What is the Shape of the Risk-Return Relation?‡

Alberto Rossi* Allan Timmermann†

*University of California San Diego
†University of California San Diego, CREATEES

June 8 2010

Abstract

Using a novel and flexible regression approach that avoids imposing restrictive modeling assumptions, we find evidence of a nonmonotonic 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. This finding helps explain the absence of a consensus in the empirical literature on the sign of the risk-return trade-off. 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 broader 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. Consumption CAPM.

‡We are grateful for comments and suggestions from seminar participants at the 2010 AFA meetings in Atlanta (especially Claudia Moise, the discussant), University of Texas at Austin, University of Houston, Rice, University of Toronto (Fields Institute), UCSD, Blackrock, the CREATEES-Stevanovich Center 2009 conference and the SoFie 2009 conference in Geneva. Allan Timmermann acknowledges support from CREATEES funded by the Danish National Research Foundation.
The existence of a systematic trade-off between risk and expected returns is central to modern finance. Yet, despite more than two decades of empirical research, there is little consensus on the basic properties of the relation between the equity premium and conditional stock market volatility. Studies such as Campbell (1987), Breen, Glosten, and Jagannathan (1989), Glosten, Jagannathan, and Runkle (1993), Whitelaw (1994), and Brandt and Kang (2004) find a negative trade-off, while conversely French, Schwert, and Stambaugh (1987), Bollerslev, Engle, and Wooldridge (1988), Harvey (1989), 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 returns is largely unrestricted with increasing, decreasing, flat, or nonmonotonic patterns all possible. Similar conclusions are drawn by studies such as Abel (1988), Gennette and Marsh (1993) and Veronesi (2000). 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 a monotonic risk-return trade-off. 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 for 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 a smooth and stable estimate of the shape of the risk-return relation. The approach is intuitive and resembles the binomial trees commonly used in financial analysis. Moreover, it allows us to track the impact and statistical significance of individual state variables.

Analysis of the risk-return relation has mostly been conducted by studying the time-series relation between the conditional mean and the conditional volatility of stock returns. Neither of these is observed, so in practice model-based proxies must be used. This introduces model and estimation errors and can bias results if the estimated models for expected returns or conditional volatility are misspecified, e.g., as a result of using overly restrictive models or including too few predictor variables.\footnote{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 volatility model is specified.}
In practice, it has proven difficult to effectively address these sources of bias because of the difficulty of maintaining a flexible functional form for the conditional mean and volatility models while also considering a large conditioning information set—a point emphasized by Ludvigson and Ng (2007). To reduce the danger of biases in the expected return and conditional volatility estimates, we make use of the ability of boosted regression trees (BRT) to form flexible estimates of expected returns and conditional volatility while dealing with large sets of predictor variables. Through out-of-sample forecasts we show that the approach does not overfit the data.

We adopt the BRT approach to empirically analyze the risk-return relation on monthly US stock returns over the period 1927-2008. We find evidence of a highly nonlinear effect of many predictor variables on both expected returns and conditional volatility, indicating that conventional linear models used to compute expected returns and conditional volatility are misspecified.

Using our improved estimates of expected returns and conditional volatility, we find strong empirical evidence of a nonmonotonic risk-return relation. At low and medium levels of conditional volatility, there is a significantly positive relation between the conditional mean and volatility of stock 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. Importantly, these results are not simply driven by a few isolated volatility spikes in the data. Formal statistical tests that account for sampling error soundly reject a monotonically increasing volatility-expected return relation.

Conditional volatility need not be an appropriate or exhaustive measure of risk. 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 the conditional volatility. Building on these models, we propose a new measure of covariance risk that is based on the high-frequency business activity index developed by Arouba, Diebold, and Scotti (2009). This index extracts daily estimates of business activity from a mixture of economic and financial variables observed at the daily, weekly, monthly and quarterly frequencies.

We first establish that there is a strongly positive correlation between changes to the economic activity index and consumption growth at the one, three, six and twelve-month horizons. We then propose to use daily changes to the economic activity measure as a proxy for the unobserved daily consumption growth. Following the approach of French, Schwert, and Stambaugh (1987) and Schwert (1989), we construct a “realized covariance” measure based on the daily changes in the economic activity index and daily stock market returns. In a final step, we model variations in the realized covariance as a function of a broad set of economic predic-

The review by Lettau and Ludvigson (2009) concludes that “the estimated risk-return relation is likely to be highly dependent on the particular conditioning variables used in any given empirical study.”
tor variables to obtain an estimate of conditional covariance risk. Consistent with the consumption CAPM we find evidence of a strongly positive relation between conditional covariance risk and expected returns. Moreover, statistical tests suggest that this relation is monotonic. 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.

These findings are robust to changing the analysis in various directions. We continue to find a nonmonotonic conditional mean-volatility trade-off but a monotonically rising mean-covariance relation at horizons of two and three months. When we follow the analysis in Ludvigson and Ng (2007) and extend our set of conditioning variables to include eight factors extracted from a larger set of more than 130 economic variables, we continue to find a nonmonotonic relation between expected returns and the conditional volatility, but a monotonically increasing relation between expected returns and the conditional covariance measure of risk. Using the Chicago Board of Trade VIX measure in place of the conditional volatility, we continue to find a nonmonotonic mean-volatility relation.

Our analysis provides a synthesis of many of the existing approaches from the literature on the risk-return trade-off. First, we consider a large set of conditioning variables to compute the conditional equity premium and volatility. 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. Second, in the spirit of the ICAPM we consider a model that includes both volatility and covariance measures of risk. Guo and Whitelaw (2006) argue that findings of a negative or insignificant risk-return relation is due to the omission of an intertemporal hedging component leading to a downward bias in the volatility coefficient. We find that our results are robust to the inclusion of both measures of risk. Third, 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. Fourth, 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).

In summary, the main contributions of our paper are as follows. First, we present a new, flexible econometric approach that reduces the risk of misspecification biases in generating 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 find that there is a nonmonotonic mean-volatility relation with expected returns first rising, then declining as the conditional volatility further increases. Third, we propose a new conditional covariance risk measure that
builds on the covariation between daily stock returns and daily economic activity. Fourth and finally, we show empirically that when the broader 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 I introduces our approach to modeling the risk-return relation. Section II describes the data and reports empirical results for the models used to generate estimates of the conditional mean and volatility. With these in place, Section III analyzes the conditional mean–volatility relation. Section IV introduces the new conditional covariance risk measure, while Section V conducts a series of robustness checks and extensions. Section VI concludes.

I. Methodology

To motivate our empirical research strategy, we first briefly review insights from the asset pricing literature on the relation between the risk premium and conditional volatility. We then discuss models for expected returns and conditional volatility and finally introduce the boosted regression tree methodology.

While the empirical literature has focused on determining the sign of the risk-return relation, theoretical analysis shows that the relation between expected returns and conditional volatility need not be linear or even monotonic. For example, Abel (1988) derives an equilibrium asset pricing model with a time-varying dividend process in which the equity premium need not be a monotonic function of volatility. Similarly, Veronesi (2000) shows that the relation between the conditional mean and conditional variance of stock returns is affected by a term that summarizes the effect of investors’ uncertainty about the economy’s growth rate on asset valuations. Changes in the economy’s level of uncertainty induce time-variations in the magnitude of this term and lead to an ambiguous relation between expected returns and conditional volatility.

Theoretical models thus do not generally constrain the risk-return relation to be linear or monotonic and so imposing such constraints on the risk-return relation is overly restrictive and can lead to biased estimates: “.. 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” (Gennaioli and Marsh (1993), page 1039).

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 when modeling the expected return and conditional volatility. We next describe an approach that accomplishes this.
I.A. Models of expected returns and conditional volatility

We initially focus on the relation between the conditional mean and the conditional volatility of stock market excess returns defined, respectively, as $\mu_{t+1|t} = E_t[\tilde{r}_{t+1}]$ and $\sigma_{t+1|t} = Var_t(\tilde{r}_{t+1})^{1/2}$, where $\tilde{r}_{t+1}$ is the stock return during period $t+1$, measured in excess of the risk-free rate. The conditional mean and variance, $E_t[\cdot]$ and $Var_t(\cdot)$, are computed conditional on information known to investors at time $t$. Both are ex ante measures that are unobserved and so empirical analysis typically relies on model-based proxies of the form

$$
\hat{\mu}_{t+1|t} = f_{\mu}(x_t|\hat{\theta}_\mu)
$$

$$
\hat{\sigma}_{t+1|t} = f_\sigma(x_t|\hat{\theta}_\sigma),
$$

where $x_t$ 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.

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

$$
\hat{\mu}_{t+1|t} = \beta'_\mu x_t
$$

$$
\hat{\sigma}_{t+1|t} = \beta'_\sigma x_t.
$$

As pointed out by Brandt and Wang (2007), there is little theoretical justification for imposing such ad-hoc restrictions on the models for expected returns and conditional volatility. Restricting the functional form to be linear can introduce model misspecification errors and bias empirical results in ways that make it difficult to interpret the results.

To address such biases, we extend the linear regression model in Eq. (2) to a class of flexible, semi-parametric models known as boosted regression trees. These have been developed in the machine learning literature and can be used to extract information about the relation between the predictor variables, $x_t$, and $\tilde{r}_{t+1}$ or $\hat{\sigma}_{t+1}$ based only on their joint empirical distribution. We do not consider conventional nonparametric or smoothing approaches since these are difficult to apply in the presence of many predictor variables and often produce poor out-of-sample forecasts.

To get intuition for how regression trees 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 predictor variables, $x_{1t}$ and $x_{2t}$ (e.g., the earnings-price ratio and the payout ratio). The functional form of the forecasting 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. This is an intuitive approach similar to the use of binomial trees in discrete time finance.

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. The process continues until some stopping criterion is reached. Boosted regression trees are additive expansions of regression trees, where each additional tree is fitted on the residuals of the previous tree. The number of trees used in the summation is also known as the number of boosting iterations.

This approach is illustrated in Figure 1, where we show boosted regression trees that use two state variables, namely the lagged values of the 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 only two terminal nodes, so every new tree stub generates two regions. The graph on the left uses only three boosting iterations. The resulting model ends up with one split along the payout ratio axis and two splits along the earnings-price ratio axis. Within each state the predicted value of stock returns is constant. The predicted mean excess return is smallest for high values of the payout ratio and low values of the earnings-price ratio, and highest when the payout ratio is small and the earnings-price ratio is high. With only three boosting iterations the model is quite coarse. This changes 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.

Figure 1 illustrates how boosted regression trees can be used to approximate the relation between the dependent variable and a set of predictors by means of a series of piece-wise constant functions. This approximation is good even in situations where, say, the true relation is linear, provided that sufficiently many boosting iterations are used. We next provide a more formal description of the methodology and explain how we implement it in our study.\footnote{Our description draws on Hastie, Tibshirani, and Friedman (2009) who provide a more in-depth coverage of the approach.}

\section{I.B. Regression trees}

Consider a sample of $T$ time-series observations on a single dependent variable, $y_{t+1}$, and $P$ predictor (state) variables, $x_t = (x_{1t}, x_{2t}, ..., x_{pt})$, 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 splitting points to use. A given splitting point may lead to $J$ disjoint sub-regions or states,
The dependent variable is modeled as a constant, \( c_j \), within each state, \( S_j \). The value fitted by a regression tree, \( T(x_t, \Theta_j) \), with \( J \) terminal nodes and parameters \( \Theta_j = \{S_j, c_j\}^J_{j=1} \) can thus be written\(^3\)

\[
T(x_t, \Theta_j) = \sum_{j=1}^{J} c_j I\{x_t \in S_j\},
\]

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, \( \sum_{t=1}^{T}(y_{t+1} - f(x_t))^2 \), the estimated constant, \( \hat{c}_j \), is the average of \( y_{t+1} \) in state \( S_j \):

\[
\hat{c}_j = \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_j\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_j\}.
\]

The 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. Furthermore, there is no guarantee that the sequential splitting algorithm leads to the globally optimal solution. To deal with these problems, we next consider a method known as boosting.

### I.C. Boosting

Boosting is a technique that 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 and has been found to improve the precision of forecasts of stock returns, see Rapach, Strauss, and Zhou (2010).

\(^3\)Only \( J - 1 \) splitting point parameters are needed to generate \( J \) states, but we suppress this to keep the notation simpler.
A boosted regression tree is simply the sum of individual regression trees:

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

where \( \mathcal{T}_b(x_t, \Theta_{J,b}) \) is the regression tree of the form (3) 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

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

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 \( j \)-th state.

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

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

Specifically, we use a shrinkage parameter, \( \lambda \), which reduces the amount by which each boosting iteration contributes to the overall fit:

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

Following common practice, we set \( \lambda = 0.001 \). In addition, 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. By fitting each tree only on a subset of the data, this method again reduces the risk of overfitting.

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 likely to be more robust to outliers.

II. Empirical Estimates of Expected Returns and Conditional Volatility

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

II.A. Data

Theoretical models generally do not offer specific guidance on which variables to use when modeling expected returns and volatility. However, empirical studies have found evidence of time variations in both expected stock market returns and volatility.\(^4\) We therefore take a broad view and consider a range of state variables from the empirical finance literature.

In particular, our empirical analysis uses a data set comprising monthly stock returns along with a set of twelve predictor variables previously analyzed in Welch and Goyal (2008) extended to cover the sample 1927-2008.\(^5\) Stock returns are tracked by the S&amp;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.

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

\[
\hat{\sigma}^2_t = \sum_{i=1}^{N_t} r_{i,t}^2.
\]

Theoretical foundations for the use of (8) can be found in Andersen, Bollerslev, Christoffersen, and Diebold (2006). The estimator in (8) is only free of measurement errors as the sampling frequency approaches infinity, so \(\hat{\sigma}^2_t\) is best thought of as a variance proxy.

\(^4\)See Lettau and Ludvigson (2009) for a survey.
\(^5\)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.
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 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 \( t + 1 \).

By considering this large set of predictor variables, we address a potentially important source of model misspecification caused by omitted variables (Ludvigson and Ng (2007)). Later we further consider economic factors extracted from an even larger set of more than 130 economic variables.

II.B. Influence of individual predictor variables

In linear regression models, 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 that do not impose linearity. As an alternative measure of influence, we consider the reduction in the forecast error every time a particular predictor 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 gives a measure of the variable’s influence (Breiman (1984)):

\[
I_p(T) = \sum_{j=2}^{J} \Delta|e(j)|I(v(j) = p),
\]

where \( \Delta|e(j)| = T^{-1} \sum_{t=1}^{T} (|e_t(j-1)| - |e_t(j)|) \), is the reduction in the size of the forecast error at the \( j' \)th node and \( v(j) \) is the variable chosen at this node, so \( I(v(j) = p) \) equals one if variable \( p \) is chosen and otherwise is zero. The sum is computed across all time periods, \( t = 1,...,T \) and nodes of the tree. 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.

Averaging over the number of boosting iterations, \( B \), and dividing by the resulting values summed across all predictor variables gives a measure of relative influence, \( \overline{RI}_p \):

\[
\overline{RI}_p = \frac{\bar{I}_p}{\sum_{p=1}^{P} I_p},
\]

where \( \bar{I}_p = \frac{1}{B} \sum_{b=1}^{B} I_p(T_b) \). This measure sums to one and can be compared across predictor variables. It does not tell if a particular predictor variable is capable of improving the forecasting performance relative to, say, a model with no predictor variables. This question is best addressed by analyzing the model’s out-of-sample forecasting performance, a point we later return to in Section V.

II.B.1. Equity premium model

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

The results suggest that two predictor variables dominate over the full sample, 1927-2008, as inflation and the earnings-price ratio both obtain weights above 17%. These are also the only individual variables whose relative influence measure are statistically significant at the 5% level.\(^6\) The de-trended T-bill rate and the long bond yield get weights close to 10% and both the top three and top five predictor variables are joint statistically significant in the full sample.

The empirical results are quite consistent over time as inflation and the earnings-price ratio are ranked first and second in both subsamples. Both of these variables, as well as the top three and top five predictors, fail to be significant in the first subsample but are statistically significant in the second subsample, 1968-2008.

II.B.2. Volatility model

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

---

\(^6\)To assess the statistical significance of the individual predictor variables, we undertake the following Monte Carlo analysis. We fix the ordering of the dependent variable and all predictor 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 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 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. We use a similar test for the joint significance of the top three and top five predictor variables.
get lower weights. Despite their small relative influence, the marginal effects of several other variables, including the default spread, past excess returns, the payout ratio, inflation and the earnings-price ratio are statistically significant. This can be attributed to the fact that the lagged volatility absorbs a large portion of the variation in the realized volatility.

With exception of the lagged volatility, no variable repeats in the top three predictors in the two subsamples. Interestingly, the dividend-earnings ratio and the default spread get weights of nearly 14% over the period 1927-67, while only lagged excess returns and, again, the lagged volatility get a weight greater than 10% in the second subsample, 1968-2008.

II.C. Marginal effect of individual state variables

To gain intuition for the boosted regression trees, it is useful to explore the relation between individual state variables and returns or realized volatility. The regression trees do not impose restrictions on the functional form of the relation between the dependent variable—returns or realized volatility—and the predictor variables. Measuring the effect of the predictor variables on the dependent variable is therefore more complicated than usual.

To address this point, we compute marginal effects by fixing the value of a particular variable, \( x_p \), and averaging out the effect of the remaining variables. Repeating this process for different values of \( x_p \) yields a partial dependence plot showing the effect a particular variable has on the predicted variable.

Figure 2 presents such plots for the three most important predictor variables in the model for stock returns. As in Figure 1, flat spots show that the mean excess return does not change in a particular range of the predictor variable. The relation between expected stock returns and the 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. Hence, in a state of deflation rising consumer prices are associated with higher mean returns. Conversely, at positive levels of inflation, higher consumer prices are associated with lower mean returns, although at very high levels of inflation there is no systematic relation between inflation and stock market performance. These effects are quite strong in economic terms: the difference between expected returns evaluated at small and large values of the inflation rate is 2% per month. Similarly, 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 of this measure. These findings suggest that linear specifications of expected returns are misspecified.

Turning to the volatility model, current realized volatility is clearly an important predictor of future volatility. The partial dependence plots in Figure 3 show that 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. However, very high values of past volatility do not translate into correspondingly high values of expected future volatility, as evidenced by the flatness of the relation at high levels of volatility.

A highly nonlinear pattern is also found in the relation between the conditional volatility and the default spread. At small or medium values of the spread, future volatility is increasing in this variable. However, at high values of the spread, the expected volatility remains constant. Past returns are also related to current volatility, but the relation is negative as higher past returns imply lower expected volatility.

III. Estimates of the Risk-Return Relation

We next use our estimates of the conditional mean and volatility of stock returns from section II to model the shape of the risk-return trade-off. We first consider linear models for this relation and then generalize the setup to allow for a general (unrestricted) mean-volatility relation. Finally, we conduct a formal test of monotonicity of the relation between conditional volatility and expected returns.

III.A. Linear risk-return model

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

$$\hat{\mu}_{t+1|t} = \alpha + \beta_1 \hat{\sigma}_{t+1|t} + \beta_2 \hat{\sigma}_{t|t-1} + \beta_3 \hat{\mu}_{t|t-1} + \epsilon_{t+1},$$  \hspace{1cm} (11)

where ‘hats’ indicate estimated values from the boosted regression trees. In a generalization of the conventional volatility-in-mean model, lags are included in order 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 2. 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 2.8. Conversely, the effect of lagged volatility is strongly negative, while the effect of lagged expected returns is strongly positive. These results carry over to the first subsample, 1927-67. In the second subsample, 1968-2008, as well as during two high-volatility periods (1927-39 and 2001-2008), the relation between the conditional mean and both the current and lagged conditional volatility is, however, insignificant and much weaker.
An alternative to using model-based estimates of the conditional volatility is to use a market-based estimate in the form of the Chicago Board Options Exchange Index, commonly known as the VIX. The VIX is effectively a market-based estimate of the conditional volatility of the S&P 500 index over the next 30 days. Data on the VIX are available over the period 1986-2008.

Panel B1 of Table 2 shows results for the linear risk-return specification based on the VIX measure of conditional volatility:

\[
\hat{\mu}_{t+1} = \alpha + \beta_1 \hat{\sigma}_{VIX} + \beta_2 \hat{\sigma}_{VIX-1} + \beta_3 \hat{\mu}_{t-1} + \varepsilon_{t+1},
\]

Similar to our finding for the second subsample in Panel A1, the coefficient on current VIX is positive but statistically insignificant.

III.B. Flexible risk-return model

The results reported so far suggest that the linear risk-return specification is not time-invariant or robust. As emphasized by Merton (1980), the risk-return relation need not be linear of course. To explore this point, we next employ the boosted regression trees to model the relation between expected returns and conditional volatility in a way that avoids imposing particular functional form assumptions. Specifically, we generalize (11) to the following model

\[
\hat{\mu}_{t+1} = f(\hat{\sigma}_{t+1}, \hat{\sigma}_{t-1}, \hat{\mu}_{t-1}) + \varepsilon_{t+1}.
\]

Panel A2 in Table 2 presents estimates of the relative influence of the three variables in this model. The relative weight on current conditional volatility is 8% for the full sample and is statistically significant at the 5% level. 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 for lagged expected returns in the linear model. Interestingly, in the first subsample, 1927-67, as well as during the high-volatility periods, 1927-1939 and 2001-2008, the current conditional volatility obtains a much greater (and significant) weight of 21%, but the weight declines to 10% in the second subsample, 1968-2008, in which it fails to be significant.

Panel B2 in Table 2 presents relative influence estimates for the flexible model in Eq. (13) when the VIX is used to measure volatility expectations. The relative weight on current conditional volatility is 18%, the weight on lagged volatility is 11% and the weight on the lagged expected return is 70%. Only the latter estimate is significant.

Turning to the shape of the risk-return relation, Figure 4 shows that the trade-off between concurrent expected returns and conditional volatility is highly nonlinear.
First consider the full sample, 1927-2008. At low-to-medium levels of volatility, a strongly positive relation emerges in which 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.

Our finding of a nonmonotonic 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.

In contrast with the analysis in Brandt and Wang (2007), we assume a constant risk-return relation. However, we can at least in part address this issue by applying our methodology to subsamples of the data. Doing this, we find that the inverted risk-return trade-off is a robust finding in the sense that it appears not to be confined to a particular historical period. This point is illustrated in the middle and right windows in Figure 4 which show the shape of the risk-return relation for the two subsamples, 1927-67 and 1968-2008. At low levels of volatility the expected return increases sharply in both subsamples as the conditional volatility rises. However, at higher volatility levels the expected return declines or remains constant as the conditional volatility rises further.

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

Our plots suggest marked nonmonotonicities in the conditional volatility-expected return relation, but they do not demonstrate that this relation is nonmonotonic in a statistically significant way. We next address this point in a more rigorous fashion.

III.C. Formal tests of monotonicity

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 (2009). To this end, we sort pairs of monthly observations into $g = 1,...,G$ groups, $\{\hat{\mu}_{t+1|t}, \hat{\sigma}_{t+1|t}\}$, ranked by the conditional volatility. A monotonic mean-volatility relation implies that, as we move from groups with low conditional

---

7One issue 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 a concern here. In the full-sample plot in Figure 4, 37% of the observations lie to the left of the steeply 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.
volatility to groups with high conditional volatility, mean returns should rise.\footnote{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. (13). Hence our analysis is based on the relation between the marginalized conditional mean and the marginalized conditional volatility.}

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

$$H_0 : E \left[ \hat{\mu}_{t+1|t} | \hat{\sigma}_{t+1|t} \right] > E \left[ \hat{\mu}_{t+1|t} | \hat{\sigma}_{t+1|t} \right], \text{ for } g = 2, \ldots, G.$$ \hspace{1cm} (14)

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

$$H_0 : \Delta \geq 0.$$ \hspace{1cm} (15)

To test this hypothesis, we use the test statistic of Wołak (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 approximate critical values can be computed via Monte Carlo simulation.

For robustness, we perform the test on different number 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, $\hat{\mu}_{t+1:t+h|t}$, and conditional volatility, $\hat{\sigma}_{t+1:t+h|t}$, and conduct tests for horizons of $h = 1, 2, 3$ months.\footnote{Going beyond the one-quarter horizon entails a significant decline in the sample size and a resulting loss in power.}

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 and quarterly horizons. These results show that a monotonically increasing relation between the conditional mean and the conditional volatility is strongly rejected at various horizons, providing strong evidence of a nonlinear mean-volatility relation.

Panel B of Table 3 reports the corresponding $p$-values for the monotonicity test when the VIX is used to measure volatility expectations. In this case we can only conduct the analysis at the one-month horizon, given that indices for options with longer expiration dates are not available. The $p$-values are below 5% in two of three
cases and below 10% in the third case. These results show that a monotonically increasing relation between the conditional mean and volatility is rejected, providing further evidence of a nonlinear mean-volatility relation.

IV. A New Conditional Covariance Risk Measure

Intuitively, we would expect to find a positive trade-off between risk and expected returns. It is possible that our findings so far simply indicate that conditional volatility is not an exhaustive measure of risk, particularly at high volatility levels. Consumption based 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. Testing such models is challenging, however, in part because high frequency consumption data is not available. To address this issue, we propose in this section a new covariance risk measure.

To motivate this risk measure, note that under assumptions of power utility of consumption, $u(C_{t+1}) = C_{t+1}^{1-\gamma}/(1-\gamma)$, $\gamma \geq 0$, and log-normally distributed consumption growth, expected excess returns on the stock market portfolio satisfy

$$E_t[r_{t+1}] \approx \gamma \text{cov}_t(\Delta c_{t+1}, r_{t+1}),$$

where $\text{cov}_t(\Delta c_{t+1}, r_{t+1})$ is the conditional covariance between consumption growth, $\Delta c_{t+1} = \log(C_{t+1}/C_t)$, 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:

$$\frac{\partial E_t[r_{t+1}]}{\partial \text{cov}_t(\Delta c_{t+1}, r_{t+1})} > 0.$$  \hspace{1cm} (17)

In situations in which 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, $\Delta EA_{t+1}$:

$$\frac{\partial E_t[r_{t+1}]}{\partial \text{cov}_t(\Delta EA_{t+1}, r_{t+1})} > 0.$$  \hspace{1cm} (18)

Intuitively, the higher the covariance between changes to economic activity and stock returns, the lower returns tend to be during economic recessions where marginal utility of consumption also 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 $\text{cov}_t(\Delta EA_{t+1}, r_{t+1})$
can be constructed from daily data on economic activity.

IV.A. Realized Covariance

Following Eq. (18), we propose an indirect test of the basic implication of consumption-based asset pricing models that there should be a monotonically increasing relation between high-frequency consumption proxies and expected stock returns. We do so by proxying high-frequency consumption by means of the ADS business conditions index proposed by Arouba, Diebold, and Scotti (2009). Daily data on this is 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. This is accomplished by using a dynamic factor model estimated through the Kalman filter. The top window in Figure 5 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 but it is important to check if this holds. 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. Results from this analysis are reported in Table 4. 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. This suggests that we can use high frequency changes to this index as a proxy for the unobserved consumption growth. Specifically, we compute monthly “realized covariances” between stock returns and changes in the ADS index, $\hat{\text{cov}}_{it}$, from observations at the daily frequency,

$$\hat{\text{cov}}_{it} = \sum_{i=1}^{N_t} \Delta ADS_{i,t} \times r_{i,t},$$

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. We use the name ‘realized covariance’ by analogy with how realized variance measures such as Eq. (8) are computed.\(^{10}\)

\(^{10}\)Since the ADS economic activity index is itself a filtered estimate of the underlying unobserved
IV.B. Variations in the conditional covariance

Armed with the new “realized covariance” risk measure, we next consider which state variables help predict time variations in the conditional covariance. As for the models of expected return and realized volatility in Eq. (1), we construct estimates of the conditional covariance from boosted regression trees:

\[ \hat{\text{cov}}_{t+1|t} = f_{\text{cov}}(x_t | \hat{\theta}_{\text{cov}}), \]

where \( x_t \) is the set of predictor variables and \( \hat{\theta}_{\text{cov}} \) are estimates of the parameters.

Empirical results for this model are shown in Panel C of Table 1. Interestingly, once again the inflation rate comes out on top with a relative influence of 13%, while the payout ratio, the long-term rate of return and the term spread also obtain weights exceeding 10%. In contrast to the volatility results, the lagged value of the covariance does not play a major role in predicting variations in the current covariance and no individual predictor variable appears to dominate the estimates.

None of the individual predictor variables is significant at the 5% level which in part can be attributed to the shorter sample used here, 1960-2008. However, the top three variables (inflation, the payout ratio and the long-term rate of return) are jointly significant at the 1% level as are the top five predictor variables. Thus, these predictor variables jointly capture time variations in the conditional covariance.

The bottom window in Figure 5 plots monthly values of the conditional covariance between the ADS index and stock returns, divided by the product of the standard deviations of returns and changes to the ADS index, respectively, so it is similar to a conditional correlation. The conditional covariance is clearly countercyclical and rises during economic recessions.

Figure 6 shows that the relation between inflation and the conditional covariance is remarkably similar to the relation between inflation and expected returns uncovered earlier. At negative values of inflation a flat relation emerges, but the realized covariance measure declines steeply as inflation turns positive, only to stabilize for mildly positive rates of inflation. Conversely, high payout ratios and high long term returns are both associated with a higher conditional covariance, although once again the relation is strongly nonlinear.

IV.C. Trade-off between the conditional covariance and expected returns

We saw earlier (Table 4) that there is a clear positive correlation between consumption growth and changes to the economic activity index. This suggests that we economic activity process, it is important to point out that this realized covariance measure is a proxy and need not have the same statistical properties as conventional realized variance measures.
should find a monotonically increasing relation between expected returns and our new conditional covariance risk measure.

To test if this holds, in parallel with Eq. (11) we estimate a linear model relating expected returns to the conditional covariance as well as lags of these variables:

$$\hat{\mu}_{t+1|t} = \alpha + \beta_1 \hat{\text{cov}}_{t+1|t} + \beta_2 \hat{\text{cov}}_{t|t-1} + \beta_3 \hat{\mu}_{t|t-1} + \varepsilon_{t+1}. \quad (21)$$

Table 5 shows that the conditional covariance measure obtains a positive and highly statistically significant coefficient in this linear risk-return specification with a $t$-statistic around eight.

Following Eq. (13) we also consider a model that is not restricted to be linear,

$$\hat{\mu}_{t+1|t} = f(\hat{\text{cov}}_{t+1|t}, \hat{\text{cov}}_{t|t-1}, \hat{\mu}_{t|t-1}) + \varepsilon_{t+1}. \quad (22)$$

For this specification, Panel B in Table 5 shows that the relative influence of the conditional covariance measure is 26% and is highly statistically significant. This compares with a relative influence measure for conditional volatility of only 7% in the earlier model.

Figure 7 shows the partial dependence plot for the current and lagged conditional covariance and lagged expected returns in Figure 7. As predicted by theory, a monotonically rising relation between the current conditional covariance of economic activity and expected stock returns emerges. Notice also the large variation in the fitted expected return (shown on the vertical axis) which varies by 1% per month. This compares with only one-third of one percent per month in the plot relating the conditional volatility to expected returns over approximately the same period (the right window in Figure 4). Our finding of a strong negative effect of the lagged conditional covariance on expected returns is similar to, but extends to the conditional covariance measure, the finding by Ludvigson and Ng (2007) of a negative lagged volatility-in-mean effect.

These findings suggest that the conditional covariance risk measure plays an important role in explaining variation in expected stock returns. Moreover, consistent with the consumption CAPM, expected returns appear to be a monotonically increasing function of the conditional covariance risk measure.

To test more formally if this holds, Panel C in Table 3 repeats the earlier monotonicity tests using the conditional covariance as our risk measure for horizons of $h = 1, 2, 3$ months. Consistent with the visual impression from Figure 7, the null hypothesis of a positive and monotonic relation between expected returns and the conditional covariance is no longer rejected for eight of nine cases. Hence, the empirical evidence is consistent with a strongly increasing relation between expected

\[^{11}\text{Similarly, Guo and Whitelaw (2006) find that expected returns are mainly driven by changes in the investment opportunity set.}\]
returns and the conditional covariance risk measure.

IV.D. Tests of the ICAPM

So far we have considered conditional volatility of stock returns or the conditional covariance between economic activity and stock returns as separate measures of risk. The latter is appropriate if the conditional covariance measure is viewed strictly as a proxy for the covariance between consumption growth and stock market returns. However, because it is based on a broad economic activity measure, the conditional covariance could also be affected by other state variables tracking time-varying investment opportunities. This suggests testing a ‘hybrid’ version of the consumption CAPM and intertemporal CAPM with both the conditional volatility and conditional covariance measures included.

Following Guo and Whitelaw (2006) we therefore consider the following ICAPM specification:

\[ E_t[r_{t+1}] = a_W \sigma_t^2(r_{t+1}) + b_W \text{cov}_t(r_{t+1}, x_{t+1}), \]  

(23)

where \( \text{cov}_t(r_{t+1}, x_{t+1}) \) represents the conditional covariance of returns with investment opportunities, tracked by the state vector, \( x_{t+1} \). The coefficients \( a_W \) and \( b_W \) depend on the representative investor’s wealth, \( W_t \), and indirect utility function, \( J(W_t, x_t, t) \).

The empirical analysis considers both linear and flexible risk-return specifications and controls for dynamic effects:

\[
\begin{align*}
\hat{\mu}_{t+1} & = \alpha + \beta_1 \hat{\tau}_{t+1} + \beta_2 \hat{\sigma}_{t+1} + \beta_3 \hat{\mu}_{t-1} + \beta_4 \hat{\sigma}_{t-1} + \beta_5 \hat{\tau}_{t-1} + \epsilon_{t+1}, \\
\hat{\mu}_{t+1} & = f(\hat{\sigma}_{t+1}, \hat{\sigma}_{t-1}, \hat{\mu}_{t-1}, \hat{\sigma}_{t-1}, \hat{\tau}_{t-1}, \epsilon_{t+1}).
\end{align*}
\]

(24)

(25)

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.\(^{12}\)

Table 6 presents estimation results for the models in Eqs. (24-25) 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 volatility 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 general specification reported in Panel B, the covariance mea-

\(^{12}\)Suppose that, as documented earlier, the economic activity index depends both on consumption growth and ICAPM state variables, \( x_{t+1} \), that track time-varying investment opportunities, i.e.,

\[ \Delta E_i A_{t+1} = \lambda_0 + \lambda_1 \Delta c_{t+1} + \lambda_2 x_{t+1}. \]

Then it follows that the specifications in (24) and (25) nest both the consumption CAPM (if \( \lambda_2 = 0 \) and conditional volatility does not matter in explaining variations in expected returns) and the ICAPM.
The conditional covariance, $\overline{\sigma}_{t+1|t}$, obtains a relative influence of 13.2% which is significantly different from zero ($p$-value of 0.0%), whereas the relative importance of the conditional volatility, $\hat{\sigma}_{t+1|t}$, is 6.6% ($p$-value of 8.8%). The formal monotonicity tests in Panel C confirm this conclusion. We fail to reject the null hypothesis of a monotonically increasing relation between the conditional covariance and expected returns but reject, in two out of three cases, that there is a monotonic relation between the conditional volatility and expected returns.$^{13}$

The partial dependence plots for the joint model shown in Figure 8 further corroborate our earlier 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. In comparison, variations in expected returns change by less than 2% per annum due to fluctuations in the conditional volatility.

V. 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.

V.A. 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.$^{14}$ In particular, 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 expand the data window to...

---

$^{13}$The case where we fail to reject compares fewer portfolios and so could excessively smooth out nonmonotonicities in the expected return-volatility relation.

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.

V.A.1. Return forecasts

As a first check of whether the boosted regression trees overfit the data, consider the top window in Figure 9 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 7 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.\(^\text{15}\)

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.

V.A.2. Volatility forecasts

Turning to the volatility forecasts, Panel A2 of Table 7 and the bottom panel in Figure 9 compare volatility forecasts from the regression tree to those from a GARCH(1,1) model or an autoregressive model that exploits the persistence in realized volatility. We also consider forecasts from a MIDAS model of the form proposed by Ghysels, Santa-Clara, and Valkanov (2005). Following their analysis,\(^\text{15}\) For the latter, we use the Bayesian information criterion to select the best specification among the 2\(^{12}\) possible linear models that use different combinations of the predictor variables.
we adopt a MIDAS estimator of the conditional variance of monthly returns:

$$\text{Var}_t(R_{t+1}) = 22 \sum_{d=0}^{D} w_d r_t^2 - d,$$  \hfill (26)

where

$$w_d(\kappa_1, \kappa_2) = \frac{\exp(\kappa_1 d + \kappa_2 d^2)}{\sum_{i=0}^{D} \exp(\kappa_1 i + \kappa_2 i^2)},$$

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

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 9 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.

### V.B. 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 7 reports the out-of-sample forecasting performance of boosted regression trees using 5,000, 10,000 and 15,000 boosting iterations.\(^{16}\) 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.

\(^{16}\)Since these are based on out-of-sample forecasts, we only report results for the mean and volatility.
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 overfitting for the volatility prediction model. This 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 $B = 1, 2, \ldots, 10,000$ boosting iterations.

Panel B1 in Table 7 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).

V.C. Expanding the information set: common factors

Ludvigson and Ng (2007) argue that the relation between expected returns and conditional volatility critically depends on the set of variables used to generate estimates of the conditional mean and volatility of stock returns. Their empirical results suggest the need to consider a broad set of macroeconomic and financial state variables that can be summarized through a small set of common factors. Following their analysis, we assume that a large set of state variables $x_{it}, i = 1, \ldots, N$ are generated by a factor model of the form

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

(27)

where $f_t$ is a vector of common factors, $\lambda_i$ is a set of factor loadings and $e_{it}$ is an idiosyncratic error. Using the common factors as predictor variables rather than the $N$ individual regressors typically 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 which are grouped into eight broad categories: output and income; labor market; housing; consumption, orders and inventories; money and credit; bond and exchange; prices; stock market. As in Ludvigson and Ng (2009), we interpret
each of the common factors through the marginal $R^2$—values obtained by regressing each of the 131 series on each of the extracted factors. This gives us a relatively straightforward way of characterizing the information content of each factor: $\hat{f}_1$ is a real activity factor that loads heavily on employment and output data, $\hat{f}_2$ loads on interest rate spreads, while $\hat{f}_3$ and $\hat{f}_4$ load on prices. $\hat{f}_5$ loads on interest rates, $\hat{f}_6$ on housing variables and $\hat{f}_7$ on measures of the money supply. Finally, $\hat{f}_8$ loads on variables related to the stock market.

Ludvigson and Ng assume a linear relation between volatility or expected returns and the extracted factors. We generalize this to allow for flexible and potentially nonlinear relations of the form

$$
\begin{align*}
\hat{\mu}_{t+1|t} &= f_{\mu}(x_t, \hat{f}_i|\theta_{\mu}) \\
\hat{\sigma}_{t+1|t} &= f_{\sigma}(x_t, \hat{f}_i|\theta_{\sigma}) \\
\hat{\text{cov}}_{t+1|t} &= f_{\text{cov}}(x_t, \hat{f}_i|\theta_{\text{cov}}).
\end{align*}
$$

Finally, we relate the factor-based estimates of expected returns, volatility and covariance using the earlier setup from Eq. (25).

Table 8 reports the relative influence of the eight factors in the return, volatility and covariance forecasting models. We also include the three variables from the earlier list deemed to be most influential over the period 1960-2007. As expected, not all factors are equally relevant. For the equity premium model, a particularly heavy weight is assigned to the factor that loads on interest rate spreads ($\hat{f}_2$). Factors appear to be less influential in explaining stock market volatility as all factors obtain weights of 4% or less. Conversely, for the covariance model, the price factor ($\hat{f}_3$) appears to be very important, obtaining a weight in excess of 15%. In all three cases, the relative influence of the top three and top five predictor variables is statistically significant at the 5% level.

Based on the results in Table 8, we generate estimates of the conditional mean, volatility and covariance and use these to model the risk-return relation in the context of the earlier ICAPM specification in Eq. (25). The left window in Figure 11 shows the shape of the risk-return relation when we use the variables in Table 8 as conditioning information. Expanding the information set to include factors seems to reinforce the nonlinearities between expected returns and conditional volatility uncovered in Figure 8. In contrast, the right window shows that, once again, a clear monotonically increasing relation emerges between the conditional covariance measure and expected stock returns.\(^{17}\) Again we see a far greater range of varia-

\(^{17}\)This is confirmed by formal statistical tests. For the volatility-expected return relation, we reject the null of a monotonically increasing volatility-expected return relation as we obtain $p$—values between 0.001 and 0.097 (depending on the number of groups used in the test). For the covariance risk measure, we obtain $p$—values between 0.19 and 0.85, so we fail to reject the null of a monotonically increasing covariance-expected return relation.
tion in expected returns due to changes in the conditional covariance risk measure compared to when the conditional volatility measures risk.

We conclude that the earlier findings on the shape of the risk-return relation remain valid when a very large conditioning information set is used. A nonmonotonic mean return-conditional volatility relation emerges, while conversely the expected return is a monotonically increasing function of the conditional covariance.

VI. Conclusion

This paper proposes a new and flexible approach to modeling the conditional equity premium and the conditional volatility 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 less likely to be 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 and why results from linear models could be sensitive to how many high volatility episodes were included in the sample under study.

The nonmonotonic trade-off between conditional volatility and expected returns uncovered in our analysis indicates some 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 broad measures of risk that account for the state of the economy.
References


This table shows the relative influence weight (in percent) and significance (Monte Carlo p-value) of the 12 predictor variables considered in the boosted regression tree model. A higher weight indicates that a predictor variable is more important. The 12 predictors are the inflation rate (infl), the log earnings-price ratio (ep), the 1-month T-bill rate relative to its three-month average (rel), the long-term government bond yield (lty), stock market volatility (vol), the log dividend-earnings ratio (de), the risk-free T-bill rate (rfree), the log dividend-price ratio (dp), the excess return on stocks (exc), the term spread (tms), the long-term rate of return (ltr) and the default spread (defspr). Results in Panel A use monthly returns on the S&P 500 index measured in excess of the T-bill rate as the dependent variable. Results in Panel B are for the realized volatility, measured as the square root of the sum of squared daily returns within a given month. Results in Panel C are for the realized covariance between changes in the ADS index and stock returns within a given month, computed using daily observations. The last two columns of each panel present the joint relative influence and joint significance of the three and five most important variables considered in each model.

<table>
<thead>
<tr>
<th></th>
<th>1927-2008</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>infl</td>
<td>ep</td>
<td>rel</td>
<td>lty</td>
<td>vol</td>
<td>de</td>
<td>rfree</td>
<td>dp</td>
<td>exc</td>
<td>tms</td>
<td>ltr</td>
<td>defspr</td>
<td>Top 3</td>
</tr>
<tr>
<td>relative influence</td>
<td>17.4%</td>
<td>17.1%</td>
<td>10.1%</td>
<td>9.47%</td>
<td>8.80%</td>
<td>7.50%</td>
<td>6.07%</td>
<td>5.84%</td>
<td>5.79%</td>
<td>5.28%</td>
<td>4.26%</td>
<td>2.62%</td>
<td>44.89%</td>
</tr>
<tr>
<td>p-value</td>
<td>0.1%</td>
<td>0.4%</td>
<td>11.5%</td>
<td>11.0%</td>
<td>29.5%</td>
<td>38.8%</td>
<td>47.5%</td>
<td>67.1%</td>
<td>64.4%</td>
<td>67.1%</td>
<td>88.0%</td>
<td>98.9%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

|                   | 1927-1967 |   |   |   |   |   |   |   |   |   |   |   |                   |
| relative influence| 12.4%     | 11.6% | 10.35% | 9.30% | 9.27% | 8.99% | 8.89% | 8.29% | 7.78% | 5.62% | 4.59% | 3.50% | 33.77% | 52.34% |
| p-value            | 11.0%     | 24.5% | 37.8% | 51.0% | 60.4% | 57.5% | 42.1% | 50.9% | 56.9% | 98.5% | 99.1% | 100% | 28.5% | 52.50% |

|                   | 1968-2008 |   |   |   |   |   |   |   |   |   |   |   |                   |
| relative influence| 16.49% | 12.54% | 12.92% | 10.97% | 8.45% | 7.31% | 6.47% | 5.80% | 5.59% | 4.99% | 4.64% | 4.18% | 41.55% | 60.97% |
| p-value            | 0.6%      | 6.3% | 7.3% | 12.6% | 47.1% | 67.7% | 70.0% | 77.4% | 87.3% | 84.5% | 97.2% | 98.3% | 0.3% | 0.0% |

|                   | 1927-2008 |   |   |   |   |   |   |   |   |   |   |   |                   |
| relative influence| 66.10% | 8.87% | 7.39% | 6.25% | 2.89% | 2.42% | 2.14% | 1.57% | 1.36% | 0.44% | 0.29% | 0.27% | 82.36% | 91.50% |
| p-value            | 0.0%      | 0.0% | 0.0% | 0.0% | 1.3% | 2.8% | 5.1% | 14.4% | 23.0% | 89.3% | 99.0% | 98.6% | 0.0% | 0.0% |

|                   | 1927-1967 |   |   |   |   |   |   |   |   |   |   |   |                   |
| relative influence| 41.99% | 13.92% | 13.61% | 16.12% | 6.09% | 5.94% | 3.07% | 2.79% | 2.48% | 2.38% | 0.91% | 0.71% | 69.52% | 81.73% |
| p-value            | 0.0%      | 0.0% | 0.0% | 0.4% | 0.1% | 0.8% | 26% | 28.5% | 35.5% | 37.4% | 99.7% | 99.6% | 0.0% | 0.0% |

|                   | 1968-2008 |   |   |   |   |   |   |   |   |   |   |   |                   |
| relative influence| 60.16% | 11.65% | 6.25% | 5.40% | 4.35% | 3.99% | 2.07% | 1.92% | 1.26% | 1.10% | 0.99% | 0.86% | 78.06% | 87.81% |
| p-value            | 0.0%      | 0.0% | 0.0% | 1.0% | 4.4% | 5.5% | 28.2% | 47.1% | 83.7% | 89.9% | 91.7% | 96.9% | 0.0% | 0.0% |

|                   | 1960-2008 |   |   |   |   |   |   |   |   |   |   |   |                   |
| relative influence| 13.27% | 12.79% | 11.16% | 16.47% | 8.60% | 8.80% | 7.04% | 6.90% | 6.19% | 5.72% | 5.17% | 4.07% | 42.02% | 60.43% |
| p-value            | 13.6%    | 16.5% | 28.9% | 30.3% | 69.1% | 70.9% | 90.8% | 92.5% | 94.8% | 95.7% | 99.9% | 99.9% | 4.1% | 0.6% |

Table 1. Relative influence of predictor variables

A. Excess returns

B. Realized volatility

C. Realized covariance
Table 2. Estimates of the risk-return trade-off

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model: ( \hat{\mu}_{t+1</td>
<td>t} = \alpha + \beta_1 \hat{\sigma}_{t+1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sub-Samples</strong></td>
<td><strong>Sub-Samples</strong></td>
</tr>
<tr>
<td></td>
<td>(t-stat)</td>
</tr>
<tr>
<td>I. Sub-samples</td>
<td></td>
</tr>
<tr>
<td>1927-2008</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(2.76)</td>
</tr>
<tr>
<td>1927-1967</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>(2.07)</td>
</tr>
<tr>
<td>1968-2008</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>(1.42)</td>
</tr>
<tr>
<td>II. High-volatility periods</td>
<td></td>
</tr>
<tr>
<td>1927-1939</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>(-0.61)</td>
</tr>
<tr>
<td>2001-2008</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>(0.82)</td>
</tr>
<tr>
<td>B1. VIX: linear model</td>
<td></td>
</tr>
<tr>
<td>Model: ( \hat{\mu}_{t+1</td>
<td>t} = \alpha + \beta_1 \hat{\sigma}_{VIX}^{t+1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sample</strong></td>
<td><strong>Sample</strong></td>
</tr>
<tr>
<td></td>
<td>(t-stat)</td>
</tr>
<tr>
<td>1986-2008</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>(1.75)</td>
</tr>
</tbody>
</table>

This table reports the effect and significance of the conditional risk, \( \hat{\sigma}_{t+1|t} \), lagged risk, \( \hat{\sigma}_{t|t-1} \) and lagged expected return, \( \hat{\mu}_{t|t-1} \), at explaining the conditional mean return \( \hat{\mu}_{t+1|t} \). In panels A1 and B1 a linear specification is employed and coefficient estimates and t-statistics are reported. In panels A2 and B2 a boosted regression tree model is employed and relative influence estimates are reported together with their significance obtained by way of Monte Carlo p-values. Relative influence measures sum to 100. In panels A1 and A2, the conditional volatility, \( \hat{\sigma}_{t+1|t} \), estimated using boosted regression trees is employed as a measure of conditional risk. In panels B1 and B2 the Chicago Board Options Exchange Volatility Index, also known as the VIX, is used as a measure of risk and is denoted by \( \hat{\sigma}_{VIX}^{t+1|t} \). Estimates of the conditional mean and volatility that are based on boosted regression trees use the twelve variables listed in Table 1 as predictors.
Table 3. Tests for a monotonically increasing risk-return relation

<table>
<thead>
<tr>
<th>Group size</th>
<th>Horizon (months)</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Volatility estimates</td>
<td>1</td>
<td>0.000</td>
<td>0.018</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.039</td>
</tr>
<tr>
<td>B. VIX-based estimates</td>
<td>1</td>
<td>0.027</td>
<td>0.041</td>
<td>0.091</td>
</tr>
<tr>
<td>C. Covariance estimates</td>
<td>1</td>
<td>0.420</td>
<td>0.984</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.320</td>
<td>0.898</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.000</td>
<td>0.960</td>
<td>0.899</td>
</tr>
</tbody>
</table>

This table presents the results of a test investigating 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, VIX or covariance), 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 non-monotonic relation. Small p-values indicate rejection of a monotonically increasing risk-return relation. In Panel A and B, conditional volatility is used as a measure of risk while the conditional covariance of stock returns with changes in the ADS economic activity index is used in Panel C. In Panels A and C, estimates of the conditional mean, volatility and covariance are based on boosted regression trees that use the twelve variables listed in Table 1 as predictors. 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. Correlation between changes to the ADS index and aggregate consumption growth

<table>
<thead>
<tr>
<th></th>
<th>Monthly</th>
<th>Quarterly</th>
<th>Semiannual</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumption</td>
<td>19.80%</td>
<td>34.06%</td>
<td>49.07%</td>
<td>54.67%</td>
</tr>
<tr>
<td>Durable Consumption</td>
<td>16.12%</td>
<td>20.11%</td>
<td>39.01%</td>
<td>45.69%</td>
</tr>
<tr>
<td>Non-Durable Consumption</td>
<td>15.81%</td>
<td>35.11%</td>
<td>45.17%</td>
<td>54.72%</td>
</tr>
</tbody>
</table>

This table presents correlations between changes to the ADS economic activity index and the percentage change in real consumption at the monthly, quarterly, semiannual and annual horizons. We use data from 1960 to 2008 and employ three different measures of consumption: 1) real personal consumption expenditures, 2) real personal consumption expenditures: durable consumption and 3) real personal consumption expenditures: non-durable consumption.
Table 5. Estimates of the expected return-conditional covariance trade-off

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\text{cov}^{t+1}$ (t-stat)</th>
<th>$\text{cov}^{t+1}$ (t-stat)</th>
<th>$\hat{\mu}^{t+1}$ (t-stat)</th>
<th>$R^2$</th>
<th>Sample</th>
<th>$\text{cov}^{t+1}$ (p-value)</th>
<th>$\text{cov}^{t+1}$ (p-value)</th>
<th>$\hat{\mu}^{t+1}$ (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960-2008</td>
<td>0.018</td>
<td>-0.007</td>
<td>0.580</td>
<td>45.34%</td>
<td>1960-2008</td>
<td>26.3%</td>
<td>9.6%</td>
<td>64.1%</td>
</tr>
<tr>
<td>(8.71)</td>
<td>(-3.37)</td>
<td>(17.23)</td>
<td></td>
<td></td>
<td>(0%)</td>
<td>(43.3%)</td>
<td>(0%)</td>
<td></td>
</tr>
</tbody>
</table>

This table reports the effect and significance of conditional covariance risk, $\text{cov}^{t+1}$, lagged covariance risk, $\text{cov}^{t+1}$, and lagged expected return, $\hat{\mu}^{t+1}$, at explaining the conditional mean return, $\hat{\mu}^{t+1}$. The covariance measure tracks daily comovements between changes in the ADS economic activity index and stock market returns. In panel A a linear specification is employed and coefficient estimates and t-statistics are reported. In panel B a boosted regression tree model is employed and relative influence estimates (in percent) are reported together with their significance obtained by way of Monte Carlo p-values. Estimates of the conditional mean and covariance are based on boosted regression trees and use the twelve variables listed in Table 1 as predictors.
Table 6. Estimates and tests of the combined model

A. Linear model estimates

\[ \hat{\mu}_{t+1|t} = \alpha + \beta_1 \hat{\sigma}_{t+1|t} + \beta_2 \hat{\text{cov}}_{t+1|t} + \beta_3 \hat{\mu}_{t|t-1} + \beta_4 \hat{\sigma}_{t|t-1} + \beta_5 \hat{\text{cov}}_{t|t-1} + \epsilon_{t+1} \]

| Sample       | $\hat{\sigma}_{t+1|t}$ (t-stat) | $\hat{\text{cov}}_{t+1|t}$ (t-stat) | $\hat{\mu}_{t|t-1}$ (t-stat) | $\hat{\sigma}_{t|t-1}$ (t-stat) | $\hat{\text{cov}}_{t|t-1}$ (t-stat) |
|--------------|----------------------------------|------------------------------------|--------------------------------|----------------------------------|------------------------------------|
| 1960-2008    | 0.026                            | 0.011                              | 0.660                         | -0.025                           | -0.006                             |
|              | (0.84)                           | (6.16)                             | (20.51)                       | (-0.80%)                         | (-3.12%)                           |

B. Flexible model estimates

Relative influence measures

\[ \hat{\mu}_{t+1|t} = f(\hat{\sigma}_{t+1|t}, \hat{\text{cov}}_{t+1|t}, \hat{\mu}_{t|t-1}, \hat{\sigma}_{t|t-1}, \hat{\text{cov}}_{t|t-1}) \]

| Sample       | $\hat{\sigma}_{t+1|t}$           | $\hat{\text{cov}}_{t+1|t}$ | $\hat{\mu}_{t|t-1}$ | $\hat{\sigma}_{t|t-1}$ | $\hat{\text{cov}}_{t|t-1}$ |
|--------------|----------------------------------|----------------------------|---------------------|------------------------|----------------------------|
| 1960-2008    | 6.61%                            | 13.15%                     | 67.44%              | 6.35%                  | 6.45%                      |
|              | (8.8%)                           | (0.0%)                     | (0.0%)              | (12.1%)                | (18.0%)                    |

C. Monotonicity tests

<table>
<thead>
<tr>
<th>Group size</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional Volatility</td>
<td>0.0000</td>
<td>0.0260</td>
<td>0.2567</td>
</tr>
<tr>
<td>Conditional Covariance</td>
<td>0.2355</td>
<td>0.2359</td>
<td>0.8242</td>
</tr>
</tbody>
</table>

This table reports estimates of the effect on the conditional mean return, $\hat{\mu}_{t+1|t}$, of the conditional volatility, $\hat{\sigma}_{t+1|t}$, the conditional covariance between stock returns and changes to economic activity, $\hat{\text{cov}}_{t+1|t}$, the lagged conditional mean return, $\hat{\mu}_{t|t-1}$, lagged volatility, $\hat{\sigma}_{t|t-1}$, and lagged covariance, $\hat{\text{cov}}_{t|t-1}$. 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 investigating whether the relation between conditional volatility (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 volatility (or conditional covariance). Small p-values indicate rejection of monotonicity. The number of observations in the small, medium and large portfolios are approximately 40, 50 and 65.
### Table 7. Out-of-sample forecasting performance

#### A1. Return models

<table>
<thead>
<tr>
<th>Forecasting Method</th>
<th>1970-2005 MSE $\times 10^3$</th>
<th>R^2</th>
<th>1970-2008 MSE $\times 10^3$</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boosted Regression Trees</td>
<td>1.964</td>
<td>1.53%</td>
<td>1.988</td>
<td>0.33%</td>
</tr>
<tr>
<td>Prevailing Mean</td>
<td>2.003</td>
<td>-0.40%</td>
<td>2.006</td>
<td>-0.58%</td>
</tr>
<tr>
<td>Multivariate Linear Model</td>
<td>2.026</td>
<td>-1.59%</td>
<td>2.037</td>
<td>-2.15%</td>
</tr>
</tbody>
</table>

#### A2. Realized volatility models

<table>
<thead>
<tr>
<th>Forecasting Method</th>
<th>1970-2005 MSE $\times 10^3$</th>
<th>R^2</th>
<th>1970-2008 MSE $\times 10^3$</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boosted Regression Trees</td>
<td>2.643</td>
<td>34.10%</td>
<td>3.405</td>
<td>40.29%</td>
</tr>
<tr>
<td>Garch(1,1)</td>
<td>3.536</td>
<td>11.82%</td>
<td>4.902</td>
<td>13.89%</td>
</tr>
<tr>
<td>MIDAS (Beta)</td>
<td>3.089</td>
<td>22.97%</td>
<td>3.693</td>
<td>35.12%</td>
</tr>
<tr>
<td>MIDAS (Exp)</td>
<td>3.199</td>
<td>20.23%</td>
<td>3.844</td>
<td>32.48%</td>
</tr>
<tr>
<td>AR(1)</td>
<td>3.089</td>
<td>22.98%</td>
<td>3.463</td>
<td>39.17%</td>
</tr>
</tbody>
</table>

#### B1. Return models: Robustness

<table>
<thead>
<tr>
<th>Forecasting Method</th>
<th>1970-2005 MSE $\times 10^3$</th>
<th>R^2</th>
<th>1970-2008 MSE $\times 10^3$</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boosted Regression Trees (5,000)</td>
<td>1.968</td>
<td>1.36%</td>
<td>1.985</td>
<td>0.45%</td>
</tr>
<tr>
<td>Boosted Regression Trees (15,000)</td>
<td>1.965</td>
<td>1.47%</td>
<td>1.992</td>
<td>0.13%</td>
</tr>
<tr>
<td>Best model selected recursively</td>
<td>1.974</td>
<td>1.06%</td>
<td>1.995</td>
<td>-0.03%</td>
</tr>
<tr>
<td>Combined Average</td>
<td>1.956</td>
<td>1.93%</td>
<td>1.977</td>
<td>0.88%</td>
</tr>
</tbody>
</table>

#### B2. Realized volatility models: Robustness

<table>
<thead>
<tr>
<th>Forecasting Method</th>
<th>1970-2005 MSE $\times 10^3$</th>
<th>R^2</th>
<th>1970-2008 MSE $\times 10^3$</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boosted Regression Trees (5,000)</td>
<td>2.792</td>
<td>30.39%</td>
<td>3.669</td>
<td>35.55%</td>
</tr>
<tr>
<td>Boosted Regression Trees (15,000)</td>
<td>2.609</td>
<td>34.95%</td>
<td>3.358</td>
<td>41.01%</td>
</tr>
<tr>
<td>Best model selected recursively</td>
<td>2.647</td>
<td>34.00%</td>
<td>3.408</td>
<td>40.13%</td>
</tr>
<tr>
<td>Combined Average</td>
<td>2.706</td>
<td>32.51%</td>
<td>3.604</td>
<td>36.69%</td>
</tr>
</tbody>
</table>

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 on the predictor variables listed in Table 1. 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.
Table 8. Relative influence of predictor variables in the factor model

A. Excess returns

<table>
<thead>
<tr>
<th>1960-2007</th>
<th>$\hat{f}_2$</th>
<th>infl</th>
<th>ep</th>
<th>$\hat{f}_6$</th>
<th>$\hat{f}_5$</th>
<th>dp</th>
<th>$\hat{f}_1$</th>
<th>$\hat{f}_7$</th>
<th>$\hat{f}_3$</th>
<th>Top 3</th>
<th>Top 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>relative influence</td>
<td>20.95%</td>
<td>10.57%</td>
<td>10.40%</td>
<td>10.34%</td>
<td>9.66%</td>
<td>7.95%</td>
<td>7.90%</td>
<td>7.46%</td>
<td>5.91%</td>
<td>5.48%</td>
<td>3.39%</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0%</td>
<td>3.2%</td>
<td>6.6%</td>
<td>10.8%</td>
<td>8.7%</td>
<td>19.7%</td>
<td>22.6%</td>
<td>33.7%</td>
<td>51.5%</td>
<td>56.8%</td>
<td>95.5%</td>
</tr>
</tbody>
</table>

B. Realized volatility

<table>
<thead>
<tr>
<th>1960-2007</th>
<th>vol</th>
<th>ep</th>
<th>exc</th>
<th>$\hat{f}_2$</th>
<th>$\hat{f}_1$</th>
<th>$\hat{f}_5$</th>
<th>$\hat{f}_6$</th>
<th>$\hat{f}_3$</th>
<th>$\hat{f}_8$</th>
<th>Top 3</th>
<th>Top 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>relative influence</td>
<td>62.52%</td>
<td>9.48%</td>
<td>8.14%</td>
<td>4.12%</td>
<td>4.06%</td>
<td>3.17%</td>
<td>3.15%</td>
<td>2.57%</td>
<td>1.1%</td>
<td>1.06%</td>
<td>0.6%</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>4.2%</td>
<td>3.2%</td>
<td>10.4%</td>
<td>7.2%</td>
<td>19.6%</td>
<td>80.1%</td>
<td>82.3%</td>
<td>99.0%</td>
</tr>
</tbody>
</table>

C. Realized covariance

<table>
<thead>
<tr>
<th>1960-2007</th>
<th>$\hat{f}_3$</th>
<th>de</th>
<th>infl</th>
<th>$\hat{f}_5$</th>
<th>$\hat{f}_1$</th>
<th>$\hat{f}_8$</th>
<th>$\hat{f}_2$</th>
<th>$\hat{f}_6$</th>
<th>$\hat{f}_4$</th>
<th>Top 3</th>
<th>Top 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>relative influence</td>
<td>15.72%</td>
<td>13.70%</td>
<td>12.60%</td>
<td>9.54%</td>
<td>8.87%</td>
<td>8.00%</td>
<td>7.40%</td>
<td>6.46%</td>
<td>6.45%</td>
<td>5.94%</td>
<td>5.32%</td>
</tr>
<tr>
<td>p-value</td>
<td>2.7%</td>
<td>6.8%</td>
<td>7.0%</td>
<td>36.6%</td>
<td>56.0%</td>
<td>66.8%</td>
<td>76.7%</td>
<td>89.2%</td>
<td>87.4%</td>
<td>94.5%</td>
<td>96.9%</td>
</tr>
</tbody>
</table>

This table shows the relative influence and significance (Monte Carlo p-value) of eight factors and three economic variables considered in the boosted regression tree model. The derived factors are obtained through principal components starting from a large dataset composed of 131 economic time series. Their interpretation is as follows: $\hat{f}_2$ is a real activity factor that loads heavily on employment and output data. $\hat{f}_1$ is a real activity factor that loads heavily on employment and output data. $\hat{f}_5$ loads on interest rate spreads, while $\hat{f}_3$ and $\hat{f}_4$ load on prices; $\hat{f}_6$ loads on interest rates, $\hat{f}_6$ on housing variables and $\hat{f}_7$ on measures of the money supply. Finally, $\hat{f}_8$ loads on variables related to the stock market. Inflation (infl), the log dividend-price ratio (dp) and the log earnings-price ratio (ep) are the additional economic variables for the equity premium model; lagged volatility (vol), lagged excess returns (exc) and the log earnings-price ratio (ep) are used in the volatility specification; inflation (infl), the log dividend-earnings ratio (de) and long-term rate of return (ltr) are employed in the covariance specification. Results in Panel A are for monthly returns on the S&P 500 index measured in excess of the T-bill rate. Results in Panel B are for the realized volatility measured as the square root of the sum of squared daily returns within a given month. Results in Panel C are for the realized monthly covariance between returns on the S&P 500 index and changes to the ADS index. The last two columns of each panel present the joint relative influence and joint significance of the three and five most important variables considered in each model.
Figure 1: Fitted values of excess returns ($e_{x}$) as a function of the log dividend-earnings ratio ($d_e$) and the log earnings-price ratio ($e_p$). Both plots are based on boosted regression trees with two nodes in each split. The panel on the left uses three boosting iterations, while the right panel uses 5,000 iterations. The scale for the log earnings-price ratio has been inverted. The plots are based on monthly data from 1927 to 2008.
The figures present partial dependence plots for the mean excess return equation based on the three predictor variables with the highest relative influence during 1927-2008, namely inflation (inf), the log earnings price ratio (ep), and the detrended T-Bill rate (rwl). 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 3: Effect of predictor variables on the conditional volatility. The figure present partial dependence plots for the realized volatility equation based on the three predictor variables with the highest relative influence during the sample 1927-2008, namely lagged stock market volatility (vol), the default spread (defspr) and the excess return on stocks (exc). The horizontal axis covers the sample support of each predictor variable, while the vertical axis tracks the change in the conditional volatility as a function of the individual predictor variables.
Figure 4: 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 the 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.
Figure 5: 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. The conditional covariances are then estimated by way of boosted regression trees that use the twelve variables listed in Table 1 as predictors.
Figure 6: Effect of predictor variables on the conditional covariance. The figure shows partial dependence plots for the realized covariance equation based on the three predictor variables with the highest relative influence over the sample 1990-2008, i.e., inflation (inf.), the log dividend earnings ratio (de), and the long-term rate of return (ltr.). The horizontal axis covers the sample support of each predictor variable, while the vertical axis tracks the change in the conditional covariance as a function of the individual predictor variables.
Figure 7: Expected return-conditional covariance trade-off. This figure shows partial dependence plots for the conditional mean return as a function of the conditional covariance between stock returns and consumption change (cov), past covariance (covlag) and the lagged conditional mean return (exclag). The plot is based on a boosted regression tree, using data over the sample 1960-2008. 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.
Figure 8: Risk-return trade-off for ICAPM. 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. 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.
Figure 9: 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 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.
Figure 11: Risk-return trade-off for ICAPM in the extended factor model. 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 1960-2007. The conditioning information is the principal components derived from a set of 131 state variables and the three most important variables selected from the twelve predictors reported in Table 1.