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

Alberto Rossi[*] Allan Timmermann[‡]

[*] *Federal Reserve Board of Governors*

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

February 13, 2012

Abstract

Using a novel and flexible regression approach that avoids imposing restrictive modeling assumptions, we find evidence of a non-monotonic relation between conditional volatility and expected stock market returns. At low and medium levels of conditional volatility there is a positive risk-return trade-off, but this relation is inverted at high levels of volatility. Conventional linear risk-return models are strongly rejected by the data. We propose a new measure of risk based on the conditional covariance between observations of a broad economic activity index and stock market returns. Using this covariance-based risk measure, we find clear evidence of a positive and monotonic risk-return trade-off.

Keywords: risk-return trade-off. Stock market volatility. Covariance risk. Boosted regression trees.

[†]We thank seminar participants at the 2010 AFA meetings in Atlanta (especially Claudia Moise, the discussant), 2010 EFA meetings in Frankfurt (especially Skander Van den Heuvel, the discussant), 2010 FMA meetings in New York (especially Qiang Kang, the discussant), Brigham Young University, University of Texas at Austin, University of Houston, Rice, University of Toronto (Fields Institute), UCSD, BlackRock, the CREATES-Stevanovich Center 2009 conference, the SoFie 2009 conference in Geneva and the 2010 FGV conference in Rio. Allan Timmermann acknowledges support from CREATES funded by the Danish National Research Foundation. The views expressed herein are those of the authors and not necessarily those of the Board of Governors of the Federal Reserve System or of the Federal Reserve System.

[*] Federal Reserve Board of Governors: alberto.g.rossi@frb.gov.

[‡] University of California San Diego, CREATES: atimmermann@ucsd.edu

1 Introduction

The existence of a systematic trade-off between market 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), Harrison and Zhang (1999), Ghysels, Santa-Clara, and Valkanov (2005), Guo and Whitelaw (2006), 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 generally imply a linear or even monotonic, risk-return relation. 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 market returns is largely unrestricted with increasing, decreasing, flat, or non-monotonic patterns all possible. Similar conclusions are drawn by studies such as Abel (1988), Gennotte and Marsh (1993) and Veronesi (2000).[1] It follows that the conventional practice of measuring the risk-return trade-off by means of a single slope coefficient in a linear model offers too narrow a perspective and can lead to biased results since it limits the analysis to monotonic or flat relations. Typically, the risk-return trade-off cannot be summarized in this manner without making strong auxiliary modeling assumptions whose validity need to be separately tested. Instead it is necessary to consider the shape of the entire risk-return relation at different levels of risk.

This paper introduces a novel and flexible regression approach that does not impose strong modeling assumptions such as linearity to analyze the shape of the risk-return relation. Our 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−we obtain smooth and stable estimates that let us map the shape of the risk return relation as well as empirically test if it is monotonic.

[1] Merton (1980), Eq. (2.1), also considers nonlinearities but assumes a monotonic mean-volatility relation.

We adopt the boosted regression tree approach to empirically analyze the risk-return relation for US monthly stock returns over the period 1927-2008 and find strong empirical evidence of a non-monotonic risk-return relation. At low and medium levels of conditional volatility, there is a strongly positive relation between the conditional mean and volatility of stock market returns. Conversely, at high levels of volatility, the relation appears to be flat or inverted, i.e., higher levels of conditional volatility are associated with lower expected returns. Formal statistical tests that account for sampling error soundly reject a monotonically increasing mean-volatility relation.

An obvious reason for the difference between these new findings and existing ones is that previous studies have generally assumed a linear model for the mean-volatility (or mean-variance) relation. Such an assumption may lead to biased estimates. We provide empirical tests which show that linear models for the risk-return relation are clearly misspecified. In fact, a simple piece-wise linear regression model shows that there is a strongly positive risk-return trade-off at low-to-medium levels of conditional volatility, but a significantly negative trade-off at high levels of conditional volatility.

A second reason why our results differ from previous estimates comes from differences in the underlying mean and volatility estimates. Empirical evidence in Glosten, Jagannathan, and Runkle (1993) and Harvey (2001) suggests that inference on the risk-return relationship can be very sensitive to how the expected return and volatility models are specified.[2] Model and estimation errors can bias results if the estimated models for expected returns or conditional volatility are misspecified as a result of using overly restrictive models or including too few predictor variables (Ludvigson and Ng (2007)). Our model specification tests show that linear models used to compute expected returns and conditional volatility are misspecified even when a large set of conditioning variables is used. The same holds for the variance estimates based on sophisticated GARCH or MIDAS models. Due to the difficulty of maintaining a flexible functional form for the conditional mean and volatility models while also considering a large conditioning information set, it has proven difficult to effectively address the resulting bias. However, we find that the boosted regression tree approach accomplishes this, passes the model specification tests, and finds evidence of a highly nonlinear effect of many predictor variables on both expected returns and conditional volatility.

[2] Lettau and Ludvigson (2009) conclude that "the estimated risk-return relation is likely to be highly dependent on the particular conditioning variables used in any given empirical study."

Dynamic asset pricing models provide economic intuition for our findings of a non-monotonic risk-return relation. In consumption asset pricing models, the conditional equity premium depends on the correlation between stock returns and the marginal rate of substitution between current and next period's consumption. Volatility of stock returns reflect changes in expectations on the entire infinite future stream of consumption and asset payoffs. In a model with regime switching in the consumption process Whitelaw (2000) shows that this difference in horizons can reduce the relation between stock market returns and the marginal rate of substitution. States with high probability of transitioning to a new regime and hence high levels of uncertainty can have low expected returns but high conditional volatility, while the traditional positive risk-return relation emerges in more "normal" and less uncertain states of the world. Combining these effects leads to a non-monotonic, inverted risk-return relation.

When the dividend and consumption processes can differ, time-varying heteroskedasticity in the dynamics of these processes can create a further wedge between marginal utility of consumption and asset payoffs and thus between expected returns and volatility. States with high uncertainty of future consumption growth and high return volatility need not have high expected returns if the correlation between consumption and dividend growth is simultaneously low. This happens when the stock market provides a partial hedge against adverse consumption states because dividends, and hence the price-dividend ratio can be high in such states. In the context of a simple asset pricing model based on the analysis in Garcia, Meddahi, and Tédongap (2008) and Bonomo, Garcia, Meddahi, and Tédongap (2011), we show how this can again lead to a non-monotonic risk-return relation similar to the one observed in the data.

Common to these models is that conditional stock market volatility is not an appropriate measure of risk. Indeed, the consumption CAPM (Breeden (1979)) suggests the covariance between returns and consumption growth as the appropriate measure of risk, while the intertemporal CAPM (Merton (1973)) adds a set of hedge factors tracking time-varying investment opportunities. To address these points, we construct a new "realized covariance" risk measure based on daily changes in the broad economic activity index developed by Arouba, Diebold, and Scotti (2009) and daily stock market returns. Consistent with the above asset pricing models we find evidence of a strongly positive and monotonic relation between conditional covariance risk and expected returns. From an economic perspective, variations in the conditional covariance lead to far greater changes in expected returns than those associated with variations in conditional volatility.

Our analysis generalizes and provides a synthesis of many existing approaches from the literature on 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. Like these authors, we consider a large set of conditioning variables to compute the conditional equity premium and market volatility. Linear mean-variance models such as Bollerslev, Engle, and Wooldridge (1988), and Ghysels, Santa-Clara, and Valkanov (2005) arise as special cases of our setup when the variance estimates are based on past (squared) returns. Guo and Whitelaw (2006) argue that findings of a negative or insignificant mean-variance relation is due to the omission of an intertemporal hedging component leading to a downward bias in the variance coefficient. We consider a model that includes both variance and covariance terms and find that our results are robust to the inclusion of both measures of risk. Following papers such as Harrison and Zhang (1999), we do not impose monotonicity on the risk-return relation, but allow its shape to be freely estimated. Our use of boosted regression trees bears similarities to forecast combinations and thus incorporates advantages of this approach for return forecasting purposes, a point recently emphasized by Rapach, Strauss, and Zhou (2010).

Our results are found to be robust in several dimensions. Overfitting is a potential concern whenever flexible methods such as boosted regression trees are adopted. To address this issue, we conduct an out-of-sample forecasting exercise that compares the precision of the boosted regression tree forecasts with forecasts from linear return models and GARCH and MIDAS volatility forecasting models. For both expected returns and volatility, we find that the boosted regression tree forecasts are more precise than those produced by conventional forecasting models, thus suggesting that overfitting is not a concern. When we account for measurement errors in the conditional volatility or covariance proxies through an instrumental variables procedure, we also find that these do not significantly impact our empirical findings. Finally, we show that our results are robust to how the boosted regression trees are implemented.

In summary, the main contributions of our paper are as follows. First, we present a new, flexible modeling approach that reduces the risk of biases in estimates of expected returns and conditional volatility. Second, we use this approach to analyze empirically the relation between the expected return and conditional volatility without imposing restrictions on the shape of this relation. Using U.S. stock returns, we present evidence of a non-monotonic mean-volatility relation with expected returns first rising, then declining as the conditional volatility further increases. Third, we use asset pricing models to gain insights into the type of economic mechanism that can induce the non-monotonic risk-return relation observed in the data. Fourth, we propose a new conditional covariance risk measure that builds on the covariation between daily stock returns and daily economic activity. Finally, we show empirically that when this broad conditional covariance is used to measure risk, a strongly increasing and monotonic risk-return relation emerges.

The remainder of the paper is organized as follows. Section 2 introduces our approach to modeling the risk-return relation. Section 3 describes the data, analyzes the risk-return relation empirically, compares our approach to existing methods and discusses reasons why our empirical findings differ from previously reported ones. Section 4 provides economic intuition for the non-monotonic risk-return relation in the context of dynamic asset pricing models. Section 5 introduces the new covariance measure of risk, while Section 6 conducts a series of robustness checks and extensions, and Section 7 concludes.

2 Empirical Methodology

Dynamic asset pricing models do not generally restrict the relation between conditional market volatility and expected returns to be linear. To quote from 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.

This section describes a new modeling approach that avoids imposing shape restrictions on the risk-return relation or on the models used to generate expected returns and conditional volatility estimates, while allowing for large-dimensional state variables. The approach uses regression trees. To get intuition for how these work and establish the appropriateness of their use in our analysis, consider the situation with a single dependent variable y_{t+1} (e.g., stock returns) and two state variables, x_{1t} and x_{2t} (e.g., the earnings-price ratio and the payout ratio), so that interest lies in modeling expected returns using ex-ante regressors. The functional form of the model mapping x_{1t} and x_{2t} into y_{t+1} is unlikely to be known, so we simply partition the sample support of x_{1t} and x_{2t} into a set of regions or "states" and assume that the dependent variable is constant within each partition.

Specifically, we first split the sample support into two states and compute the mean of y in each state. We choose the state variable (x_1 or x_2) and the splitting point to achieve the best fit. Next, one or both of these states is split into two additional states. Boosted regression trees are additive expansions of regression trees, where each additional tree is fitted on the residuals of the previous tree until some stopping criterion is reached. The number of trees used in the summation is known as the number of boosting iterations.

An economic illustration of the approach is provided in Figure 1, which shows boosted regression trees that use the (lagged) log payout ratio (i.e., the dividend-earnings ratio) and the log earnings-price ratio to predict excess returns on the S&P500 index. Each iteration fits a tree with 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. With only three boosting iterations the model is quite coarse and shows that the expected return is at its lowest (highest) when the payout ratio is high (low) and the earnings-price ratio is low (high). The fit improves as more boosting iterations are added. As an illustration, the figure on the right is based on 5,000 boosting iterations. Now the plot is much smoother, but clear similarities between the two graphs remain.

We next provide a more formal description of the methodology and explain how we implement it in our analysis. Our description draws on Hastie, Tibshirani, and Friedman (2009) who provide a more in-depth coverage of the approach.

2.1 Regression Trees

Consider a time-series with T observations on a single dependent variable, y_{t+1}, and P predictor (state) variables, $x_t = (x_{t1}, x_{t2}, ..., x_{tp})$, for $t = 1, 2, ..., T$. As illustrated in Figure 1, implementing a regression tree requires deciding, first, which predictor variables to use to split the sample space and, second, which split points to use. A given split 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. For example, in Figure 1 there are $J = 2$ nodes at each split point. For the resulting states c_j corresponds to the value of the flat spots on the vertical (expected return) axis. The value fitted by a regression tree, $\mathcal{T}(x_t, \Theta_J)$, with J nodes and parameters $\Theta_J = \{S_j, c_j\}_{j=1}^J$ can thus be written

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

where the indicator variable $I\{x_t \in S_j\}$ equals one if $x_t \in S_j$ 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, the estimated constant, $\widehat{c}_j$, is the average of y_{t+1} in state S_j:

$$\widehat{c}_j = \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_j\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_j\}. \tag{2}$$

Optimal splitting points are more difficult to determine, particularly in cases where the number of state variables, P, is large, but sequential algorithms have been developed for this purpose.

Regression trees are very flexible and can capture local features of the data that linear models overlook. Moreover, they can handle cases with large-dimensional data. This becomes important when modeling stock returns because the identity of the best predictor variables is 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 overfitting. There is also no guarantee that the sequential splitting algorithm leads to the globally optimal solution. To deal with these problems, we next consider a method known as boosting.

2.2 Boosting

Boosting is based on the idea that combining a series of simple prediction models can lead to more accurate forecasts than those available from any individual model. Boosting algorithms iteratively re-weight data used in the initial fit by adding new trees in a way that increases the weight on observations modeled poorly by the existing collection of trees. By summing over a sequence of trees, boosting performs a type of model averaging that increases the stability of the forecasts.

A boosted regression tree (BRT) is simply the sum of individual regression trees:

$$f_B(x_t) = \sum_{b=1}^{B} \mathcal{T}_b(x_t; \Theta_{J,b}), \tag{3}$$

where $\mathcal{T}_b(x_t, \Theta_{J,b})$ is the regression tree of the form (1) used in the b-th boosting iteration and B is the total number of boosting iterations. Given the previous model, $f_{B-1}(x_t)$, 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

$$\hat{\Theta}_{J,B} = \arg\min_{\Theta_{J,B}} \sum_{t=0}^{T-1} \left[e_{t+1,B-1} - \mathcal{T}_B(x_t, \Theta_{J,B})\right]^2, \tag{4}$$

where $e_{t+1,B-1} = y_{t+1} - f_{B-1}(x_t)$ is the forecast error remaining after $B-1$ boosting iterations. The solution 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. Figure 1 shows that as the number of boosting iterations increases, the area covered by individual states shrinks and the fit becomes better.

Boosting makes it more attractive to employ small trees at each boosting iteration, thus reducing the risk that the regression trees will overfit. Our estimations therefore use $J = 2$ nodes and follow the stochastic gradient boosting approach of Friedman (2001) and Friedman (2002). The baseline implementation employs B =10,000 boosting iterations. Robustness analysis reveals that the results are not sensitive to this choice.

The literature on ensemble learning (e.g. Dietterich (2000)) suggests various ways in which the learning rate of the BRTs can be controlled, and we adopt three common refinements to the basic regression tree methodology, namely (i) shrinkage, (ii) subsampling, and (iii) minimization of absolute errors. These techniques are all known to decrease the rate at which the average forecast errors are minimized on the training data and hence reduce the risk of overfitting (e.g., Hastie, Tibshirani, and Friedman (2009)).

Specifically, following Friedman (2001) we use a small shrinkage parameter, i.e., $\lambda = 0.001$, that reduces the amount by which each boosting iteration contributes to the overall fit:

$$f_B(x_t) = f_{B-1}(x_t) + \lambda \sum_{j=1}^{J} c_{j,B} I\{x_t \in S_{j,B}\}. \tag{5}$$

This procedure reduces the ability of the algorithm to overfit individual outlier observations such as October 1987.

Each tree is fitted 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. Again this reduces the risk of overfitting. Importantly, since individual trees are fitted on different subsets of the data, the starting point of the algorithm, or more specifically the particular sequence of splits selected by the regression trees, has very little effect on our results.

Finally, the empirical analysis minimizes mean absolute errors. We do this in light of a large literature suggesting that squared-error loss places too much weight on observations with large 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 more robust to outliers.

To see how the resulting BRTs can flexibly approximate a range of true relations by means of a series of piece-wise constant functions, Figure 2 plots the fitted values for three different shapes−linear, inverse-V, and linear-quadratic−using one, five and 10,000 boosting iterations. With only a single boosting iteration and two nodes, the BRTs simply classify the data into high and low values. For the inverse-V or linear-quadratic shapes, the BRTs capture the nonlinear relation with only five iterations and with 10,000 iterations the fit gets very good, with only peak values missing out. In contrast, the linear model is clearly misspecified suggesting that the BRT estimates are more robust and can capture a wide range of patterns.

2.3 Measuring the Effect of Individual Variables

In a linear model the importance of a particular state variable can be measured through the magnitude and statistical significance of its slope coefficient. This measure is not applicable to regression trees since these do not impose linearity. As an alternative measure of influence, we instead consider the reduction in the forecast error every time a particular variable, x_p, is used to split the tree. Summing the reductions in forecast errors (or improvements in fit) across the nodes in the tree and across boosting iterations gives a measure of each variable's influence (Breiman (1984)). The more frequently a variable is used for splitting and the bigger its effect on reducing the forecast errors, the greater its influence. If a variable never gets chosen to conduct the splits, its influence will be zero. Finally, the resulting measure of influence is divided by the summed influence across all variables to get a measure of relative influence. This sums to one and can be compared across predictor variables.

Similarly, we can compute the marginal effect of one state variable, X_p, on the dependent variable by fixing the value of the state variable and averaging out the effect of the remaining variables. Repeating this process for different values of X_p yields a partial dependence plot that shows the effect individual state variables have on the dependent variable.

3 Empirical Risk-Return Estimates

This section presents estimates of the risk-return relation, contrasting results from conventional linear models with those obtained using the boosted regression tree approach described in Section 2. We first present the data used in our empirical analysis and then report empirical results.

3.1 Data

Our empirical analysis of the risk-return trade-off relies on proxies for the conditional expectation of stock returns and the conditional volatility. Following the extensive empirical literature on time variations in both expected stock market returns and volatility (e.g., Lettau and Ludvigson (2009)), these are constructed using a broad range of state variables.

Specifically, our empirical analysis uses a data set comprising monthly stock returns along with a set of predictor variables previously analyzed in Welch and Goyal (2008) extended to cover the sample 1927-2008.[3] Stock market 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 fall into three broad categories. First, there are valuation ratios capturing some measure of 'fundamental' value to market value such as the log dividend-price

[3]A few variables were excluded from the analysis since they were not available up to 2008. We also excluded the CAY variable since this is only available quarterly since 1952.

ratio and the log earnings-price ratio. Second, there are bond yield measures capturing the level or slope of the term structure or measures of default risk such as the three-month T-bill rate, the de-trended T-bill rate, i.e., the T-bill rate minus a three-month moving average, the yield on long term government bonds, the term spread measured by the difference between the yield on long-term government bonds and the three-month T-bill rate, and the default yield spread measured by the yield spread between BAA and AAA rated corporate bonds. Third, there are estimates of equity risk and returns such as the lagged excess return, long term (bond) returns, and stock variance, i.e., a volatility estimate based on daily squared returns. Finally, we also consider the dividend payout ratio measured by the log of the dividend-earnings ratio and the inflation rate 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 $t+1$.

Following the analysis in Ludvigson and Ng (2007), we also consider a much larger information set. Specifically, suppose that a large set of state variables $z_{it},\ i = 1, ..., N$ are generated by a factor model of the form $z_{it} = \lambda_i' f_t + e_{it}$,where f_t is a vector of common factors, λ_i is a set of factor loadings, and e_{it} is an idiosyncratic error. Using common factors as predictor variables rather than the N individual regressors achieves a substantial reduction in the dimension of the information set. We follow Ludvigson and Ng (2007) and extract factors through the principal components method. Their data contain $N = 131$ economic time series for the period 1960-2007. By considering this large set of predictor variables, we address a potentially important source of model misspecification caused by omitted variables.

Market variance is unobserved, so we follow a large recent literature in proxying it through the realized variance. 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 that month. Following, e.g., French, Schwert, and Stambaugh (1987) and Schwert (1989) we construct the realized variance measure

$$\hat{\sigma}_t^2 = \sum_{i=1}^{N_t} r_{i,t}^2. \tag{6}$$

This estimator is only free of measurement errors as the sampling frequency approaches infinity, so $\hat{\sigma}_t^2$ is best thought of as a variance proxy.

3.2 Estimates of Expected Return and Conditional Volatility

Let r_{t+1} be the market return during period $t+1$, measured in excess of the risk-free rate. Estimates of the expected excess return, $\mu_{t+1|t} = E_t[r_{t+1}]$, and the conditional volatility, $\sigma_{t+1|t} = Var_t(r_{t+1})^{1/2}$, are computed conditional on information known to investors at time t. Both are unobserved and so empirical analysis typically relies on model-based proxies of the form

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

Here x_t is 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. We provide further details of these estimates below, but first turn to the risk-return model.

3.3 Linear Estimates of the Risk-Return Relation

Following Ludvigson and Ng (2007) we first consider a reduced-form relation that models the conditional equity premium as a linear function of the conditional volatility:

$$\hat{\mu}_{t+1|t} = \alpha + \beta_1\hat{\sigma}_{t+1|t} + \beta_2\hat{\sigma}_{t|t-1} + \beta_3\hat{\mu}_{t|t-1} + \varepsilon_{t+1}, \tag{8}$$

where 'hats' indicate estimated values from the boosted regression trees. In a generalization of the conventional volatility-in-mean model, lags are included to account for the complex lead-lag relation between the conditional mean and volatility, see, e.g., Whitelaw (1994) and Brandt and Kang (2004).

Empirical estimates of this model are shown in Panel A1 of Table 1. 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 of 2.7. Conversely, the effect of lagged volatility is strongly negative, while the effect of lagged expected returns is strongly positive. Although these results carry over to the first subsample, 1927-67, they are not stable. In the second subsample, 1968-2008, the relation between the conditional mean and both the current and lagged conditional volatility is insignificant and much weaker. Similar findings hold for the comparable sample, 1960-2007, used to obtain factor-based estimates of the mean and volatility.

While Ludvigson and Ng (2007) use the conditional volatility as their risk measure, it is more common to use the conditional variance, so we also consider the following linear mean-variance specification

$$\hat{\mu}_{t+1|t} = \alpha + \beta_1 \hat{\sigma}^2_{t+1|t} + \beta_2 \hat{\sigma}^2_{t|t-1} + \beta_3 \hat{\mu}_{t|t-1} + \varepsilon_{t+1}. \tag{9}$$

For this specification, Panel B1 shows that the estimated slope coefficient on the conditional variance term is positive but not statistically significant in any of the samples or when common factor estimates are used to obtain the underlying moments.[4]

To see if the linear mean-volatility or mean-variance models are correctly specified, we undertake a set of non-parametric specification tests. These results, reported in Panel A in Table 2, show that linearity of the mean-volatility or mean-variance relation is strongly rejected for both the baseline variance specification that uses the full sample from 1927-2008 and for the factor model based on the shorter sample 1960-2007. Linearity of the mean-variance relation continues to be rejected in the full sample, though not always in the shorter sample, when the variance estimates are based on GARCH (Bollerslev, Engle, and Wooldridge (1988)) or MIDAS (Ghysels, Santa-Clara, and Valkanov (2005)) models. Both of these approaches model current conditional variance as a function of past return shocks measured either at the monthly or daily horizon.

As a simple first measure of the nature of the misspecification of the linear risk-return relation, we adopt the threshold regression approach of Hansen (2000) and estimate a piece-wise linear model relating the expected return to the conditional variance:

$$\hat{\mu}_{t+1|t} = \left\{ \begin{array}{ll} \begin{array}{ccc} -0.003 & + & 25.538 \\ (-1.49) & & (6.23) \end{array} \hat{\sigma}^2_{t+1|t} & \text{for } \hat{\sigma}^2_{t+1|t} \leq (0.0253)^2 \\ \begin{array}{ccc} 0.009 & - & 0.542 \\ (20.73) & & (-5.22) \end{array} \hat{\sigma}^2_{t+1|t} & \text{for } \hat{\sigma}^2_{t+1|t} > (0.0253)^2 \end{array} \right. \tag{10}$$

At low levels of the conditional variance (corresponding to annualized volatility levels below 9%), a strongly positive and significant mean-variance relation emerges. In contrast, for higher values of the conditional variance, the relationship is negative and strongly significant. A linearity

[4] These are OLS estimates and ignore measurement errors in the proxies for the conditional variance. Section 6.3 shows how to deal with this problem using instrumental variables estimation and demonstrates that this has only a minor effect on the results.

test for equal slope of the two segments is rejected with a p-value well below 0.01%.[5] This demonstrates the limitations of linear and monotonic models for the relation between the conditional mean and variance and indicates that the risk-return relation is inverted at high levels of volatility.

An alternative to using model-based estimates of the conditional volatility is to adopt a market-based estimate of conditional volatility, namely the Chicago Board Options Exchange Index, commonly known as the VIX. The VIX is effectively a market-based estimate of the volatility of the S&P 500 index over the next 30 days. Data on the VIX are available over the period 1986-2008. Table 1 shows results for the linear risk specification based on the VIX measure of conditional volatility or its square. In both cases the coefficient on current VIX is positive but statistically insignificant.

3.4 Flexible Risk-Return Model

Given the evidence that the linear model is clearly misspecified, we next turn to the more flexible risk-return model based on the boosted regression trees which do not impose particular functional form assumptions. Specifically, consider the following generalization of Eqs. (8-9):

$$\hat{\mu}_{t+1|t} = f(\hat{\sigma}_{t+1|t}, \hat{\sigma}_{t|t-1}, \hat{\mu}_{t|t-1}), \tag{11}$$

where now f is estimated by means of the BRTs. As noticed in Section 2.3, we can no longer measure the importance of the explanatory variables through their slope coefficients, so instead Panels A2 and B2 in Table 1 present estimates of the relative influence of the three variables in this model using either $\hat{\sigma}_{t+1|t}$ or $\hat{\sigma}^2_{t+1|t}$ as the measure of risk.[6] The relative weight on current conditional volatility is 8% for the full sample which is statistically significant at the 5% level.[7]

[5] For simplicity we have omitted the lagged expected return and conditional variance from Eq. (10), but we continue to reject the linear specification when these terms are included. Nearly identical results were obtained when we used the conditional volatility in place of the conditional variance in Eq. (10).

[6] For the flexible risk-return model it should make no difference whether we use $\hat{\sigma}_{t+1|t}$ or $\hat{\sigma}^2_{t+1|t}$ as the measure of risk since the squaring can be undone by the BRT. This is consistent with what we find in Table 1 with only very minor differences (due to random sampling) between the results in Panels A2 and B2.

[7] To assess the statistical significance of the relative influence of individual variables, we undertake the following analysis. We fix the ordering of the dependent variable and all explanatory variables except for one variable, whose values are redrawn randomly in time. We then calculate the relative influence measure for the data with the reshuffled variable. Because any relation between the randomized variable and expected returns is broken, we would expect to find a lower value of its relative influence, any results to the contrary reflecting random sampling variation. Repeating

This weight is similar to that of the lagged volatility (9%). The weight on the lagged expected return (83%) is higher, which is unsurprising since the expected return is quite persistent and so its lagged value is likely to be important in this model. This finding is also consistent with the larger coefficient and t-statistic on lagged expected returns in the linear model. Interestingly, in the first subsample, 1927-67, the current conditional volatility obtains a large (and significant) weight of 21%, but the weight declines to 10% in the second subsample, 1968-2008, where it fails to be significant.

Panel A in Table 2 reports model specification tests for the BRTs fitted on the risk-return data. In contrast with the linear model, we find no evidence of misspecification for the BRT model, indicating that this approach captures the shape of the risk-return relation much better and suggesting that the true risk-return relation is nonlinear.

Figure 3 plots the marginal effect of the conditional volatility on expected returns, obtained by integrating across the lagged values of volatility and expected returns. The figure confirms that the trade-off between concurrent expected returns and conditional volatility is highly nonlinear in all three samples. At low-to-medium levels of volatility, a strongly positive relation emerges where higher conditional volatility is associated with higher expected returns. As volatility rises further, the relation flattens out and, at high levels of conditional volatility, it appears to be inverted so higher conditional volatility is associated with declining expected returns. Section 6.2 shows how a Bayesian modification to the boosted regression trees can be used to provide standard errors on these plots.

Our analysis assumes a constant risk-return relation. However, we can at least in part address this issue by applying our methodology to subsamples of the data. As shown in the middle and right plots in Figure 3, the inverted risk-return trade-off at medium-high levels of volatility is a robust finding in the sense that it appears not to be confined to a particular historical period.[8]

A possible concern with these findings is that the flat and decreasing parts of the risk-return plot could be driven by relatively few observations. However, this does not seem to be the case. In the full-sample plot in Figure 3, 37% of the observations lie to the left of the steeply

this experiment a large number of times and recording how often the randomized relative influence measure exceeds the estimated empirical value from the actual data, we obtain a p-value for the significance of the individual variables.

[8]The non-monotonic risk-return relation is related to the finding by Brandt and Wang (2007) that, while the risk-return relation is mostly positive, it varies considerably over time and is negative for periods around the oil price shocks of the early 1970s, the monetarist experiment, 1979-81, and again around the recession of 2000-2001. Those are all periods associated with greater than normal volatility and so these findings are closely related to our results.

increasing part, while 63% lie on the flat and declining parts. For the first subsample, 75% of the observations lie to the left of the peak of the graph while for the second subsample, 65% of the observations lie to the left of the peak. These numbers do not suggest that the shape of the graphs are driven by a few outliers.

3.5 Comparison with Existing Approaches

Differences between the findings reported here versus earlier estimates can be attributed either to differences in how the risk-return relation is modeled or to the use of different estimates of the conditional mean and variance of market returns. We already addressed the first point and now turn to the second point.

Differences in estimates of market variance and expected returns turn out to be very important for the risk-return relation. This finding is consistent with Glosten, Jagannathan, and Runkle (1993), Harvey (2001) and Lettau and Ludvigson (2009) who conclude that empirical analysis of the risk-return is highly sensitive to changes in the mean-volatility estimates. Indeed, in unreported results we find that the coefficient in a regression of the conditional mean on the conditional volatility is highly sensitive to whether the first-stage conditional mean and volatility estimates are based on a linear model−in which case the coefficient is significant and negative−or on the boosted regression tree, which leads to a positive and significant coefficient.

It is therefore important to avoid using misspecified models when generating estimates of the expected return and conditional volatility. To see the relevance of this point, Panel B in Table 2 reports a set of diagnostic tests for whether the first-stage estimates of the conditional mean and variance are misspecified. Linear models for the expected return and conditional variance (or volatility) are clearly misspecified. In contrast, the BRTs do not appear to be misspecified for the expected return and conditional variance. Furthermore, both the GARCH variance model adopted by Bollerslev, Engle, and Wooldridge (1988) and the MIDAS model of Ghysels, Santa-Clara, and Valkanov (2005) appear to be misspecified.

To understand why linear models for the volatility and expected returns are misspecified, we study the BRT estimates of the conditional mean and volatility in more detail. These are unconstrained and so are able to reveal the nature of any deviations from linearity. The top row in Figure 4 presents partial dependence plots for the three most important predictor variables in the BRT model for expected returns, namely inflation, the earnings-price ratio and the relative return. The relation between expected stock returns and these predictor variables is highly nonlinear. At negative levels of inflation the relation between the rate of inflation and expected returns is either flat or rising. Conversely, at positive levels of inflation, higher consumer prices are associated with lower mean returns. Although the relation 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.

Turning to the volatility plots in the middle row, the predicted volatility quadruples from roughly 2% to 8% per month as the lagged realized volatility increases over its historical support. The relation between current and past volatility is basically linear for small or medium values of past volatility but very high values of past volatility do not translate into corresponding high values of expected future volatility, as evidenced by the flatness of the relation at high levels of volatility. A highly nonlinear pattern is also found in the relation between the conditional volatility and the default spread or past returns.

This evidence indicates that conventional linear risk-return models may get rejected not only because they are intrinsically misspecified but also because they rely on misspecified proxies for the conditional mean and volatility. Hence, it is important to use a flexible modeling approach in both stages of the analysis.

3.6 Tests of Monotonicity

The results reported so far suggest that expected market returns rise when the conditional volatility goes from low to medium levels. The opposite finding holds for periods with medium-to-high levels of conditional volatility, where rising volatility is associated with constant or declining expected returns. While the plots in Figure 3 suggest marked non-monotonicities in the mean-volatility relation, they do not demonstrate that this relation is non-monotonic in a statistically significant way.

To formally test if the relation between the conditional volatility and expected returns is monotonic in a statistical sense, we use the approach in Patton and Timmermann (2010). We sort pairs of monthly observations into $g = 1, .., G$ groups, $\{\hat{\mu}^g_{t+1|t}, \hat{\sigma}^g_{t+1|t}\}$ and then rank them by the conditional volatility estimate. A monotonic mean-volatility relation implies that, as we move from groups with low conditional volatility to groups with high conditional volatility, mean returns should rise.[9]

We seek to test whether the conditional expected return increases when ranked by the associated value of $\hat{\sigma}^g_{t+1|t}$:

$$H_0 : E\left[\hat{\mu}^g_{t+1|t}|\hat{\sigma}^g_{t+1|t}\right] \geq E\left[\hat{\mu}^{g-1}_{t+1|t}|\hat{\sigma}^{g-1}_{t+1|t}\right], \text{ for } g = 2, .., G. \tag{12}$$

Because $\hat{\sigma}^g_{t+1|t} > \hat{\sigma}^{g-1}_{t+1|t}$, this hypothesis says that the expected return associated with observations where the conditional volatility is high exceeds the expected return associated with observations with lower conditional volatility. Defining $\Delta_g \equiv E\left[\hat{\mu}^g_{t+1|t}|\hat{\sigma}^g_{t+1|t}\right] - E\left[\hat{\mu}^{g-1}_{t+1|t}|\hat{\sigma}^{g-1}_{t+1|t}\right]$, for $g = 2, ..., G,$ and letting $\Delta{=}(\Delta_2, \Delta_3, ..., \Delta_G)'$, the null hypothesis can be rewritten as[10]

$$H_0 : \Delta \geq 0. \tag{13}$$

To test this hypothesis, we use the test statistic of Wolak (1989). The null that the conditional mean increases monotonically in the level of conditional volatility is 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 conditional mean and conditional volatility. The test statistic has a distribution that, under the null, is a weighted sum of chi-squared variables whose critical values can be computed via Monte Carlo simulation.

For robustness, we perform the test on different numbers of groups, G, chosen so that there are 40, 50 and 65 observations per group. Furthermore, because it could be of interest to study the results across different forecast horizons, we compound the monthly returns and compute the associated estimates of the h-month conditional mean and conditional volatility and conduct tests for horizons of $h = 1, 2, 3$ months.[11]

Test results are reported in Panel A of Table 3. At the one-month horizon, we get p-values below 5% irrespective of the number of groups, G. Similar results are obtained for the bimonthly

[9]Since we are interested only in the relation between the concurrent conditional mean and volatility, we integrate out the effects of the lagged variables in Eq. (11). Hence our analysis is based on the relation between the marginalized conditional mean and the marginalized conditional volatility.

[10]Since rankings by $\hat{\sigma}_{t+1|t}$ are identical to rankings by $\hat{\sigma}^2_{t+1|t}$, the tests for monotonicity in the relation between expected returns and conditional volatility also apply to the conditional variance measure of risk.

[11]Going beyond the one-quarter horizon entails a significant decline in sample size and a resulting loss in power.

and quarterly horizons. Panel B shows similar findings for the VIX measure of volatility. These results demonstrate that a monotonically increasing relation between the conditional mean and the conditional volatility is strongly rejected, providing evidence of a nonlinear mean-volatility or mean-variance relation.

4 Why is there a Non-monotonic Risk-Return Relation?

Simple intuition suggests a positive trade-off between stock market volatility and expected returns, so our empirical finding of a non-monotonic relation may at first seem puzzling. In fact, this section shows that many dynamic asset pricing models can generate a non-monotonic or even negative relation between expected stock returns and market volatility.

In many dynamic asset pricing models, expected returns depend not only on the conditional variance of next-period returns but also on how returns are correlated with future shocks to investment opportunities so that the equity premium contains an intertemporal hedging component. For example, in a log-linearized asset pricing model with Epstein-Zin preferences, Campbell (1993) shows that the expected market excess return takes the form:

$$E_t[r_{t+1}] = (\gamma - 0.5)\sigma^2_{t+1|t} + \left[\gamma - 1 - \frac{\theta\kappa}{\psi}\right] Cov_t(r_{t+1}, [E_{t+1} - E_t]\sum_{j=1}^{\infty}\rho^j r_{t+1+j},). \tag{14}$$

Here γ is the coefficient or relative risk aversion, ψ is the elasticity of intertemporal substitution, $\theta = (1-\gamma)/(1-\ \psi^{-1})$, ρ is a linearization constant, and κ measures the sensitivity of consumption with respect to changes in the expected market return. The last term in Eq. (14) measures the covariance between the single period market return, r_{t+1}, and revisions to expectations of all future discounted market returns. The constant in front of this term can be positive or negative, with θ/ψ measuring the market price of consumption risk. A non-monotonic mean-variance relation can arise if the covariance term depends on the level of the variance of the market return, $\sigma^2_{t+1|t}$. Suppose, for example, that during periods with high market volatility the covariance between stock returns and revisions to long-run market return expectations is higher than normal. Since agents do not like to be exposed to this uncertainty, they increase their precautionary savings and lower their consumption. If the market price of investment opportunity set risk $(\theta\kappa/\psi)$ is sufficiently high, this can lead to a non-monotonic shape of the risk-return relation.

As a specific example, in a simple asset pricing model with power utility, Whitelaw (2000) shows that switches between two regimes with large differences in consumption growth can induce a complex, nonlinear relation between expected returns and conditional market volatility. High conditional return volatility is induced by high levels of uncertainty about future states caused by high probabilities of switching to a new regime. Such regime switches can also reduce the correlation between stock returns and the marginal rate of substitution between current and next period's consumption and may lower the equity premium. In states of the world where the stock market portfolio acts as a hedge against adverse shocks to consumption−e.g., when the price-dividend ratio is high in economic downturns−the equity risk premium can be low even when the conditional market volatility is high.[12]

A non-monotonic risk-return relation can alternatively arise because of the dynamics in the fundamentals of the economy and learning effects. David and Veronesi (2009) derive a model in which investors' learning about the unknown state of the economy leads to a V-shaped relation between return volatility and valuation measures such as the price-earnings ratio. Uncertainty and thus return volatility is low in "normal" states of the world since these are highly persistent and thus are associated with only a small probability of large future shifts in the economy. Conversely, "bad" and "good" states are much less persistent than the normal states and so are surrounded by much greater uncertainty about the near future. Asset prices react more strongly to directional information and so are relatively low in the bad state and high in the good state. Conversely, volatility is high in both of these outlier states, giving rise to a V-shaped relation between asset prices and volatility. An inverted V-shaped relation between expected returns and volatility arises in this model if, as one would expect, expected returns are highest in the bad state (where marginal utility is high) and lowest in the good state with low marginal utility. Simulations from the model confirm that this holds as there is a negative relation between the price-earnings ratio and expected returns.[13]

We next provide a more detailed example of these effects using a model that matches the

[12]Similarly, in the simple dynamic exchange economy analyzed by Backus, Gregory, and Zin (1989) and Backus and Gregory (1993), the sign of the risk-return relation can shift from being positive at low-to-medium levels of volatility to becoming negative at medium-to-high volatility levels, taking an inverse V-shaped form. This effect is driven by the volatility properties of the endowment process and the associated expected returns or "risk prices".

[13]We are grateful to Alex David for confirming this point.

non-monotonic risk-return relation observed here.

4.1 Illustration from a Simple Model

As an illustration of how a non-monotonic risk-return relation may arise, we adopt the regime switching model for consumption and dividends proposed by Garcia, Meddahi, and Tédongap (2008) and Bonomo, Garcia, Meddahi, and Tédongap (2011). In common with papers such as Bansal and Yaron (2004), this model allows for state-dependent volatility and distinguishes between consumption and dividends, the former being the payoff on the market portfolio, while dividend payoffs are received by equity owners.

Log consumption and dividend growth, Δc, Δd, follow a process with four states characterized by their mean (μ_c, μ_d), volatility (σ_c, σ_d) and correlation (ρ):

$$\begin{aligned} \Delta c_{t+1} &= \mu_c'\zeta_t + (\sigma_c'\zeta_t)\ \epsilon_{c,t+1} \\ \Delta d_{t+1} &= \mu_d'\zeta_t + (\sigma_d'\zeta_t)\ \epsilon_{d,t+1}, \end{aligned} \tag{15}$$

where the innovations $(\epsilon_{c,t+1},\, \epsilon_{d,t+1})$ are drawn from a normal distribution with zero mean, unit variance and correlation $\rho'\zeta_t$. The state vector, ζ_t, follows a Markov process with transition probabilities collected in a matrix P. $\zeta_t = e_i$ if state i occurs at time t, where e_i is a column vector with zeroes everywhere except in the i^{th} position, which takes the value one.

A representative investor is assumed to have Epstein-Zin preferences as characterized by the following continuation value of investor utility, V_t:

$$V_t = \left\{ (1-\delta) C_t^{1-\frac{1}{\psi}} + \delta \left[\left(E_t \left(V_{t+1}^{1-\gamma} \right) \right)^{\frac{1}{1-\gamma}} \right]^{(1-\frac{1}{\psi})} \right\}^{\frac{1}{1-\frac{1}{\psi}}}. \tag{16}$$

Here γ is the coefficient of relative risk aversion, $\psi \neq 1$ measures the elasticity of intertemporal substitution, and δ is a subjective discount factor. With this in place, Garcia, Meddahi, and Tédongap (2008) show that the conditional mean and variance of excess returns are given by:

$$E_t[R_{t+1}^e] = \left[(\lambda_2'\zeta_t) \exp(\mu_d'\zeta_t + \sigma_d^{2\prime}\zeta_t/2)(\lambda_1 + e)'P - \lambda_{2f}' \right] \zeta_t, \tag{17}$$

$$Var_t[R_{t+1}^e] = (\theta_{2d}'\zeta_t) \left(\left[(\lambda_1 + e) \odot (\lambda_1 + e) \right]' P\zeta_t \right), \tag{18}$$

where $\{\lambda_1, \lambda_2, \theta_{2d}\}$ are vectors of constants that depend on the underlying parameters of the economy, $(\mu_c, \mu_d, \sigma_c, \sigma_d, \rho, \gamma, \delta, \psi, P)$.

Following the analysis in Garcia, Meddahi, and Tédongap (2008), we consider a four-state model with parameters for the consumption and dividend growth processes calibrated to match the empirical moments reported by Lettau, Ludvigson, and Wachter (2008). Specifically, the means and volatilities of the consumption and dividend growth processes across the four states (in percent/year) are $\mu_c = \begin{pmatrix} 0.62 & 0.62 & -0.32 & -0.32 \end{pmatrix}$; $\mu_d = \begin{pmatrix} 2.80 & 2.80 & -1.45 & -1.45 \end{pmatrix}$; $\sigma_c = \begin{pmatrix} 0.75 & 0.40 & 0.75 & 0.40 \end{pmatrix}$; $\sigma_d = \begin{pmatrix} 3.36 & 1.82 & 3.36 & 1.82 \end{pmatrix}$. Volatility of consumption and dividend growth is highest in states one and three and lowest in states two and four, while the mean growth rates are highest in states one and two and lowest in states three and four. Consistent with a configuration in Bonomo, Garcia, Meddahi, and Tédongap (2011), preference parameters are set to $\psi = 0.5$, $\gamma = 30$, $\delta = 0.9925$. The only additional parameters in our analysis are the pairwise correlations between dividend and consumption growth which we set to $\rho = \begin{pmatrix} 3/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$, so that the first state has a higher correlation parameter than the other states.

The left graph in Figure 5 plots the conditional expected excess return against the conditional volatility implied by the model. Expected returns and conditional volatility are lowest in the second state (which has low volatility and high mean dividend and consumption growth) and both increase as we move through states four and one which, compared with state two, experience either higher volatility or lower mean consumption and dividend growth. While return volatility is highest in state three, the expected return is lower in this state than in state one. This happens because of the higher correlation between dividend and consumption growth in state one compared with the third state. Although return volatility is very high in state three, marginal utility of consumption and stock returns are less strongly correlated in this state than in state one, which means that the market portfolio acts as a hedge in state three. Figure 5 shows how this translates into a non-monotonic relation between the conditional volatility and expected returns in a way that matches our empirical findings.

This analysis suggests that the conventional positive link between conditional volatility and expected returns can be broken when the relation between marginal utility and asset payoffs is weakened. One risk measure that is less subject to this issue because it more closely tracks the covariance between excess returns and marginal utility is the conditional covariance between consumption growth and excess returns. In the present model this can be shown to be given by

$$Cov_t(R^e_{t+1}, \Delta c_{t+1}) = (\lambda_2' \zeta_t)\left((\lambda_1 + e)' P \zeta_t\right)\rho' \zeta_t. \tag{19}$$

The right graph in Figure 5 plots the expected return against the conditional covariance of excess returns and consumption growth. For this measure of risk, a monotonic relation is obtained. The conditional covariance increases as we move from state two through states four, three and state one. Expected returns increase in the same order. This motivates an extension of the volatility or variance risk measures to a measure that incorporates covariance risk.

5 Conditional Covariance Risk

Up to now we have used conditional market volatility as our proxy for risk. However, consumption asset pricing models suggest that the covariance between returns and consumption growth would be a more appropriate measure of risk (Breeden (1979)), while the ICAPM (Merton (1973)) suggests including further state variables tracking time-varying investment opportunities. For example, Merton derived a relationship between the conditional expectation of excess returns on the market portfolio, $E_t[r_{t+1}]$, its conditional variance, $\sigma^2_{t+1|t}$, and the conditional covariance between market returns and state variables capturing time variations in the investment opportunity set, $Cov_{t+1|t}$:

$$E_t[r_{t+1}] = a_W \sigma^2_{t+1|t} + b_{Wx} Cov_{t+1|t}. \tag{20}$$

Here a_W measures the representative investor's relative risk aversion and b_{Wx} depends on the sensitivity of the investor's indirect utility function with respect to wealth (W) and state variables (x). Leaving out the conditional covariance term from Eq. (20) could lead to omitted variable bias and so must be addressed (Guo and Whitelaw (2006)).

Consumption based asset pricing models lead to similar suggestions. For example, when the representative investor has power utility, $u(C_{t+1}) = C_{t+1}^{1-\gamma}/(1-\gamma)$, $\gamma \geq 0$, and consumption growth is log-normally distributed, expected excess returns on the stock market portfolio satisfy

$$E_t[r_{t+1}] \approx \gamma cov_t(\Delta c_{t+1}, r_{t+1}), \tag{21}$$

where $cov_t(\Delta c_{t+1}, r_{t+1})$ is the conditional covariance between consumption growth, Δc_{t+1}, and stock returns. A broader result is obtained under weaker assumptions requiring only concave utility and a positive relation between consumption growth and stock returns:

$$\frac{\partial E_t[r_{t+1}]}{\partial cov_t(\Delta c_{t+1}, r_{t+1})} > 0. \tag{22}$$

When consumption growth is unobserved, this result is not very useful. However, a similar result holds if an economic activity variable is used to proxy for consumption growth, provided that there is a monotonically increasing relation−not necessarily a linear one−between consumption growth and changes in economic activity, ΔA_{t+1} :

$$\frac{\partial E_t[r_{t+1}]}{\partial cov_t(\Delta A_{t+1}, r_{t+1})} > 0. \tag{23}$$

Intuitively, the higher the covariance between changes to economic activity and stock market returns, the lower returns tend to be during economic recessions where marginal utility of consumption is high, suggesting that stocks are a poor hedge against shocks to marginal utility. Hence, investors must be offered a higher expected return to induce them to hold stocks. We next show how an estimate of $cov_t(\Delta EA_{t+1}, r_{t+1})$ can be constructed from daily data on economic activity.

5.1 Realized Covariance

There is no proxy for the covariance term in Eq. (20) equivalent to the realized variance measure in Eq. (6). To overcome this, we construct a new measure based on the covariance between stock market returns and a high frequency proxy for economic activity measured by means of the ADS business conditions index proposed by Arouba, Diebold, and Scotti (2009). Daily data on this are available back to 1960.

The ADS index is designed to track high frequency (daily) business conditions. Its underlying economic indicators (daily spreads between 10-year and 3-month Treasury yields, weekly initial jobless claims, monthly payroll employment, industrial production, personal income less transfer payments, manufacturing and trade sales, and quarterly real GDP) optimally blend high- and low-frequency information and stock and flow data. The top window in Figure 6 plots the ADS index over the period 1960-2008. The index displays a clear cyclical pattern with distinct declines during economic recessions.

The ADS index is a broad measure of economic activity so it seems reasonable to expect that consumption growth is positively correlated with this index. Because daily consumption data is not available, we consider instead the correlation between changes to the ADS index and real consumption growth at monthly, quarterly, semi-annual and annual horizons. Correlations are uniformly positive and increase with the horizon, rising from 0.15-0.20 at the monthly horizon to 0.40-0.50 at the semi-annual and 0.50 at the annual horizon, irrespective of whether durable or nondurable real consumption is used.

These findings are consistent with a monotonically increasing relation between consumption growth and changes to the ADS index and suggest that we can use high frequency changes to this index as a proxy for the unobserved consumption growth or, alternatively, as a proxy for time-varying investment opportunities. Specifically, we compute monthly "realized covariances" between stock returns and changes in the ADS index from observations at the daily frequency,

$$\widehat{cov}_t = \sum_{i=1}^{N_t} \Delta ADS_{i,t} \times r_{i,t}, \tag{24}$$

where $\Delta ADS_{i,t}$ is the change in the ADS index on day i during month t, and $r_{i,t}$ is the corresponding stock market return.

The bottom window in Figure 6 plots monthly (scaled) values of the conditional covariance between changes to the ADS index and stock returns. The conditional covariance is distinctly countercyclical and rises during economic recessions.

5.2 Empirical Results

We next extend the risk-return model to include estimates of our new conditional covariance measure in addition to the earlier measure of conditional volatility. We estimate both linear and flexible risk-return specifications that control for dynamic effects:

$$\hat{\mu}_{t+1|t} = \alpha + \beta_1 \widehat{\sigma}^2_{t+1|t} + \beta_2 \widehat{cov}_{t+1|t} + \beta_3 \widehat{\mu}_{t|t-1} + \beta_4 \widehat{cov}_{t|t-1} + \beta_5 \widehat{\sigma}^2_{t|t-1} + \varepsilon_{t+1}, \tag{25}$$

$$\hat{\mu}_{t+1|t} = f(\widehat{\sigma}^2_{t+1|t}, \widehat{cov}_{t+1|t}, \widehat{\mu}_{t|t-1}, \widehat{cov}_{t|t-1}, \widehat{\sigma}^2_{t|t-1}) + \varepsilon_{t+1}. \tag{26}$$

To the extent that the consumption CAPM is valid and our conditional covariance measure proxies well for time-variations in consumption betas, we would expect only the covariance terms to be significant. Conversely, if the ICAPM better describes the data, both the conditional volatility and the covariance should be significant.[14]

Table 4 presents estimation results for the models in Eqs. (25-26) using data over the sample, 1960-2008, for which the covariance measure can be estimated. For the linear model shown in Panel A, the coefficient on the conditional variance is insignificant with a t-statistic below one. In contrast, the coefficient on the conditional covariance is positive and highly significant with a t-statistic above six.

Turning to the flexible specification reported in Panel B, the covariance measure is most important in explaining variations in expected returns. The conditional covariance, $\widehat{cov}_{t+1|t}$ obtains a relative influence of 13.4% which is significantly different from zero (p-value of 0.0%), whereas the relative importance of the conditional variance, $\widehat{\sigma}^2_{t+1|t}$, is 6.1% ($p$-value of 14%).

Panel C shows that the null hypothesis of a monotonically increasing relation between the conditional covariance and expected returns is not rejected. Conversely, we reject, in two out of three cases, that there is a monotonic relation between the conditional volatility and expected returns.[15]

Figure 7 shows that the partial dependence plots for the joint model in Eq. (26) further corroborate these findings. Expected returns increase monotonically in the conditional covariance, whereas the expected return-conditional volatility relation rises at first but then declines at higher levels of volatility. Moreover, expected returns vary by approximately 5% per annum due to variations in the conditional covariance but change by less than 2% per annum due to variations in the conditional variance.

Guo and Whitelaw (2006) also include a covariance estimate of time-varying investment opportunities in their analysis of the risk-return relation. In common with much of the literature,

[14]Suppose that the economic activity index depends both on consumption growth and state variables, x_{t+1}, that track time-varying investment opportunities, i.e.,

$$\Delta EA_{t+1} = \lambda_0 + \lambda_1 \Delta c_{t+1} + \lambda_2 x_{t+1}.$$

Then it follows that the specifications in Eqs. (25) and (26) nest both the consumption CAPM (if $\lambda_2 = 0$ and conditional volatility does not matter in explaining variations in expected returns) and the ICAPM (if $\lambda_2 \neq 0$).

[15]The case where we fail to reject compares fewer portfolios and so could excessively smooth out nonmonotonicities in the expected return-variance relation.

they assume that f is a linear function of both $\widehat{\sigma}^2_{t+1|t}$ and $\widehat{cov}_{t+1|t}$ and they compute $\widehat{\sigma}^2_{t+1|t}$ and $\widehat{cov}_{t+1|t}$ from linear projections on observable state variables. However, the plots in the bottom row in Figure 4 show that the conditional covariance is a highly nonlinear function of the most important state variables such as inflation, the payout rate and long term returns, suggesting that linear models for the covariance are misspecified. Corroborating this, the bottom rows of Panel B in Table 2 show that linear specifications for the conditional covariance are clearly misspecified. While there is evidence that the BRT covariance estimates are also misspecified, this is driven by a single observation following the default of Lehmann (October 2008) which represents an extreme outlier for the realized covariance measure. Without this single observation, the BRT estimates appear to be correctly specified, while the linear estimates are not.

6 Extensions and Robustness Analysis

We finally report some results that shed light on the robustness of our findings with regard to the underlying economic predictor variables and the implementation of the boosted regression tree methodology. To address the concern that our methodology could overfit the return and volatility data, however, we first provide out-of-sample forecasting results.

6.1 Out-of-sample forecasting performance

Our analysis suggests that a range of predictor variables from the finance literature capture time variations in expected returns, volatility and covariance. Moreover, the effect of these variables appears to be nonlinear. However, the boosted regression trees could be more prone to estimation error and overfitting than more tightly parameterized linear regressions. It is therefore far from certain that our approach provides a useful way to generate estimates of expected returns and conditional volatility.

To explore this point, we follow the literature on out-of-sample forecasting and estimate forecasting models recursively through time.[16] We use data up to 1969:12 to fit the first regression tree. We then predict returns or volatility for the following month, 1970:01. The next month we

[16]See, e.g., Pesaran and Timmermann (1995), Bossaerts and Hillion (1999), Campbell and Thompson (2008), and Welch and Goyal (2008) for mean returns and Engle, Ghysels, and Sohn (2006) and Paye (2010) for volatility forecasting.

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.

We limit our analysis of out-of-sample forecasts to returns and realized volatility. For these variables we have a sufficiently long data sample that we can estimate the forecasting models with reasonable precision at the start of the out-of-sample period. In contrast, this is an issue for the realized covariance measure since the economic activity data only begin in 1960, leaving too short an out-of-sample period for a meaningful model comparison.

6.1.1 Return forecasts

As a first check of whether the boosted regression trees overfit the data, consider the top window in Figure 8 which shows return forecasts over the period 1985-2008. While there is a visible relation between actual and predicted returns, there is no tendency for the model to fit outliers. Indeed, the fitted values are confined to a far narrower range than actual returns.

Next consider the out-of-sample forecasting performance of the boosted regression trees. Panel A1 in Table 5 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 multivariate linear regressions.[17]

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, although the difference is reduced in the period that includes the recent financial crisis.

6.1.2 Volatility forecasts

Turning to the volatility forecasts, Panel A2 of Table 5 and the bottom panel in Figure 8 compare volatility forecasts from the regression tree to those from a GARCH(1,1) model or an autoregres-

[17]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.

sive 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 a MIDAS estimator of the conditional variance of monthly returns:

$$Var_t(R_{t+1}) = 22 \sum_{d=0}^{D} w_d r_{t-d}^2, \tag{27}$$

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 is the maximum lag length which is set to 250 days following Ghysels, Santa-Clara, and Valkanov (2005). The bottom window in Figure 8 plots the fitted volatility levels associated with the GARCH(1,1), MIDAS and the boosted regression trees.

In the shorter sample that ends in 2005, the best volatility forecasts are generated by the boosted regression trees which produce smaller forecast errors and an out-of-sample R^2-value of 34% 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 high volatility levels at the end of this period. However, the boosted regression trees continue to generate the best volatility forecasts.

In summary, this out-of-sample analysis shows that the boosted regression tree estimates of the conditional mean and volatility are not overfitting the data. Moreover, because they do not make restrictive assumptions on the shape of the relation 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 suggests that our estimates of the conditional equity premium and conditional volatility are better suited for analyzing the risk-return trade-off than estimates from linear models.

6.2 Confidence Interval for the Fitted Risk-Return Relation

It is natural to ask how precisely the regression function relating the risk and return estimates is estimated. We are unaware of methods for constructing confidence intervals on the estimates from boosted regression trees. However, using Bayesian additive regression trees (BART), which is a nonparametric regression method based on boosting, we can construct interval estimates for the unknown regression function. While BRTs and BARTs are similar in many respects, the latter have two distinctive features. First, the contribution of each tree to the final BART model is reduced by imposing a prior that regularizes its fit, and not a shrinkage parameter as in BRT. Second, the iterative construction and fit of successive BART residuals is performed by way of Bayesian backfitting MCMC instead of the steepest descent used in BRT. Our implementation employs the prior forms specified in Chipman, George, and McCulloch (1998), and the standard choice of hyper-parameters contained in Chipman, George, and McCulloch (2010).

Figure 9 presents confidence intervals on the resulting regression trees for the model that includes volatility and covariance as risk measures. Although the confidence bands are quite wide for the graph linking expected returns to the conditional volatility, they are tight enough to suggest a clear non-monotonic risk-return pattern. Conversely, the relation between expected returns and the conditional covariance measure appears markedly monotonic after accounting for parameter estimation error.

6.3 Estimation Errors

In common with most studies in the literature on the risk-return relation, our analysis has so far ignored the fact that we use estimates in place of 'true' values for expected return and volatility. This could potentially bias parameter estimates. To deal with this, Ludvigsson and Ng (2007) estimate the linear risk-return model in Eq. (8) using two-stage least squares, instrumenting $\hat{\sigma}_{t|t-1}$ through the lagged variables $\hat{\mu}_{t-1|t-2}$, $\hat{\mu}_{t-2|t-3}$, $\hat{\sigma}_{t-1|t-2}$, $\hat{\sigma}_{t-2|t-3}$, r_{t-1}, and $\hat{\sigma}_{t-1}$.

To see if estimation errors affect our results, we conduct a similar analysis. In the first step, we obtain a BRT estimate of $\hat{\sigma}_{t|t-1}$, denoted $\tilde{\sigma}_{t|t-1}$, using the same instruments as in Ludvigson and Ng (2007). We then use the BRT to estimate the expected return with the instrumented value, $\tilde{\sigma}_{t|t-1}$, as a predictor variable. The resulting plot relating the conditional mean and conditional volatility estimate is very similar to that reported earlier in Figure 3 and is therefore omitted.

We also cast the two-stage least square estimate as a GMM regression, approximating the non-linear risk-return relation through a second-order polynomial with instruments identical to those used by Ludvigson and Ng. The resulting instrumented estimate of the coefficient on $\tilde{\sigma}_{t|t-1}$is positive and highly significant, while the coefficient on $\tilde{\sigma}^2_{t|t-1}$ is negative and highly significant, giving rise to a risk-return relation that increases for low values, but decreases for high values of the conditional volatility.

Finally, we adopted the GMM approach with the Ludvigson-Ng instruments for $\hat{\sigma}_{t|t-1}$ using the earlier threshold regression from Eq. (10). The result obtained with the instrumental variables approach is again very similar to the earlier one:

$$\hat{\mu}_{t+1|t} = \begin{cases} \begin{array}{ccl} -0.006 & + \ 29.532 & \hat{\sigma}^2_{IV,t+1|t} \text{ for } \hat{\sigma}^2_{IV,t+1|t} \leq (0.0253)^2 \\ (-2.50) & (7.04) & \\ 0.009 & - \ 0.562 & \hat{\sigma}^2_{IV,t+1|t} \text{ for } \hat{\sigma}^2_{IV,t+1|t} > (0.0253)^2 \\ (23.77) & (-5.80) & \end{array} \end{cases} \tag{28}$$

In summary, these results suggest that errors-in-variables problems in the conditional mean and variance estimates do not affect our results in any meaningful manner.

6.4 Robustness of boosting regression tree results

Our benchmark analysis uses 10,000 boosting iterations to estimate the regression trees. We next explore the sensitivity of the results to different ways of choosing the number of boosting iterations. As a first robustness exercise, Panel B1 of Table 5 reports the out-of-sample forecasting performance of boosted regression trees using 5,000, 10,000 and 15,000 boosting iterations.[18] The results are not particularly sensitive to this choice. On the whole, the regression tree forecasts outperform the benchmarks listed in Panel A 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 stock returns there is evidence only of a very slow decay in forecasting accuracy beyond 7,000 boosting iterations. There are no signs of over-fitting for the volatility prediction model. This

[18]Since these are based on out-of-sample forecasts, we only report results for the mean and volatility.

stability across different numbers of boosting iterations, B, makes the choice of the number of boosting iterations of little significance 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 only based on model performance up to time t. Second, we use forecast combinations as a way to lower the sensitivity of our results to the choice of B by using the simple average of the forecasts from regression trees with B =1, 2, ..., 10,000 boosting iterations.

Panel B1 in Table 5 shows that the combined average is particularly effective in generating precise return predictions. Selecting the best model on the basis of recent performance appears to be more effective for volatility prediction (Panel B2).

7 Conclusion

This paper proposes a new and flexible approach to modeling the risk-return relation that avoids imposing strong functional form assumptions. The approach can handle large sets of state variables and is not prone to overfitting the data. Hence it is not subject to the misspecification and omitted variable biases that have been a big concern in empirical studies of the risk-return trade-off.

Using this approach on US stock return data, our empirical analysis finds that there is a positive trade-off between conditional volatility and expected returns at low or medium levels of conditional volatility, but that the relation is flat or inverted during periods with high volatility. These findings make it easier to understand why so many empirical studies differ in their findings on the sign and magnitude of the conditional volatility-mean return relation.

The non-monotonic trade-off between conditional volatility and expected returns uncovered in our analysis indicates limitations of conditional volatility as a measure of risk. To address this, we develop a high-frequency risk measure that captures the covariance between a broad economic activity index and stock returns. Changes to the economic activity index are shown to be strongly positively correlated with consumption growth at horizons of one month or longer, and have the further advantage that they are measured at the daily frequency. This enables us to compute 'realized covariances' and facilitates estimation of conditional covariance risk. We find strong and significant evidence of a monotonically increasing relation between expected stock returns and conditional covariance risk. This suggests that there is indeed a positive risk-return trade-off, but that it is important to use a broad measure of risk that account for the state of the economy.

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.

Arouba, S., F. Diebold, and C. Scotti (2009): "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 27(4).

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.

Bansal, R., and A. Yaron (2004): "Risks for the long run: A potential resolution of asset pricing puzzles," *Journal of finance*, 59(4), 1481–1509.

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.

Bonomo, M., R. Garcia, N. Meddahi, and R. Tédongap (2011): "Generalized Disappointment Aversion, Long-run Volatility Risk, and Asset Prices," *Review of Financial Studies*, 24(1), 82–122.

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.

Brandt, M. W., and L. Wang (2007): "Measuring the Time-Varying Risk-Return Relation from the Cross-Section of Equity Returns," *Unpublished Working Paper, Duke University.*

Breeden, D. (1979): "An intertemporal asset pricing model with stochastic consumption and investment opportunities," *Journal of Financial Economics*, 7(3), 265–296.

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

——— (1993): "Intertemporal Asset Pricing without Consumption Data," *American Economic Review*, 83(3), 487–512.

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

Chipman, H., E. George, and R. McCulloch (1998): "Bayesian CART model search," *Journal of the American Statistical Association*, pp. 935–948.

——— (2010): "BART: Bayesian additive regression trees," *The Annals of Applied Statistics*, 4(1), 266–298.

David, A., and P. Veronesi (2009): "What Ties Return Volatilities to Price Valuations and Fundamentals?," *NBER working paper*.

Dietterich, T. (2000): "Ensemble methods in machine learning," *Multiple classifier systems*, pp. 1–15.

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

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

Garcia, R., N. Meddahi, and R. Tédongap (2008): "An analytical framework for assessing asset pricing models and predictability," *Université de Montréal Working Paper*.

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.

Guo, H., and R. Whitelaw (2006): "Uncovering the risk return relation in the stock market," *Journal of Finance*, 2(61), 1433–1463.

Hansen, B. (2000): "Sample splitting and threshold estimation," *Econometrica*, 68(3), 575–603.

Harrison, P., and H. Zhang (1999): "An investigation of the risk and return relation at long horizons," *Review of Economics and Statistics*, 81(3), 399–408.

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

——— (2001): "The specification of conditional expectations," *Journal of Empirical Finance*, 8(5), 573–637.

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

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

Lettau, M., S. C. Ludvigson, and J. A. Wachter (2008): "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?," *Review of Financial Studies*, 21(4), 1653–1687.

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

Merton, R. (1973): "An intertemporal capital asset pricing model," *Econometrica*, 5(41), 867–887.

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

Patton, A., and A. Timmermann (2010): "Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts," *Journal of Financial Economics*, 98(3), 605–625.

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

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

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

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.

——— (2000): "Stock market risk and return: An equilibrium approach," *Review of Financial Studies*, 13(3), 521.

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

Table 1. Estimates of the risk-return trade-off

A1. Linear volatility model

Model: $\hat{\mu}_{t+1|t} = \alpha + \beta_1\ \hat{\sigma}_{t+1|t} + \beta_2\ \hat{\sigma}_{t|t-1} + \beta_3\ \hat{\mu}_{t|t-1} + \epsilon_{t+1}$

Sub-Samples	$\hat{\sigma}_{t+1\|t}$ (t-stat)	$\hat{\sigma}_{t\|t-1}$ (t-stat)	$\hat{\mu}_{t\|t-1}$ (t-stat)	R^2
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)	
	Common Factor Model			
1960-2007	0.007	-0.011	0.607	36.8%
	(0.13)	(-0.21)	(17.65)	
	VIX Volatility Proxy			
1986-2008	0.053	-0.042	0.656	42.35%
	(1.75)	(-1.33)	(14.07)	

A2. Flexible volatility model

Model: $\hat{\mu}_{t+1|t} = f(\hat{\sigma}_{t+1|t}, \hat{\sigma}_{t|t-1}, \hat{\mu}_{t|t-1}) + \epsilon_{t+1}$

Sub-Samples	$\hat{\sigma}_{t+1\|t}$ (p-value)	$\hat{\sigma}_{t\|t-1}$ (p-value)	$\hat{\mu}_{t\|t-1}$ (p-value)
1927-2008	7.6%	9.4%	83.0%
	(1.8%)	(0.0%)	(0.0%)
1927-1967	21.1%	19.5%	59.4%
	(0.0%)	(0.0%)	(0.0%)
1968-2008	9.9%	9.9%	80.2%
	(70.6%)	(53.9%)	(0.0%)
	Common Factor Model		
1960-2007	15.2%	12.1%	72.8%
	(9.6%)	(48.6%)	(0.0%)
	VIX Volatility Proxy		
1986-2008	18.3%	11.3%	70.4%
	(55.3%)	(88.8%)	(0.0%)

B1. Linear variance model

Model: $\hat{\mu}_{t+1|t} = \alpha + \beta_1\ \hat{\sigma}^2_{t+1|t} + \beta_2\ \hat{\sigma}^2_{t|t-1} + \beta_3\ \hat{\mu}_{t|t-1} + \epsilon_{t+1}$

Sub-Samples	$\hat{\sigma}^2_{t+1\|t}$ (t-stat)	$\hat{\sigma}^2_{t\|t-1}$ (t-stat)	$\hat{\mu}_{t\|t-1}$ (t-stat)	R^2
1927-2008	0.242	-0.415	0.653	44.10%
	(1.56)	(-2.68)	(26.60)	
1927-1967	0.135	-0.507	0.567	41.10%
	(0.72)	(-2.71)	(14.92)	
1968-2008	0.412	-0.131	0.660	43.00%
	(1.45)	(-0.44)	(19.36)	
	Common Factor Model			
1960-2007	0.015	-0.004	0.607	36.8%
	(0.02)	(-0.01)	(17.71)	
	VIX^2 Volatility Proxy			
1986-2008	0.223	-0.107	0.651	42.2%
	(1.47)	(-0.66)	(13.98)	

B2. Flexible variance model

Model: $\hat{\mu}_{t+1|t} = f(\hat{\sigma}^2_{t+1|t}, \hat{\sigma}^2_{t|t-1}, \hat{\mu}_{t|t-1}) + \epsilon_{t+1}$

Sub-Samples	$\hat{\sigma}^2_{t+1\|t}$ (p-value)	$\hat{\sigma}^2_{t\|t-1}$ (p-value)	$\hat{\mu}_{t\|t-1}$ (p-value)
1927-2008	7.6%	8.8%	83.6%
	(3.3%)	(0.2%)	(0.0%)
1927-1967	20.2%	19.6%	60.2%
	(0.0%)	(0.0%)	(0.0%)
1968-2008	10.5%	9.6%	79.9%
	(59.1%)	(57.8%)	(0.0%)
	Common Factor Model		
1960-2007	18.7%	9.3%	72.0%
	(0.6%)	(82.1%)	(0.0%)
	VIX^2 Volatility Proxy		
1986-2008	18.1%	11.3%	70.6%
	(9.0%)	(89.9%)	(0.0%)

This table reports estimates of the relationship between expected returns, $\hat{\mu}_{t+1|t}$, and conditional volatility, $\hat{\sigma}_{t+1|t}$, or conditional variance, $\hat{\sigma}^2_{t+1|t}$, as well as lags of these. Panels A1 and B1 use linear specifications and report coefficient estimates and t-statistics. Panels A2 and B2 employ boosted regression tree models and report relative influence estimates together with their significance (p-values) obtained through Monte Carlo simulations. Relative influence measures sum to 100. Panels A1 and A2, use conditional volatility, $\hat{\sigma}_{t+1|t}$, while panels B1 and B2, use conditional variance, $\hat{\sigma}^2_{t+1|t}$, as a measure of risk. The last rows in both panels use the Chicago Board Options Exchange (CBOE) Volatility Index, VIX, or its square, as a measure of risk. Estimates of the conditional mean and volatility or variance are based on boosted regression trees and use the state variables described in Section 3.1, except for the common factor estimates which are based on a wider set of 131 economic variables.

Table 2. Model Specification Tests

A. Risk-Return Relation

Model	Baseline Results Coeff.	p-val	Model with Factors Coeff.	p-val
	F-test	**p-val**	**F-test**	**p-val**
Linear (Volatility)	8.807	0.000	1.961	0.069
Linear (Variance)	52.297	0.000	4.244	0.000
BRT (Volatility)	0.919	0.508	0.533	0.851
BRT (Variance)	0.579	0.815	0.298	0.975
GARCH(1,1)	25.143	0.000	2.113	0.050
MIDAS	7.703	0.000	0.719	0.635

B. Moment Estimates

B.I. Mean

Model	Baseline model F-test	p-val	Model with Factors F-test	p-val
Linear	4.538	0.001	2.210	0.067
BRT	0.964	0.509	0.982	0.486

B.II. Variance

Model	Baseline model F-test	p-val	Model with Factors F-test	p-val
Linear	10.226	0.000	6.168	0.000
BRT	0.933	0.550	0.873	0.632
GARCH(1,1)	10.968	0.000	—	—
MIDAS	3.759	0.000	—	—

B.III. Covariance

Model	Baseline model F-test	p-val	Model with Factors F-test	p-val
Linear	4.605	0.000	2.084	0.125
BRT	3.355	0.000	1.213	0.229

This table presents Ramsey RESET specification tests applied to different models for the risk-return relation (Panel A) or the underlying conditional mean (Panel B.I), variance (Panel B.II) and covariance (Panel B.III). The null is that a model is correctly specified, so a small p-value (i.e., below 0.05) indicates misspecification. The linear and boosted regression tree (BRT) models use the state variables described in Section 3.1 as predictors, while the MIDAS and GARCH(1,1) models are based on past returns. The baseline results are based on the full sample (1927-2008), while the factor results use a shorter period (1960-2007) for which data on 131 underlying economic variables are available.

Table 3. Tests for a monotonically increasing risk-return relation

	Group size		
Horizon (months)	**Small**	**Medium**	**Large**
A. Volatility estimates			
1	0.000	0.018	0.010
2	0.000	0.000	0.017
3	0.000	0.000	0.039
B. VIX-based estimates			
1	0.027	0.041	0.091

This table presents the results of a test of whether the relationship between conditional risk and expected returns is monotonic after marginalizing out the effect of lagged risk and lagged expected returns. The test uses pairs of expected return, risk observations that are sorted on the basis of the conditional risk measure (volatility or VIX), yielding groups of observations corresponding to different levels of risk. The number of monthly observations in the small, medium and large groups or "portfolios" are approximately 40, 50 and 65. We then test if the associated mean return is monotonically increasing as we move from low to high risk observations. 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 decreasing or non-monotonic relation. Small p-values indicate rejection of a monotonically increasing risk-return relation. In panel A, estimates of the conditional mean and volatility are based on boosted regression trees that use the state variables from Section 3.1. The Chicago Board Options Exchange Volatility Index, also known as the VIX, is used as a proxy for the conditional volatility in Panel B.

Table 4. Estimates and tests of the covariance model

A. Linear model

$$\hat{\mu}_{t+1|t} = \alpha + \beta_1\hat{\sigma}^2_{t+1|t} + \beta_2\widehat{cov}_{t+1|t} + \beta_3\hat{\mu}_{t|t-1} + \beta_4\hat{\sigma}^2_{t|t-1} + \beta_5\widehat{cov}_{t|t-1} + \epsilon_{t+1}$$

Sample	$\hat{\sigma}^2_{t+1\|t}$ (t-stat)	$\widehat{cov}_{t+1\|t}$ (t-stat)	$\hat{\mu}_{t\|t-1}$ (t-stat)	$\hat{\sigma}^2_{t\|t-1}$ (t-stat)	$\widehat{cov}_{t\|t-1}$ (t-stat)
1960-2008	0.207 (0.74)	0.011 (6.16)	0.660 (20.70)	-0.112 (-0.37%)	-0.006 (-3.20%)

B. Flexible model

Relative influence measures

Model: $\hat{\mu}_{t+1|t} = f(\hat{\sigma}^2_{t+1|t}, \widehat{cov}_{t+1|t}, \hat{\mu}_{t|t-1}, \hat{\sigma}^2_{t|t-1}, \widehat{cov}_{t|t-1})$

Sample	$\hat{\sigma}^2_{t+1\|t}$	$\widehat{cov}_{t+1\|t}$	$\hat{\mu}_{t\|t-1}$	$\hat{\sigma}^2_{t\|t-1}$	$\widehat{cov}_{t\|t-1}$
1960-2008	6.16% (14.0%)	13.45% (0.0%)	67.50% (0.0%)	6.37% (11.4%)	6.52% (15.9%)

C. Monotonicity tests

	Group size		
	Small	**Medium**	**Large**
Conditional Variance	0.0000	0.0260	0.2567
	Small	**Medium**	**Large**
Conditional Covariance	0.2355	0.2359	0.8242

This table reports estimates of the effect on the conditional mean return, $\hat{\mu}_{t+1|t}$, of the conditional variance, $\hat{\sigma}^2_{t+1|t}$, the conditional covariance between stock returns and changes to economic activity, $\widehat{cov}_{t+1|t}$, the lagged conditional mean return, $\hat{\mu}_{t|t-1}$, lagged variance, $\hat{\sigma}^2_{t|t-1}$, and lagged covariance, $\widehat{cov}_{t|t-1}$. Panel A reports estimates and t-statistics from a linear model. Panel B shows relative influence estimates (in percent) from the boosted regression tree model. In parentheses we present the significance of the relative influence estimates by way of Monte Carlo p-values. Panel C presents the results of a test of whether the relation between conditional variance or conditional covariance) and expected returns is monotonic after marginalizing out the effect of the other variables in the model. 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 variance (or conditional covariance). Small p-values indicate rejection of a monotonically increasing relation. The number of observations in the small, medium and large portfolios are approximately 40, 50 and 65.

Table 5. Out-of-sample forecasting performance

A1. Return models

Forecasting Method	1970-2005 MSE $\times 10^3$	1970-2005 $\mathbf{R^2}$	1970-2008 MSE $\times 10^3$	1970-2008 $\mathbf{R^2}$
Boosted Regression Trees	1.964	1.53%	1.988	0.33%
Prevailing Mean	2.003	-0.40%	2.006	-0.58%
Multivariate Linear Model	2.026	-1.59%	2.037	-2.15%

A2. Realized volatility models

Forecasting Method	1970-2005 MSE $\times 10^3$	1970-2005 $\mathbf{R^2}$	1970-2008 MSE $\times 10^3$	1970-2008 $\mathbf{R^2}$
Boosted Regression Trees	2.643	34.10%	3.405	40.20%
Garch(1,1)	3.536	11.82%	4.902	13.89%
MIDAS (Beta)	3.089	22.97%	3.693	35.12%
MIDAS (Exp)	3.199	20.23%	3.844	32.48%
AR(1)	3.089	22.98%	3.463	39.17%

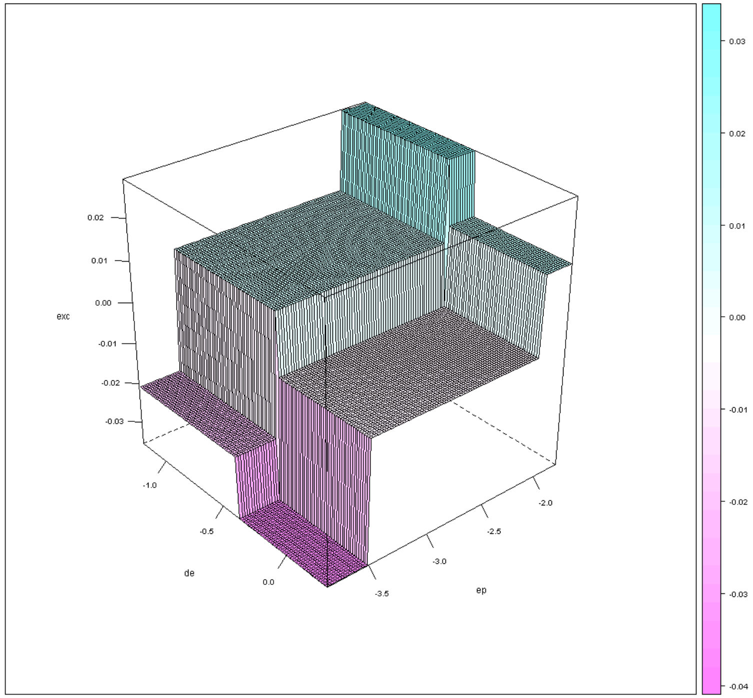
B1. Return models: Robustness

Forecasting Method	1970-2005 MSE $\times 10^4$	1970-2005 $\mathbf{R^2}$	1970-2008 MSE $\times 10^4$	1970-2008 $\mathbf{R^2}$
Boosted Regression Trees (5,000)	1.968	1.36%	1.985	0.45%
Boosted Regression Trees (15,000)	1.965	1.47%	1.992	0.13%
Best model selected recursively	1.974	1.06%	1.995	-0.03%
Combined Average	1.956	1.93%	1.977	0.88%

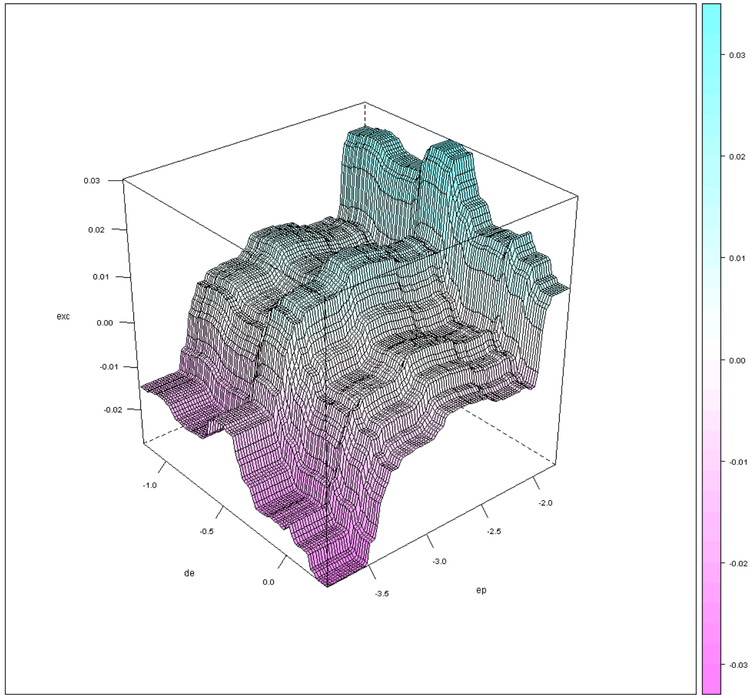
B2. Realized volatility models: Robustness

Forecasting Method	1970-2005 MSE	1970-2005 $\mathbf{R^2}$	1970-2008 MSE	1970-2008 $\mathbf{R^2}$
Boosted Regression Trees (5,000)	2.792	30.39%	3.669	35.55%
Boosted Regression Trees (15,000)	2.609	34.95%	3.358	41.01%
Best model selected recursively	2.647	34.00%	3.408	40.13%
Combined Average	2.706	32.51%	3.604	36.69%

This table reports the mean squared error (MSE) and the out-of-sample R^2-value for the benchmark boosted regression tree model based on 10,000 boosting iterations used to forecast monthly stock returns (Panel A1) and realized volatility (Panel A2). For comparison, in Panel A1 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 Bayesian Information Criterion. Panel A2 reports the results for GARCH(1,1), AR(1) and MIDAS models (Ghysels et al. (2005)). Panels B1 and B2 show the out-of-sample forecasting performance of the boosted regression trees under different rules for determining the number of boosting iterations. The first two lines show results using 5,000 and 15,000 boosting iterations. The third 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 B1 covers the return prediction model and Panel B2 covers the realized volatility prediction model. The out-of-sample performance is computed from 1970. 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.

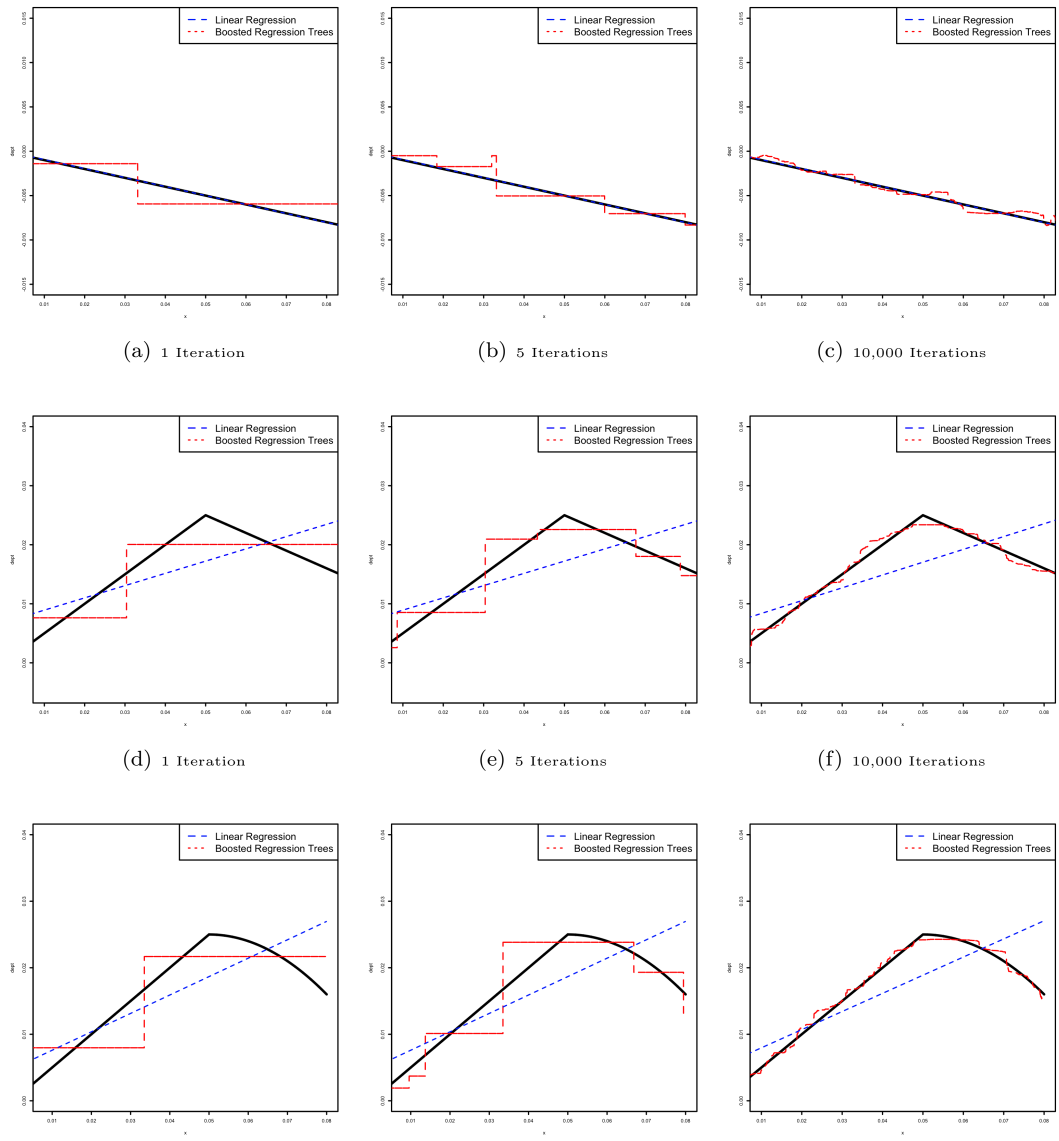


(a) 3 Iterations



(b) 5000 Iterations

Figure 1: **Fitted values of market 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 plots are based on monthly data from 1927 to 2008.



(a) 1 Iteration (b) 5 Iterations (c) 10,000 Iterations

(d) 1 Iteration (e) 5 Iterations (f) 10,000 Iterations

(g) 1 Iteration (h) 5 Iterations (i) 10,000 Iterations

Figure 2: **Fitted values from linear regression and boosted regression trees.** The top row assumes the true relation is linear, the middle row assumes an inverted V-shaped relation, while the bottom row assumes the true relation is first linear, then quadratic. The number of boosting iterations is set to one (left column), five (middle column), or 10,000 (right column).

1927-2008 **1927-1967** **1968-2008**

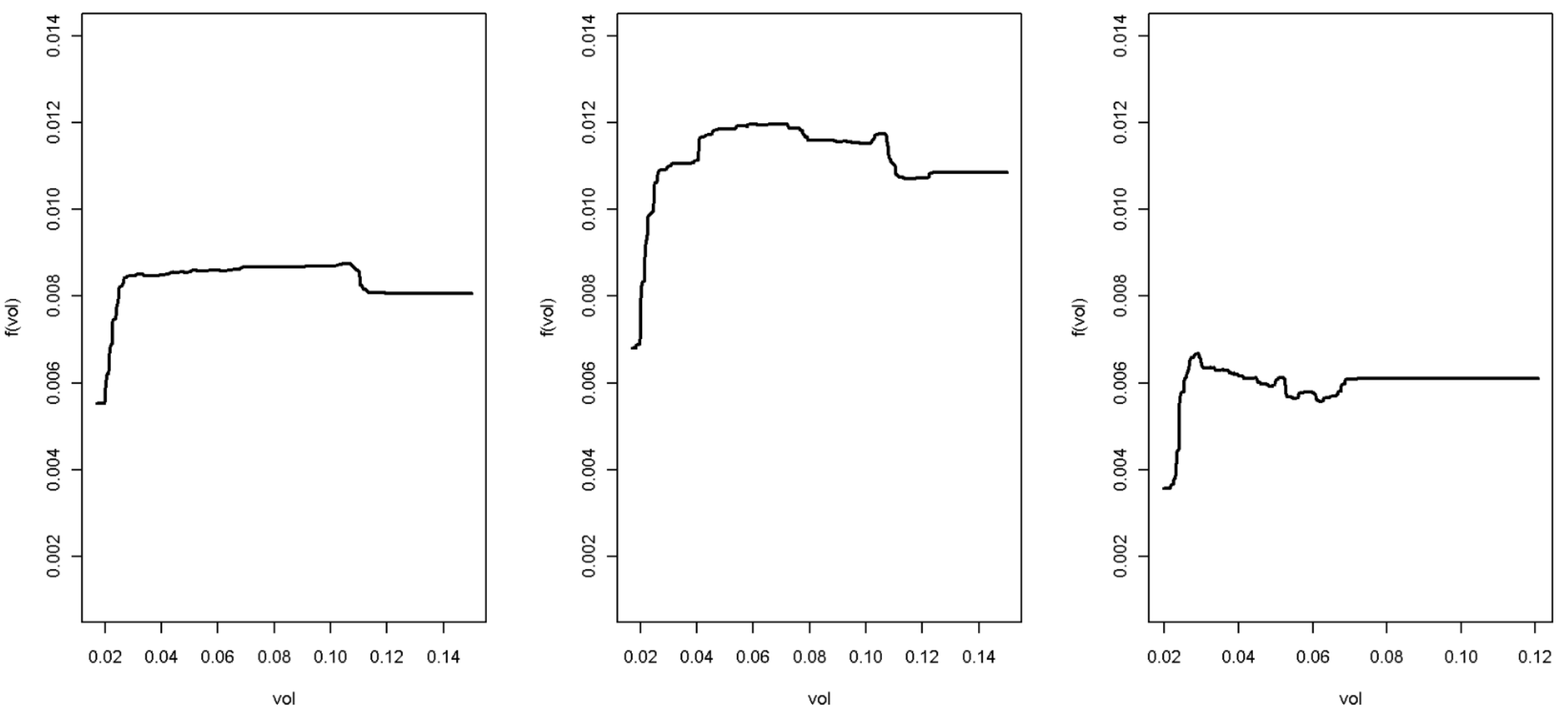
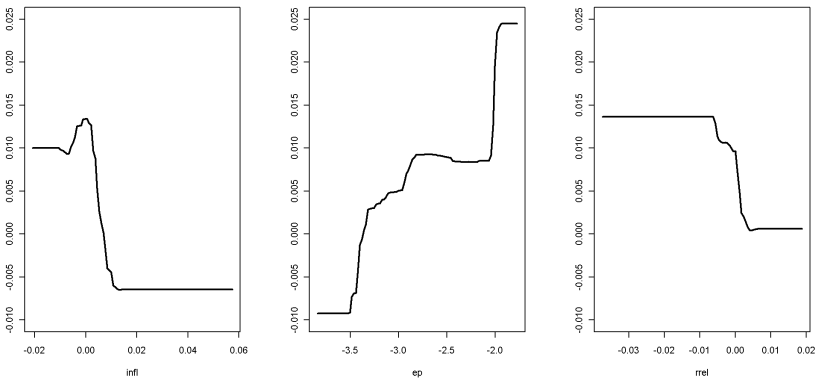
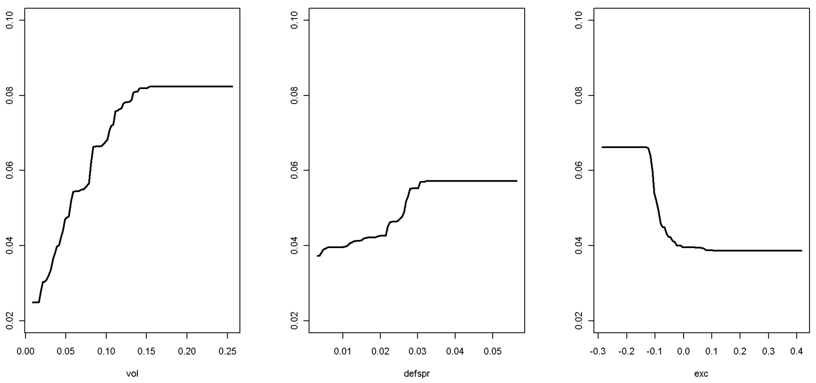


Figure 3: **Expected return-conditional volatility trade-off.** The figures show partial dependence plots for the conditional equity premium as a function of the conditional volatility (vol). The plots are based on boosted regression trees, using data for three samples, 1927-2008 (left panel), 1927-1967 (middle panel) and 1968-2008 (right panel). The horizontal axis covers the sample support of the conditional volatility, while the vertical axis tracks the resulting change in the conditional mean as a function of the conditional volatility.

Panel A. Expected Returns



Panel B. Volatility



Panel C. Covariance

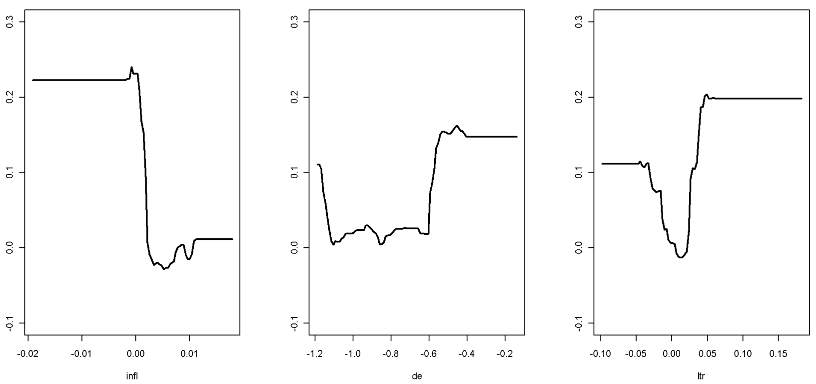


Figure 4: **Effect of predictor variables on conditional moments.** The figures present partial dependence plots for the mean excess return (Panel A), the conditional volatility (Panel B) and the conditional covariance (Panel C) based on the three predictor variables with the highest relative influence during 1927-2008, namely inflation (infl), the log earnings price ratio (ep), and the detrended T-Bill rate (rrel) for returns; the lagged stock market volatility (vol), the default spread (defspr) and the excess return on stocks (exc) for volatility; inflation (infl), the log dividend earnings ratio (de), and the long-term rate of return (ltr) for the covariance. The horizontal axis covers the sample support of the individual predictor variables, while the vertical axis tracks the change in the conditional equity premium as a function of the individual predictor variables.

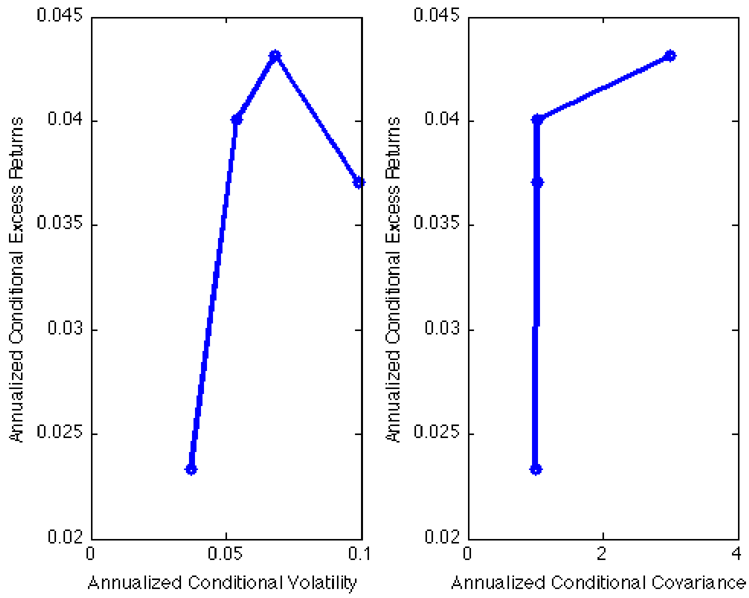


Figure 5: **Risk-Return relation in a dynamic asset pricing model.** This figure plots the conditional expected excess returns against the conditional volatility (left graph) and the conditional covariance (right graph) implied by the four-state regime switching model proposed by Garcia et al. (2008). The left plot shows a non-monotonic relation between conditional volatility and expected returns. The state with the highest conditional volatility has a weaker correlation between consumption and dividend growth than the other states. This means that the market portfolio provides a partial hedge against adverse shocks to consumption in this state, resulting in a reduced equity premium. Conversely, the right plot shows that there is a monotonically increasing relation between expected returns and the conditional covariance between consumption growth and stock returns.

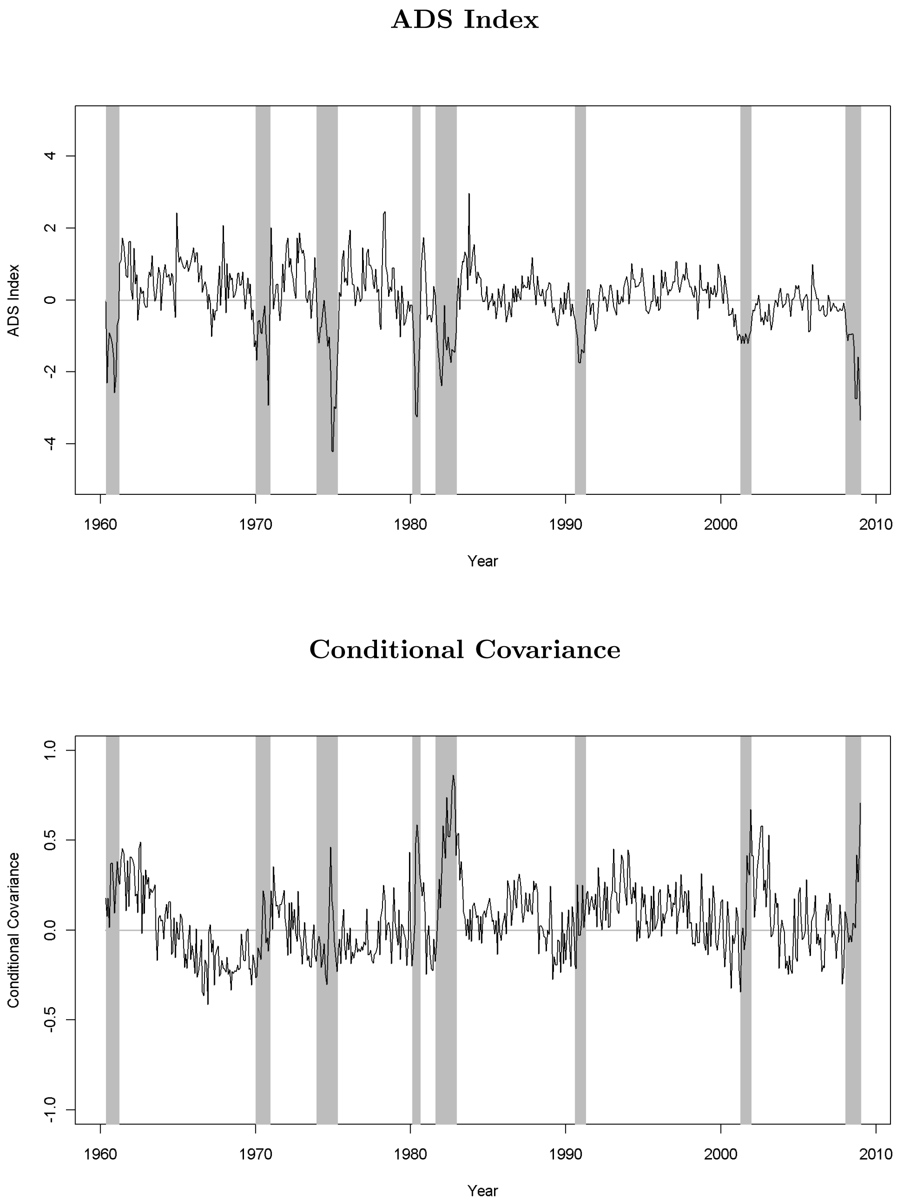
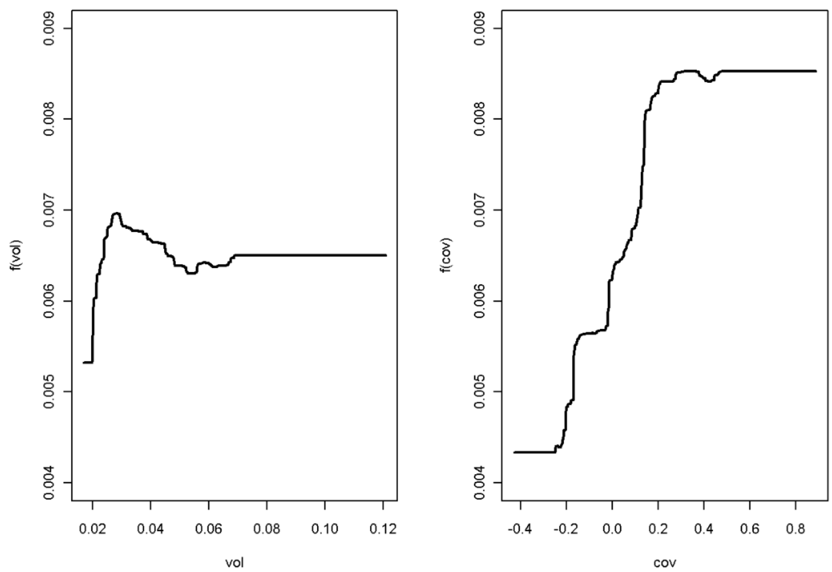


Figure 6: **ADS index and the conditional covariance measure.** This figure plots the ADS index at the monthly frequency in the top panel and the scaled conditional covariance between changes in the ADS index and stock returns in the bottom panel. The scaled conditional covariances are obtained as follows. First, monthly realized covariances between changes in the ADS index and stock returns are obtained using observations at the daily frequency. The scaled changes in the ADS Index are obtained by dividing the change to the ADS Index by the standard deviation of returns times the standard deviation of changes to the index. Finally, the conditional covariances are estimated by way of boosted regression trees.

Panel A. Baseline Model



Panel B. Model With Factors

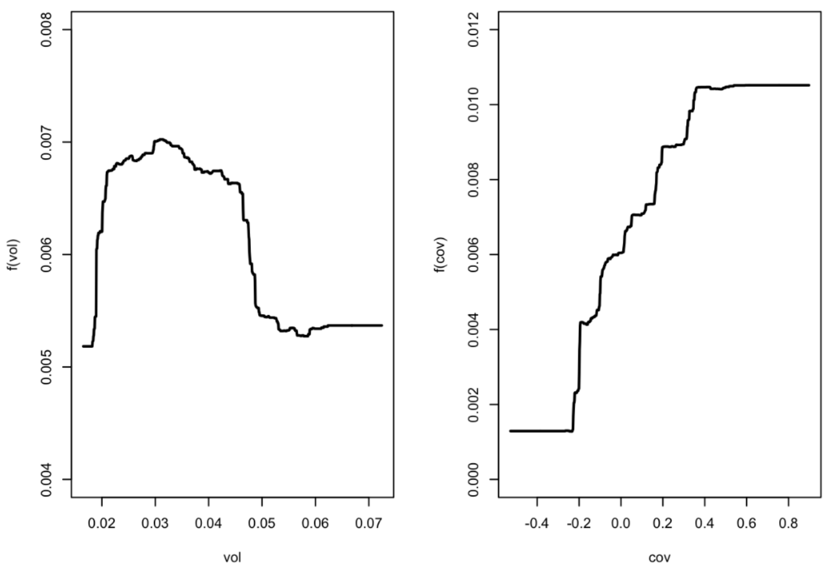
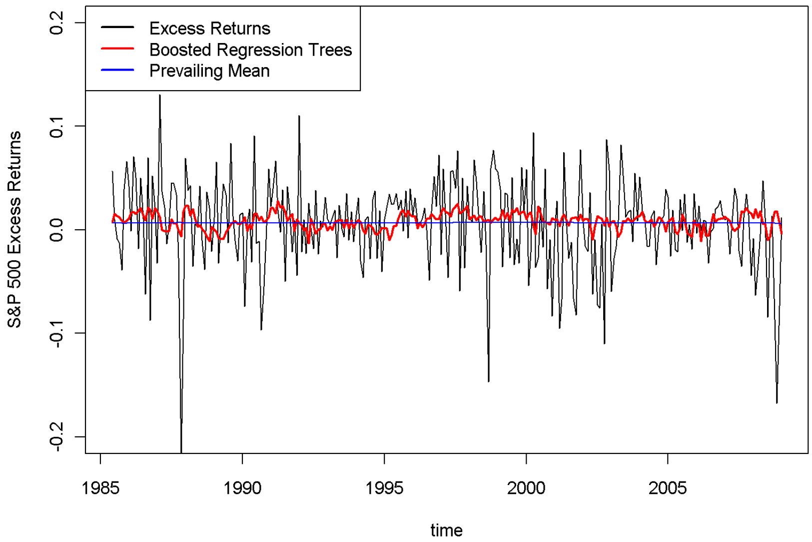


Figure 7: **Risk-return trade-off in the model with the conditional covariance risk measure.** The figure shows partial dependence plots for the conditional mean return as a function of the conditional volatility (vol) and the conditional covariance between stock returns and changes to economic activity (cov). The plot is based on a boosted regression tree, using data over the sample 1960-2008 in Panel A and 1960-2007 in Panel B. The horizontal axis covers the sample support of each predictor variable, while the vertical axis tracks the change in the conditional mean as a function of the individual state variables. In Panel A, the conditioning information is the predictor variables described in Section 3.1. In Panel B the conditioning information is the principal components derived from a set of 131 state variables and the three most important variables selected from the predictors in Section 3.1.

Excess Returns



Realized Volatility

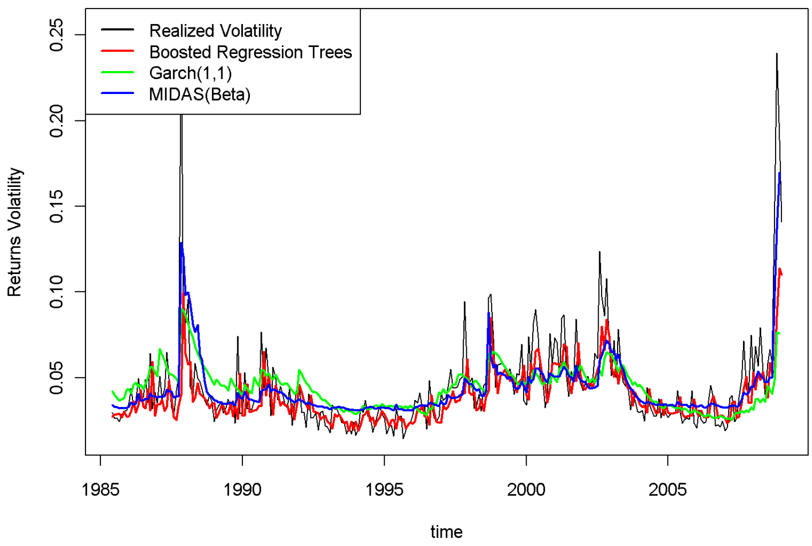


Figure 8: Actual versus out-of-sample predicted values of returns and realized volatility. The top graph shows the time series of excess returns plotted against out-of-sample predicted values from a boosted regression tree model and the prevailing mean. The bottom graph plots the realized volatility against the predicted values from a boosted regression tree, a GARCH(1,1) model and a MIDAS model with beta weights.

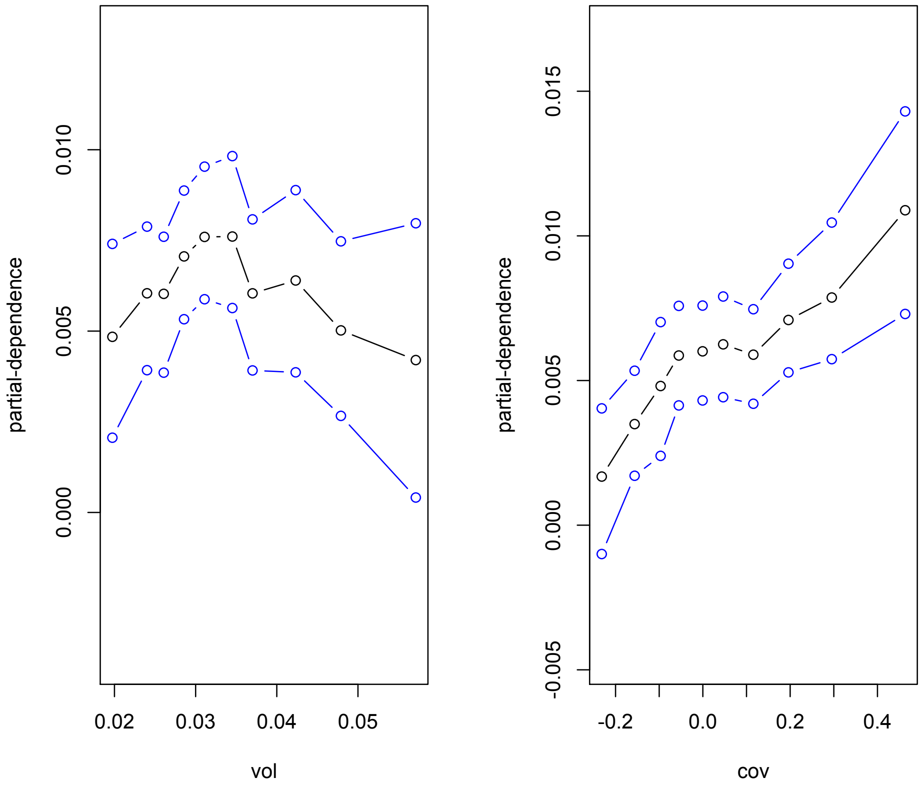


Figure 9: **Confidence bands on the risk-return relation for the model with the conditional covariance risk measure.** The figure shows partial dependence plots for the conditional mean return as a function of the conditional volatility (vol) and the conditional covariance between stock returns and changes to economic activity (cov). The plot is based on a Bayesian Additive Regression Tree (BART), using data over the sample 1960-2007. The horizontal axis covers the sample support of each predictor variable, while the vertical axis tracks the change in the conditional mean as a function of the individual state variables as well as its 90% confidence interval. Each circle represents one of the 10 quantiles of the volatility and covariance distributions (from the 5th to the 95th) at which the confidence intervals are computed. The conditioning information is the principal components derived from a set of 131 state variables and the three most important variables selected from the predictors in Section 3.1.

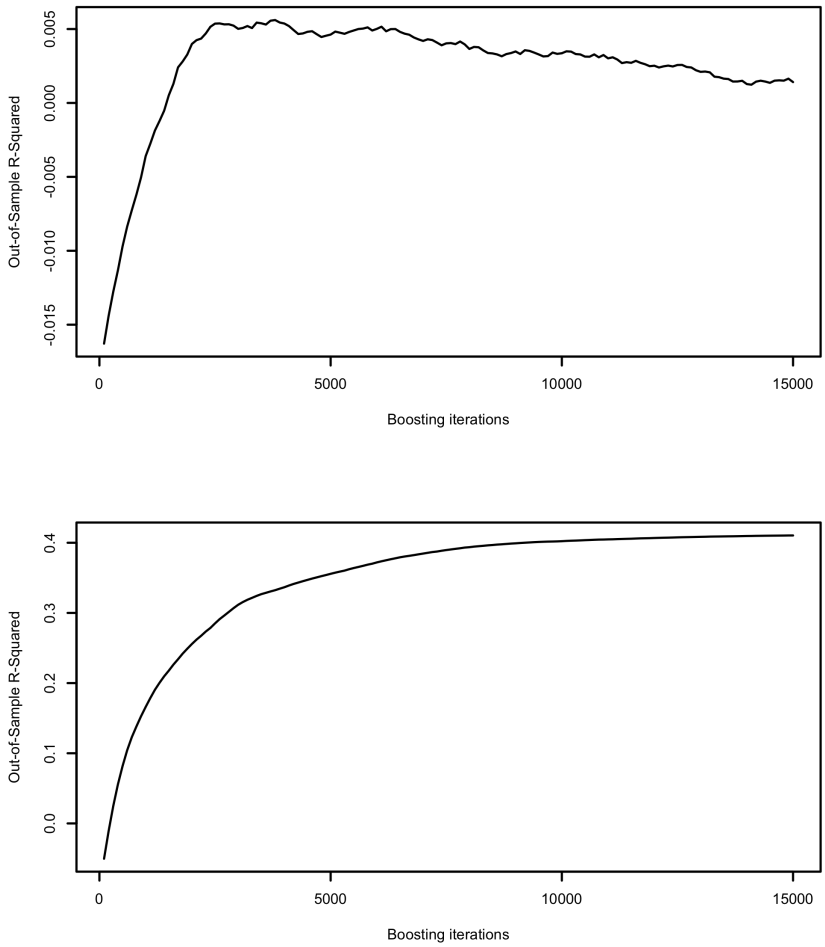


Figure 10: Out-of-sample forecasting performance (R-squared) of the boosted regression trees as a function of the number of boosting iterations (listed on the horizontal axis). The top panel shows results for excess returns, while the bottom panel covers the realized volatility.