What is the Shape of the Risk-Return Relation?[†]

Alberto Rossi[*] Allan Timmermann[†]

[*] *University of California San Diego*

[†] *University of California San Diego, CREATES*

July 1, 2009

Abstract

Using a flexible modeling approach that avoids imposing restrictive parametric assumptions, we find evidence of a clear non-monotonic relation between risk and expected returns: At low-to-medium levels of conditional volatility there is a positive trade-off between risk and expected returns, but this relationship gets inverted at high levels of conditional volatility as observed during the recent financial crisis. Our finding of a non-monotonic risk-return relation helps resolve why some empirical studies find a negative risk-return relation, while others find a positive risk-return trade-off.

Keywords: risk-return trade-off, time-varying expected returns, conditional volatility, boosted regression trees.

[†]We are grateful for comments and suggestions from seminar participants at the CREATES-Stevanovich Center conference and the SoFie 2009 conference in Geneva. Allan Timmermann acknowledges support from CREATES funded by the Danish National Research Foundation.

1 Introduction

The notion of a systematic trade-off between risk and expected returns is central to modern finance. Yet, despite more than two decades of empirical research, there is little consensus on the basic properties of the relation between the equity premium and conditional stock market volatility. Studies such as Campbell (1987), Breen, Glosten, and Jagannathan (1989), Glosten, Jagannathan, and Runkle (1993), Whitelaw (1994), and Brandt and Kang (2004) find a negative trade-off, while conversely French, Schwert, and Stambaugh (1987), Bollerslev, Engle, and Wooldridge (1988), Harvey (1989), Ghysels, Santa-Clara, and Valkanov (2005) and Ludvigson and Ng (2007) find a positive trade-off. While these studies use different methodologies and sample periods, it remains a puzzle why empirical results vary so much.

Theoretical asset pricing models do not in general imply a linear or even monotonic, risk-return relationship. For example, in the context of a simple endowment economy, Backus and Gregory (1993) show that the shape of the relation between the risk premium and the conditional variance of stock returns is largely unrestricted with increasing, decreasing, flat, or non-monotonic patterns all being possible. In fact, the risk-return trade-off depends in a complicated way on agents' preferences and the probabilities of transitioning across states. Similar results are derived in Abel (1988), Gennotte and Marsh (1993) and Veronesi (2000).

Inference on the risk-return relation is mostly conducted by studying the relation between the conditional mean and the conditional volatility of stock returns. This introduces two possible sources of bias. First, it follows from the ambiguity of the risk-return relation suggested by asset pricing models that empirical analysis based on restrictive assumptions (such as linearity) on the risk-return trade-off may produce biased results. Second, neither the conditional mean nor the conditional volatility of stock returns is observed, so in practice model-based proxies must be used for these. This introduces model and estimation errors and opens up another possibility, namely that the functional form of the models used to generate expected returns and conditional volatility is misspecified. Again this might lead to biases in the subsequent risk-return analysis.

Both explanations have received little attention in the literature. The main reason is that it is difficult to entertain a flexible functional form for the conditional mean and volatility models while simultaneously considering a reasonably large conditioning information set−a point recently emphasized by Ludvigson and Ng (2007). For example, non-parametric regression methods would quickly run into difficulties for even small subsets of the potentially relevant state variables considered in models of the expected return or variance.

We address this concern in the present paper by adopting a flexible regression approach to model the risk-return trade-off. The approach uses regression trees to carve out the state space through a sequence of piece-wise constant models that approximate the unknown shape of the risk-return relation. By using additive expansions of simple regression trees−a process known as boosting− a smooth and stable estimate of the shape of the risk-return relation can be obtained. Such boosted regression trees are capable of dealing with large sets of predictor variables without imposing strong assumptions such as linearity on the risk-return relation or on the underlying models used to generate estimates of expected returns and conditional volatility. The approach accomplishes variable selection and model estimation by assigning weights to the predictor variables used to model time-variations in the conditional mean and variance.[1] Due to its flexibility, this modeling approach is less likely to lead to biased inference than linear model specifications.

We adopt this approach to analyze the risk-return relation using monthly data on US stock returns over the period 1927-2008. The most important empirical findings from our analysis are as follows. First, we find that only few of the variables considered in the literature on predictability of stock returns (see, e.g., Welch and Goyal (2008) for a recent analysis) have predictive power, with inflation, the earnings-price ratio and the T-bill rate all playing important roles. Moreover, in many cases the effect of the predictor variables on the expected return is non-linear. For example, starting from a state with deflation, rising consumer prices are associated with higher expected returns. This turns into a steep negative relationship at positive levels of inflation. Hence, rising consumer prices are good news in a deflationary state but bad news when the economy is already running inflation.

Past volatility dominates the forecasts of realized volatility. This is unsurprising since our analysis is based on realized volatility which is quite persistent through time (Schwert (1989)). Nevertheless, variables such as the dividend payout ratio, past returns and the default spread also help improve the volatility forecasts. Again we see a pronounced non-linear relation between the realized volatility and the state variables. This typically shows up in the form of dampened effects on future volatility when the predictor values are unusually small or large. For example, higher current volatility leads to higher predicted volatility only up to a certain point beyond which further increases in the conditional volatility have no additional effect on the volatility forecast.

Overfitting is a potential hazard when applying flexible modeling approaches. To verify that this is not a concern for our analysis, we consider the out-of-sample forecasting performance of the return and volatility models. Boosted regression trees are found to produce more precise forecasts of both returns and realized volatility than competing approaches such as linear regressions, GARCH models or models based on past realized volatility.

Turning to the empirical risk-return relation, we find strong evidence of a non-linear relationship between variations in the conditional mean and volatility. At high levels of volatility, the relationship appears to be flat or inverted−i.e., higher levels of conditional volatility are associated with lower expected returns. Conversely, at low-to-medium levels of conditional volatility, there is a significant positive relationship between the conditional mean and volatility of stock returns.

Importantly, the inverted risk-return relation is not confined to subsamples such as the Great Depression which saw very high levels of conditional volatility but also

[1]This stands in contrast to conventional regression analysis that uses a binary approach to variable selection− either a variable is significant and gets included or it is insignificant and gets excluded.

experienced periods with low expected returns. In particular, we also find a distinct negative trade-off between the conditional mean and volatility during the recent financial crisis which has seen similarly high volatility levels.

Our paper is part of a recent literature that has attempted to address biases in models of the risk-return trade-off. Ludvigson and Ng (2007) argue that most studies consider too few conditioning variables and provide a factor-based approach that parsimoniously summarizes information from a large cross-section of variables. Once the conditioning information set is expanded in this way, they find evidence of a positive risk-return trade-off. Rather than conditioning on a specific set of variables, Brandt and Kang (2004) instead propose a latent variable vector autoregression for exploring the trade-off between the conditional mean and variance, which they estimate to be negative. Ghysels, Santa-Clara, and Valkanov (2005) refine the estimated conditional volatility based on a slowly decaying weighted average of squared returns over the previous year. Using this volatility proxy, they find evidence of a positive trade-off between risk and returns. Common to these studies is, however, that they use a linear framework to characterize the risk-return relation. By construction they are therefore limited to studying monotonic risk-return trade-offs. However, it is important to go beyond this setup in view of the indeterminacy of the risk-return relation highlighted by theoretical asset pricing models.

The remainder of the paper is organized as follows. Section 2 offers a brief review of the theoretical literature in order to motivate our empirical modeling strategy which is introduced in Section 3. Section 4 describes the data on returns, realized volatility and predictor variables and reports empirical results for the models used to generate estimates of the conditional mean and volatility, while Section 5 reports on the forecasts generated by the risk and return models. With these in place, Section 6 analyzes the conditional mean−volatility relation. Section 7 conducts a series of robustness checks and extensions, while Section 8 concludes.

2 The Risk-Return Trade-off: Theoretical Insights

To motivate our empirical research strategy, this section provides a brief review of the theoretical literature on the relation between the risk premium and conditional volatility.

While the empirical literature has focused on the sign of the risk return relation, theoretical analysis shows that the relationship between expected returns and conditional volatility need not be linear or even monotonic. For example, Abel (1988) derives an equilibrium asset pricing model in which both the risk premium and the volatility of stock returns vary over time as a result of time-variations in the dividend process. Abel finds that the risk premium need not, in general, be a monotonic function of risk. Similarly, the analysis in Gennotte and Marsh (1993) implies that the equity risk premium is generally a non-linear function of the volatility of equity returns.

This conclusion is consistent with the analysis of Veronesi (2000). Using a Lucas economy, Veronesi shows that the relation between the conditional mean and conditional variance of stock returns is affected by a term that summarizes investors' degree of uncertainty about the economy's unknown growth rate and the effect of such uncertainty on asset valuations. Because of changes in the economy's level of uncertainty, the magnitude of this term varies over time. This means that the relationship between expected returns and conditional volatility is ambiguous and depends on the level of uncertainty facing investors.

2.1 A Simple Endowment Model

As a simple illustration of these points, we follow Backus, Gregory, and Zin (1989) and Backus and Gregory (1993) and examine the risk-return relation in the context of a simple dynamic exchange economy. We first derive analytical expressions for the risk premium and conditional variances and then numerically compute the conditional risk and return values implied by the model.

Consider a stochastic endowment economy with a representative agent who consumes a single good. The endowment, e, varies over time according to the process: $e_{t+1} = x_{t+1}e_t$ where x represents the economy's growth rate. This takes on a finite number of values that evolve according to a stationary first-order Markov chain with transition probabilities

$$\pi_{i,j} = Pr\left(x_{t+1} = \lambda_j | x_t = \lambda_i\right). \tag{1}$$

The representative agent is assumed to have a utility function that allows for habit persistence:[2]

$$u(d) = \frac{d^{(1-\alpha)} - 1}{(1-\alpha)}, \tag{2}$$

where $d_t = c_t - \delta c_{t-1}$, c_t measures consumption, $\alpha > 0$ is the coefficient of relative risk aversion, $0 < \beta < 1$ is the subjective discount factor and δ determines the agent's habit persistence. Under the assumption that the agent maximizes expected utility $U = E_t \sum_{k=0}^{\infty} \beta^k u(d_{t+k})$, an equilibrium in this economy with complete markets is a set of state-contingent prices for which, at all dates and in all states, consumption equals the endowment, i.e., $c_t = e_t$ for every t. Relative prices are then computed by equating them to marginal rates of substitution evaluated at the equilibrium values. For example, if the current state at time t is (e_t, λ_i), then the relative price of a state-contingent claim that pays $e_{t+1} = \lambda_j e_t$ if state j occurs at time $t+1$, is

$$\pi_{i,j}\beta\frac{u'(e_{t+1} - \delta e_t)}{u'(e_t - ey_{t-1})} \times \frac{e_{t+1}}{e_t} = \pi_{i,j}\beta\lambda_j\left(\frac{\lambda_j - \delta}{1 - \frac{\delta}{\lambda_i}}\right)^{-\alpha}.$$

Following Mehra and Prescott (1985), the price per unit of output, denoted by $\mathbf{w}$, is

$$\mathbf{w} = \left(\mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{d}, \tag{3}$$

[2]Allowing for habit persistence makes it easier to generate sizeable variations in risk premia across states in this particular setup. However, non-monotonic risk-return relations can also be generated under the more conventional power utility which arises when $\delta = 0$.

where

$$\mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_I \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_{1,1} & ... & a_{1,J} \\ \vdots & \ddots & \vdots \\ a_{I,1} & ... & a_{I,J} \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} d_1 \\ \vdots \\ d_I \end{pmatrix},$$

while $a_{i,j} = \pi_{i,j}\beta\ \lambda_j \left(\frac{\lambda_j - \delta}{1 - \frac{\delta}{\lambda_i}}\right)^{-\alpha}$ and $d_i = \sum_{j=1}^{J} a_{i,j}$. One share of equity is a claim to the entire stream of future endowments. If the economy transitions from state i to state j, the equity return is

$$r_{i,j} = \frac{\lambda_j(w_j + 1)}{w_i} - 1. \tag{4}$$

Similarly, the conditional *equity premium*, starting from state i, is the expected equity return in excess of the single-period risk-free rate, rf_i

$$E[r_{i,j}|x_t = \lambda_i] - rf_i = \left[\sum_{j=1}^{J} \pi_{i,j} r_{i,j}\right] - rf_i, \tag{5}$$

where the risk-free rate is $rf_i = \left[\sum_{j=1}^{J} \pi_{i,j}\beta \left(\frac{\lambda_j - \delta}{1 - \frac{\delta}{\lambda_i}}\right)^{-\alpha}\right]^{-1}$.

Finally, the conditional variance of stock returns, given that we start from state i, can be computed by:

$$var[r_{i,j}|e_t = \lambda_i] = \sum_{j=1}^{J} \pi_{i,j} \left(r_{i,j} - E[r_{i,j}|x_t = \lambda_i]\right)^2. \tag{6}$$

For a given set of model parameters we can then use (5) and (6) to compute the conditional equity premium and the conditional volatility of stock returns and map out the risk-return relationship across the states.

2.2 Numerical Illustration

To keep the analysis simple, we consider a three-state model. Our parameters are chosen to be sensible in view of empirical findings reported in the literature, i.e., $\alpha = 5$; $\beta = 0.98$, $\delta = .7$ and $\lambda = [.985, 1.00, 1.03]$. Hence in bad states, the economy contracts by 1.5% while in good states it expands by 3%. Moreover, the state transition matrix $\mathbf{\Pi} = [\pi_{ij}]$ is assumed to be symmetric with persistence measured by a single parameter, p:

$$\pi_{ij} = \begin{cases} p & \text{for } i = j \\ \frac{1-p}{2} & \text{for } i \neq j \end{cases}.$$

Figure 1 plots the conditional variance and the associated conditional equity premium under this parameterization, letting p take on values $(0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$. Very different shapes of the risk-return relation emerge from this example, including monotonically increasing, monotonically decreasing as well as hump- and v-shaped patterns.

We conclude from this brief analysis that theoretical models do not constrain the risk-return relation to be linear or monotonic. The expected return and volatility pattern depends in a complicated manner on the interaction between habit persistence (δ), state transition probabilities (π_{ij}) and the growth rates for the individual states (λ_i), making it difficult to draw general conclusions.

These results suggest that imposing linearity on the risk-return relation is too restrictive a framework. To quote Gennotte and Marsh (1993, page 1039), ".. in a general equilibrium framework, the market risk premium is a complicated function of the cash flow uncertainty, implying that the simple regression and time series fits of the relation between equity risk premiums and asset price volatility are likely to be misspecified."

To avoid biases that follow from restricting the shape of the risk-return trade-off, it is therefore important to adopt an empirical modeling approach that is flexible, yet as emphasized by Ludvigson and Ng (2007) can simultaneously deal with large sets of predictor variables. We next describe an approach that accomplishes this.

3 Methodology

We are interested in analyzing the relation between the conditional mean and the conditional volatility of stock returns defined, respectively, as $\mu_{t+1|t} = E_t[r_{t+1}]$ and $\sigma_{t+1|t} = Var_t(r_{t+1})^{1/2}$, where r_{t+1} is the stock return during period $t+1$, measured in excess of the risk-free rate, while $E_t[\cdot]$ and $Var_t(\cdot)$ are computed conditional on investors' information at time t. Both are ex-ante measures that are unobserved and so empirical analysis typically relies on model-based proxies of the form

$$\begin{aligned} \hat{\mu}_{t+1|t} &= f_\mu(x_t|\hat{\theta}_\mu) \\ \hat{\sigma}_{t+1|t} &= f_\sigma(x_t|\hat{\theta}_\sigma), \end{aligned} \tag{7}$$

where x_t represents a set of publicly available predictor variables and $\hat{\theta}_\mu$ and $\hat{\theta}_\sigma$ are estimates of the parameters of the expected return and volatility models, respectively.

Theoretical asset pricing models have been used to suggest broad categories of predictor variables tracking risk premia or levels of uncertainty in the economy. However, they typically do not identify the functional form of the relationship between economic state variables, x_t, and expected returns or volatility in (7). Assuming that a proxy for the unobserved volatility, $\hat{\sigma}_{t+1}$, can be obtained, it is common to simply base estimates of $\mu_{t+1|t}$ and $\sigma_{t+1|t}$ on linear models of the form

$$\begin{aligned} r_{t+1} &= \beta_\mu' x_t + \varepsilon_{rt+1} \\ \hat{\sigma}_{t+1} &= \beta_\sigma' x_t + \varepsilon_{\sigma t+1}. \end{aligned} \tag{8}$$

This gives rise to proxies $\hat{\mu}_{t+1|t} = \hat{\beta}_\mu' x_t$ and $\hat{\sigma}_{t+1|t} = \hat{\beta}_\sigma' x_t$. There is little theoretical justification for imposing such restrictions on empirical forecasting models. Moreover, restricting the functional form to be linear may introduce model misspecification errors and could bias empirical results in ways that makes it difficult to interpret the results.

To address this issue, we extend the basic linear regression model to a class of more flexible and potentially more accurate 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} or $\hat{\sigma}_{t+1}$

based only on their joint empirical distribution. We do not consider conventional non-parametric or smoothing approaches since these are difficult to apply in the presence of many predictor variables and often produce poor forecasts.

To get intuition for how regression trees work and establish the appropriateness of their use 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 earnings-price ratio and the payout ratio). The functional form of the forecasting model mapping X_1, 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, to carve out the state space we use only binary partitions and limit ourselves to lines that are parallel to axes tracking X_1 and X_2. Hence, 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 2, where we show boosted regression trees that use two state variables, namely the lagged values of the log payout ratio (i.e., the dividend-earnings ratio) and the log earnings-price ratio, to predict excess returns on the S&P500 portfolio. Each iteration fits a tree with only two terminal nodes, so every new tree "stub" generates two regions. The graph on the left uses only three boosting iterations. The resulting model ends up with one split along the payout ratio axis and two splits along the earnings-price ratio axis. Within each state the predicted value of stock returns is constant. The predicted value of excess returns is smallest for high values of the payout ratio and low values of the earnings-price ratio, and highest when the payout ratio is small and the earnings-price ratio is high. With only three boosting iterations the model is quite coarse. This changes when more boosting iterations are added. To illustrate this, the figure on the right is based on 5,000 boosting iterations. Now the plot is much more smooth, but clear similarities between the two graphs remain.

Figure 2 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. We next provide a more formal description of the methodology.[3]

3.1 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 $x_t = (x_{t1}, x_{t2}, ..., x_{tp})$.

[3]Our description is in part based on Hastie, Tibshirani, and Friedman (2009) who provide a more in-depth coverage of this approach.

As illustrated in Figure 2, regression trees require deciding two things, namely (i) which predictor variables to use to split the sample space; and (ii) which split points to use. A given splitting point may lead to J disjoint sub-regions or states, $S_1, S_2, ..., S_J$, and the dependent variable is modeled as a constant, c_j, within each state, S_j. The fitted value implied by a regression tree, $\mathcal{T}(x, \Theta_J)$, with J terminal nodes and parameters $\Theta_J = \{S_j, c_j\}_{j=1}^J$ can thus be written

$$\mathcal{T}(x, \Theta_J) = \sum_{j=1}^{J} c_j I(x \in S_j), \tag{9}$$

where $I(x \in S_j)$ is an indicator variable that equals one if $x \in S_j$ and is zero otherwise.

Estimates of S_j and c_j can be obtained as follows. Under the conventional objective of minimizing the sum of squared forecast errors, $\sum_{t=1}^{T}(y_{t+1} - f(x_t))^2$, the estimated constant, $\widehat{c}_j$, is the average of y_{t+1} in state S_j:

$$\widehat{c}_j = \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_j\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_j\}. \tag{10}$$

The globally optimal splitting point is more difficult to determine, particularly in cases where the number of state variables, P, is large. Often a sequential algorithm is used to split the sample space. For a given starting model, this 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:

$$\min_{p,s} \left[\min_{c_1} \sum_{x_t \in S_1(p,s)} (y_{t+1} - c_1)^2 + \min_{c_2} \sum_{x_t \in S_2(p,s)} (y_{t+1} - c_2)^2 \right]. \tag{11}$$

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

$$\begin{aligned} \widehat{c}_1 &= \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_1(p, s)\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_1(p, s)\}, \\ \widehat{c}_2 &= \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_2(p, s)\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_2(p, s)\}. \end{aligned} \tag{12}$$

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 forecast so the choice of splitting variable effectively performs variable selection.

Regression trees are very flexible and can capture local features of the data which linear models may overlook. Moreover, they can handle cases with large-dimensional data. This becomes important when modeling stock returns since the identity of the best predictor variables is generally unknown and so must be determined empirically. 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.

3.2 Boosting

Boosting is a technique that can be used to identify which of a large number of possible variables help to improve forecasting performance. It 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.

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

$$f_B(x) = \sum_{b=1}^{B} \mathcal{T}_b(x; \Theta_{J,b}), \tag{13}$$

where $\mathcal{T}_b(x; \Theta_{J,b})$ is the regression tree of the form (9) used in the b-th boosting iteration and B is the number of boosting iterations. Given the previous model, $f_{B-1}(x)$, the subsequent boosting iteration seeks to find parameters $\Theta_{J,B} = \{S_{j,B}\,, c_{j,B}\}_{j=1}^{J}$ for the next tree to solve a problem of the form

$$\hat{\Theta}_{J,B} = \arg\min_{\Theta_{J,B}} \sum_{t=0}^{T-1} \left[y_{t+1} - (f_{B-1}(x_t) + T_B(x_t; \Theta_{J,B}))\right]^2 . \tag{14}$$

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

$$\begin{aligned} \hat{c}_{j,B} &= \arg\min_{c_{j,B}} \sum_{x_t \in S_{j,B}} \left[y_{t+1} - (f_{B-1}(x_t) + c_{j,B})\right]^2 \\ &= \arg\min_{c_{j,B}} \sum_{x_t \in S_{j,B}} \left[e_{t+1,B-1} - c_{j,B}\right]^2 , \end{aligned} \tag{15}$$

where $e_{t+1,B-1} = y_{t+1} - f_{B-1}(x_t)$ is the forecast 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 $t = 1, ..., \tau$. Then the forecast of $y_{\tau+1}$ is based on the model estimates and the value of the predictor variable at time τ, x_τ.

The most common explanation for why boosting is effective at increasing the forecast accuracy of regression trees are the following. First, 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. Second, by summing over a sequence of trees, boosting performs a type of model averaging that increases the stability of the forecasts.

3.3 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. In the robustness analysis (Section 7) we show that the results are not very sensitive to this choice.

We adopt three refinements to the basic regression tree methodology, namely (i) shrinkage, (ii) subsampling and (iii) minimization of absolute forecast errors. These are all known to decrease the rate at which the objective function is minimized on the training data. Controlling the "learning rate" in this manner reduces the risk of overfitting.

In particular, we use a shrinkage parameter, $0 < \lambda < 1$, which determines how much each boosting iteration contributes to the overall fit:

$$f_B(x) = f_{B-1}(x) + \lambda \sum_{j=1}^{J} c_{j,B} I\{x \in S_{j,B}\}. \tag{16}$$

Following common practice we set $\lambda = 0.001$.

In addition, each tree is fit on a randomly drawn subset of the training data, whose length is set at one-half of the full sample, the default value most commonly used. By fitting the tree only on a subset of the data, this method reduces the risk of overfitting.

Finally, the empirical analysis minimizes mean absolute errors, i.e. $T^{-1} \sum_{t=1}^{T} |y_{t+1} - f(x_t)|$. Under this objective, the optimal forecast is the conditional median of y_{t+1} rather than the conditional mean used under squared error objectives. 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 particular 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.

4 Empirical Estimates

Our empirical analysis of the risk-return trade-off relies on proxies for the conditionally expected stock return and the conditional volatility. This section shows how we generate estimates of these, using the boosted regression tree approach described in Section 3. We first present the data used in our empirical analysis and then report results from the boosted regression trees fitted to expected returns and stock market volatility.

4.1 Data

There is a shortage of theoretical models offering guidance on which variables to use when modeling expected returns and volatility. However, empirical studies have found some evidence of countercyclical patterns in volatility (e.g., Schwert (1989), Engle, Ghysels, and Sohn (2006)). In general, however, we do not know *a priori* which, if any, of a set of economic state variables is capable of predicting stock returns or volatility. We take a broad view and consider a range of state variables that been considered in the finance literature.

In particular, our empirical analysis uses a data set comprising monthly stock returns along with a set of twelve predictor variables previously analyzed in Welch and Goyal (2008). Stock returns are tracked by the S&P 500 index and include dividends. A short T-bill rate is subtracted to obtain excess returns. For brevity we refer to these simply as the returns. The predictor variables from the Welch-Goyal analysis are available during 1927-2005 and we extend their sample up to the end of 2008.[4]

The predictor variables fall into three broad categories:

- Valuation ratios capturing some measure of 'fundamental' value to market value
  - log dividend-price ratio (symbol: dp);
  - log earnings-price ratio (ep);
- Bond yield measures capturing the level or slope of the term structure or measures of default risk
  - three-month T-bill rate (Rfree);
  - de-trended T-bill rate, i.e. the T-bill rate minus a three-month moving average (rrel);
  - yield on long term government bonds (lty);
  - term spread measured by the difference between the yield on long-term government bonds and the three-month T-bill rate (tms);
  - default yield spread measured by the yield spread between BAA and AAA rated corporate bonds (defspr);
- Estimates of equity risk and returns
  - lagged excess return (exc);
  - long term return (ltr);
  - stock variance, i.e. a volatility estimate based on daily squared returns (vol);

We also consider the dividend payout ratio measured by the log of the dividend-earnings ratio (de) and the inflation rate (infl) measured by the rate of change in the consumer price index. 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 $t+1$.

[4] 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.

4.1.1 Realized Volatility

Market volatility is unobserved, so we follow a large recent literature in proxying it through the square root of the realized variance. More specifically, let $r_{i,t}$ be the daily return on day i during month t and let N_t be the number of trading days during this month. Following Schwert (1989), Ludvigson and Ng (2007) and others, we construct the realized variance measure as follows:

$$\hat{\sigma}_t^2 = \sum_{i=1}^{N_t} r_{i,t}^2. \tag{17}$$

The realized volatility, $\hat{\sigma}_t$, is simply the square root of this. Theoretical foundations for the use of (17) can be found in Andersen, Bollerslev, Christoffersen, and Diebold (2006) and Barndorff-Nielsen (2002). It should be recalled that the estimator in (17) is only free of measurement errors as the sampling frequency approaches infinity, so $\hat{\sigma}_t$ is best thought of as a volatility proxy.

4.2 Importance of individual predictor variables

For linear regression models, the conventional measure of a particular state variable's importance is the magnitude and statistical significance of its slope coefficient. This measure is not applicable to regression trees which do not impose linearity. As an alternative measure of influence, we consider the reduction in the squared forecast error every time a particular predictor variable, X_p, is used to split the tree. Summing the reductions in squared forecast errors (or improvements in fit) across the nodes in the tree gives a measure of the variable's influence (Breiman (1984)):

$$I_p(\mathcal{T}) = \sum_{j=2}^{J} \Delta|e(j)| I(v(j) = p), \tag{18}$$

where $\Delta|e(j)| = T^{-1} \sum_{t=1}^{T} (|e_t(j-1)| - | e_t(j)|)$, is the reduction in the (absolute) forecast error at the $j'th$ node and $v(j)$ is the variable chosen at this node, so $I(v(j) = p)$ equals one if variable p is chosen and otherwise is zero. 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 of influence is simple: At each node, one predictor variable gets selected to partition the sample space into two sub-states. The particular variable chosen at node j achieves the greatest reduction in the mean absolute forecast error for the model fitted up to node $j - 1$. The importance of each predictor variable, X_p, is the sum of the reductions in the forecast errors computed over all internal nodes for it was chosen as the splitting variable. If a variable never gets chosen to conduct the splits, its influence will be zero. Conversely, the more frequently a variable is used for splitting and the bigger its effect on reducing the forecast errors, 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:

$$\bar{I}_p = \frac{1}{B}\sum_{b=1}^{B} I_p(\mathcal{T}_b). \tag{19}$$

This is best interpreted as a measure of relative influence that can be compared across predictor variables. In the empirical analysis we therefore report the following measure of relative influence, $\overline{RI}_p$, which sums to one:

$$\overline{RI}_p = \bar{I}_p / \sum_{p=1}^{P} \bar{I}_p. \tag{20}$$

Since it captures relative influence, this measure does not tell if a particular predictor variable is capable of improving the forecasting performance relative to, say, a model with no predictor variables. This question is best addressed by analyzing the model's out-of-sample forecasting performance, a point we shall later return to in Section 5.

4.2.1 Return Model

Panel A of Table 1 shows estimates of the relative influence of the predictor variables when the boosted regression trees are applied to stock returns. We report results for the full sample, 1927-2008, in addition to results based on splitting the sample in halves, i.e. 1927-1967 and 1968-2008.

The results suggest that two predictor variables dominate over the long sample, 1927-2008, as inflation and the earnings-price ratio both obtain weights above 17%. In addition, the de-trended T-bill rate and the long bond yield get weights close to 10%. If each forecasting variable was equally important, we would expect them to get a weight of 1/12 or 8%.

The results are quite consistent over the two subsamples as inflation and the earnings-price ratio are ranked first and second in both subsamples. The de-trended t-bill rate also continues to be relatively important in both subsamples.

4.2.2 Realized Volatility Model

Panel B of Table 1 shows that the lagged volatility is the dominant predictor variable in the model for realized volatility, obtaining a weight close to 70% in the full sample. The default spread and lagged return obtain weights around 8% percent, while the remaining variables get lower weights.

With exception of the lagged volatility, no variable repeats in the top three most important predictors in the two subsamples. Interestingly, the dividend-earnings ratio and the default spread get weights of nearly 14% over the period 1927-67, while only excess returns (and, again, the lagged volatility) get a weight greater than 10% in the second subsample, 1968-2008. This indicates that the volatility forecasting model is less stable than the excess return forecasting model.

Caution is due when interpreting these results, however, since they only measure the *relative* influence of individual variables. Hence they do not show whether or not stock returns and realized volatility are in fact predictable. To address this, we present out-of-sample forecasting results in the next section.

4.3 Effect of Individual State Variables

To gain economic intuition for the boosted regression trees, it is clearly of interest to explore the relation between individual state variables and expected returns or realized volatility. The regression trees do not impose any restrictions on the functional form of the relationship between the dependent variable, Y−returns or realized volatility−and the predictor variables, X. Measuring the effect of the predictor variables on the dependent variable is therefore more complicated than usual.

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, $Y = f(X_p, X_{-p})$

$$f_p(X_p) = E_{X_{-p}} f(X_p, X_{-p}). \tag{21}$$

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

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

$$\bar{f}_p(X_p) = \frac{1}{T} \sum_{t=1}^{T} f(X_p, x_{t,-p}), \tag{22}$$

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

4.3.1 Conditional Equity Premium

Using this measure, Figure 3 presents partial dependence plots for the three most important predictor variables in the model for stock returns. In each case the x-axis is chosen so it covers the empirical support of the predictor variable during the corresponding sample period. As in Figure 2, flat spots show that the mean excess return does not vary in a particular range of the variable.

Figure 3 shows that the relationship between returns and the three most important predictor variables is highly non-linear. At negative levels of inflation the relationship between the rate of inflation and returns is either flat or rising. Thus in a state of deflation, rising consumer prices are good news for stocks. Conversely, at positive levels of inflation, higher consumer prices become bad news for stocks, although at very high levels of inflation there is no systematic relation between inflation and stock market performance. These effects are quite strong in economic terms: the difference between expected returns evaluated at small and large values of the inflation rate is two percent per month.

Similarly, although the relationship between expected stock returns and the log earnings-price ratio is always positive, it is strongest at low or high levels of this ratio, and gets weaker at medium levels of this measure. Again the variation in expected returns as a function of the earnings-price ratio is economically large, spanning a range of 3.5% per month.

Movements in the de-trended T-bill rate seem mainly to indicate reduced expected returns when this variable becomes slightly negative, suggesting that stock returns are most sensitive to small reductions in interest rates.

Figure 4 repeats the plot in Figure 3 for the three most important variables in the two subsamples, 1927-67 and 1968-2008. The figure suggests that the non-linear relationship between both expected returns and inflation and expected returns and the earnings-price ratio or other payout ratios is preserved across the two sub-samples. Viewed together, these findings suggest that models of expected returns based on linear specifications are likely to be misspecified.

4.3.2 Conditional Volatility

Turning to the volatility model, current realized volatility is clearly an important predictor of future volatility. This can be seen from the partial dependence plots in Figure 5 by noting that the predicted volatility quadruples from roughly two percent to eight percent per month as the lagged realized volatility increases over its historical support. The relationship between current and past volatility is basically linear for small or medium values of past volatility. However, very high values of past volatility do not translate into correspondingly high values of expected future volatility, as evidenced by the flatness of the relationship at high levels of volatility.

A highly nonlinear pattern is also found in the relation between the conditional volatility and the default spread: At small or medium values of the spread, future volatility is increasing in this variable. However, at high values of the spread, the expected volatility remains constant. Past returns are also clearly related to current volatility, but the relation is negative as higher past returns imply lower expected volatility. Volatility varies mostly with past returns as these move in a range between minus fifteen percent and zero per month.

5 Out-of-sample Forecasting Performance

Our analysis so far suggests that some predictor variables from the finance literature capture time-variations in expected returns and volatility. Moreover, the effect of these variables appears to be nonlinear. The boosted regression trees may, however, be more prone to estimation error and overfitting than more tightly parameterized linear regressions. It is therefore far from certain that this approach provides a useful way to generate forecasts.

To explore this point, this section follows the literature on real-time forecasting (e.g., Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Campbell and Thompson (2008) and Welch and Goyal (2008)) and estimates the forecasting models recursively through time. In particular, we use data up to 1969:12 to fit the first regression tree. We then predict excess returns or realized volatility for the following month, 1970:01. The next month we expand the data window to 1970:01 and produce forecasts for 1970:02. This procedure continues to the end of the sample in 2008:12.

5.1 Return Forecasts

As a first way to check whether the boosted regression trees overfit the data, consider Figure 6 which shows return forecasts over the period 1985-2008. While there is a visible relationship between the actual and predicted returns (top panel), there is no tendency for the model to fit outliers. Indeed, the fitted values are confined to a relatively narrow range between -3 and 5 percent per month which is far smaller than the corresponding range for actual returns.

We next consider the out-of-sample forecasting performance of the boosted regression trees. Table 2 compares the performance of the forecasts of returns from the boosted regression trees to that from the prevailing mean, a benchmark advocated by Welch and Goyal (2008), and a multivariate linear regression model. For the latter, we use the Bayesian information criterion to select the best specification among the 2^{12} possible linear models that use different combinations of the predictor variables.

We present results separately up to 2005 (the end of Welch and Goyal's original sample) and for the sample extended up to 2008. This serves to illustrate the substantial deterioration in forecasting performance during the very volatile period, 2007-2008. Both the prevailing mean and the multivariate linear regression model generate negative out-of-sample R^2-values. The boosted regression tree model generates more precise out-of-sample forecasts than the prevailing mean model as witnessed by its smaller sum of squared forecast errors and its positive R^2-values in both subsamples.

5.2 Volatility Forecasts

Turning to the volatility forecasts, we would expect to find better out-of-sample performance since the realized volatility series is considerably more persistent than excess returns. The bottom panel in Figure 6 and Table 3 shows that this is what we find. The table compares volatility forecasts from the regression tree to those from a GARCH(1,1) model or an autoregressive model that exploits the persistence in realized volatility. We also consider forecasts from a MIDAS model of the form proposed by Ghysels, Santa-Clara, and Valkanov (2005). Following their analysis, we adopt the following MIDAS estimator of the conditional variance of monthly returns:

$$Var_t(R_{t+1}) = 22 \sum_{d=0}^{D} w_d r_{t-d}^2, \tag{23}$$

where

$$
\begin{aligned}
w_d(\kappa_1, \kappa_2) &= \frac{\exp(\kappa_1 d + \kappa_2 d^2)}{\sum_{i=0}^{D} \exp(\kappa_1 i + \kappa_2 i^2)}, \\
w_d(\kappa_1, \kappa_2) &= \frac{\left(\frac{d}{D}\right)^{\kappa_1 - 1} \left(1 - \frac{d}{D}\right)^{\kappa_2 - 1}}{\sum_{i=0}^{D} \left(\frac{i}{D}\right)^{\kappa_1 - 1} \left(1 - \frac{i}{D}\right)^{\kappa_2 - 1}}
\end{aligned}
$$

for the models that use exponential and beta weights, respectively. "D" represents the maximum lag length which is set to 250 days following Ghysels, Santa-Clara, and Valkanov (2005). Figure 6 plots the fitted volatility levels associated with the GARCH(1,1), MIDAS and the boosted regression trees.

Using the MIDAS approach to forecast monthly volatility yields results that strongly improve on the GARCH model. In the shorter sample that ends in 2005, the out-of-sample R^2-value is 21-24% for the MIDAS models, and only 8% for the GARCH model. However, the best forecasts are generated by the boosted regression trees which produce smaller forecast errors and an out-of-sample R^2-value of 35% which is substantially higher than for the other models.

In the longer sample, 1970-2008, that includes the recent financial crisis, we see larger forecast errors but also greater out-of-sample R^2-values, reflecting the persistently higher volatility levels at the end of this period. The boosted regression trees continue to generate the best volatility forecasts, closely followed by the AR(1) model fitted to lagged realized volatility and the MIDAS models.

In summary, this out-of-sample analysis suggests that the boosted regression tree estimates of the conditional mean and volatility are not overfitting the data. Moreover, as they do not make restrictive assumptions on the shape of the relationship between predictor variables and the conditional mean or volatility, such estimates are less likely to be biased than conventional estimates based on linear regression models. This makes our estimates of the conditional equity premium and conditional volatility well suited for an analysis of the risk-return trade-off.

6 Models of the Risk-Return Relation

This section uses the estimates of the conditional mean and conditional volatility of stock returns from the previous section to explore the shape of the risk-return trade-off. We first consider linear models for this relation and then generalize the setup to cover the case with a general (unknown) risk-return relation. We study the risk-return relation at various horizons and implement a formal test of the monotonic relation between conditional volatility and expected returns.

6.1 Linear Risk-Return Model

We initially follow Ludvigson and Ng (2007) and consider the following reduced-form relation that models the conditional equity premium as a function of the conditional volatility and lags of both volatility and expected returns:

$$\hat{\mu}_{t+1|t} = \alpha + \beta_1 \hat{\sigma}_{t+1|t} + \beta_2 \hat{\sigma}_{t|t-1} + \beta_3 \hat{\mu}_{t|t-1} + \varepsilon_{t+1}. \tag{24}$$

We use 'hats' to indicate that estimates of the conditional expected return and volatility are used. In a generalization of the conventional volatility-in-mean model, lags are included in order to account for the complex lead-lag relationship between the conditional mean and volatility, see, e.g., Whitelaw (1994) and Brandt and Kang (2004).

Empirical results of this analysis are shown in Table 4. For the full sample, 1927-2008, we find evidence of a positive and significant linear relation between the contemporaneous volatility and expected returns with a t-statistic around 2.8. Conversely, the effect of lagged volatility is strongly negative, while the effect of lagged expected returns is strongly positive. These effects carry over to the first subsample 1927-67. In the second subsample, 1968-2008, the relationship between the conditional mean and both the current and lagged conditional volatility is, however, insignificant and much weaker, suggesting that the linear risk-return specification is perhaps not robust.

Further evidence that the positive and significant linear risk-return relation is not particularly robust arises when we estimate the model separately for two high-volatility periods. During 1927-39 (which includes the Great Depression) and 2001-2008 (which includes the recent financial crisis), there is no significant evidence of a linear relation between the conditional mean return and current or lagged volatility.

6.2 Unknown Shape of the Risk-Return Relation

The results reported so far call the linear risk-return specification into question. The risk-return relationship need not be linear of course, a point emphasized also by Merton (1980). For this reason we next employ the boosted regression trees to provide a flexible model of the relationship between expected returns and conditional volatility that avoids imposing particular functional form assumptions. More specifically, we generalize (24) to the following model

$$\hat{\mu}_{t+1|t} = g(\hat{\sigma}_{t+1|t}, \hat{\sigma}_{t|t-1}, \hat{\mu}_{t|t-1}) + \varepsilon_{t+1}. \tag{25}$$

Table 5 presents estimates of the relative influence of the three variables in this model. The relative weight on current conditional volatility is 8% for the full sample which is similar to the weight on lagged volatility (9%). The weight on the lagged expected return (83%) is somewhat higher, which is unsurprising since the expected return is quite persistent and so its lagged value is likely to have large explanatory power in this model. This finding is also consistent with the much larger t-statistic and greater coefficient estimates observed for lagged expected returns in the linear model. Interestingly, in the first sub-sample, 1927-67, the current conditional volatility obtains a much greater weight of 21%, while the weight declines to 10% in the second sub-sample, 1968-2008.

For the periods with the highest realized volatility, 1927-39 and 2001-2008, we see that, at 23-24%, the relative influence of the conditional volatility on the conditional mean is larger than during calmer markets.

Turning to the shape of the risk-return relation, Figure 7 shows that the trade-off between concurrent expected returns and conditional volatility is highly non-linear in the full sample, 1927-2008: At low-to-medium levels of volatility, we find a strong positive relationship, where higher volatility is associated with higher expected returns. In contrast, at high levels of conditional volatility, the relationship appears to be inverted so higher volatility is associated with declining expected returns.

The inverted risk-return trade-off is a robust finding in the sense that it appears not to be confined to a particular historical period. This point is illustrated in Figure 8 which shows the shape of the risk-return relation for the two sub-samples, 1927-67 and 1968-2008. For small values of the initial volatility the expected return increases sharply in both subsamples as the conditional volatility rises. However, at higher volatility levels the expected return declines or remains constant as the conditional volatility rises further.

In summary, our results show that expected returns tend to decline during periods with high conditional volatility. The opposite finding holds for periods with low or medium levels of conditional volatility, where rising volatility levels lead to a systematic increase in the expected return.

6.3 Risk-Return Relation at Longer Horizons

The analysis presented above focused on the risk-return relation at the monthly horizon. However, as pointed out by Ghysels et al (2005), it is of interest to study how similar the results are across different horizons. To accomplish this we next present results for bimonthly, quarterly and semi-annual frequencies and show that the non-linearities in the risk-return relation become more pronounced, the longer the horizon under consideration.

The first step in the analysis entails aggregating monthly returns and volatility figures into their lower frequency counterparts. We do it in the following standard fashion:

$$r_{t+1:t+h} = \prod_{j=1}^{h} (1 + r_{t+j}) - 1 \tag{26}$$

$$\hat{\sigma}_{t+1:t+h} = \left(\sum_{j=1}^{h} \hat{\sigma}_{t+j}^2 \right)^{1/2} .$$

Note that in this notation, the earlier results for the one-month horizon represent the special case $h = 1$. After obtaining estimates of the h-month conditional mean $\hat{\mu}_{t+1:t+h|t}$ and conditional volatility $\hat{\sigma}_{t+1:t+h|t}$, we estimate the following variant of (25):

$$\hat{\mu}_{t+1:t+h|t} = g(\hat{\sigma}_{t+1:t+h|t}, \hat{\sigma}_{t:t+h-1|t-1}, \hat{\mu}_{t:t+h-1|t-1}) + \varepsilon_{t+1:t+h}. \tag{27}$$

Figure 9 presents the shape of the risk-return relation for $h = 2, 3$ and 6 months. Confirming the findings of Figure 7 and 8, at low-to-medium levels of volatility, we find a strong positive relationship, where higher volatility is associated with higher expected returns. Even more so than at the monthly frequency, at high volatility levels we find a negative relation between volatility and expected returns. Furthermore, the negative relation becomes stronger the longer the horizon under consideration, suggesting that the time dimension plays an important role in shaping the trade-off between risk and return.

While the estimates of the risk-return relation suggest marked non-monotonicities at various horizons, they do not necessarily imply that the relationship between conditional volatility and expected returns is non-monotonic in a statistical sense. In the next section we take on this point in a more rigorous fashion.

6.4 Formal Test of Monotonicity

To test formally if the relationship between the conditional volatility and expected returns $\{\hat{\mu}_{t+1:t+h|t}, \hat{\sigma}_{t+1t+h|t}\}$ is monotonic, we use the approach advocated in Patton and Timmermann (2009). To keep the notation simple, we will illustrate the procedure for $h = 1$ even though we compute the test separately for various horizons.

The monotonicity test proceeds as follows. Since our model uses the lagged conditional mean and lagged conditional volatility as control variables, but we are interested in the relationship between the concurrent conditional mean and volatility, we first 'integrate out' the effects of the lagged variables. In particular, we first compute the marginal effect of the lagged conditional volatility and lagged conditional mean on the current mean and then study the relationship between the marginalized conditional mean, $\hat{\mu}^m_{t+1|t}$, and the conditional volatility, $\hat{\sigma}_{t+|t}$. We then sort pairs of monthly observations, $\{\hat{\mu}^m_{t+1|t}, \hat{\sigma}_{t+|t}\}$ ranked by the conditional volatility and group them into N 'portfolios'. Monotonicity would imply that, as we move from groups associated with low conditional volatility to groups with high conditional volatility, the expected value of the conditional mean return should rise.

More specifically, we seek to test whether the "marginalized" conditional expected return increases when ranked by the associated value of $\hat{\sigma}_{t+1|t}$:

$$E\left[\hat{\mu}^m_{t+1|t,i}\right] > E\left[\hat{\mu}^m_{t+1|t,i-1}\right], \text{ for } i = 2, .., N. \tag{28}$$

Following Patton and Timmermann (2009) and defining $\Delta_i \equiv E\left[\hat{\mu}^m_{t+1|t,i}\right] - E\left[\hat{\mu}^m_{t+1|t,i-1}\right]$, for $i = 2, ..., N$, this can be re-written as

$$\Delta_i > 0 \quad \text{for } i = 2, ..., N. \tag{29}$$

To test this hypothesis, we follow the approach of Wolak (1989) which has (weak) monotonicity under the null hypothesis, so that, for $\boldsymbol{\Delta} = (\Delta_2, \Delta_3, ..., \Delta_N)'$:

$$H_0 : \boldsymbol{\Delta} \geq \mathbf{0}, \tag{30}$$

whereas the alternative is of a non-monotonic relationship, i.e. $H_1 : \boldsymbol{\Delta}$ unrestricted. Under this test, the null of a monotonically increasing "marginalized" conditional mean gets rejected if there is sufficient evidence against it. Conversely, a failure to reject the null implies that the data is consistent with a monotonically increasing relation between the "marginalized" conditional mean and conditional volatility. Wolak (1989) shows that the test statistic has a distribution which, under the null, is a weighted sum of chi-squared variables, $\sum_{i=1}^{N} \omega(N,i)\chi^2(i)$, with $\omega(N,i)$ being the weights and $\chi^2(i)$ a chi-squared variable with i degrees of freedom. Approximate critical values are calculated through Monte Carlo simulation.

We have 982 observations when the horizon $h = 1$ and we perform the test under a variety of different values of N. In particular, we divide the data into $N = 15, 20$ and 25 groups and conduct the test for horizons $h = 1, 2, 3$ and 6. The results are reported in Table 6. At the one-month horizon, we get p-values below 2% irrespective of the number of portfolios N. Similar results are obtained for the bimonthly and quarterly horizons. The only exception is the six-month horizon, where the number of portfolios seem to play a role given that we cannot reject the null of monotonicity when the number of portfolios, $N = 15$.

Overall, these results show that a monotonically increasing relation between the conditional mean and conditional volatility at various horizons is strongly rejected, providing further evidence of a non-linear risk-return relation.

7 Robustness and Extensions

In this section we conduct robustness tests that shed light on the sensitivity of our results with regard to parameterizations of the boosted regression trees. We also consider an extension to the basic analysis that allows for a larger set of conditioning state variables.

7.1 Robustness Analysis

Our benchmark analysis uses 10,000 boosting iterations to estimate the regression trees. Since it is possible that the results could be sensitive to this choice, we next explore the sensitivity of our results to alternative ways of choosing the number of boosting iterations. As a first robustness exercise, Table 7 reports the performance of boosted regression trees using 5,000, 10,000 and 15,000 boosting iterations. The results are not particularly sensitive to this choice. On the whole, the regression tree forecasts outperform the benchmarks listed in Tables 2 and 3 for both the return and volatility series.

To further corroborate these results, Figure 10 presents out-of-sample R^2-values as the number of boosting iterations is varied from 100 to 15,000. Signs of overfitting would take the form of a declining R^2-value as the number of boosting iterations rises beyond a certain point. For the stock return model there is evidence only of a very slow decay in forecasting accuracy beyond 7,000 boosting iterations. The lower graph does not show any signs of over-fitting for the volatility prediction model. This stability across different numbers of boosting iterations, B, makes the choice of number of boosting iterations less critical to our analysis.

We finally consider two alternative ways for selecting the number of boosting iterations 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 based on model performance up to time $t-1$. Second, we use forecast combinations as a way to lower the sensitivity of our results to the choice of B by aggregating forecasts across different boosted regression trees by using the simple average of the forecasts from regression trees with 1, 2, ..., 10,000 boosting iterations.

Table 7 shows that the combined average is particularly effective in generating precise return predictions while selecting the best model on the basis of recent performance is most effective for volatility prediction.

7.2 Expanding the Information Set: Common Factors

Ludvigson and Ng (2007) argue that the relationship between expected returns and conditional volatility critically depends on the variables included in the conditioning information set used to generate estimates of time-variations in the conditional mean and volatility of stock returns. Their analysis suggests the need to include a broad set of macroeconomic and financial state variables that can be summarized through a small set of common factors.

Following their analysis, we assume that a large set of state variables $x_{it}, i = 1, ..., N$ are generated by a factor model of the form

$$x_{it} = \lambda_i' f_t + e_{it}, \tag{31}$$

where f_t is a vector of common factors, λ_i is a set of variable-specific factor loadings and e_{it} is an idiosyncratic error assumed to be cross-sectionally uncorrelated. Using the common factors as predictor variables rather than the N individual regressors will achieve a substantial reduction in the dimension of the information set provided that the dimension of f_t is much smaller than N.

We follow Ludvigson and Ng (2007) and extract factors through the principal components method. The data contains N =131 economic time series for the period 1960-2007 which are grouped into 8 broad categories:[5] 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders and Inventories, 5. Money and Credit, 6. Bond and Exchange, 7. Prices, 8. Stock Market. The factors are simply the first principal components in each of these variable groupings, yielding a total of eight factors.

We follow Ludvigson and Ng (2009) and interpret each of the common factors by way of the marginal R^2-values obtained by regressing each of the 131 series on each of the extracted factors. This gives us a relatively straightforward way of characterizing the information content of each factor: $\widehat{f}_1$ is a real activity factor that loads heavily on employment and output data. $\widehat{f}_2$ loads on interest rate spreads, while $\widehat{f}_3$ and $\widehat{f}_4$ load on prices. $\widehat{f}_5$ loads on interest rates, $\widehat{f}_6$ on housing variables and $\widehat{f}_7$ on measures of the money supply. Finally, $\widehat{f}_8$ loads on variables relating to the stock market.

Ludvigson and Ng's (2007) analysis assumes a linear relation between volatility or

[5]These are the same time series used in Ludvigson and Ng (2009)

expected returns and the extracted factors. We generalize this to allow for a flexible and potentially non-linear relation of the form

$$\mu_t = f_\mu(x_t, \hat{f}_t|\theta_\mu) \quad (32)$$
$$\sigma_t = f_\sigma(x_t, \hat{f}_t|\theta_\sigma). \quad (33)$$

In a final step we relate the factor-based estimates of expected returns and volatility using the setup from (25).

Table 8 reports the relative influence of the eight factors in the return and volatility forecasting models. We also include the three variables from the earlier list deemed to be most influential over the period 1960-2007. For the equity premium model this is inflation, the log dividend-price ratio and the log earnings-price, while for the volatility model this is the lagged volatility, lagged excess returns and the log earnings-price ratio for volatility. As expected, not all factors are equally relevant. For the equity premium model, a particularly heavy weight is assigned to the factors that load on interest rate spreads ($\widehat{f_2}$) and stock market variables ($\widehat{f_8}$). Factors appear to be less influential in explaining stock market volatility as all factors obtain weights of 4% or less.

Based on the results in Table 8, we generated estimates of the conditional mean and conditional volatility and used these in the general model for the risk-return trade-off (25). The relative weight on current conditional volatility is 18% while the weights on lagged volatility and lagged expected returns are 14% and 68%.

Figure 11 shows the shape of the risk-return relation when we use the variables in Table 8 as conditioning information. Expanding the information set to include factors seems to reinforce the non-linearities between expected returns and conditional volatility uncovered in Figures 7 and 8.

We conclude that the weight on the current conditional volatility in the risk-return model nearly doubles as a result of expanding the information set to include eight factors. Moreover, the earlier conclusions on the non-monotonic, hump-shape of the risk-return relation continue to hold.

8 Conclusion

We propose a new and flexible approach to modeling the conditional equity premium and the conditional volatility that avoids imposing strong assumptions on the functional form of the model used to generate estimates of the conditional mean and volatility. Imposing invalid restrictions could lead to biased estimates and erroneous inference. We then use this approach to model the risk-return trade-off.

Our empirical analysis suggests that there is a positive risk-return trade-off during periods with low or 'normal' levels of volatility, but that the relationship is flat or inverted during periods with high volatility, as observed, for example, during the recent financial crisis. These findings make it easier to understand why so many studies differ in their results regarding the sign and magnitude of the empirical risk-return relationship and why results from linear models appear not to be robust to the sample period used in the analysis. In particular, studies that are conducted over periods without bouts of high (conditional) volatility are more likely to find a positive trade-off between risk and returns, while studies that include such episodes are more likely to find a negative or insignificant risk-return trade-off. In both cases, the conclusions can only be viewed as partial evidence that fails to cover the global behavior of the risk-return relation.

The non-monotonic risk-return trade-off uncovered here naturally raises the question of which type of asset pricing model is required to generate this pattern. Models where the representative consumer has constant relative risk aversion preferences are also able to generate non-monotonic patterns, but it is not clear that they can match the magnitude of the effect observed empirically. Our brief analysis indicated that a model with habit persistence in which the risk premium varies because the risk aversion varies across states shows promise in generating a sizeable non-monotonic risk-return pattern. We leave it to future research to explore this important issue in more depth.

References

Abel, A. B. (1988): "Stock prices under time-varying dividend risk : An exact solution in an infinite-horizon general equilibrium model," *Journal of Monetary Economics*, 22(3), 375–393.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006): "Volatility and Correlation Forecasting," *Handbook of Economic Forecasting*, 1, 777–878.

Backus, D. K., and A. W. Gregory (1993): "Theoretical Relations between Risk Premiums and Conditional Variances," *Journal of Business & Economic Statistics*, 11(2), 177–85.

Backus, D. K., A. W. Gregory, and S. E. Zin (1989): "Risk premiums in the term structure : Evidence from artificial economies," *Journal of Monetary Economics*, 24(3), 371–399.

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

Barndorff-Nielsen, O. (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 253–280.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988): "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy*, 96(1), 116–31.

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.

Brandt, M. W., and Q. Kang (2004): "On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach," *Journal of Financial Economics*, 72(2), 217–257.

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.

Campbell, J., 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.

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

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

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987): "Expected stock returns and volatility," *Journal of Financial Economics*, 19(1), 3–29.

Friedman, J. (2002): "Stochastic gradient boosting," Technical report, Department of Statistics, Stanford University.

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

Gennotte, G., and T. A. Marsh (1993): "Variations in economic uncertainty and risk premiums on capital assets," *European Economic Review*, 37(5), 1021–1041.

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.

Glosten, L., R. Jagannathan, and D. Runkle (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48(5), 1779–1801.

Harvey, C. R. (1989): "Time-varying conditional covariances in tests of asset pricing models," *Journal of Financial Economics*, 24(2), 289–317.

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

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

Ludvigson, S., and S. Ng (2009): "Macro Factors In Bond Risk Premia," *Forthcoming in The Review of Financial Studies.*

Mehra, R., and E. C. Prescott (1985): "The equity premium: A puzzle," *Journal of Monetary Economics*, 15(2), 145–161.

Merton, R. (1980): "On Estimating the Expected Return on the Market," *Journal of Financial Economics*, 8(4), 323–361.

Patton, A., and A. Timmermann (2009): "Monotonicity in Asset Returns: New Tests with Applications to the Term Structure, the CAPM and Portfolio Sorts," *Manuscript, Oxford and UCSD.*

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

Schwert, G. W. (1989): "Why Does Stock Market Volatility Change over Time?," *Journal of Finance*, 44(5), 1115–53.

Veronesi, P. (2000): "How Does Information Quality Affect Stock Returns?," *The Journal of Finance*, 55(2), 807–837.

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.

Whitelaw, R. F. (1994): "Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns," *Journal of Finance*, 49(2), 515–41.

Wolak, F. A. (1989): "Testing inequality constraints in linear econometric models," *Journal of Econometrics*, 41(2), 205–235.

Table 1. Relative influence of predictor variables

A. Excess Returns

1927-2008	infl 17.48	ep 17.31	rrel 10.10	lty 9.47	vol 8.30	de 7.50	rfree 6.07	dp 5.82	exc 5.79	tms 5.28	ltr 4.26	defspr 2.62
1927-1967	infl 12.41	ep 11.01	de 10.35	dp 9.30	exc 9.27	ltr 8.99	rrel 8.89	lty 8.29	rfree 7.78	vol 5.62	tms 4.59	defspr 3.50
1968-2008	infl 16.49	ep 12.54	dp 12.52	rrel 10.97	exc 8.45	vol 7.31	lty 6.47	tms 5.80	ltr 5.59	defspr 4.99	de 4.68	rfree 4.18

B. Realized Volatility

1927-2008	vol 66.10	defspr 8.87	exc 7.39	de 6.25	infl 2.89	ep 2.42	dp 2.14	rfree 1.57	rrel 1.36	lty 0.44	ltr 0.29	tms 0.27
1927-1967	vol 41.99	de 13.92	defspr 13.61	dp 6.12	lty 6.09	exc 5.94	ltr 3.07	infl 2.79	tms 2.48	rrel 2.38	ep 0.91	rfree 0.71
1968-2008	vol 60.16	exc 11.65	ep 6.25	dp 5.40	infl 4.35	rfree 3.99	defspr 2.07	rrel 1.92	de 1.26	ltr 1.10	lty 0.99	tms 0.86

This table shows the relative influence (in percent) of the 12 predictor variables considered in the boosted regression tree model. A higher number indicate that a predictor variable is more important. The 12 predictors are the inflation rate (infl), the log earnings-price ratio (ep), the 1-month T-bill rate relative to its three-month average (rrel), the long-term government bond yield (lty), stock market volatility (vol), the log dividend-earnings ratio (de), the risk-free T-bill rate (rfree), the log dividend-price ratio (dp), the excess return on stocks (exc), the term spread (tms), the long-term rate of return (ltr) and the default spread (defspr). Results in Panel A use monthly returns on the S&P 500 index measured in excess of the T-bill rate as the dependent variable. Results in Panel B are for the realized volatility, measured as the square root of the sum of squared daily returns within a given month.

Table 2. Forecasting performance of the return models

	1970-2005		1970-2008	
Forecasting Method	**MSE** $\times 10^3$	$\mathbf{R^2}$	**MSE** $\times 10^3$	$\mathbf{R^2}$
Boosted Regression Trees	1.945	1.56%	1.970	0.30%
Prevailing Mean	1.987	-0.58%	1.992	-0.79%
Multivariate Linear Model	2.003	-1.37%	2.015	-1.97%

This table reports the mean squared error (MSE) and the out-of-sample R^2-value for a boosted regression tree model based on 10,000 boosting iterations used to forecast monthly stock returns. For comparison we also present results for the prevailing mean model proposed by Welch and Goyal (2008) and for a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Table 1. Results are showed separately for the period pre-dating the recent financial crisis (1970-2005) and the period that includes this episode (1970-2008). The parameters of the forecasting models are estimated recursively through time.

Table 3. Forecasting performance of the realized volatility models

	1970-2005		1970-2008	
Forecasting Method	**MSE** $\times 10^4$	$\mathbf{R^2}$	**MSE** $\times 10^4$	$\mathbf{R^2}$
Boosted Regression Trees	2.550	35.51%	3.314	40.84%
Garch(1,1)	3.664	7.69%	4.810	14.13%
MIDAS (Beta)	3.012	23.90%	3.602	35.69%
MIDAS (Exp)	3.116	21.28%	3.746	33.12%
AR(1)	3.001	24.42%	3.369	39.85%

This table reports the mean squared error (MSE) and the out-of-sample R-squared value for a boosted regression tree model based on 10,000 boosting iterations used to forecast the monthly value of realized volatility. For comparison we report the results for GARCH(1,1), AR(1) and MIDAS models (Ghysels et al (2005)). Results are presented separately for the period pre-dating the recent financial crisis (1970-2005) and the period that includes this episode (1970-2008). The parameters of the forecasting models are estimated recursively through time.

Table 4. Least-squares estimates of linear models for the risk-return trade-off

Model: $\mu_{t+1|t} = \alpha + \beta_1 \sigma_{t+1|t} + \beta_2 \sigma_{t|t-1} + \alpha \mu_{t|t-1} + \epsilon_{t+1}$

Period	$\sigma_{t+1\|t}$ (t-stat)	$\sigma_{t\|t-1}$ (t-stat)	$\mu_{t\|t-1}$ (t-stat)	R^2
		A. Sub-samples		
1927-2008	0.058	-0.079	0.659	44.40%
	(2.76)	(-3.79)	(26.78)	
1927-1967	0.056	-0.104	0.584	41.54%
	(2.07)	(-3.92)	(15.41)	
1968-2008	0.048	-0.025	0.662	42.93%
	(1.42)	(-0.72)	(19.29)	
		B. High-Volatility Periods		
1927-1939	-0.024	-0.052	0.43	34.46%
	(-0.61)	(-1.34)	(5.82)	
2001-2008	0.063	-0.067	0.510	25.77%
	(0.82)	(-0.81)	(5.64)	

This table reports estimates and t-statistics from ordinary least squares regressions of the estimated conditional mean return ($\mu_{t+1|t}$) on the conditional volatility ($\sigma_{t+1|t}$), the lagged volatility ($\sigma_{t|t-1}$) and the lagged conditional mean ($\mu_{t|t-1}$). Estimates of the conditional mean and volatility used in the analysis are based on boosted regression tree models that use the 12 variables listed in Table 1 as predictors.

Table 5. Relative influence of variables in risk-return model

Model: $\mu_{t+1|t} = f(\sigma_{t+1|t}, \sigma_{t|t-1}, \mu_{t|t-1})$

Sub-samples	$\sigma_{t+1\|t}$	$\sigma_{t\|t-1}$	$\mu_{t\|t-1}$
A. Sub-samples			
1927-2008	7.6%	9.4%	83.0%
1927-1967	21.1%	19.5%	59.4%
1968-2008	9.9%	9.9%	80.2%
B. High-Volatility Periods			
1927-1939	24.5%	35.4%	40.1%
2001-2008	23.0%	20.6%	56.4%

This table shows the relative influence in the expected return specification of the current conditional volatility, $\sigma_{t+1|t}$, the lagged volatility, $\sigma_{t|t-1}$, and the lagged expected return, $\mu_{t|t-1}$. Relative influence estimates sum to 100. Estimates of the conditional mean and volatility are based on boosted regression trees that use the 12 predictor variables listed in Table 1 as predictors.

Table 6. Monotonicity test for the risk-return relation

	Number of Portfolios		
Horizon (months)	**15**	**20**	**25**
1	0.0103	0.0180	0.0000
2	0.0166	0.0000	0.0000
3	0.0386	0.0000	0.0000
6	0.6092	0.0015	0.0077

This table presents the results of a test investigating whether the relationship between conditional volatility and expected returns after marginalizing out the effect of lagged volatility and lagged expected returns is monotonic. Marginalized expected returns and conditional volatilities $\{\hat{\mu}^m_{t+1:t+h|t}, \hat{\sigma}_{t+1t+h|t}\}$ are grouped into N portfolios based on the value of the conditional volatility. Each entry reports the p-value of a Wolak test that entertains the null hypothesis of an increasing and monotonic relation between expected returns and conditional volatility against the alternative of a non-monotonic relationship. Small p-values indicate rejection of monotonicity. For robustness purposes, the test is performed for different time horizons and using different numbers of portfolios.

Table 7. Robustness of the forecasting performance of the boosted regression trees.

A. Return Prediction

	1970-2005		**1970-2008**	
Forecasting Method	**MSE** $\times 10^3$	$\mathbf{R^2}$	**MSE** $\times 10^3$	$\mathbf{R^2}$
Boosted Regression Trees (5,000)	1.947	1.47%	1.965	0.53%
Boosted Regression Trees (10,000)	1.945	1.56%	1.970	0.30%
Boosted Regression Trees (15,000)	1.946	1.49%	1.975	0.06%
Best model selected recursively	1.959	0.86%	1.978	-0.01%
Combined Average	1.942	1.72%	1.960	0.76%

B. Volatility Prediction

	1970-2005		**1970-2008**	
Forecasting Method	**MSE** $\times 10^4$	$\mathbf{R^2}$	**MSE** $\times 10^4$	$\mathbf{R^2}$
Boosted Regression Trees (5,000)	2.701	31.76%	3.571	36.24%
Boosted Regression Trees (10,000)	2.550	35.51%	3.314	40.84%
Boosted Regression Trees (15,000)	2.525	36.22%	3.275	41.53%
Best model selected recursively	2.568	35.13%	3.318	40.75%
Combined Average	2.701	31.74%	3.656	34.73%

This table shows the out-of-sample forecasting performance of the boosted regression trees under different rules for determining the number of boosting iterations. The first three lines show results using 5,000, 10,000 and 15,000 boosting iterations. The fourth line (best model selected recursively) chooses the number of boosting iterations recursively based on data up to the previous month. The combined average uses the simple average of forecasts from regression trees with 1, 2, ..., 10,000 boosting iterations. Panel A covers the return prediction model, while Panel B covers the realized volatility prediction model. Results are showed separately for the period pre-dating the recent financial crisis (1970-2005) and for the period that includes this episode (1970-2008).

Table 8. Relative influence of predictor variables in the model with factors

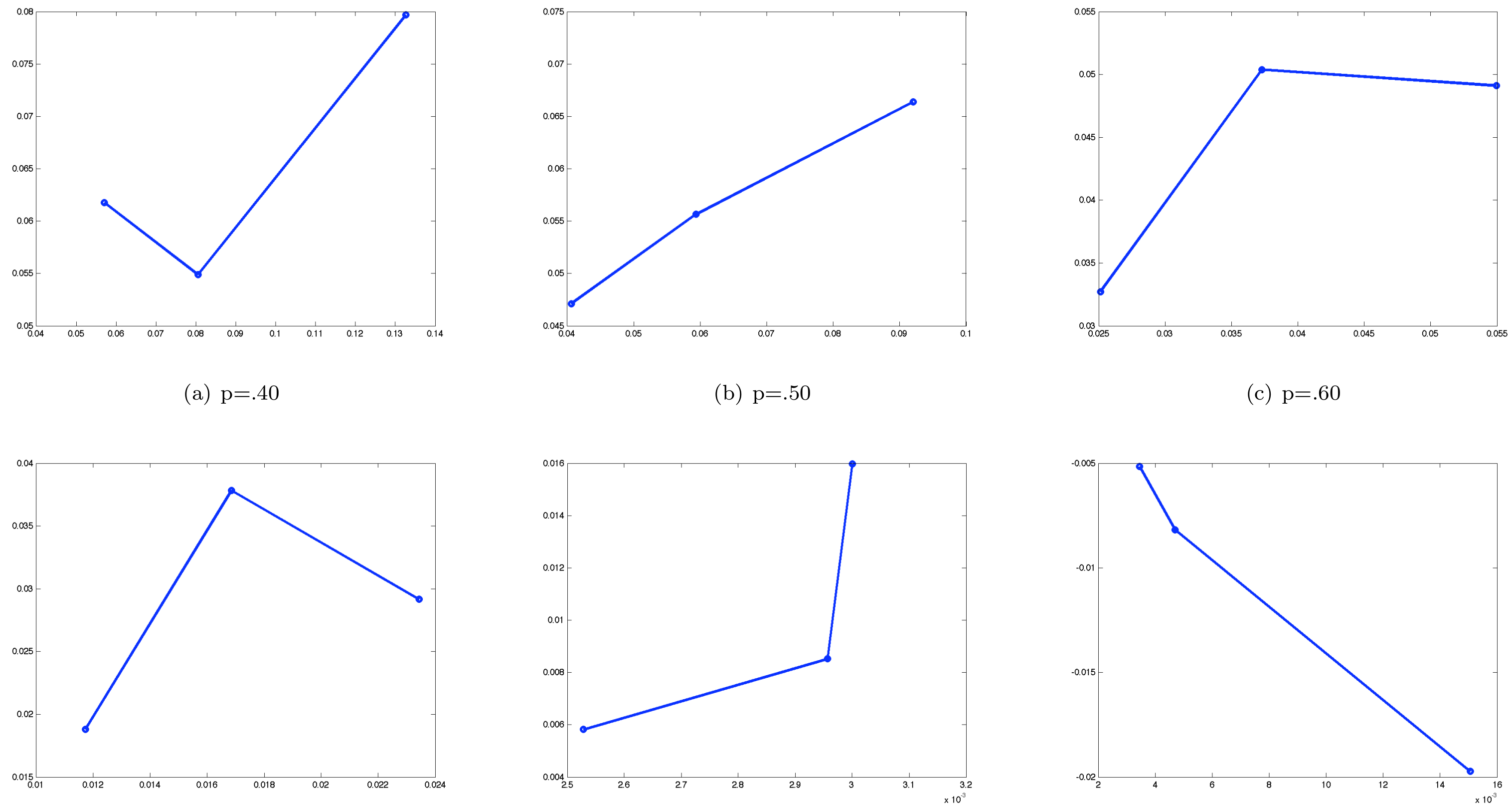
A. Excess Returns

1960-2007	$\widehat{f}_2$	infl	ep	$\widehat{f}_8$	$\widehat{f}_5$	dp	$\widehat{f}_1$	$\widehat{f}_7$	$\widehat{f}_6$	$\widehat{f}_4$	$\widehat{f}_3$
	20.95%	10.57%	10.40%	10.34%	9.66%	7.95%	7.90%	7.46%	5.91%	5.48%	3.39%

B. Realized Volatility

1960-2007	vol	ep	exc	$\widehat{f}_7$	$\widehat{f}_1$	$\widehat{f}_2$	$\widehat{f}_6$	$\widehat{f}_4$	$\widehat{f}_3$	$\widehat{f}_8$	$\widehat{f}_5$
	62.52%	9.48%	8.14%	4.12%	4.06%	3.17%	3.15%	2.57%	1.1%	1.06%	0.6%

This table shows the relative influence of the 8 factors and 3 predetermined variables considered in the boosted regression tree model. The derived factors are obtained through principal components starting from a large dataset composed of 131 economic time series. Their interpretation is as follows: $\widehat{f}_1$ is a real activity factor that loads heavily on employment and output data. $\widehat{f}_2$ loads on interest rate spreads, while $\widehat{f}_3$ and $\widehat{f}_4$ load on prices; $\widehat{f}_5$ loads on interest rates, $\widehat{f}_6$ on housing variables and $\widehat{f}_7$ on measures of the money supply. Finally, $\widehat{f}_8$ loads on variables relating to the stock market. Inflation (infl), the log dividend-price ratio (dp) and the log earnings-price ratio (ep) are the predetermined variables for the equity premium model; lagged volatility (vol), lagged excess returns (exc) and the log-earnings price ratio (ep) are used in the volatility specification. Results in Panel A are for monthly returns on the S&P 500 index measured in excess of the T-bill rate. Results in Panel B are for the realized volatility measured as the square root of the sum of squared daily returns within a given month.



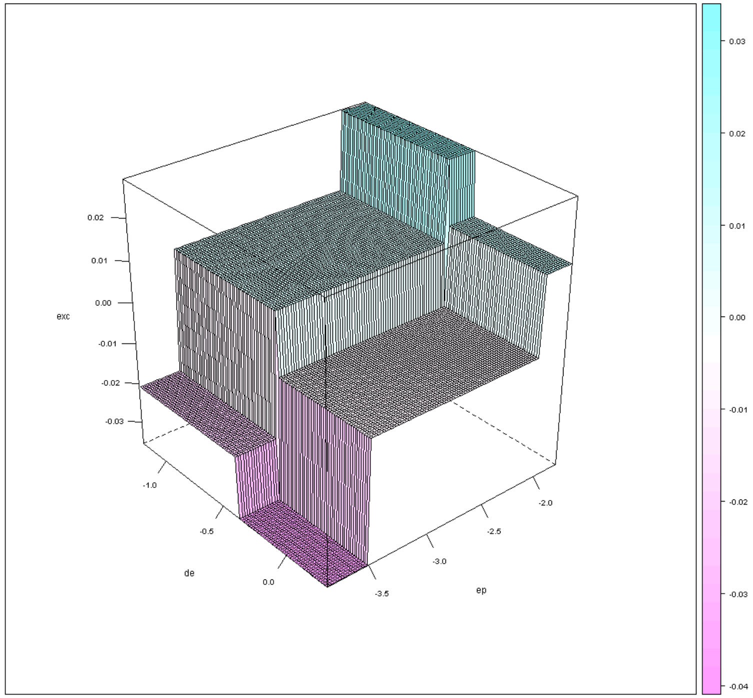
(a) p=.40 (b) p=.50 (c) p=.60

(d) p=.70 (e) p=.80 (f) p=.90

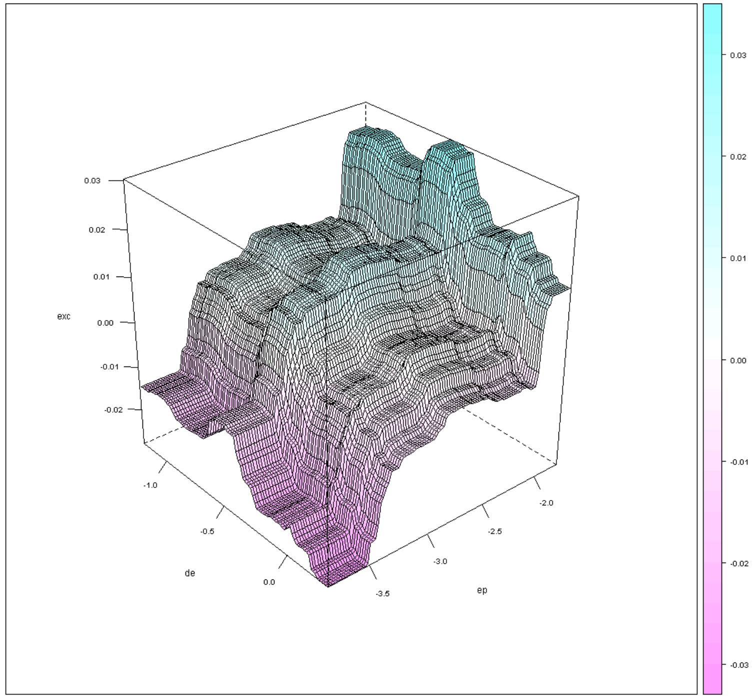
Figure 1: This figure plots the the equity premium (y-axis) against the conditional variance of expected returns (x-axis) for a three-state endowment economy whose growth rate λ follows a stationary Markov chain with state transition probabilities given by

$$\pi_{i,j} = p \quad for \quad i = j \qquad and \qquad \pi_{i,j} = \frac{(1-p)}{2} \quad for \quad i \neq j$$

The representative agent is parameterized by a risk-aversion coefficient α, a discount factor β and a parameter δ that captures habit persistence. The plots above are obtained by varying the value of p, while leaving all the other parameters at the following base values: $\alpha = 5$; $\beta = 0.98$, $\delta = .7$ and $\lambda = [.985, 1.00, 1.03]$.



(a) 3 Iterations



(b) 5000 Iterations

Figure 2: Fitted values of excess returns (exc) as a function of the log dividend-earnings ratio (de) and the log earnings-price ratio (ep). Both plots are based on boosted regression trees with two nodes in each split. The panel on the left uses three boosting iterations, while the right panel uses 5,000 iterations. The scale for the log earnings-price ratio has been inverted. The plots are based on monthly data from 1927-2008.

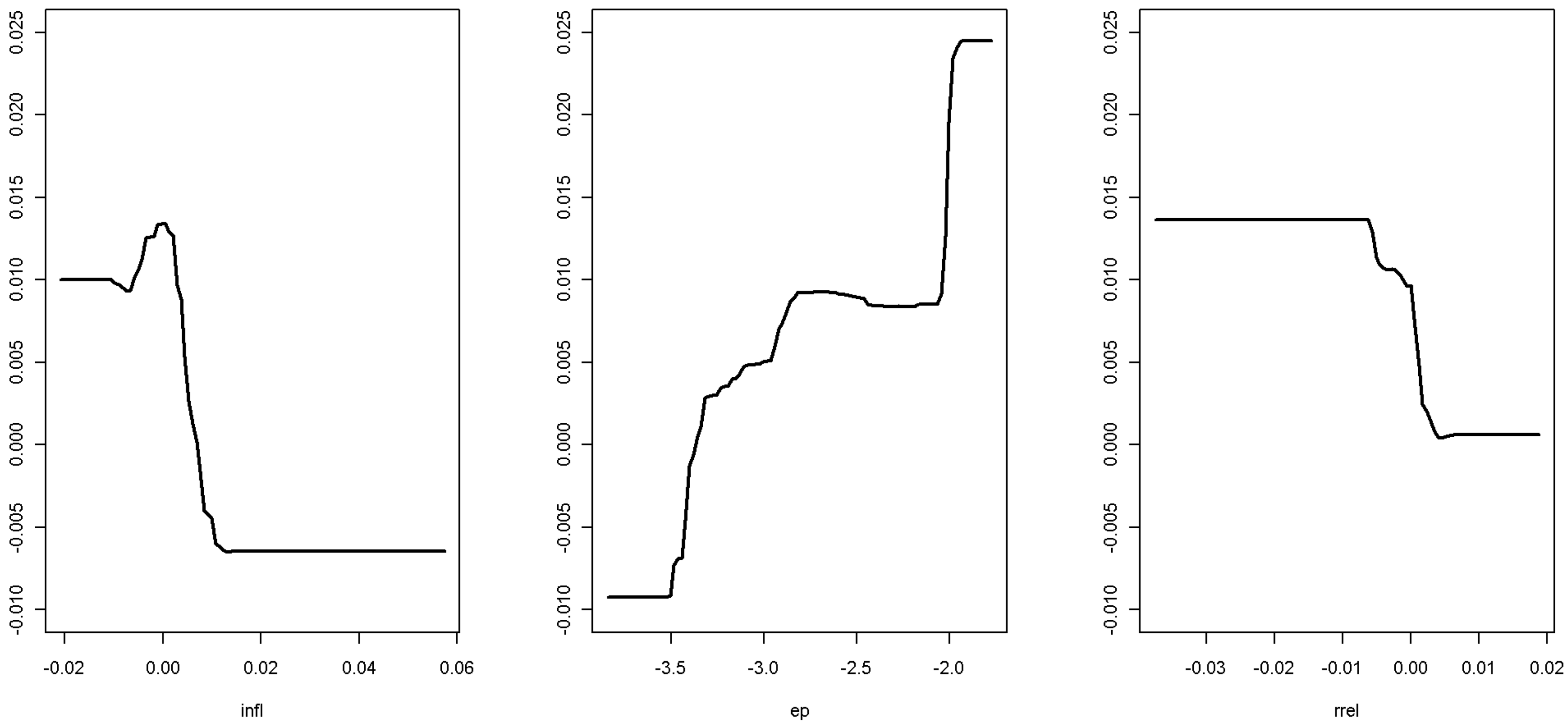


Figure 3: Partial dependence plots for the mean excess return equation based on the three predictor variables with the highest relative influence during the full sample, 1927-2008, namely inflation (infl), the log earnings price ratio (ep) and the detrended T-Bill rate (rrel). The x-axis covers the sample support of the predictor variable, while the y-axis tracks the change in the conditional equity premium as a function of the individual predictor variables.

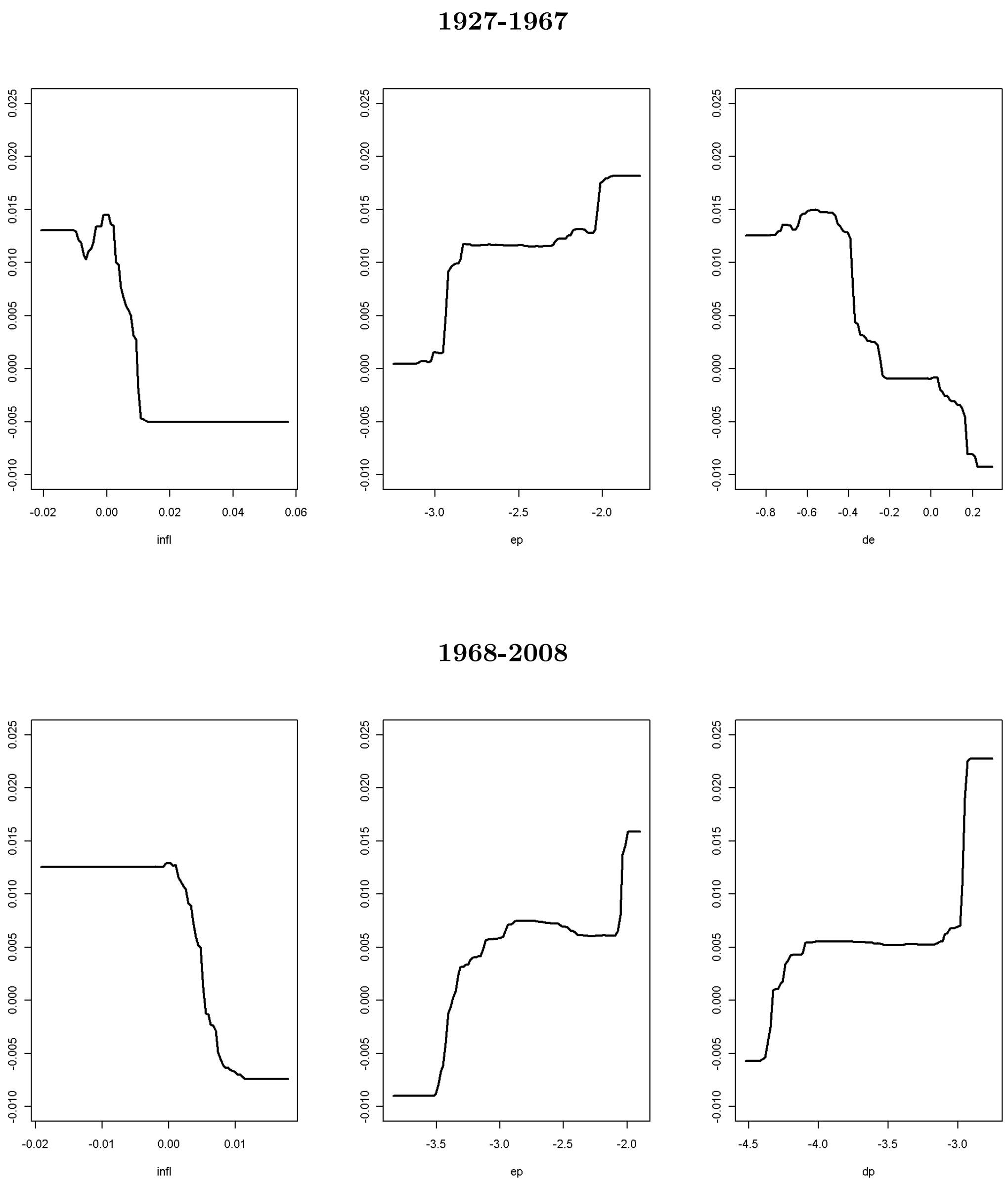


Figure 4: Partial dependence plots for the mean excess return equation based on the three predictor variables with the highest relative influence during the sub-samples, 1927-1967 (top graphs) and 1968-2008 (bottom graphs), namely, inflation (infl), the log earnings price ratio (ep), the log dividend earnings ratio (de), and the log dividend price ratio (dp). The x-axis covers the sample support of the individual predictor variables, while the y-axis tracks the change in the conditional equity premium as a function of the individual predictor variables.

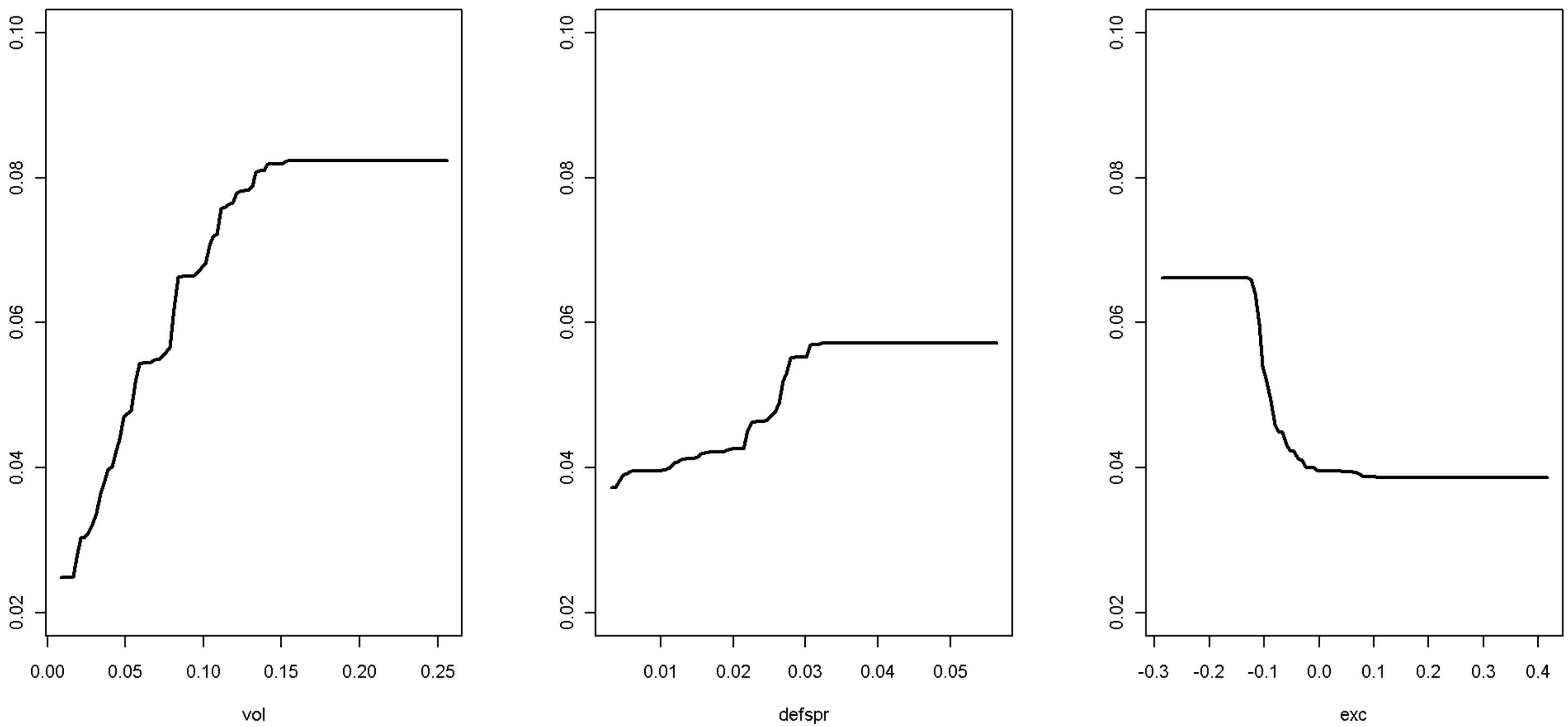
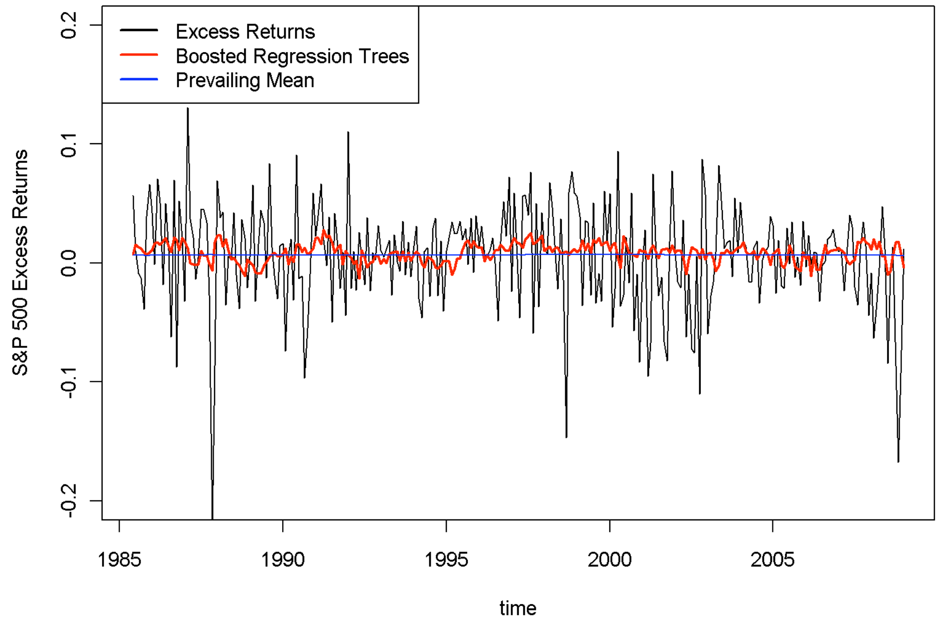


Figure 5: Partial dependence plots for the realized volatility equation based on the three predictor variables with the highest relative influence during the full sample, 1927-2008, namely stock market volatility (vol), the default spread (defspr) and the excess return on stocks (exc). The x-axis covers the sample support of each predictor variable, while the y-axis tracks the change in the conditional volatility as a function of the individual predictor variables.



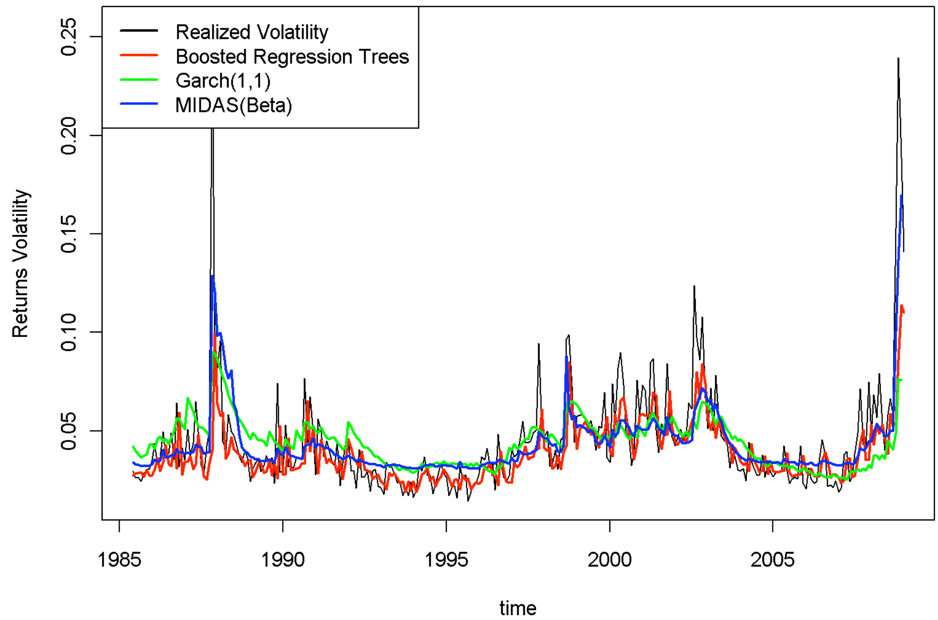


Figure 6: Actual versus predicted values of returns and realized volatility. The top graph shows the time-series of excess returns plotted against the out-of-sample predicted values from the boosted regression tree models. The bottom graph plots the realized volatility against the predicted values from a boosted regression tree, a GARCH(1,1) model and a MIDAS model with beta weights.

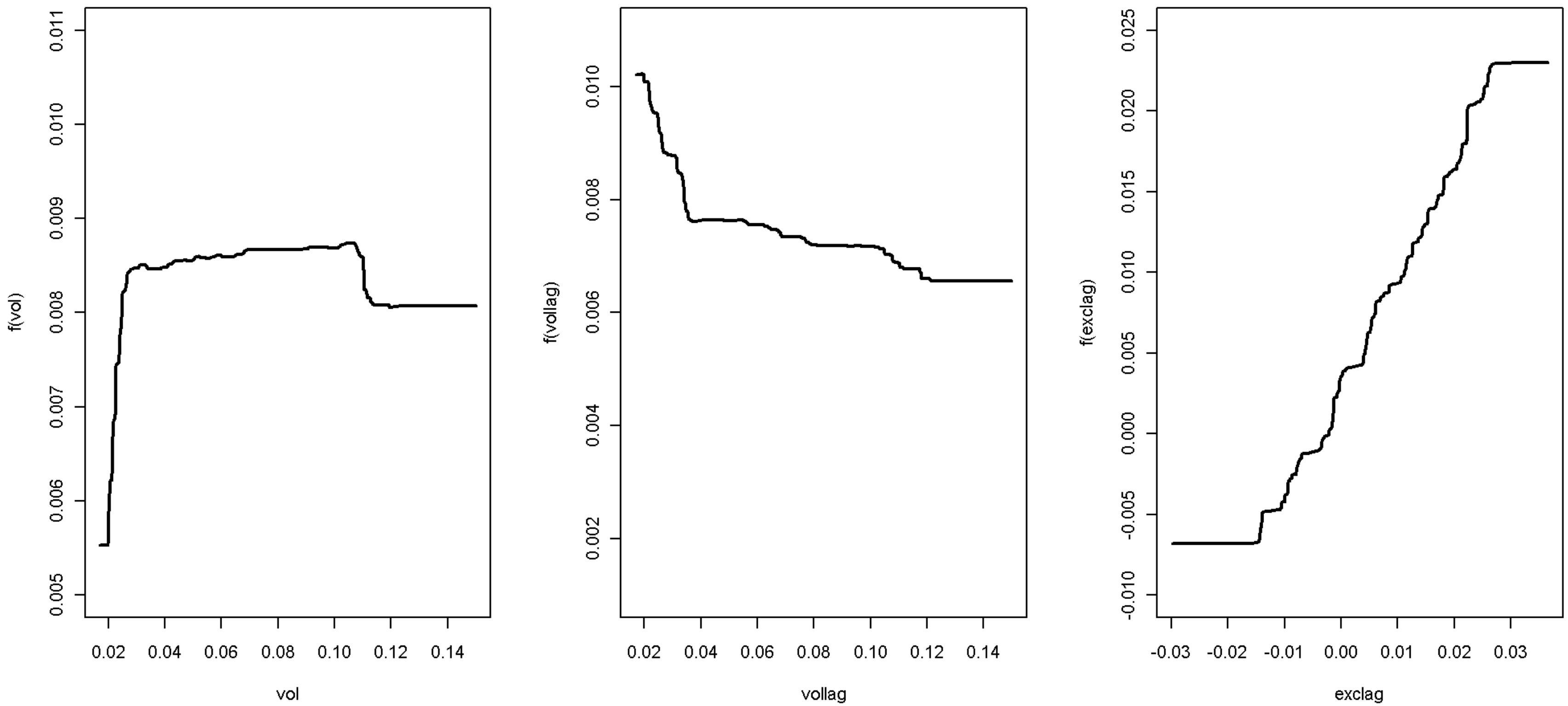
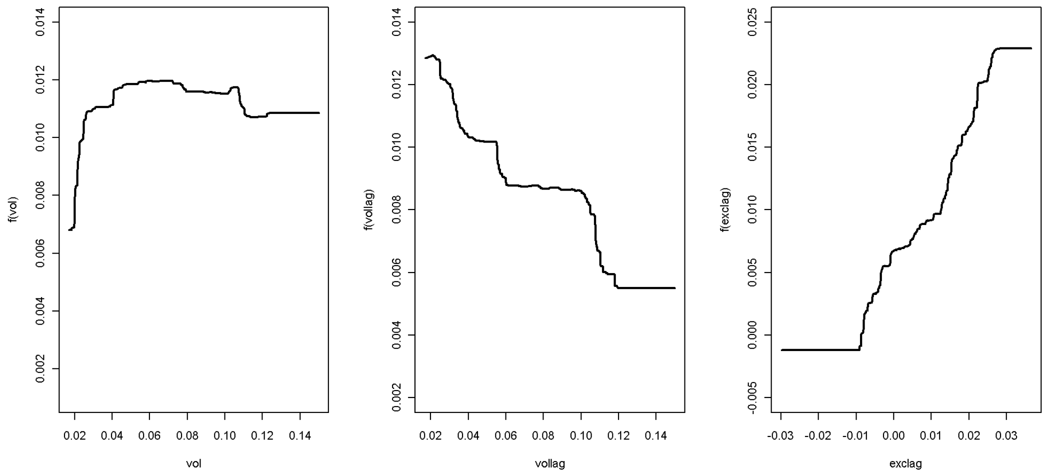


Figure 7: Partial dependence plots showing the conditional equity premium as a function of the conditional volatility (vol), past volatility (vollag) and the lagged conditional mean return (exclag). The plot is based on a regression tree with 10,000 boosting iterations, using data for the period 1927-2008. The x-axis covers the sample support of each predictor variable, while the y-axis tracks the change in the conditional mean as a function of the individual state variables.

1927-1967



1968-2008

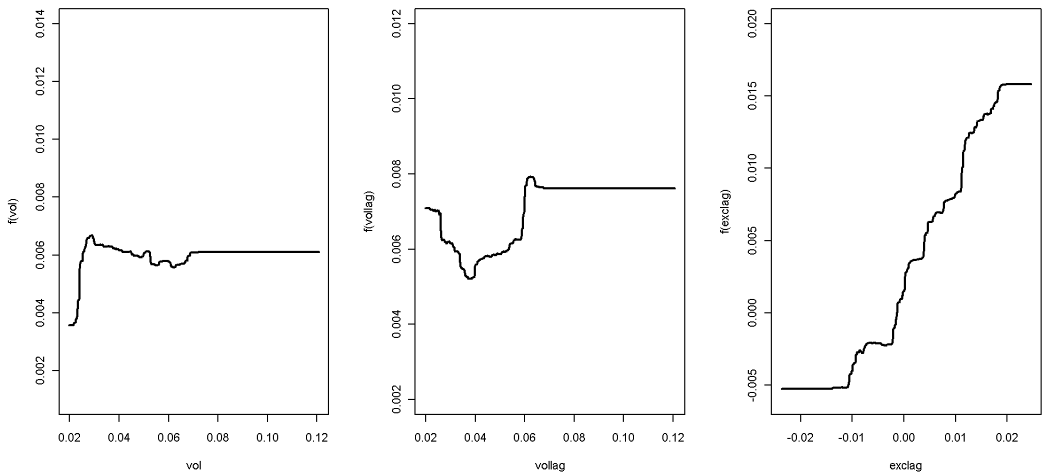
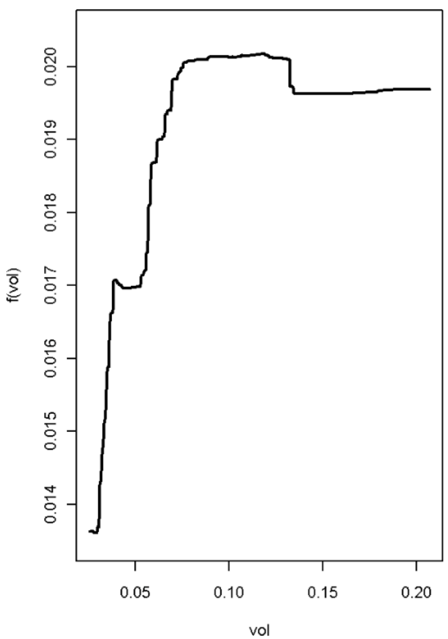
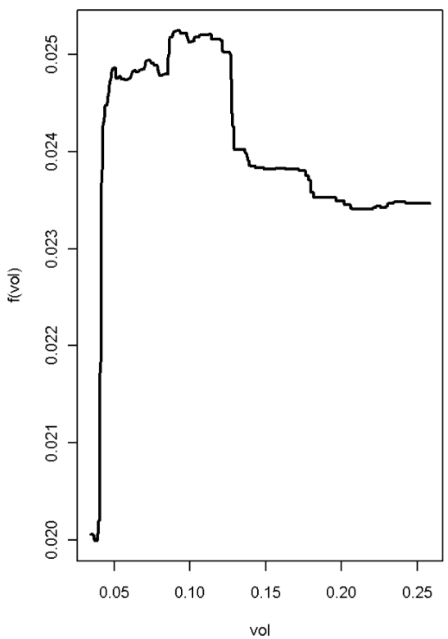


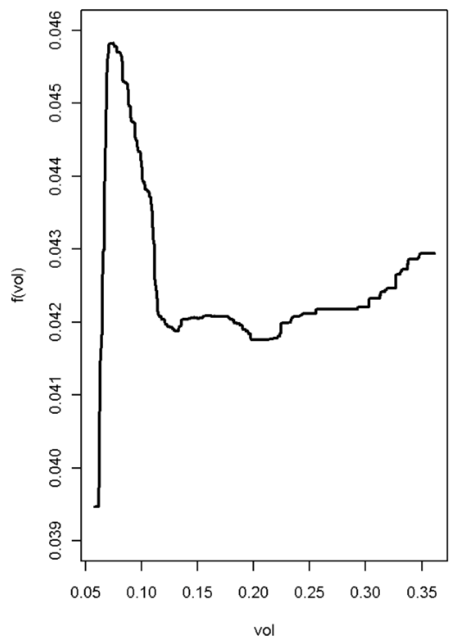
Figure 8: Partial dependence plots for the conditional equity premium as a function of the conditional volatility (vol), past volatility (vollag) and the lagged conditional mean return (exclag). The plot is based on a Boosted Regression Tree with 10,000 boosting iterations, using data for the two sub-samples, 1927-1967 (top graphs) and 1968-2008 (bottom graphs). The x-axis covers the sample support of each predictor variable, while the y-axis tracks the change in the conditional mean as a function of the individual state variables.



(a) **horizon: 2 months**

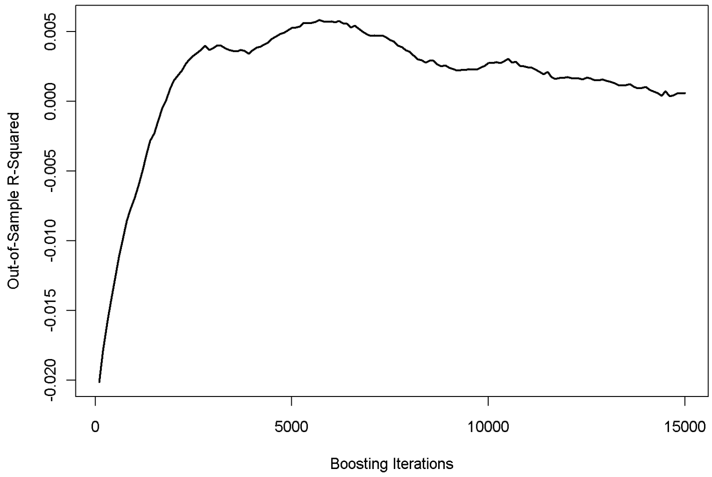


(b) **horizon: 1 quarter**



(c) **horizon: 6 months**

Figure 9: Partial dependence plots for the conditional mean return as a function of the conditional volatility (vol) for different horizons: 2 months (panel (a)); 3 months (1 quarter, panel (b)) and 6 months (panel (c)). All plots are based on regression trees with 10,000 boosting iterations, using data from 1927-2008. The x-axis covers the sample support of each predictor variable, while the y-axis tracks the change in the conditional mean as a function of the individual state variables.



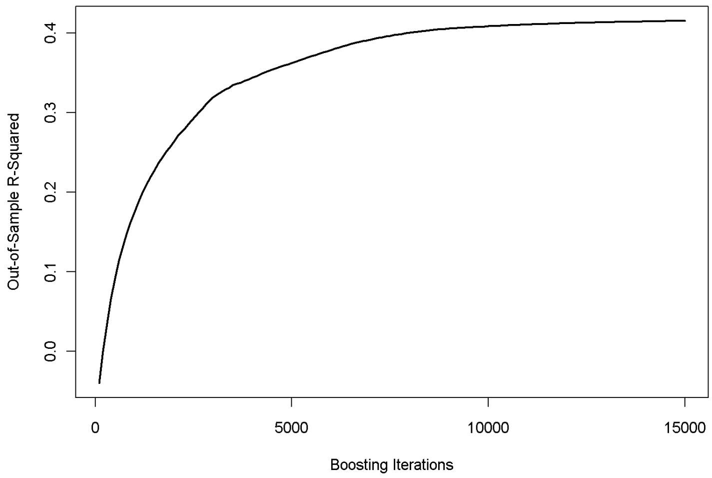


Figure 10: Out-of-sample forecasting performance of the boosted regression trees (measured by the R-squared) as a function of the number of boosting iterations (listed on the x-axis). The top panel shows results for excess returns, while the bottom panel covers the realized volatility. In both cases, results are based on the sample 1970-2008.

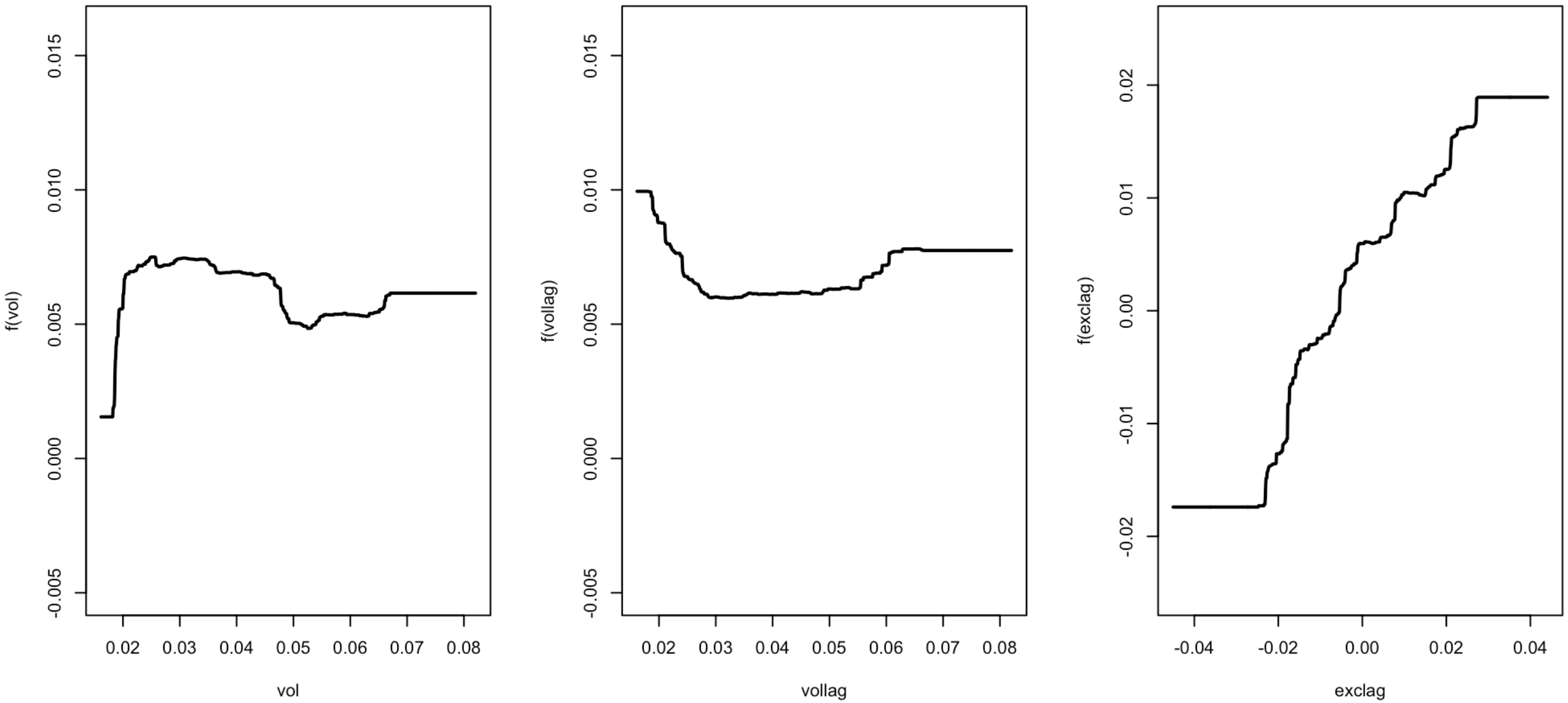


Figure 11: Partial dependence plots for the conditional mean return as a function of the conditional volatility (vol), past volatility (vollag) and the lagged conditional mean return (exclag). The plot is based on a regression tree with 10,000 boosting iterations, using data from 1960-2007. The conditioning information is the principal components derived from a set of 131 state variables and the three most important predetermined variables selected from the 12 predictors reported in Table 1.