_{t+1} is the actual return on the market between t and $t+1$ and σ_{t+1}^2 is the variance of the market returns estimated using daily observations. The average utility level is interpreted as the certain return, in percentage per month, that provides the same utility to the investor as the risky strategy. Also, the difference between the average realized utilities delivered by two portfolio allocation strategies can be interpreted as the maximum monthly fee an agent would be willing to pay to hold one instead of the other.

Another issue that has been ignored until now is that of estimation uncertainty. Expected returns and volatilities are not known but estimated from data. Even though we used the longest possible time-series available, noise may be responsible for making the portfolio weights too volatile over time. A number of approaches have been proposed to mitigate the effect of estimation uncertainty, see, for example, Ter Horst, De Roon, and Werker (2000) and Maenhout (2004). Here, for simplicity, we follow Marquering and Verbeek (2005) and estimate the optimal portfolio weights using a "pseudo-risk aversion coefficient" that is double the true risk-aversion coefficient of the agent.

In Table 5 we report the results for the average realized utilities without accounting for estimation uncertainty (Panel A) and accounting for estimation uncertainty (Panel B). Panel C presents the average realized utilities for an investor that places 100% of his wealth in the market portfolio at all times. The results are presented for coefficients of risk-aversion ranging from 2 to 10. The constrained investment strategies and the ones that are based on return predictability only deliver greater average realized utilities compared to the passive benchmark strategy for all levels of risk-aversion. The results hold whether we use one or two-step BRT models and whether we correct for estimation uncertainty or not. The unconstrained investment strategies are more interesting as the portfolio allocations that are not corrected for estimation uncertainty deliver lower realized utilities than the benchmark strategy for risk-aversion coefficients below 5. The portfolio allocations that correct for estimation uncertainty provide instead very high average realized utilities for low coefficients of risk-aversion. For example, the unconstrained investment strategy based on a two-step BRT model that corrects for estimation uncertainty delivers an average monthly realized utility of 1.19% to an agent with coefficient of risk-aversion equal to 2. The average realized utility for the one-step counterpart is 0.90%. Both are much larger than the 0.60% delivered by the passive strategy. Overall, our results indicate that unconstrained portfolio weights that do not account for estimation uncertainty result in too volatile portfolio allocations.

Finally, we compute the maximum monthly portfolio fees that a representative investor would be willing to pay in order to have his wealth invested in one of the investment strategies we propose as opposed to holding the market portfolio. In Figure 6 we present the results for the one and two-step BRT models that account for estimation uncertainty. The top panel presents the break-even monthly portfolio fees for the unconstrained, constrained and "return predictability only" strategies for the two-step BRT models with zero, low and high transaction costs (left, middle and right plot respectively). The bottom panel repeats the exercise for the one-step BRT model. Each figure reports the break-even portfolio fees (in percentage) on the y-axis and the risk-aversion coefficient (ranging from 1 to 10) on the x-axis. When transaction costs are not accounted for, the monthly break-even portfolio fees for the two-step BRT models are quite sizable and range from 0.47% to 1.20% for the unconstrained investment strategy, from 0.22% to 0.91% for the constrained investment strategy and from 0.18% to 0.47% for the investment strategy that exploits only the predictability in returns. Similar, but somewhat lower, are the results for the one-step BRT model. The results remain very strong and are almost unaffected by low transaction costs, but they deteriorate for high transaction costs. For example, an investor with a coefficient of risk aversion equal to 2 would rather hold the market portfolio than the unconstrained investment strategy formulated using the one or two-step BRT models. All the other risk-aversion, transaction costs and investment strategy combinations deliver positive and sizable monthly break-even portfolio fees, implying that the investment strategies proposed are economically valuable.

5 Conclusions

We present new evidence on the predictability of stock returns and volatility at the monthly frequency using Boosted Regression Trees (BRT). BRT is a novel semi-parametric statistical method that generates forecasts on the basis of large sets of conditioning information without imposing strong parametric assumptions such as linearity or monotonicity. Our forecasts outperform those generated by benchmark models in terms of both mean squared error and directional accuracy. They also generate profitable portfolio allocations for mean-variance investors even when market frictions are accounted for. Finally, our analysis shows that the relation between the predictor variables constituting the conditioning information set and the investors' optimal portfolio allocation to risky assets is, in most cases, nonlinear and nonmonotonic.

References

Ait-Sahalia, Y., and M. Brandt (2001): "Variable Selection for Portfolio Choice," *The Journal of Finance*, 56(4), 1297–1351.

Bai, J., and S. Ng (2009): "Boosting diffusion indices," *Journal of Applied Econometrics*, 24(4), 607–629.

Baker, M., and J. Wurgler (2000): "The Equity Share in New Issues and Aggregate Stock Returns," *The Journal of Finance*, 55(5), 2219–2257.

Bossaerts, P., and P. Hillion (1999): "Implementing statistical criteria to select return forecasting models: what do we learn?," *Review of Financial Studies*, 12(2), 405–428.

Breen, W., L. Glosten, and R. Jagannathan (1989): "Economic significance of predictable variations in stock index returns," *Journal of Finance*, 44(5), 1177–89.

Breiman, L. (1984): *Classification and regression trees Regression trees The Wadsworth statistics/probability series*. Wadsworth International Group.

——— (1996): "Bagging predictors," *Machine learning*, 24(2), 123–140.

Campbell, J. (1988): "Stock Returns and the Term Structure," *NBER Working Paper*.

Campbell, J., and R. Shiller (1988a): "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies*, 1(3), 195–228.

Campbell, J., and R. Shiller (1988b): "Stock prices, earnings, and expected dividends," *Journal of Finance*, 43(3), 661–676.

Campbell, J., and M. Yogo (2006): "Efficient tests of stock return predictability," *Journal of Financial Economics*, 81(1), 27–60.

Campbell, J. Y. (1987): "Stock returns and the term structure," *Journal of Financial Economics*, 18(2), 373–399.

Campbell, J. Y., and S. Thompson (2008): "Predicting excess stock returns out of sample: Can anything beat the historical average?," *Review of Financial Studies*, 21(4), 1509–1531.

Cumby, R., and D. Modest (1987): "Testing for Market Timing Ability: A Framework for Forecast Evaluation," *Journal of Financial Economics*, 19(1), 169–189.

Dangl, T., and M. Halling (2008): "Predictive regressions with time-varying coefficients," *Working Paper, Vienna University of Technology.*

Engle, R., E. Ghysels, and B. Sohn (2006): "On the Economic Sources of Stock Market Volatility," *Manuscript, New York University.*

Engle, R., and J. Rangel (2005): *The Spline GARCH Model for Unconditional Volatility and Its Global Macroeconomic Causes*. Czech National Bank.

Fama, E., and K. French (1988): "Dividend yields and expected stock returns," *Journal of Financial Economics*, 22(1), 3–25.

Fama, E., and K. French (1989): "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics*, 25(1), 23–49.

Fama, E., and G. Schwert (1977): "Asset Returns and Inflation," *Journal of Financial Economics*, 5(2), 115–146.

Fleming, J., C. Kirby, and B. Ostdiek (2001): "The Economic Value of Volatility Timing," *Journal of Finance*, 56, 329–352.

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

——— (2002): "Stochastic gradient boosting," *Computational Statistics and Data Analysis*, 38(4), 367–378.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005): "There is a Risk-Return Tradeoff After All," *Journal of Financial Economics*, 76(3), 509–548.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006): "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, 131(1-2), 59–95.

Goyal, A., and I. Welch, I. (2003): "Predicting the Equity Premium with Dividend Ratios," *Management Science*.

Hansen, P., and A. Lunde (2005): "A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)?," *Journal of Applied Econometrics*, 20(7), 873–889.

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

Henriksson, R., and R. Merton (1981): “On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills,” *Journal of Business*, 54(4), 513.

Hodrick, R. (1992): “Dividend yields and expected stock returns: alternative procedures for inference and measurement,” *Review of Financial Studies*, 5(3), 357–386.

Johannes, M., A. Korteweg, and N. Polson (2009): “Sequential learning, predictive regressions, and optimal portfolio returns,” *Mimeo, Columbia University*.

Keim, D., and R. Stambaugh (1986): “Predicting returns in the stock and bond markets,” *Journal of Financial Economics*, 17(2), 357–390.

Kothari, S., and J. Shanken (1997): “Book-to-market, dividend yield, and expected market returns: A time-series analysis,” *Journal of Financial Economics*, 44(2), 169–203.

Lamont, O. (1998): “Earnings and Expected Returns,” *The Journal of Finance*, 53(5), 1563–1587.

Leitch, G., and J. Tanner (1991): “Economic Forecast Evaluation: Profits Versus the Conventional Error Measures,” *American Economic Review*, 81(3), 580–590.

Lettau, M., and S. Ludvigson (2009): “Measuring and Modeling Variation in the Risk-Return Tradeoff,” *Handbook of Financial Econometrics, edited by Y. Ait-Sahalia and L.P. Hansen. North Holland.*

Lewellen, J., S. Nagel, and J. Shanken (2010): “A skeptical appraisal of asset pricing tests,” *Journal of Financial Economics*, 96(2), 175–194.

Lo, A., and A. MacKinlay (1990): “Data-snooping biases in tests of financial asset pricing models,” *Review of Financial Studies*, 3(3), 431–467.

Ludvigson, S. C., and S. Ng (2007): “The empirical risk-return relation: A factor analysis approach,” *Journal of Financial Economics*, 83(1), 171–222.

Maenhout, P. (2004): “Robust Portfolio Rules and Asset Pricing,” *Review of Financial Studies*, 17(4), 951–983.

Marquering, W., and M. Verbeek (2005): “The Economic Value of Predicting Stock Index Returns and Volatility,” *Journal of financial and Quantitative Analysis*.

Nelson, C., and M. Kim (1993): “Predictable stock returns: The role of small sample bias,” *Journal of Finance*, 48(2), 641–661.

Paye, B. (2010): "Do Macroeconomic Variables Predict Aggregate Stock Market Volatility?," *Working Paper Series*.

Pesaran, H., and A. Timmermann (2005): "Real-time econometrics," *Econometric Theory*, 21(01), 212–231.

Pesaran, M., and A. Timmermann (1995): "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance*, 50(4), 1201–1228.

Pesaran, M., and A. Timmermann (2000): "A Recursive Modelling Approach to Predicting UK Stock Returns 'I. Working; Paner No. 9625. Department of Applied Economics, Universitv of Cambridge," *Economic Journal*, 110, Issue 460.

Pontiff, J., and L. Schall (1998): "Book-to-market ratios as predictors of market returns," *Journal of Financial Economics*, 49(2), 141–160.

Rapach, D., J. Strauss, and G. Zhou (2010): "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy," *Review of Financial Studies*, 23(2), 821.

Rossi, A., and A. Timmermann (2010): "What is the Shape of the Risk-Return Relation?," *UCSD Discussion Paper*.

Rozeff, M. (1984): "Dividend yields are equity risk premiums," *Journal of Portfolio Management*, 11(1), 68–75.

Stambaugh, R. (1999): "Predictive Regressions," *Journal of Financial Economics*, 54(3), 375–421.

Sullivan, R., A. Timmermann, and H. White (1999): "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap," *The Journal of Finance*, 54(5), 1647–1691.

Ter Horst, J., F. De Roon, and B. Werker (2000): "Incorporating Estimation Risk in Portfolio Choice," *CentER Working Papers Series*.

Welch, I., and A. Goyal (2008): "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21(4), 1455–1508.

Table 1. Out-of-Sample Forecasting Accuracy

A. Return Models

	L^2 Norm	Directional Accuracy
	$\mathbf{R^2}$	**% Accuracy**
BRT	0.30%	57.35%
Prevailing Mean	-0.79%	55.26%
Linear Model	-2.28%	52.17%

B. Realized Volatility Models

	L^2 Norm	Directional Accuracy
	$\mathbf{R^2}$	**% Accuracy**
BRT	40.84%	75.98%
Semi-Par. MIDAS	34.95%	76.40%
MIDAS (Beta)	35.69%	77.02%
GARCH (1,1)	14.13%	62.31%

C. Optimal Weight Prediction Models

	L^2 Norm	Directional Accuracy
	$\mathbf{R^2}$	**% Accuracy**
BRT	0.52%	56.11%
Prevailing Mean	-6.24%	55.26%

This table reports the out-of-sample R^2 and directional accuracy for a boosted regression tree model with 10,000 boosting iterations used to forecast monthly stock returns, realized volatility and the optimal portfolio weight in the risky asset for the mean-variance investor. For return prediction (Panel A) we also present results for the prevailing mean model proposed by Welch and Goyal (2008) and a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2 of the paper. For volatility prediction (Panel B) we report the results for a semi-parametric MIDAS specification described in Section 2 of the paper, a MIDAS (Beta) model (Ghysels et al (2005)) and a GARCH (1,1). For optimal portfolio allocation prediction (Panel C) we also report the results for the prevailing mean. The out-of-sample performance is computed over the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Table 2. Tests of Out-Of-Sample Forecasting Accuracy

A. Return Models

	HM Test	CM Test	BH Test
BRT	0.106 (2.53)	0.010 (2.33)	0.537 (2.85)
Linear Model	-0.017 (-0.46)	0.003 (0.56)	0.148 (0.56)

B. Volatility Models

	HM Test	CM Test	BH Test
BRT	0.519 (13.48)	0.022 (11.65)	1.181 (19.69)
Semi-Par. MIDAS	0.528 (13.65)	0.021 (10.85)	1.141 (16.55)
MIDAS (Beta)	0.541 (14.16)	0.021 (11.11)	1.041 (16.39)
GARCH (1,1)	0.247 (6.55)	0.014 (5.85)	0.845 (9.84)

C. Models for the Optimal Portfolio Weight

	HM Test	CM Test	BH Test
BRT	0.059 (1.62)	0.009 (1.84)	0.537 (2.86)

This table reports forecasting accuracy tests for a boosted regression tree model with 10,000 boosting iterations used to forecast monthly stock returns in Panel A, realized volatility in Panel B and the optimal portfolio weight in the risky asset for the mean-variance investor in Panel C. Each panel presents the results for three tests: the Henriksson and Merton (HM), the Cumbey and Modest (CM) and the Bossaerts and Hillion (BH) test. We report the coefficient of interest, α_1, whose significance determines the directional predictive ability of each forecasting framework. The numbers in brackets are t-test values. For comparison, in Panel A we also present results for a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2 of the paper; in Panel B we report the results for a semi-parametric MIDAS specification described in Section 2 of the paper, a MIDAS (Beta) model (Ghysels et al (2005)) and a GARCH (1,1). The tests are performed for the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Table 3. Portfolio Allocation Performance

A. Unconstrained Weights

	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
Two-Step BRT	2.10%	10.19%	15.81%	1.37% (3.09)	4.484 (3.93)
One-Step BRT	2.02%	10.38%	14.81%	1.12% (2.77)	4.939 (4.79)
Linear Model & GARCH (1,1)	0.62%	4.84%	2.70%	-0.07% (-0.35)	0.397 (0.81)
Linear Model & MIDAS (Beta)	0.69%	6.35%	3.18%	-0.07% (-0.31)	0.213 (0.34)

B. Short-Selling and Borrowing Constraints

	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
Two-Step BRT	0.93%	3.17%	14.01%	0.26% (2.77)	0.931 (3.90)
One-Step BRT	0.93%	3.57%	12.59%	0.22% (2.45)	0.825 (3.54)
Linear Model & GARCH (1,1)	0.64%	2.86%	5.52%	-0.02% (-0.21)	0.039 (0.19)
Linear Model & MIDAS (Beta)	0.69%	3.07%	6.69%	0.01% (0.17)	0.126 (0.58)

C. Return Predictability Only

	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
Two-Step BRT	0.94%	3.55%	12.83%	0.24% (2.44)	0.729 (2.89)
One-Step BRT	0.91%	3.78%	11.18%	0.18% (1.93)	0.741 (3.15)
Linear Model	0.80%	3.91%	8.09%	0.05% (0.60)	0.070 (0.31)

D. Passive Strategy

	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
100% Market	0.83%	4.45%	7.71%	—	—

This table reports the performance of an investment portfolio based on a boosted regression tree model with 10,000 boosting iterations. The first row of each panel presents the results obtained by forecasting stock market returns and volatility separately and computing the optimal mean-variance portfolio allocation for an agent with risk-aversion coefficient of 4 (two-step BRT model). In the second row we forecast the optimal portfolio allocation directly on the basis of the conditioning variables listed in Section 2 of the paper (one-step BRT model). For comparison, we also present results for investment portfolios constructed using stock return predictions from a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables described in Section 2 of the paper and volatility predictions from alternatively a GARCH (1,1) or a MIDAS (Beta) model (Ghysels et al (2005)). For each investment portfolio we report the average monthly return, standard deviation, Sharpe ratio as well as its Jensen's *Alpha* and Treynor-Mazuy market timing measure. In brackets are t-test values. Panel A reports the results for unconstrained portfolio allocations. Panel B imposes shortselling and borrowing constraints. Panel C exploits only return predictability. Panel D report results for a passive investment strategy that places 100% of wealth in the market portfolio. The performance is computed over the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Table 4. Portfolio Performance with Transaction Costs

A. Two-Step BRT

	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
			Low Transaction Costs		
Unconstrained	1.94%	10.19%	14.33%	1.22%	4.523
				(2.75)	(3.97)
Short. and Borr. Constraints	0.91%	3.16%	13.28%	0.23%	0.934
				(2.53)	(3.92)
Return Predictability Only	0.92%	3.55%	12.24%	0.22%	0.738
				(2.23)	(2.92)
			High Transaction Costs		
Unconstrained	1.34%	10.19%	8.35%	0.61%	4.679
				(1.37)	(4.11)
Short. and Borr. Constraints	0.81%	3.17%	10.34%	0.14%	0.947
				(1.52)	(3.97)
Return Predictability Only	0.84%	3.56%	9.87%	0.13%	0.771
				(1.36)	(3.05)

B. One-Step BRT

	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
			Low Transaction Costs		
Unconstrained	1.91%	10.37%	13.77%	1.00%	4.925
				(2.50)	(4.78)
Short. and Borr. Constraints	0.92%	3.57%	12.07%	0.20%	0.829
				(2.24)	(3.56)
Return Predictability Only	0.89%	3.78%	10.73%	0.16%	0.746
				(1.74)	(3.18)
			High Transaction Costs		
Unconstrained	1.48%	10.36%	9.57%	0.57%	4.866
				(1.42)	(4.74)
Short. and Borr. Constraints	0.84%	3.58%	9.98%	0.13%	0.846
				(1.41)	(3.63)
Return Predictability Only	0.82%	3.78%	8.91%	0.09%	0.767
				(0.98)	(3.27)

This table reports the performance of an investment portfolio based on a boosted regression tree model with 10,000 boosting iterations accounting for transaction costs. Panel A presents the results obtained by forecasting stock market returns and volatility separately and computing the optimal mean-variance portfolio allocation for an agent with risk-aversion coefficient of 4 (two-step BRT model), while accounting for low and high transaction costs. For each transaction costs level we report the performance of the unconstrained optimal portfolio, the optimal portfolio with shortselling and borrowing constraints as well as an investment strategy that invests 100% of wealth in the risky asset if expected excess returns are greater than zero and 0% of wealth in the risky asset otherwise. In Panel B we conduct the same exercise, but we forecast the optimal portfolio allocation directly on the basis of the conditioning variables listed in Section 2 of the paper (one-step BRT model). The tests are performed on the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Table 5. Average Realized Utilities

A. Without Accounting for Estimation Uncertainty

	$\gamma = 2$	$\gamma = 4$	$\gamma = 6$	$\gamma = 8$	$\gamma = 10$
			Two-Step BRT		
Unconstrained	0.10%	0.29%	0.36%	0.39%	0.41%
Short. and Borr. Constraints	0.81%	0.70%	0.62%	0.58%	0.54%
Returns Predictability Only	0.78%	0.63%	0.47%	0.31%	0.15%
			One-Step BRT		
Unconstrained	-0.95%	-0.23%	0.01%	0.13%	0.20%
Short. and Borr. Constraints	0.75%	0.63%	0.52%	0.43%	0.38%
Return Predictability Only	0.74%	0.56%	0.39%	0.22%	0.04%

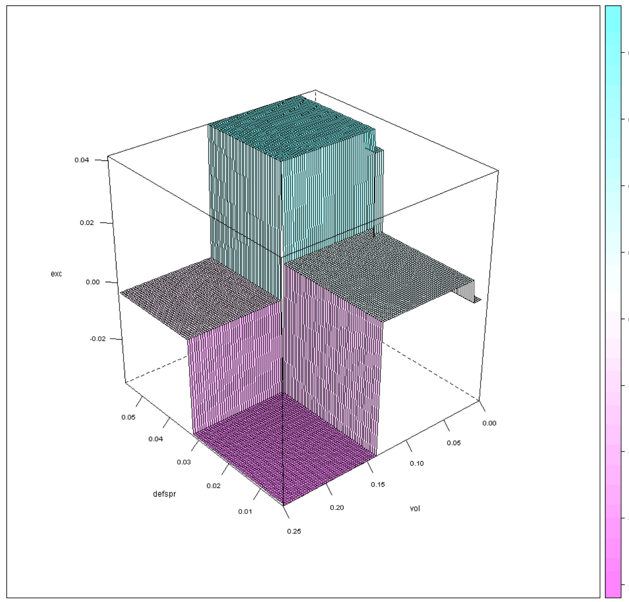
B. Accounting for Estimation Uncertainty

	$\gamma = 2$	$\gamma = 4$	$\gamma = 6$	$\gamma = 8$	$\gamma = 10$
			Two-Step BRT		
Unconstrained	1.19%	0.84%	0.72%	0.66%	0.63%
Short. and Borr. Constraints	0.82%	0.73%	0.66%	0.62%	0.60%
Return Predictability Only	0.78%	0.63%	0.47%	0.31%	0.15%
			One-Step BRT		
Unconstrained	0.90%	0.69%	0.62%	0.59%	0.57%
Short. and Borr. Constraints	0.78%	0.67%	0.60%	0.57%	0.55%
Return Predictability Only	0.74%	0.56%	0.39%	0.22%	0.04%

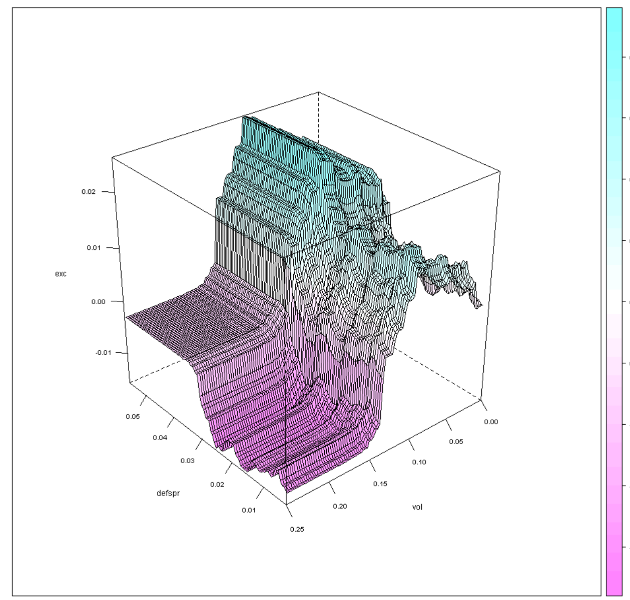
C. Passive Strategy

	$\gamma = 2$	$\gamma = 4$	$\gamma = 6$	$\gamma = 8$	$\gamma = 10$
100% Market	0.60%	0.37%	0.14%	-0.09%	-0.32%

This table reports the realized utility of an investment portfolio based on a boosted regression tree model with 10,000 boosting iterations for different coefficients of risk-aversion. The top rows of Panel A present the results obtained by forecasting stock market returns and volatility separately and computing the optimal mean-variance portfolio allocation in a second step (two-step BRT model). The bottom rows of Panel A present the results obtained by forecasting the optimal portfolio allocation directly on the basis of the conditioning variables listed in Section 2 of the paper (one-step BRT model). In Panel B we conduct the same exercise using a "pseudo risk-aversion coefficient" that doubles the investor's genuine risk-aversion γ to adjust portfolio weights for estimation uncertainty. For comparison, in Panel C we report results for a passive investment strategy that places 100% of wealth in the market portfolio. The analysis is conducted over the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.



(a) 3 Iterations



(b) 5000 Iterations

Figure 1: Fitted values of excess returns (exc) as a function of volatility (vol) and the default spread (defspr). Both plots are based on boosted regression trees with two terminal nodes. The panel on the left uses three boosting iterations, while the right panel uses 5,000 iterations. The scale for volatility has been inverted. The plots are based on monthly data from 1927-2008.

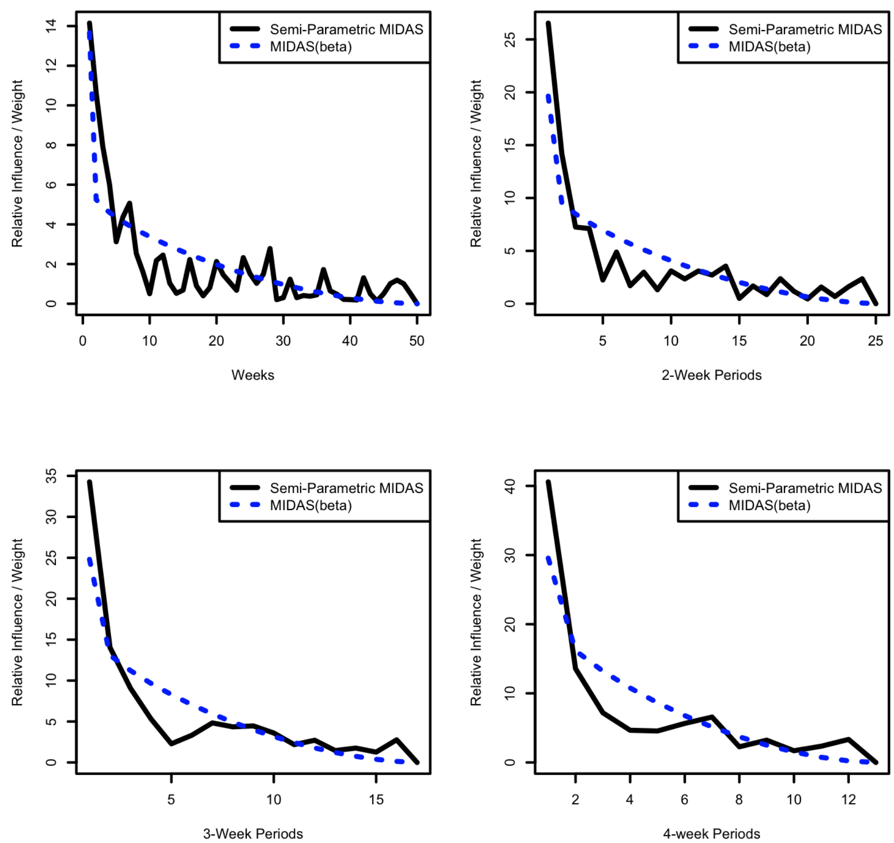


Figure 2: This figure presents the relative influence (in percent) of lagged squared returns at predicting monthly volatility for both the MIDAS (beta) model (Ghysels et al (2005)) and the Semi-Parametric MIDAS model. Monthly volatility is predicted using 250 lagged squared daily returns as conditioning information. For ease of interpretation, in each panel we aggregate individual days' relative influence in 1, 2, 3, and 4 week periods.

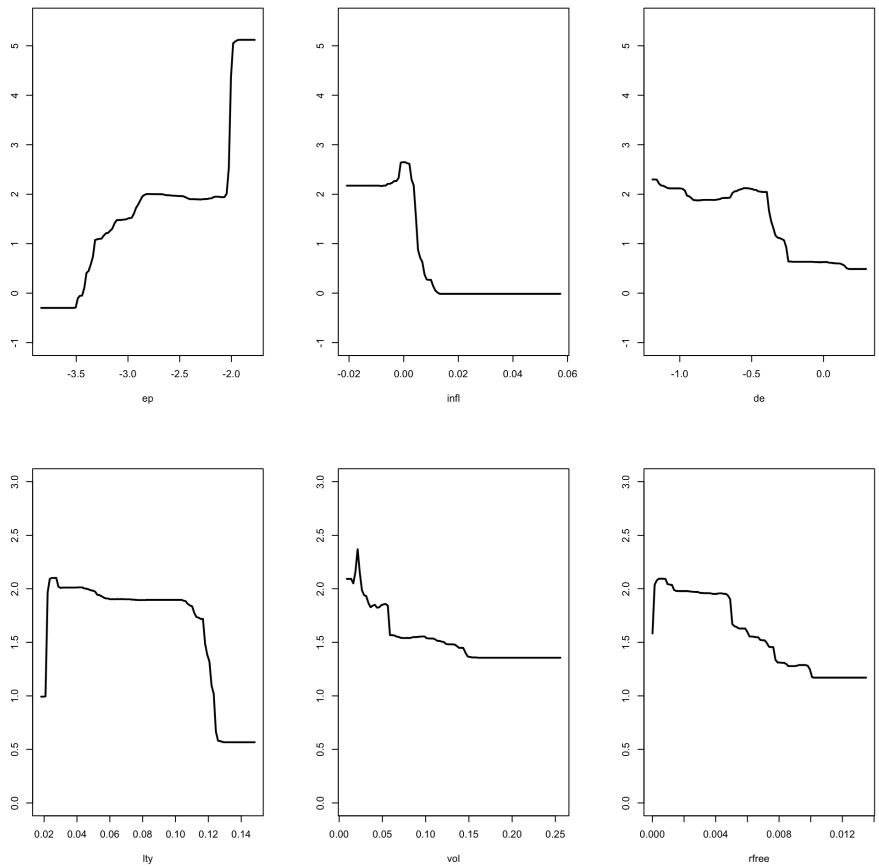


Figure 3: This figure presents the partial dependence plots for the optimal allocation to risky assets as a function of the six predictor variables with the highest relative influence over the sample 1927-2008. The most influential predictors are the log earnings price ratio (ep), inflation (infl), the log dividend earnings ratio (de), the yield on long term government bonds (lty), stock market volatility (vol) and the three-month T-bill rate (Rfree) that have a relative influence of 16.63%, 15.88%, 13.94%, 9.38%, 7.23% and 6.67%, respectively. The horizontal axis covers the sample support of each predictor variable, while the vertical axis tracks the change in the optimal weight in the market portfolio as a function of the individual predictor variables.

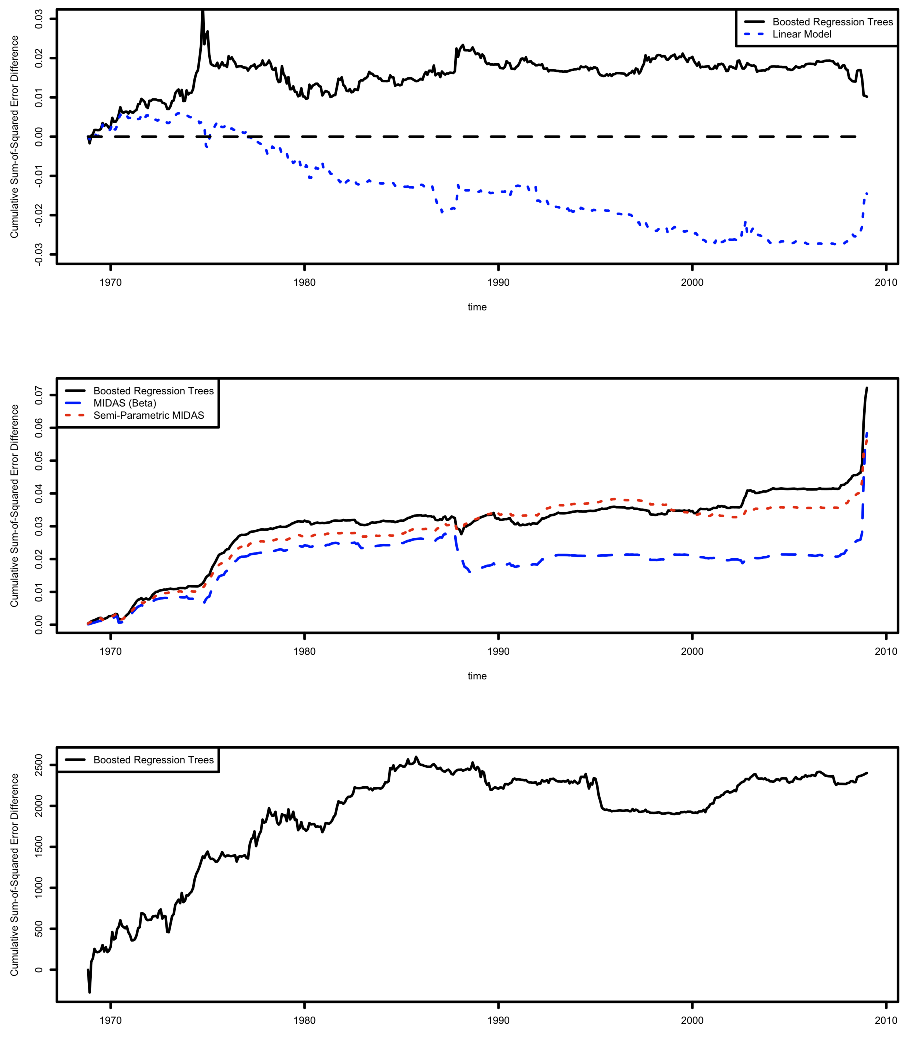


Figure 4: This figure presents cumulative out-of-sample sum-of-squared error differences between a boosted regression tree and benchmark forecasting models for returns, volatility and portfolio weight predictions. The top graph performs the analysis for stock returns: it uses the prevailing mean proposed by Goyal and Welch (2008) as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations and a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2 of the paper. The middle graph performs the analysis for volatility: it uses a GARCH (1,1) as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations, a MIDAS (beta) model (Ghysels et al (2005)) and a semi-parametric MIDAS specification described in Section 2 of the paper. The bottom graph performs the analysis for portfolio weight predictions: it uses the prevailing mean as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations.

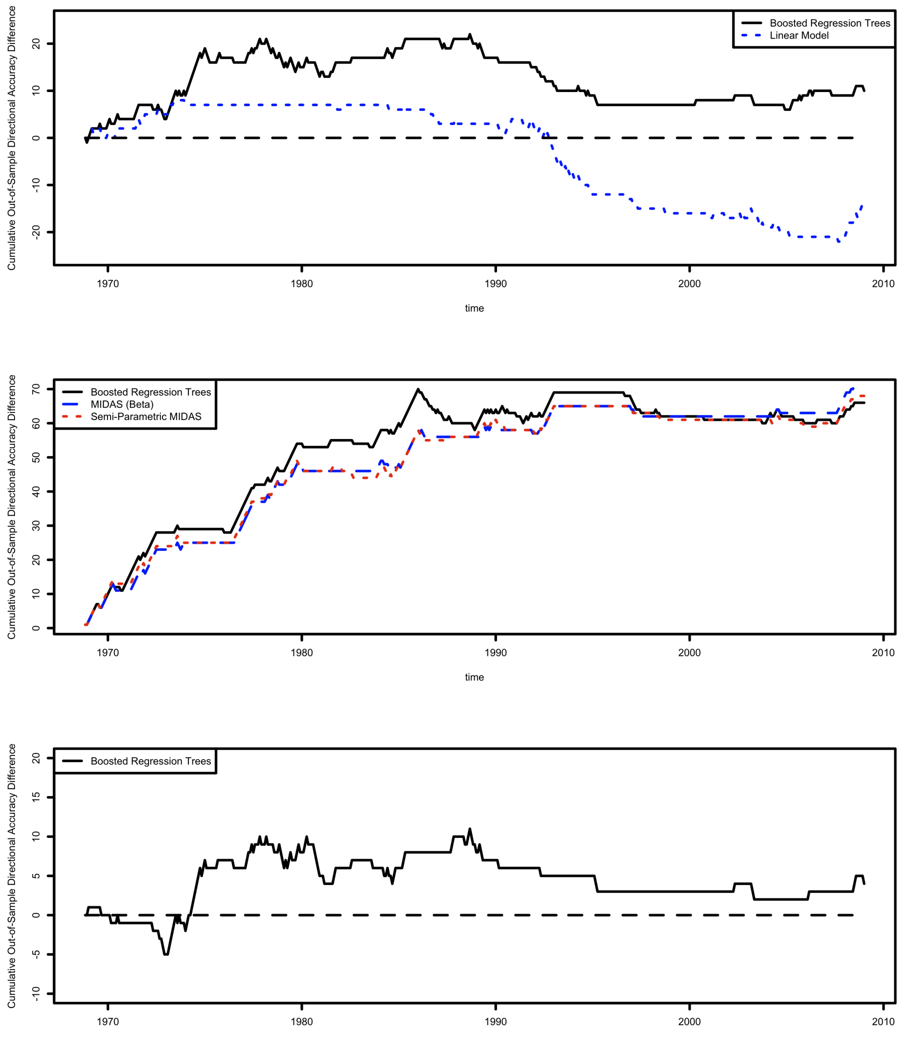


Figure 5: This figure presents cumulative out-of-sample directional accuracy differences between a boosted regression tree and benchmark forecasting models for returns, volatility and portfolio weight predictions. The top graph performs the analysis for stock returns: it uses the prevailing direction as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations and a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2 of the paper. The middle graph performs the analysis for volatility: it uses a GARCH (1,1) as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations, a MIDAS (beta) model (Ghysels et al (2005)) and a semi-parametric MIDAS specification described in Section 2 of the paper. The bottom graph performs the analysis for portfolio weight predictions: it uses the prevailing direction as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations.

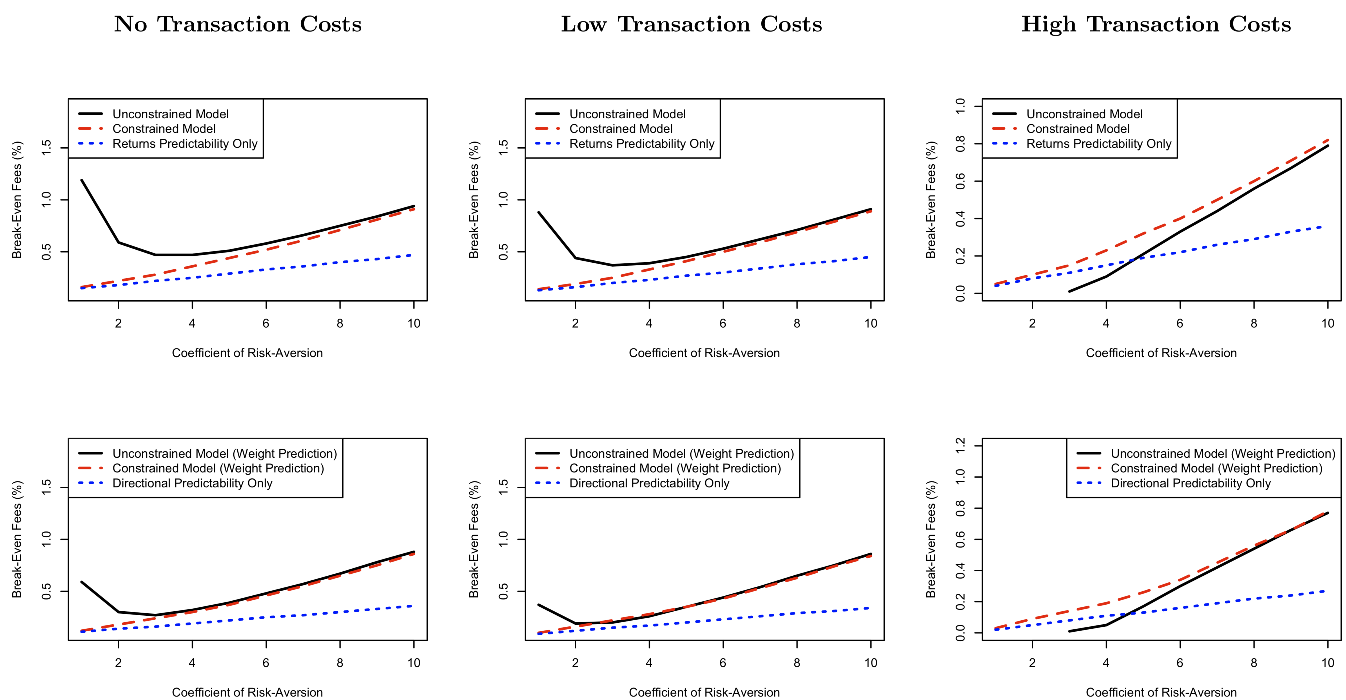


Figure 6: This figure plots break-even portfolio fees that a representative investor would be willing to pay in order have his wealth invested in one of the investment strategies we propose as opposed to investing 100% of his wealth in the market portfolio. All investment strategies proposed are based on a boosted regression tree model with 10,000 boosting iterations. The analysis is conducted for different coefficients of risk-aversion. Each plot presents the results for an unconstrained model, a model with short-selling and borrowing constraints and a model where the portfolio allocation is decided on the basis of return predictability only. The top panel presents the results obtained by forecasting stock market returns and volatility separately and computing the optimal mean-variance portfolio allocation in a second step (two-step BRT model). The bottom panel presents the results obtained by forecasting the optimal portfolio allocation directly on the basis of the conditioning variables listed in Section 2 of the paper (one-step BRT model). In each panel, the left, middle and right plots report the results for zero, low and high transaction costs. In both panels, the optimal portfolio allocations are computed using a "pseudo risk-aversion coefficient" that doubles the investor's genuine risk-aversion γ to adjust portfolio weights for estimation uncertainty. The analysis is conducted over the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.