Title
Essays on time-varying investment opportunities and investors' asset allocation

Permalink
https://escholarship.org/uc/item/5bx650p5

Author
Rossi, Alberto Gianluca Paolo

Publication Date
2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Essays on Time-Varying Investment Opportunities and Investors'
Asset Allocation**

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Alberto Gianluca Paolo Rossi

Committee in charge:

> Professor Allan Timmermann, Chair
> Professor Richard Carson
> Professor Graham Elliott
> Professor Jun Liu
> Professor Rossen Valkanov

2011

The dissertation of Alberto Gianluca Paolo Rossi is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2011

DEDICATION

To Julie, Paolo, Margherita and Stefano

EPIGRAPH

"Prediction is very difficult,
especially if it's about the future"

— Niels Bohr

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

1982	Born, Milano, Italy
2004	B.A. International Economics and Management, Bocconi University, Italy
2005	M.Sc. Economics, Tilburg University, The Netherlands
2011	Ph.D. Economics, University of California, San Diego, USA

ABSTRACT OF THE DISSERTATION

**Essays on Time-Varying Investment Opportunities and Investors'
Asset Allocation**

by

Alberto Gianluca Paolo Rossi

Doctor of Philosophy in Economics

University of California, San Diego, 2011

Professor Allan Timmermann, Chair

This dissertation presents three stand-alone contributions to the fields of theoretical and empirical asset pricing. The first chapter presents a theoretical model in which the attention investors pay to the stock market varies over time. This feature is obtained by introducing information costs into a continuous-time model of asset allocation with time-varying investment opportunities. My model explains why investors do not trade uniformly through time. It also rationalizes why agents do not modify their portfolio allocations gradually with the arrival of new information, but rather alternate extended periods of inertia (no-trade) with brief moments of action where asset allocations are updated according to the current state of the economy.

The second chapter analyzes the role of information in the context of financial market predictions. I employ a novel semi-parametric method known as Boosted Regres-

sion Trees (BRT) to forecast stock returns and volatility at the monthly frequency. The framework allows me to generate forecasts on the basis of a large set of predictor variables without incurring over-fitting related problems. My results indicate that expanding the conditioning information set results in greater out-of-sample predictive accuracy compared to the standard models proposed in the literature and that the forecasts generate profitable portfolio allocations even when market frictions are considered.

The third chapter (co-authored with Allan Timmermann) analyzes the limitations of parametric models in evaluating the relation between risk and return. By taking advantage of the flexible and semi-parametric nature of Boosted Regression Trees, 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.

1 Towards a Theoretical Explanation of Time-Varying Trading

Abstract

Investors do not continuously modify their portfolio allocations with the arrival of new information. Rather, extended periods of inertia are alternated with brief moments of action where asset allocations are updated according to the current state of the economy. We explain this pattern of behavior by developing a theoretical framework that introduces information costs into a continuous-time model of asset allocation with time-varying investment opportunities. Our model explains why individuals' attention to their investment portfolio is high when the expected gains from investing in the stock market are large, such as during recessions. It also explains why investors' portfolio adjustments are lumpy in nature. The model implies a weak and time-varying co-variation between consumption growth and equity returns and so helps explain why the equity premium can be high despite the low correlation between stock returns and consumption changes. The microeconomic implications of our analysis are tested empirically against competing models of asset allocation using duration models estimated on Odean (1999)'s brokerage account data and through panel data estimates based on the PSID. Our macroeconomic predictions are supported by tests based on both high and low frequency economic and financial time-series.

1.1 Introduction

Most theoretical models of asset allocation envision a frictionless environment where consumption and investment decisions are made by agents who continuously exploit all the available information. Yet, a pervasive finding in empirical finance is that investors revise their portfolio allocations and consumption decisions rather infrequently. For example, using no-load mutual fund data, Johnson (2010) estimates that investors trade once every 20 months on average. Recent research also shows that financial agents do not trade uniformly through time, implying that investors' attention to the stock market is time-varying. This gap between theory and empirical findings is the motivation of our work.

This paper introduces information costs into a continuous-time asset pricing framework with time-varying investment opportunities and proposes a model that contributes to the financial economics literature by explaining investors' time-varying trading and inertia in their asset allocation. When the cost of inattention is low as in those states of the world characterized by low expected returns on the stock market, investors choose to trade as infrequently as once every two or three years. Conversely, when the expected gains from investing in the stock market are high, investors pay close attention to the stock market and place numerous trades within a given month. Time-varying attention to the stock market implies a co-variation between returns and consumption that is generally weak, but increases sharply during market booms and crashes. In this respect our model contributes also to the macro-finance literature that provides a theoretical explanation for the equity premium puzzle.

We formally characterize the optimal inattentive behavior of a representative agent who pays information costs on his investment portfolio. The cost is proportional to the contemporaneous value of the portfolio and is intended to summarize the financial and non-financial costs associated with processing the information needed to formulate the optimal investment and consumption plans. Once information costs are introduced, it is no longer optimal for the investor to make consumption and portfolio decisions on a continuous basis and periods of inattention to the stock market become optimal. Our analysis shows that the length of optimal inattention depends on the economic conditions faced by the agent at the time of action: if the agent expects his investment portfolio to be very profitable, he will pay close attention to the stock-market; he will instead disregard it for a long time if expected returns are low. The mechanism is generated

by a cash-in-advance constraint that forces the investor to store part of his wealth in a low-yield account (a transactions account) to finance his every-day consumption. When expected returns on the market are high, such as during recessions, the agent has the incentive to keep very little of his wealth in the transactions account and frequently transfer wealth from the investment to the transactions account. The opposite holds true if expected returns on the market portfolio are low. Recognizing explicitly that the agent's optimal behavior depends on the economic conditions he faces at the time of action is one of the key insights of our model. Conversely, by ignoring time-variations in investment opportunities, extant partial equilibrium analyses in this literature are essentially static: the agent always faces the same economic conditions, always makes the same portfolio and consumption decisions and chooses the same optimal inattention interval.

The theoretical contributions of our paper can be summarized as follows. First, consistent with the results in Abel, Eberly, and Panageas (2007, 2009) as well as Gabaix and Laibson (2002), our analysis shows that small information costs generate lengthy periods of inattention. Second, we show that the period of optimal inattention is not fixed but depends on the prevailing risk-return trade-off as summarized by the market Sharpe ratio. The value of the Sharpe ratio at the time of action determines the optimal portfolio allocation, the expected return on the investment portfolio and consequently the amount of wealth in the transactions account. Third, in our model the co-variation between returns and consumption is weak and time-varying. Because the correlation between aggregate consumption and market returns is generally low, but increases sharply during market booms and crashes, our model entails that standard CCAPM estimates are affected by a time-varying bias. Fourth, the inattention periods followed by sudden awakenings give rise to lumpy portfolio adjustments, a feature that theoretical models have difficulties in explaining. Finally, financial economists have explored exact solutions to optimal non-myopic strategic asset allocation for agents with different investment horizons, but the investment horizon has always been taken as an exogenous parameter.[1] In our model the horizon is endogenously determined by the trade-off between the opportunity cost of keeping wealth in the transactions account and the cost incurred every time the value of the portfolio is observed and funds transferred from the investment to the transactions account.

---

[1]see for example Kim and Omberg (1996), Brennan (1998), Brennan and Xia (2000, 2001, 2002), Wachter (2003), Sangvinatsos and Wachter (2005) and Liu (2007)

Allowing for a multi-agent extension of the standard framework implies a number of additional results. First, our model explains the cross-sectional dispersion of portfolio allocations between the agents populating the economy. Identical agents who differ only in their first day of trading are shown to hold different portfolio allocations over time. Furthermore, the dispersion of portfolio allocations across agents is shown to be time-varying and positively related to the absolute value of the conditional Sharpe ratio. Second, even if all agents are endowed with the same wealth on their first trading day, different investment decisions translate into different wealth processes for different agents so that over a given investment horizon, some see their wealth increase dramatically while others end up poorer. Finally, the model implies that the aggregate trading behavior is time-varying and is a function of the Sharpe ratio of the aggregate market.

Introducing time-variations in the investment opportunity set is important in light of the large literature assessing the predictability of aggregate stock returns and market volatility. Numerous studies, see Lettau and Ludvigson (2009) for a comprehensive treatment, show that excess stock market returns can be predicted using financial and economic time-series. The same holds for volatility, that has been shown to be forecastable at horizons ranging from one day to six years. While classic asset pricing models such as the CAPM imply a constant price of risk, the conditional Sharpe ratio, which is defined as the ratio of conditional excess return and volatility, has been shown empirically to be strongly countercyclical and highly volatile. Our modeling strategy implies that the testable implications we derive depend only on the conditional Sharpe ratio, as the latter is the only state variable of our model. Once estimates of this quantity are obtained, empirical tests can easily be computed. We conduct four distinct empirical exercises that estimate quantitatively the micro and macroeconomic predictions of our framework.

The first tests whether the conditional Sharpe ratio affects individuals' trading frequencies using a duration model estimated on Odean (1999)'s brokerage account data. Our findings indicate that time-varying investment opportunities are a fundamental determinant of time-varying trading and that a one-standard deviation increase in the conditional Sharpe ratio leads to a 20% increase in households' trading probabilities. Our work is the first to analyze the impact of the investment opportunity set on households' trading. While the number of empirical studies in the field of household finance has grown rapidly in the recent past, the majority are concerned with assessing behav-

ioral biases in investors' trading. Odean (1998) uses account-level data to assess whether investors are reluctant to realize losses and Odean (1999) employs a similar dataset to study whether investors' trade too much. Huberman and Jiang (2006) analyze whether the structure of 401(k) plans influence investors' positions in risky assets. A number of papers analyze the socio-economic determinants of individuals' portfolio allocation and stock market participation. Ameriks and Zeldes (2002) analyze how equity positions in investors' portfolios vary with age. Vissing-Jorgensen (2002) proposes a model explaining the non-participation puzzle, i.e. the empirical finding that a large fraction of US households do not participate in the stock market, and tests it using data from the PSID. The analyses closest to ours are the ones undertaken in Johnson (2004) and Johnson (2010). The first uses a proprietary panel dataset in one no-load mutual fund family to show that observable shareholder characteristics can be exploited to determine their investment horizon. Using the same dataset, the second shows that mutual fund shareholders do not trade uniformly through time and that while demographic traits are important trade predictors, time effects are much more important.

To estimate the extent to which the conditional Sharpe ratio affects investors' positions in risky assets our second empirical exercise uses panel data estimates from the PSID. Our findings show that changes in investment opportunities are important determinants of investors' portfolio allocations: a 10% change in the average conditional Sharpe ratio increases by approximately 4% the proportion of risky assets held by investors. Our analysis nests that of Brunnermeier and Nagel (2008) who use the same dataset to show that wealth fluctuations are not associated with changes in investors' risky asset positions, as habit-persistence models predict.

Turning to the model's implications for the macroeconomy, we test the time-varying covariance between stock returns and aggregate consumption using a number of macroeconomic and financial time-series. After obtaining a high frequency proxy for consumption using the Arouba-Diebold-Scotti Index of economic activity, we compute the realized covariance between consumption and stock returns at the monthly frequency using daily data. We then show that the conditional covariance between consumption and stock returns is time-varying and a function of the conditional Sharpe ratio.

Our fourth empirical exercise shows that aggregate volumes on the S&P 500 are a function of the investment opportunity set. Periods characterized by large conditional Sharpe ratios are also characterized by large trading volumes.

The rest of the paper is organized as follows. Section 1.2 briefly surveys the theoretical literature relevant to our work, presents the key components of the model and solves it. Section 1.3 analyzes its static and dynamic properties. In Section 1.4 we conduct the main empirical exercises. Section 1.5 concludes.

1.2 The Model

Standard continuous-time models (e.g. Merton (1969)) study the optimal investment and consumption decisions of an agent that has free access to unlimited information and does not face transactions costs. This setup entails that agents are able (and willing) to continuously re-optimize their decisions. While economists recognize the elegance of this framework, they are aware of its limitations at explaining investors' actual behavior. Agents do not continuously take advantage of all available information, they change their investment positions infrequently and do not modify their consumption instantaneously after making a profitable (or unprofitable) trade in the market. To explain the divergence between standard theory and investors' behavior, economists devised a wide array of explanations. Relevant to our paper are the ones that focus on the role played by transactions costs and agents' limited information-processing capacity.

Papers under the first heading find their roots in Baumol (1952) and Tobin (1956) and share the common feature that agents are willing to hold on to cash as opposed to riskless bonds (that provide a higher rate of return) because only cash can be exchanged for consumption. The idea of employing "cash-in-advance" constraints became particularly popular in the real business cycle literature thereafter (e.g. Cooley and Hansen (1989)), but the emphasis was generally placed on the effect of such constraints on the dynamic behavior of wages, prices and capital accumulation in aggregate economies.

The second line of research has proceeded in two directions. The first, which includes Sims (2003) and Moscarini (2004), characterizes individuals' rational inattention by way of entropy, a concept developed in the field of information theory. In these models rational inattention is generally intended as agents' decision to focus on a specific source or type of information out of the many available ones. The second direction comprises models that impose a cost of obtaining and processing information. Examples of this literature are Abel, Eberly, and Panageas (2007), Abel, Eberly, and Panageas (2009), Gabaix and Laibson (2002) and Duffie and Sun (1990). These papers require consumption to be purchased with a liquid asset, i.e. cash, and include information costs

so that the consumer does not observe continuously the value of the stock market. Our research is closely related to this literature and generalizes the results obtained so far by allowing for a time-varying investment opportunity set.

In this paper we present a partial equilibrium model for a representative agent that lives in an economy with three assets: a liquid risk-less asset, an illiquid risk-less asset and an illiquid risky asset. The liquid asset will be referred to as the *transactions account* and can be thought of as a checking account whose rate of return is constant and known. Only the wealth stored in the transactions account can be exchanged for consumption. The illiquid assets constitute what we will refer to henceforth as the *investment portfolio*. Within the investment portfolio, wealth is allocated on the basis of the expected return and volatility of the risky asset as well as the (known) return on the risk-less asset. To capture parsimoniously the time-varying behavior of expected return and volatility we let the market expected Sharpe ratio follow an Ornstein-Uhlenbeck process. This characterization is motivated by the strong evidence of a significant cyclical variation in the market Sharpe ratio uncovered, among others, by Kandel and Stambaugh (1990), Whitelaw (1997) and Perez-Quiros and Timmermann (2000).[2] It is also convenient as many aspects of the model can be solved in closed-form using standard techniques. The agent can observe the current value of his investment portfolio and the value of the conditional market Sharpe ratio by paying an information cost that is proportional to the value of the invested wealth.[3] Furthermore, the agent can transfer assets from the investment to the transactions account and can change the composition of the investment portfolio only by paying the information costs. We will refer to the instant right after information costs are paid as the "moment of action" and the interval between two moments of action as an "optimal inattention period". We next provide a formal description of the model.

Assume a consumer with power utility and infinite horizon that maximizes

$$E_t \left\{ \int_0^\infty \frac{1}{1-\alpha} c_{t+s}^{1-\alpha} e^{-\rho s} ds \right\} \tag{1.1}$$

where $0 < \alpha \neq 1$ and $\rho > 0$.

The agent has access to three assets. A liquid riskless asset that pays a rate of return r^L, a non-liquid riskless bond trading at price B with return r and a non-liquid

---

[2]See also to Lettau and Ludvigson (2009) for a survey.

[3]This assumption is imposed for analytical convenience: at the expense of losing the closed-form solutions we could model information costs in terms of utility. On the other hand, in the presence of a market for information investors could purchase the knowledge required to update their consumption and portfolio decisions from an external source and pay a monetary amount for it.

risky asset P with time-varying expected return μ_t and volatility σ_t. The only asset that can be exchanged for consumption is the liquid riskless asset that the agent keeps in the transactions account. We assume that $0 \leq r^L < r$ to reflect the liquidity premium associated with the transactions account and to rule out arbitrage opportunities. The return processes for the riskless bond and risky asset are:

$$\frac{dB}{B} = r \ dt \tag{1.2}$$

$$\frac{dP}{P} = \mu_t \ dt + \sigma_t \ dZ \tag{1.3}$$

where Z_t is a standard Brownian motion. The drift rate μ_t and volatility σ_t of the risky asset are both diffusion processes. Define

$$M_t = \frac{\mu_t - r}{\sigma_t} \tag{1.4}$$

to be the conditional Sharpe ratio of the risky asset, which is assumed to follow an Ornstein-Uhlenbeck process

$$dM = -\lambda_M(M - \bar{M})dt + \sigma_M dZ_M, \tag{1.5}$$

where λ_M, σ_M, and $\bar{M}$ are positive constants and $Z_{M,t}$ is a second standard Brownian motion. The correlation between the asset's return and Sharpe ratio processes is given by $E\{dZdZ_M\} = \rho_{sM}dt$. The opportunity set has three stochastic variables μ_t, σ_t and M_t linked by the definition of the Sharpe ratio: $\mu_t = r + \sigma_t M_t$, but we know from Kim and Omberg (1996) that in this specific setting, the Sharpe ratio provides all the currently available information on present and future investment opportunities.

Wealth, W_t, is the sum of invested wealth $W_{I,t}$ and the funds in the transactions account, denoted by X_t: $W_t = W_{I,t} + X_t$. The agent can observe the value of the Sharpe ratio and the investment portfolio only by paying a fraction θ, $0 \leq \theta \leq 1$, of his invested wealth. Furthermore, funds can be transferred from the investment account to the transactions account (and vice-versa) only when the value of the invested portfolio is observed. However, because $r^L < r$ the agent arrives at the end of the optimal inattention period with no funds in the transactions accounts. This means that transfers will only occur from the investment account to the transactions account.[4] It follows that for an investor that observes the value of his invested portfolio at the time of action "t_j",

---

[4]Abel, Eberly, and Panageas (2009) show that this result holds under more general specifications for information and transaction costs.

the t_j^+ invested wealth is simply the current wealth, minus the amount deposited in the transactions account multiplied by $(1 - \theta)$: i.e. $W_{I,t_j^+} = (W_{t_j} - X_{t_j})(1 - \theta)$.

At the time of action t_j, the consumer chooses the timespan of optimal inattention, τ, the amount of wealth to be deposited in the transactions account X and the fraction ϕ of the investment portfolio to be held in the risky asset. The agent's optimal behavior is the following. At the moment of action the investor observes the current Sharpe ratio. Because its evolution and its relation to the risky asset are known, the agent's expectations regarding the return of the risky asset at any horizon incorporate all the currently available information. For any horizon the agent can calculate the optimal investment portfolio, the expected end-of-period wealth[5] and therefore the expected utility from the investment portfolio. He is also able to calculate the utility deriving from the wealth stored in the transactions account. The trade-off he faces is the following: on the one hand, he would like to keep very little wealth in the transactions account, because the investment portfolio offers a higher rate of return; on the other hand, he would like to transfer funds from the investment portfolio to the transactions account as infrequently as possible to avoid paying the costs of information. The economic conditions faced at the time of action affect fundamentally this decision.

1.2.1 Model Solution

We solve the model by first characterizing the dynamics of consumption. We then derive the optimal portfolio allocation and the optimal proportion of wealth to be deposited in the transactions account. Finally, we obtain the optimal inattention period, τ^*. Proofs of propositions and theorems are provided in the Appendix.

Consumption dynamics

We derive a rule for optimal consumption that holds for any inattention period τ and any amount X_{t_j} deposited in the transactions account at time t_j. Given X_{t_j} and τ, the agent chooses consumption to maximize his utility:

$$U_{t_j}(\tau) \equiv \max_{\{c_{t+s}\}_{s=0}^{\tau}} \int_0^{\tau} \frac{1}{1-\alpha} c_{t_j+s}^{(1-\alpha)} e^{-\rho s} \, ds \qquad (1.6)$$

---

[5]In Abel et al (2007) and in our paper, the investment portfolio is managed by a portfolio manager that continuously rebalances the portfolio, i.e. a mutual fund, following the rule dictated by the agent. This allows for greater tractability of the problem and excludes the re-balancing motive as a reason for the agent to observe the value of his portfolio.

subject to the constraint $X_{t_j}(\tau) = \int_0^\tau c_{t_j+s} e^{-r^L s} ds$.

***Proposition 1* (Consumption Rule and Maximized Utility from Consumption)**: *Given the wealth deposited in the transactions account, X_{t_j}, and the inattention period τ, the maximized utility of consumption is:*

$$U_{t_j}(\tau) = \frac{1}{1-\alpha} X_{t_j}^{(1-\alpha)} h(\tau)^\alpha. \tag{1.7}$$

Over the interval of inattention, consumption evolves according to the following rule:

$$c_{t_j+s} = c_{t_j}\ e^{-\frac{(\rho-r^L)}{\alpha}s}, \quad for\ \ 0 \le s \le \tau. \tag{1.8}$$

Eq. 1.7 describes how the maximized utility of consumption from period t_j to $t_j + \tau$ depends on the utility derived by the sum deposited $[1/(1-\alpha)]X_{t_j}^{(1-\alpha)}$ and a scaling factor $h(\tau)^\alpha$ that depends on the degree of risk aversion α, the subjective discount factor ρ and the rate of return r^L paid by the liquid risk-free asset. Eq. 1.8 characterizes how consumption over the inattention period increases or decreases depending on the relative values of ρ and r^L. If the agent discounts future consumption more than the rate of interest paid by the transactions account, consumption decreases over time and vice-versa, i.e. $c_{t_j+s} \gtreqless c_{t_j}$ if $r^L \gtreqless \rho$. Under power utility the risk aversion coefficient α is the inverse of the elasticity of inter-temporal substitution and so determines the rate at which consumption increases or decreases over time.

To build intuition, Figure 1.1 plots the dynamics of consumption (Top Panel) as well as the dynamics of the wealth stored in the transactions account (Bottom Panel) between periods of inattention. Consumption changes in a lumpy fashion at the time of action and varies very little during the periods of inattention. In fact, if $r^L = \rho$ consumption is constant between the times of action. Furthermore, the behavior of consumption during the inattention periods is independent from the dynamics of invested wealth. The wealth stored in the transactions account is exhausted completely when the period of optimal inattention ends. This is optimal because the transactions account pays a lower rate of returns compared to the expected returns paid by the investment portfolio.

Portfolio Choice

So far we have derived a rule for consumption that holds for any wealth X_{t_j} stored in the transactions account and any inattention interval. We next show how the

agent calculates the optimal allocation within the investment portfolio. At t_j, the agent observes the value of his invested wealth W_{t_j} (which coincides with his total wealth) and the value of the current Sharpe Ratio M_{t_j}. Conditional on these, the value function $V(W_{t_j}, M_{t_j}, \tau)$ satisfies:

$$V(W_{t_j}, M_{t_j}, \tau) = \max_{X_{t_j}, \phi} U_{t_j}(\tau) + e^{-\rho\tau} E_{t_j} \left\{ V((W_{t_j} - X_{t_j})(1-\theta)R(t_j, t_j + \tau)) \mid M = M_{t_j} \right\}. \quad (1.9)$$

Conjecture the solution:

$$V(W_{t_j}, M_{t_j}, \tau) = \gamma(M_{t_j}, \tau) \frac{W_{t_j}^{1-\alpha}}{1-\alpha} \quad (1.10)$$

and substitute (1.7) and (1.10) into (1.9) to obtain

$$\frac{1}{1-\alpha}\gamma W_{t_j}^{1-\alpha} = \max_{X_{t_j}, \phi} \frac{1}{1-\alpha}X_{t_j}^{1-\alpha}[h(\tau)]^\alpha \quad (1.11)$$

$$+ e^{-\rho\tau} \frac{1}{1-\alpha}\gamma(W_{t_j} - X_{t_j})^{1-\alpha}(1-\theta)^{1-\alpha} E_{t_j} \left\{ [R(t_j, t_j + \tau)]^{1-\alpha} \mid M = M_{t_j} \right\}.$$

The proposition that follows characterizes the optimal portfolio allocation for the investment portfolio and the associated expected returns.

Proposition 2 (**Optimal Portfolio Allocation and Maximized Expected Returns**): *Given the Sharpe ratio value M_{t_j} and investment horizon τ, the optimal fraction of the investment portfolio in the risky asset is:*

$$\phi^* = \frac{M_{t_j}}{\alpha\sigma_{t_j}} + \frac{(C(\tau)M_{t_j} + B(\tau))\rho_{sM}\sigma_M}{\alpha\sigma_{t_j}}, \quad (1.12)$$

The associated expected returns on the investment portfolio from t_j to $t_j + \tau$ can be written as:

$$E_{t_j} \left\{ [R(t_j, t_j + \tau)]^{1-\alpha} \mid M = M_{t_j} \right\} = \exp\left[r(1-\alpha)\tau + A(\tau) + B(\tau)M + C(\tau)\frac{M^2}{2} \mid M = M_{t_j} \right]. \quad (1.13)$$

It is not straightforward to compute closed form comparative statics of Eqs. 1.12 and 1.13 because the terms $A(\tau)$, $B(\tau)$ and $C(\tau)$ are rather intractable. We can however build intuition regarding the optimal portfolio allocation and expected returns by focusing on the "normalized return process", which is the return obtained by buying $1/\sigma_t$ of the risky asset financed by going short in the risk-free asset:

$$dR_t = M_t \, dt + dZ. \quad (1.14)$$

The normalized return process has four parameters: $\bar{M}$, λ_M, σ_M and ρ_{sM}. The first three measure the Sharpe ratio's long-run average, the strength of its mean-reversion

and its volatility, respectively. The fourth is the cross-sectional correlation between the risky asset's return and the Sharpe ratio. As shown by Kim and Omberg (1996), to analyze the optimal portfolio allocation, it suffices to understand the role of $\bar{M}$, ρ_{sM} and that of the derived parameter $k^* = (\sigma_M/\lambda_M)^2 + 2\rho_{sM}(\sigma_M/\lambda_M)$. $\bar{M}$ controls the long-run value for the Sharpe ratio. To understand the sign and magnitude of ρ_{sM}, imagine that the current price for the risky asset is P_0 and that the agent expects future prices to follow a certain trajectory over the interval $[0, T]$. Now imagine that negative news regarding the risky asset lowers the price at time 0 to P_0' and shifts downward the entire price trajectory over the interval $[0, T]$. If the fall in the current price is large compared to the fall in expected future prices, the trajectory will rotate counter-clockwise, causing the current Sharpe ratio to rise. This would produce a negative correlation between the asset's return and the Sharpe ratio, resulting in a bias towards higher expected returns after a series of price decreases and vice-versa. If small changes in the current price are associated with changes in expected future prices in the same direction, the correlation between the risky asset's return and the Sharpe ratio will be positive, leading to a bias towards higher expected returns after price rises and vice-versa. To sum up: the magnitude of the correlation ρ_{sM} depends on how much expected future prices are influenced by information uncorrelated with the current price, while the sign of ρ_{sM} determines the sign of the intertemporal correlation in normalized asset returns over discrete time-periods.[6] The derived parameter k^* determines the long-run variance of normalized return process (Eq. 1.14) as summarized in the proposition that follows.

Proposition 3 (**Variance of the Normalized Return Process**) (**Kim and Omberg (1996)**): *The variance of the normalized return process (Eq. 1.14) can be written as:*

$$Var\{R(\tau)\} = \tau + \left(\frac{\sigma_M}{\lambda_M}\right)^2 \left(\tau + \frac{2e^{-\lambda_M\tau}}{\lambda_M} - \frac{e^{-2\lambda_M\tau}}{2\lambda_M} - \frac{3}{2\lambda_M}\right) \qquad (1.15)$$
$$+ \left(\frac{2\rho_{sM}\sigma_M}{\lambda_M}\right)\left(\tau + \frac{e^{-\lambda_M\tau}}{\lambda_M} - \frac{1}{\lambda_M}\right),$$

from which it can be shown that $\lim_{\tau\to 0}\frac{\partial Var\{R(\tau)\}}{\partial \tau} = 1$ and $\lim_{\tau\to\infty} = 1 + k^*$.

The variance of the normalized return process determines how appealing the risky asset is to the representative investor, therefore affecting the portfolio allocation

---

[6]This is an important subtlety: while the normalized returns process $dR_t = M_t dt + dZ$ has unbounded variation and no serial correlation in Z_t, asset returns are inter-temporally correlated over discrete intervals due to the cumulative impact of past return dZ on future risk-premia M, via the cross-sectional correlation between dZ and dM combined with the dynamics of M_t.

and the optimal inattention interval. For a positive ρ_{sM}, the long-run variance of the normalized process is positively related to σ_M and negatively related to λ_M. A negative correlation ρ_{sM} is a necessary but not a sufficient condition for $k^* < 0$: in fact the necessary and sufficient condition is $\rho_{sM} < 0$ and $\frac{\sigma_M}{\lambda_M} \in (0, -2\rho_{sM})$. The minimum value for the normalized return process is $k^*_{min} = -\rho^*_{sM} \geq -1$ and is attained for the values $\sigma_M/\lambda_M = -\rho_{sM}$. The intuition is that with $\rho_{sM} < 0$ and $\sigma_M/\lambda_M > 0$, low returns are associated with high Sharpe ratios. It follows that the variance of the return process decreases as the horizon increases as long as the variations in the Sharpe ratio are not too large.

Wealth in the Transactions Account and Optimal Inattention

Substituting Eq. 1.13 into 1.11 and differentiating the resulting expression w.r.t X_{t_j} we obtain the optimal wealth deposited in the transactions account:

Theorem 1 (**Optimal Wealth Deposited in the Transactions Account**)*: Given the Sharpe ratio value M_{t_j}, wealth W_{t_j} and investment horizon τ, the optimal amount of wealth deposited in the transactions account is:*

$$X^*_{t_j} = \frac{K(M_{t_j}, \tau)}{K(M_{t_j}, \tau) + 1} \, W_{t_j}, \tag{1.16}$$

where $K(M_{t_j}, \tau) = \chi^{-1} e^{S(M_{t_j}, \tau)} - 1$; $\chi \equiv (1-\theta)^{\frac{1-\alpha}{\alpha}}$ and

$$S(M_{t_j}, \tau) = \left[\frac{1}{\alpha} \left((\rho - r(1-\alpha))\tau - A(\tau) - B(\tau)M - C(\tau)\frac{M^2}{2} \right) \; | \; M = M_{t_j} \right].$$

Furthermore, the analytical expression for the $\gamma(M_{t_j}, \tau)$ conjectured in Eq. 1.10 is:

$$\gamma(M_{t_j}, \tau) = \left(\frac{1 - e^{-\omega\tau}}{1 - \chi e^{-S(M_{t_j}, \tau)}} \right)^\alpha \omega^{-\alpha}. \tag{1.17}$$

$X^*_{t_j}$ represents the optimal amount of funds to be deposited in the transactions account for any given optimal inattention period τ and current value of the Sharpe Ratio M_{t_j}. It depends on current wealth W_{t_j}, the expected returns on the market $S(M_{t_j}, \tau)$, the risk-aversion coefficient α and the information cost θ. Having obtained a closed-form expression for $\gamma(M_{t_j}, \tau)$ allows us to characterize the solution because choosing τ to maximize the value function (Eq. 1.10) is equivalent to solving the problem:

$$F(M_{t_j}) \equiv \max_\tau \; \frac{\gamma(M_{t_j}, \tau)}{1 - \alpha} \tag{1.18}$$

$$s.t. \quad \chi e^{-S(M_{t_j}, \tau)} < 1 \tag{1.19}$$

where the constraint guarantees that consumption is positive at all times.

We solve the problem numerically because a closed form expression would be very hard, if not impossible, to obtain. It also would not be very informative. For θ equal to 0.1 basis points[7] the optimal period of inattention is 1.80 years, while it is 1.93 years for θ equal to 1 basis point. Both values are close to the 20 months of inattention estimated empirically by Johnson (2010).

1.3 Analysis of the Model

In this section we present an analysis of the static and dynamic properties of the model. The results are computed for the following parameterization: $\sigma_M = 1.06$, $\lambda_M = 0.99$, $\rho_{s_M} = 0.1$, $\alpha = 5$; $\bar{M} = 0.5$; $M = 0.5$, $\rho = 0.01$, $r^L = 0.01$, $r = 0.02$, $\theta = 0.0001$. The long-run mean of the Sharpe ratio $\bar{M}$ and the volatility of its process σ_M are calibrated using S&P 500 realized Sharpe ratio estimates from 1960 to 2008. We assume that the process is strongly mean reverting and it has a small and positive correlation with the returns process. We also assume that that we are in the long-run equilibrium by imposing $M = \bar{M}$. The parameter values for the return on the liquid risk-free asset, the illiquid risk-free asset, the risk-aversion coefficient, the subjective discount factor and the cost of information are chosen to match those of Abel, Eberly, and Panageas (2007).

Under this baseline parameterization we find that that the optimal inattention span is approximately 1.80 years and the fraction of invested wealth in the risky asset, given the optimal horizon, is 0.47. These numbers provide the benchmark for the analysis we report next: we allow each parameter to vary while keeping the others fixed and track the optimal inattention span as well as the optimal portfolio allocation. We first analyze the effect of the conditional Sharpe ratio on the optimal inattention and optimal portfolio allocation. We then discuss the effect of information costs and agents' characteristics.

1.3.1 Current Value of the Sharpe ratio

The advantage of incorporating a state variable in our framework is that we are able to characterize the agent's optimal behavior conditional on the economic environment he faces at the time of action. This is one of the key contributions of our paper

---

[7]For the other parameters we employ the standard values reported at the beginning of Section 1.3.

because it allows for an analysis of investors' optimal behavior over the course of the business cycle.

Figure 1.2 plots the optimal inattention span (continuous line) and the optimal risky asset allocation (dashed line) for different values of the annualized conditional Sharpe ratio M. The y-axis on the left reports the years of optimal inattention, while the right hand side reports the fraction of wealth in the investment portfolio allocated to the risky asset. The values for the conditional Sharpe ratios reported on the x-axis are chosen to match the empirical estimates reported in Lettau and Ludvigson (2009).

In periods characterized by high expected returns on the stock market, represented by a high value of M_t, the agent's optimal strategy is to allocate more than 100% of his invested wealth to the risky asset, deposit very little in the transactions account and observe the value of the portfolio very frequently. The same holds when expected returns on the stock market are large and negative: the agent shorts the risky asset and reaps the benefits of the negative expected returns. Interestingly, in sluggish economic environments characterized by a Sharpe ratio around zero, it is optimal for the agent to invest very little in the risky asset and opt for very long periods of inattention. The phenomenon deserves a closer inspection.

For a myopic investor, it is the magnitude of the Sharpe ratio rather than its sign that determines the attractiveness of the risky asset. But the non-myopic investor also considers the asymmetry in the dynamics of the Sharpe ratio process: i.e. the reversion to a positive long run Sharpe ratio $\bar{M}$. Very large Sharpe ratios offer two advantages: high expected returns in the short run, combined with an ultimate return to the long run Sharpe ratio through a region of high expected returns. Very negative Sharpe ratios offer the same type of short-run profitability (recall that the agent is allowed to short the risky asset), combined with the disadvantage of the ultimate return to a long-run Sharpe ratio through a region of low expected returns. The two effects counter each other but it can be shown that for a sufficiently low Sharpe ratio, there is a net advantage for even lower values.

The mechanism described above explains the "sail" shape of the optimal inattention depicted in Figure 1.2: for very negative values of the current Sharpe ratio, the agent shorts the risky asset and reaps the large benefits coming from the negative returns.[8] The transactions account has a large opportunity cost, so it is optimal for the

---

[8]Negative conditional Sharpe ratios can arise in general equilibrium asset pricing models with time-varying covariance between consumption growth and the stochastic discount factor. See, for example,

agent to deposit very little wealth in it and pay attention to the stock market very often. For the same reason the optimal inattention period is very small for large positive values of the Sharpe ratio. For Sharpe ratios small in absolute value, the low expected returns from the investment portfolio translates into a low opportunity cost for the transactions account and large periods of inattention.

1.3.2 Comparative Statics with Respect to the Remaining Parameters

We summarize next the results for the numerical comparative statics analysis for the remaining parameters.

(Comparative Statics Results): *Under the baseline parameterization of Section 1.3 :*

a) *An increase in the persistence λ_M of the Sharpe ratio process increases the optimal fraction of invested wealth in the risky asset and decreases the optimal inattention period (Figure 1.3 (a)).*

b) *An increase in the volatility σ_M of the Sharpe ratio process decreases the optimal fraction of invested wealth in the risky asset and increases the optimal inattention period (Figure 1.3 (b)).*

c) *An increase in the correlation ρ_{sM} between the asset return and the Sharpe ratio processes decreases the optimal fraction of invested wealth in the risky asset and decreases the optimal inattention period (Figure 1.3 (c)).*

d) *An increase in information costs θ decreases the optimal fraction of invested wealth in the risky asset and increases the optimal inattention period (Figure 1.3 (d))*

e) *An increase in the risk-aversion coefficient "α" or the subjective discount factor "$1 - \rho$" decreases the optimal fraction of invested wealth in the risky asset and increases the optimal inattention period (Figure 1.3 (e) and (f))*

In Figure 1.3 (a) we allow the persistence "λ_M" of the state variable to vary from 0.01 to 1 while keeping all the other parameters constant. The continuous line represents the optimal inattention period and its values are reported in years on the left hand side. The right hand side reports the optimal portfolio allocation, intended as the fraction of

---

Boudoukh, Richardson, and Whitelaw (1997) and Whitelaw (2000).

wealth in the investment portfolio allocated to the risky asset. From the definition of k^* we know that for positive values of ρ_{sM}, the more persistent the state variable the lower the long-run variance of the risky asset: i.e. $\partial k^*/\partial \lambda_M = -2\left(\frac{\sigma_M}{\lambda_M^2}\right)\left(\frac{\sigma_M}{\lambda_M} + \rho_{sM}\right) < 0$. This results in a more profitable investment portfolio and increases the opportunity cost of keeping funds in the transactions account. The agent opts for a lower inattention period, places less funds in the transactions account and pays more often the information cost θ.

In Figure 1.3 (b) we report the results for different volatility levels "σ_M" of the state variable. Contrary to the persistence case, for positive values of ρ_{sM} an increase in the volatility of the state variable increases the variance of the risky-asset, making it less desirable: i.e. $\partial k^*/\partial \sigma_M = \frac{2}{\lambda_M}\left(\frac{\sigma_M}{\lambda_M} + \rho_{sM}\right) > 0$. This results in a lower opportunity cost for the transactions account. The optimal inattention increases and the wealth allocated to the risky asset decreases.

Turning to the correlation ρ_{sM}, we know from the definition of k^* that given our base choice of σ_M and λ_M, the long-run risk of the stock is inversely related to ρ_{sM}. The intuition is the one explained in Section 1.2.1 and repeated here for convenience: the tendency for lower returns to be followed by higher Sharpe ratios lowers the overall risk of the stock as the horizon increases. The strong positive relation between the expected return on the investment portfolio and the investment horizon motivates the optimal choice of the agent: as ρ_{sM} varies from -1 to 1, both the optimal inattention and investment in the risky asset decrease. This relation is illustrated in Figure 1.3 (c).

In Figure 1.3 (d) we evaluate the effect of information costs on the optimal inattention interval. We allow the information costs to vary from 1 to 20 basis points. The relationship between optimal inattention and information costs is positive. Greater information costs translate into a lower opportunity cost for the transactions account leading to longer optimal inattention periods. The optimal portfolio allocation is not affected directly by the information cost, it is only affected indirectly through the horizon effect: because the risk of the stock increases with the investment horizon,[9] the weight in the risky asset decreases with it.

The agent's preferences for risk and inter-temporal consumption are controlled by two parameters: the degree of risk-aversion α and the subjective discount factor $1-\rho$. Intuition dictates that α has an impact on the fraction of invested wealth allocated to

[9]To see this, recall that $\rho_{sM} = 0.1$ in our base case and use the definition of k^*.

the risky asset as higher risk-aversion leads to more conservative investment portfolios. This is confirmed by Figure 1.3 (e). The effect of α on the length of optimal inattention is unclear a priori as the parameter represents both the degree of risk-aversion and the inverse of the elasticity of inter-temporal substitution. The relation is essentially flat and mildly upward sloping. The relationship between the discount factor and the optimal inattention is clear-cut as depicted in Figure 1.3 (f). As $1 - \rho$ increases, the agent discounts future consumption less, implying that he consumes less in the short-run. The diminished short-run consumption implies that a given amount stored in the transactions account will finance consumption for a longer time-span, allowing for longer optimal inattention periods. The optimal portfolio allocation is only mildly negatively affected through the horizon effect.

The comparative statics analysis reported in this section allows us to get a first feel for the implications of our theoretical framework. The real appeal of our model, though, is that it allows us to explore how the time-varying investment opportunity set affects the dynamic behavior of single and multiple-agents economies. We explore these implications next.

1.3.3 Dynamic Behavior of the Representative Agent

It is widely known that agents do not update their portfolio holdings continuously, but it is difficult to account for time-varying lumpiness in portfolio adjustments within the context of standard continuous-time models. Our model is an exception in this respect.

In this section we simulate the process for the Sharpe ratio using the parameterization specified in Section 1.3 and describe the portfolio allocation and trading activity for a representative investor. The top panel of Figure 1.4 reports the evolution of the agent's normalized consumption[10] and portfolio allocation. The consumption process is represented by the black continuous line and the portfolio allocation by the red dashed line. The y-axis on the left reports consumption values, while the one on the right reports the weight of the risky asset in the investment portfolio. The middle panel presents the process for the agent's invested wealth and the bottom panel plots the simulated process for the Sharpe ratio. The vertical dotted lines in each panel represent the periods at which the agent updates his consumption and portfolio decisions. An easy way to

---

[10]We normalize to 1 the average consumption of the representative agent over the sample.

capture the essence of Figure 1.4 is to think of it as the dynamic counterpart of Figure 1.2.

The agent enters the three year interval presented in the figure with an invested wealth of $85,000. Approximately around year 1.3, the agent ends his period of inattention. He pays the information costs and observes the value of his investment portfolio as well as the market Sharpe ratio. Given that expected returns are relatively high, the agent increases his proportion of invested wealth in the risky asset as shown in the top panel and increases his per-period consumption. The opportunity cost of the transactions account is high, so it is optimal for the agent to store relatively little wealth in it and choose a relatively short period of inattention. Over the following half year, the agent breaks his inattention seven more times and updates his portfolio allocation and consumption according to his wealth and the current investment opportunity set. Around year 1.7 the investment opportunity set worsens and it is optimal for the agent to reduce his investments in risky assets, decrease his consumption and choose a 5 months period of inattention. When the agent "wakes-up" the investment opportunities have worsen even more, so the agent chooses an even longer period of inattention of approximately one year. He then observes a rather high conditional Sharpe ratio and so trades repeatedly (approximately nine times) over the following six months. At the end of this intense trading period, the investment opportunities worsen and the investor reduces his exposure to the stock market by placing a small fraction of his invested wealth in the risky asset, storing a large portion of his wealth in the transactions account and choosing a very long period of inattention.

Changes in consumption co-vary strongly with changes in investment opportunities and wealth only during the periods characterized by intense trading, i.e. years 1.3-1.7 and years 3.3-3.7 in Figure 1.4, because only in those periods the agent updates his consumption decisions to reflect the current economic environment he faces. Consumption is instead uncorrelated with the investment opportunity set over years 2.2-3.3, because the agent opts for a long period of inattention. These findings show that models of inattention that do not account for time-varying investment opportunities can be severely mis-specified in at least two respects. First, by featuring constant periods of inattention, they imply a constant bias in CCAPM estimates while, in fact, the bias is time-varying. Second, they imply that consumption is only a function of current wealth and not of the current investment opportunities and both should matter as shown by

our extended model.

A unique feature of our model is that the investor acts on periods characterized by good investment opportunities only if he happens to end his inattention during one of them. For example, by observing the bottom panel of Figure 1.4 it is clear that the investment opportunity set over the first half of year 2 is not very different from the one that characterize the first half of year 3 and yet the investor acts very differently. He leaves untouched his investments and consumption paths in the first and he trades repeatedly in the second. The behavior of the investor at a given point in time depends on the current Sharpe ratio only if he observes it by placing a trade and paying the associated information costs.

The peculiar features described in this section sets our model apart from the ones that have been proposed so far in the literature. Standard continuous time models entail that investors' portfolio allocations and consumption paths change continuously through time. Models that introduce transactions costs in the standard framework imply that the agent behaves in a similar fashion if similar investment opportunities recur over time. Finally, models that incorporate inattention, but do not allow for time-varying investment opportunities imply constant portfolio allocations over time as well as a constant covariance between consumption and market returns. In the empirical section we test the merits of our theoretical framework using a duration model on brokerage account data provided by Odean (1999). Our tests show that individuals' behavior is consistent with the trading pattern entailed by our theoretical model, but inconsistent with the one entailed by the others.

Next we present the results for a multi-agent extension of our baseline model, with the intent of describing the dynamic behavior of an aggregate economy populated by a large number of inattentive consumers.

1.3.4 Dynamics in an Economy with Multiple Agents

In this section we extend the results obtained above by increasing the number of agents in the economy while maintaining a partial equilibrium perspective. The approach we follow is in the same spirit of Lynch (1996) and Gabaix and Laibson (2002) and it is aimed at showing that rather different portfolio allocations and inattention intervals can coexist across different agents, if only the initial time at which they join the economy is allowed to differ.

We introduce 1000 agents in the economy, we divide the first year in 1000 equally spaced intervals and we use these to determine when each agent is allowed to observe the Sharpe ratio for the first time. The top panel of Figure 1.5 reports the percentage of the population trading at a given point in time. The bottom panel depicts the Sharpe ratio process.

A low value for the Sharpe ratio implies a long period of inattention for all those agents that update their portfolio decision at that time. The opposite holds true if the value of the Sharpe ratio is high. It follows that high Sharpe ratio periods are characterized by a rather intense trading activity, while little trading takes place in low Sharpe ratio periods. This is a reflection of the simple fact that while information on the current investment opportunities is always available in this economy, it is not always optimal for the agents to exploit it. Agents take advantage of it only in those periods when the gains from actively investing in the market are high. In our example intense trading takes place during the Bull market periods, i.e. approximately years 1.3-1.8, 2.2-2.4 and 3.3-3.6, while little trading takes place in low Sharpe-ratio periods, i.e. years 1.8-2.0, 2.7-3.2 and 3.7-4.0.

The top panel of Figure 2.4.5 plots the average portfolio allocation in the economy as well as the 20th and 80th percentiles of the portfolio allocations for a given period. The bottom panel plots the value of the Sharpe ratio over time.

The average exposure to risky assets in this economy is not constant, but time-varying. Between bear and bull market periods the average portfolio allocation in risky assets varies from approximately 45% to 60%, indicating that accounting for changes in investment opportunities have the potential of explaining time-variations in investors' risky assets positions. Furthermore, because of inattention the change in the average portfolio allocation over time is not a straightforward function of the current Sharpe ratio as implied by a standard models, but a complex function of the present and past Sharpe ratios and agents' inattention levels.

Also due to inattention, not all investors hold the same portfolio allocations at a given point in time and the dispersion in portfolio allocations is time-varying. High Sharpe ratio periods are characterized by a proportion of the population that takes advantage of the information of high expected returns, i.e. those whose optimal inattention end sometime during the high Sharpe ratio period, and a proportion of the population that doesn't. Because only a fraction of the population reacts to the new information

available, the portfolio allocation dispersion across agents is rather large. Furthermore, given that the agents that take advantage of the high expected returns have short optimal inattention periods while the value of the Sharpe ratio stays high, but go back to long periods of inattention when the Sharpe ratio is low, the cross-sectional dispersion of portfolio allocation shrinks over low Sharpe ratio periods.

Overall, the results reported in this section show that information costs and time-varying investment opportunities result in an economy where different agents face and exploit different information sets. In this respect our model is unique, because it explains rather simply the cross-sectional dispersion in investment positions across agents that have otherwise identical risk-profiles and investment attitudes. The rest of the paper is dedicated to testing empirically the theoretical implications of the model.

1.4   Empirical Framework

Our theoretical model provides testable implications along a number of dimensions. The first is that investors do not trade uniformly through time and that individuals' trading and the investment opportunity set are closely related: periods characterized by high (absolute) Sharpe ratios should be characterized by intense trading. Unlike standard continuous-time models of asset allocations, our model does not imply that agents trade continuously through time. It also does not imply that agents trade uniformly through time as implied by models of optimal inattention that do not allow for time-varying investment opportunities. Finally, unlike frameworks that account only for transactions costs, it implies that current investment opportunities should affect investors' behavior only at times when they pay attention to the stock market. We test empirically these predictions by estimating a duration model on the very detailed brokerage account data provided by Odean (1999).[11]

The second testable implication of our model is that agents' portfolio allocations vary over time as a function of the conditional Sharpe ratio. Unlike standard asset pricing models that entail agents reacting instantaneously to changes in investment opportunities, our model implies that portfolio allocations should be characterized by a considerable amount of inertia due to agents' inattention. We test this implication using panel data estimates from the PSID.

The model carries macroeconomic testable implications as well. We have shown

---

[11]We are grateful to Terrance Odean for providing this data.

in Section 1.3 that economy-wide trading activity should be a function of the Sharpe ratio. By the same token, the covariance between consumption and returns is a function of the length of inattention among the agents populating the economy. Shorter periods of inattention imply a greater covariance between aggregate consumption and stock returns and the opposite holds true for longer periods of inattention. We test these economy-wide implications of our model using high and low frequency macroeconomic and financial time-series.

Our theoretical framework has a number of testable implications that rest on the econometrician's knowledge of the conditional market Sharpe ratio. This quantity is unobservable, however. To circumvent this issue, we adopt the remedy commonly employed in the literature of obtaining conditional returns and volatility series using observable economic and financial time-series and construct conditional Sharpe ratio series in a second step.

A second obstacle that makes the empirical estimates hard to implement is that in our theoretical model the agent knows the process for the Sharpe ratio, so it is enough for him to know its current value to form expectations regarding the return on the risky asset in the subsequent periods. It is hard to replicate empirically this mechanism, as the assumptions we imposed on the process for the Sharpe ratio may be too simplistic. The empirical strategy we employ to side-step this problem lies in obtaining conditional expected returns and volatility series at various horizons and use these in our empirical tests. The next section shows how we extract conditional returns and volatility series from financial and economic variables.

1.4.1 Constructing Conditional Sharpe Ratio Series

In order to empirically test our framework, we need to construct conditional Sharpe ratio series from hard data, i.e. we need to construct conditional excess returns $E_t\{r_{t+1:t+k}\}$ and volatility $Vol_t\{r_{t+1:t+k}\}$ series at horizon k on the basis of investors' conditioning information set at time t. Depending on the empirical study at hand we allow the horizon k to range from 1 week to 12 months.

As evidenced by the returns predictability literature, a large number of conditioning variables have the potential to explain time-variations in expected returns at monthly, quarterly and yearly horizons. Welch and Goyal (2008) show that, if included in linear regression frameworks, none of the variables commonly employed in the literature can

consistently outperform the prevailing mean out-of-sample at the monthly or quarterly frequency. More recently, Rossi (2010) and Rossi and Timmermann (2010) show that the same predictor variables incorporated in a new forecasting method also known as Boosted Regression Trees (BRT) provide accurate predictions in- and out-of-sample.

At the monthly frequency, we construct conditional expected returns and volatility series according to the following model specifications:

$$E_t\{r_{t+1:t+k}\} = f_\mu(x_t|\hat{\theta}_\mu) \tag{1.20}$$

$$\widehat{Vol}_t\{r_{t+1:t+k}\} = f_\sigma(x_t|\hat{\theta}_\sigma), \tag{1.21}$$

where x_t represents a set of publicly available predictor variables, while $\hat{\theta}_\mu$ and $\hat{\theta}_\sigma$ are parameter estimates obtained via Boosted Regression Trees.[12] We use BRT at the monthly frequency because of their ability to incorporate a large amount of conditioning information without over-fitting the training dataset.

Our data comprises stock returns along with a set of twelve predictor variables previously analyzed in Welch and Goyal (2008). Stock returns are tracked by the S&P 500 index and include dividends. A short T-bill rate is subtracted to obtain excess returns. The covariates from the Welch-Goyal analysis are available during 1927-2005 and we extend their sample up to the end of 2008.[13] The twelve predictors are the lagged returns, the long-term returns, the volatility, the log dividend-price ratio, the log earnings-price ratio, the log dividend-earnings ratio, the three-month T-bill rate, the T-bill rate minus a three-month moving average, the yield on long term government bonds, the yield spread between BAA and AAA rated corporate bonds, the term spread and the inflation rate. Market volatility is unobserved, so we follow a large recent literature in proxying it through the square root of the realized variance obtained from daily data.

At the weekly horizon we use instead linear estimates of the form:

$$E_t\{r_{t+1:t+k}\} = \beta'_\mu x_t \tag{1.22}$$

$$\widehat{Vol}_t\{r_{t+1:t+k}\} = \beta'_\sigma x_t \tag{1.23}$$

where x_t is a set of publicly available predictor variables at the weekly frequency, i.e. the dividend yield, the earnings-price ratio, the default spread and the T-bill rate, while β_μ

---

[12]For a detailed introduction to Boosted Regression Trees and their application in financial economics, please refer to Rossi (2010) and Rossi and Timmermann (2010).

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

and β_σ vectors of parameter estimates obtained via ordinary least squares. We employ linear models for conditional estimates at the weekly horizon, because at that frequency we have very few covariates available and the benefits of using BRT would not be substantial. Furthermore, BRT generally work well when estimated on long time-series, but observations at the weekly frequency are available only for the most recent years. Armed with conditional Sharpe ratio estimates we can test the micro and macroeconomic implications of our model.

1.4.2 Empirical Tests Based on Micro-Data

Next we conduct two empirical exercises on micro-data. The first evaluates investors' trading activity as a function of the investment opportunity set by estimating duration models on Odean (1999)'s brokerage account data. The second evaluates the extent to which changes in the conditional Sharpe ratios affect investors' positions in risky assets using panel data estimates on the PSID.

Individuals' trading activity and the investment opportunity set

As mentioned in Sections 1.3, our model implies that individuals' trading and the investment opportunity set are closely related: periods characterized by high (absolute) conditional Sharpe ratios should also be characterized by intense trading. By taking advantage of the very detailed data of Odean (1999), we test this prediction via a duration model.

In a continuous-time model of asset allocation with no optimal inattention (and no transactions costs), agents should continuously change their portfolio allocations to incorporate new information as it becomes available. In a model that incorporates transactions costs, agents should trade only if the changes in the portfolio allocations deliver an increase in the expected portfolio returns that are large enough to outweigh the transactions costs incurred to place the trade. Finally, in a model of optimal inattention without time-varying investment opportunities, agents' inattention should not be a function of the conditional Sharpe ratio.

For each household in our dataset, we have access to all the transactions undertaken between January 1991 and November 1996: i.e. we have detailed information on the various accounts held by the household and the transactions undertaken therein. In accordance with our theoretical model, we use the time between transactions as a natural

measure of inattention to the stock market. In Odean (1999)'s dataset,[14] the transactions are recorded at the daily frequency and this would in principle require conditional Sharpe ratio estimates at the daily frequency as well. We compute them instead at the weekly frequency, because the latter strikes a good balance between the need to capture quick changes in investment opportunities and the need to condition our estimates on a sufficient number of covariates. Our weekly conditional returns and volatility series are obtained from the following linear models:

$$r_{\{t+1:t+5\}} = \alpha + \beta \, dy_t + \gamma \, ep_t + \delta \, defspr_t + \theta \, rfree_t + \zeta \, vol_{\{t-4:t\}} + \epsilon_{\{t+1:t+5\}}$$
(1.24)

$$vol_{\{t+1:t+5\}} = \alpha + \beta \, dy_t + \gamma \, ep_t + \delta \, defspr_t + \theta \, rfree_t + \zeta \, vol_{\{t-4:t\}} + \epsilon_{\{t+1:t+5\}}$$
(1.25)

where $r_{\{t+1:t+5\}}$ are the (excess) returns on the S&P 500 index over a weekly horizon and $vol_{t+1:t+5}$ is the realized volatility estimate for the S&P 500 index obtained from daily data.

After computing the conditional returns and volatility estimates $\widehat{r}_{t+1:t+5}$ and $\widehat{vol}_{t+1:t+5}$, we construct conditional Sharpe ratios

$$\widehat{Sh}_{\{t+1:t+5\}} = \frac{\widehat{r}_{\{t+1:t+5\}}}{\widehat{vol}_{\{t+1:t+5\}}}$$
(1.26)

based on observables. We employ the latter interacted with an indicator variable I_{trade} indicating whether the household placed a trade over the previous calendar year as the covariate of interest in our baseline model specification:

$$\lambda(t,x) = \lambda_0(t) \exp\{\alpha + \beta \, I_{\{trade\}} \times \widehat{Sh}_{\{t+1:t+5\}} + \gamma x\},$$
(1.27)

where t is the number of days since the last transaction and x is a vector of constant household covariates. The vector contains information on the "Investment Objective", "Knowledge", "Experience", "Tax Rate", "Income" and "Net Worth" of each household. These covariates should affect how often households trade and are included to obtain more precise estimates of the effect of the conditional Sharpe ratio on time-varying trading.

There are four investment objectives: conservative, income, growth and speculation. Individuals are allowed to choose more than one investment objective and all

---

[14]Please refer to Odean (1999) and related papers for a more precise description of the data at hand

possible combinations appear in the dataset. In Table 1 Panel A we report the categories, together with their frequencies to give a picture of the type of investors populating the dataset. Only 5% of the investors have conservative investment objectives and approximately 70% of the households have growth, income or both as their investment objectives. There exists some cases of rather contradictory investment objectives like the "conservative and speculation" category, for example. Fortunately, the category comprises only 0.15% of the households. Furthermore, recall that households generally hold more than one investment accounts, so it is not inconceivable that different accounts serve different investment strategies, making *prima facie* contradictory categories more plausible. "Knowledge" and "experience" have four categories each: none, limited, good and extensive. In Table 1 Panels B and C we show the frequency of each category. Approximately 9% of the households claim to have no financial sophistication, while 33% believe to have limited financial expertise. The "good" and "extensive" categories are chosen by 46% and 12% of the households, respectively. The frequencies for the "experience" categories are very similar. In fact "knowledge" and "experience" have a correlation coefficient of 0.8 on our sample. Because of the strong collinearity between the two measures, we only use the first in our estimates. Including that latter instead does not change our findings.

The distributions of income and net worth are, as expected, right-skewed so we log both variables before performing our estimations. Finally, households are categorized as "active traders" if they perform 48 or more commissioned trades in a year, "affluent household" if their invested wealth exceeds $100,000, and "general brokerage" otherwise. We first estimate our duration model for the three household categories separately to assess whether households' behavior vary substantially across household types. We then pool our estimates over the whole dataset.

Over our sample, January 1991 to November 1996, the conditional Sharpe ratio is always positive, so our model implies a positive and monotonic relation between the agent's attention to the stock market and the Sharpe ratio. Table 2 Panel A reports the results for a Cox semi-parametric proportional hazard model estimating Eq. 1.27. We report the hazard ratios and the z-statistics for the conditional Sharpe ratios as well as the constant control variables described above. The results indicate that higher Sharpe ratios are associated with higher trading probabilities for the investor. The hazard ratios for the conditional Sharpe ratio in "affluent", "general brokerage" and "active trader"

households are 1.465, 1.744 and 1.228, respectively and the pooled coefficient estimate across household types is 1.628. The coefficients' display a high degree of statistical significance as indicated by the z-statistics as well as high economic significance. For example, an estimated hazard ratio equal to 1.628 implies that a standard deviation increase in the conditional Sharpe ratio (0.25 in our sample) increases the probability of a trade by 16.7% for the average household.

Panel B repeats the estimates of Panel A without including the recent trade indicator variable: i.e. it estimates

$$\lambda(t, x) = \lambda_0(t) \exp\{\alpha + \beta \, \widehat{Sh}_{\{t+1:t+5\}} + \gamma x\} \tag{1.28}$$

instead of Eq. 1.27. We estimate the two models in an effort to distinguish between our model of inattention and alternative models of asset pricing that incorporate transactions costs, but not information costs. In our model, financial agents do not observe the Sharpe ratio unless they pay information costs. The implication is that over a period of inattention, the investor would not be able to reap the benefits from periods of high conditional Sharpe ratios because he literally would not be aware of them. In a model that incorporates transactions costs but not information costs, on the other hand, the agent would always know the value of the Sharpe ratio and act accordingly. Similar coefficient estimates across the two model specifications would rule in favor of a model with transactions costs rather than information costs. On the other hand, larger coefficient estimates for the model that includes the recent trade indicator variable would indicate that a model with information costs is a better characterization of agents' time-varying trading behavior. The results reported in Panel B show that the conditional Sharpe ratio does not seem to have an impact on trading for "affluent households" and has a slightly negative and economically weak impact on trading for "general brokerage" households. The only household group where the conditional Sharpe ratio has a positive impact are the active traders, which is expected, given that they are the ones most likely to track closely the stock market at all times.

Comparing Panels A and B is instructive because it shows that the conditional Sharpe ratio increases the investor's trading probability only if the agent has been recently paying attention to the stock market. It therefore provides evidence supporting our model of information costs against those featuring only transactions costs. Models of inattention that do not allow for time-varying investment opportunities cannot be consistent with our findings, because they imply that trading probabilities are constant

over time and not a function of the conditional Sharpe ratio. Finally, standard models of time-varying asset allocation cannot be consistent with our findings, because they require that agents trade continuously over time.

Turning to the constant covariates, the results show that compared to the base group that has no financial sophistication, limited knowledge of the stock market generally implies a lower probability of trading as the hazard ratios average approximately 0.9 across the three household groups. For affluent and general brokerage, good and extensive knowledge of financial markets is associated with greater trading probabilities with hazard ratios averaging approximately 1.15 and 1.35 across the two household groups. In the active traders group, instead, agents with extensive knowledge of financial markets trade as often as agents with no financial expertise, while agents with good knowledge trade significantly less. The non-monotonic relation between trading and financial expertise is likely the outcome of two competing effects. On the one hand, agents with more experience and knowledge of financial markets realize rather quickly that stock picking and market timing skills are very hard to develop and buy-and-hold strategies are often more remunerative. On the other hand, the higher the degree of financial sophistication, the greater the number of trading signals an agent is likely to receive; it is also possible that over-confidence could play a role in the active investment strategies pursued by this group.

The reason why active traders behave differently from the other two groups is probably due to the fact that the competing effects mentioned above have different relative strengths in this group. Furthermore, it is hard to believe that individuals that open an investment account claiming to place more than 48 trades a year (approximately one trade a week!) do not have any degree of financial sophistication.

Turning to the investment objectives, affluent and general brokerage accounts generally have higher trading probabilities if their investment objectives contain growth or speculation. "Income" is instead associated with less trading. The table separates the large investment objectives groups that characterize at least 1% of the sample, from the small ones. The latter groups tend to have estimated coefficients that are less significant and more erratic in general, reflecting the smaller number of observations they are estimated on. The results for active traders are slightly out of line compared to the ones displayed by the other two household groups: speculation is associated with more frequent trading, while growth and income are associated with longer duration between

transactions.

Tax and income tend to have statistically and economically insignificant coefficients. As far as taxes are concerned, agents have greater incentives not to realize gains if their tax rate is high and the opposite holds true for realizing losses. Given that we do not make a distinction between buys and sales, the two effects are probably canceling each other out. Income is either insignificant or has a negative impact on trading. The coefficient is surprising at first sight, but it is probably due to the collinearity between income and net worth, which is instead significant across the board with hazard ratios equal to 1.054, 1.018 and 1.077 for affluent households, general brokerage households and active traders, respectively.

To check the robustness of our results, in Table 3 Panel A we repeat the analysis reported above by modeling parametrically the $\lambda_0(t)$ function that was estimated non-parametrically in Table 2. We model it as an exponential function because the exponential distribution is suitable for modeling data with constant (unconditional) hazard. We report only the hazard ratios and the z-statistics for the conditional Sharpe ratio and omit the results for the control variables because similar to the ones reported in Table 2.

The results are in line with the ones reported in Table 2, with the important difference that the hazard ratios are greater compared to the ones estimated using the Cox semi-parametric model. This holds true for both the specification that includes the recent trade indicator, Panel A, and the one that does not, Panel B. Economically the results are more significant. For example, a hazard ratio equal to 1.883, as estimated for the general brokerage households, implies that the a one-standard deviation increase in the conditional Sharpe ratio increases, on average, the probability of a trade by 22% if the household has placed a trade some time over the course of the previous calendar year.

In Panels B and C of Table 3 we present the coefficient estimates from a logistic regression model specification. To perform the analysis, we first construct an indicator variable denoting whether a trade occurred in a given week and convert the dataset into

weekly observations. We then estimate the following two model specifications:

$$Pr(trade = 1|\widehat{Sh}_{\{t+1:t+5\}}, x) = \frac{exp\{\alpha + \beta \ I_{\{trade\}} \times \widehat{Sh}_{\{t+1:t+5\}} + \gamma x\}}{1 + exp\{\alpha + \beta I_{\{trade\}} \times \widehat{Sh}_{\{t+1:t+5\}} + \gamma x\}} \tag{1.29}$$

$$Pr(trade = 1|\widehat{Sh}_{\{t+1:t+5\}}, x) = \frac{exp\{\alpha + \beta \widehat{Sh}_{\{t+1:t+5\}} + \gamma x\}}{1 + exp\{\alpha + \beta \widehat{Sh}_{\{t+1:t+5\}} + \gamma x\}}, \tag{1.30}$$

where the covariates are the same as the ones specified in the duration models described above. The results are reported in Panel B. In Panel C we estimate the above expression without the constant covariates and introducing fixed effects in order to capture the constant households' heterogeneity in a different fashion. In all specifications and across all household types, the coefficient for the conditional Sharpe ratio is positive and significant. Including the recent trade indicator results in more statistically and economically significant coefficients, as implied by our model of time-varying trading. Qualitatively, the logistic regression specifications with constant covariates yield similar results to the ones that employ fixed effects. Quantitatively, their coefficient estimates are smaller for the specifications that includes the recent trade indicator. They are instead uniformly larger for the specifications that does not include the recent trade indicator.

In unreported results we construct an alternative measure of time-varying investment opportunities based only on past realized Sharpe ratio estimates of the form

$$Real\_Sharpe_t = \frac{\prod_{i=1}^{T}(1 + r_{t-i})}{\left(\sum_{i=1}^{T} r_{t-i}^2\right)^{1/2}} \tag{1.31}$$

at the weekly, bi-weekly, monthly and bi-monthly frequencies. In the second step of our analysis we use this quantity instead of the conditional Sharpe ratio to estimate the duration and logistic regression models. The results for the realized Sharpe ratios are qualitatively consistent with the conditional ones, but smaller in magnitude.

Overall, the results reported in this section show that the changes in time-varying investment opportunities as captured parsimoniously by the conditional Sharpe ratio play an important role at explaining time-variations in investors' trading as implied by our model. They are instead inconsistent with the models of time-varying asset pricing proposed so far in the literature.

Another important prediction of our model is that changes in investment opportunities should be associated with time-variations in investors' exposure to risky assets. Unfortunately, the time-series dimension of Odean's dataset is too short to test this hypothesis. We decide instead to employ panel data estimates from the PSID, because the

latter dataset extends for a longer time-span even though the data is sampled at a lower frequency.

Investors' positions in risky assets and the investment opportunity set

We test whether changes in the investment opportunity set affect investors' portfolio allocations using panel data estimates from the PSID. The PSID is a longitudinal study that tracks family units and their offspring over time. The financial addenda relevant to our analysis were collected from 1984 to 1999 every 5 years. For a more precise description of the data, refer to Brunnermeier and Nagel (2008) as our analysis is based on the same dataset and data-cleaning procedure performed by them.

We define financial assets as the sum of holdings of stocks and mutual funds plus riskless assets. The riskless assets are defined as the sum of cash-like assets and holding of bonds. The share of risky assets held by each household is computed as the sum of stocks and mutual funds, divided by the households' total financial assets. We abstract from issues related to stock market participation and we include in our analysis only stock market participants.[15] In order to analyze whether an increase/decrease in the conditional Sharpe ratio is associated with an increase/decrease in risky assets held in individuals' portfolios, we first denote for each household the change in the proportion of risky assets over a k years horizon as $\Delta_k \phi_{i,t} = \phi_{i,t} - \phi_{i,t-k}$, where $k = 5$ and use it as a regressand in a linear panel data model. Our model predicts that investors' positions in risky assets are a function of the conditional Sharpe ratio, but also that inattention introduces inertia in investors' portfolio weights. In order to capture this idea in a reduced form model, we use yearly conditional Sharpe ratio estimates obtained from BRT and define the average conditional Sharpe ratio over a five year horizon preceding (and including) time t as a regressor and denote it by $\widehat{AVGt\_Sh}_{t-(k-1):t}$. We use $t-k+1$ instead of $t-k$, because the conditional Sharpe ratio at time $t-k$ likely to affect the change in portfolio allocation at time $t-k$. Given that the PSID data collection is carried out throughout the year, but we do not know when each household is interviewed, we use the conditional Sharpe ratios from June of each year in our analysis. Given that our covariate of interest is the change in the average conditional Sharpe ratio, we then compute $\Delta_k \widehat{AVG\_Sh}_{t-(k-1):t}$.

While the change in the conditional Sharpe ratio is the key quantity of our

---

[15]For analyses of the determinants of households' stock market participation, please refer to Brunnermeier and Nagel (2008) and Vissing-Jorgensen (2002).

theoretical model, decades of research in financial economics have identified a multitude of socioeconomic variables that determine investors' positions in risky assets. In order to account for their effects we follow Brunnermeier and Nagel (2008) and condition our estimates on a vector q_{t-k} of household characteristics that includes a broad range of variables related to the life-cycle: age and age^2; indicators for completed high school and college education, respectively, and their interaction with age and age^2; marital status and health status; the number of children in the household and the number of people in the household. It also includes variables related to the household's employment and financial situation: a number of dummy variables for any unemployment in the k years leading up to and including year $t - k$; the coverage of the household head's job by a union contract; the log of the equity in the vehicles owned by the household; the log family income at time $t - k - 4$; the two-year growth in log family income at $t - k$ and $t - k - 2$ and a variable for inheritances received in the k years leading up to and including year $t - k$.

In addition to the life-cycle variables reported above, we include preference shifters that control for changes in some household characteristics between $t - k$ and t and denote them by $\Delta_k h_t$: changes in family size; changes in the number of children; sets of dummies for house ownership, business ownership and non-zero labor income at t and $t - k$. The idea to control for house ownership is motivated by the body of research showing that households might save towards buying a house by way of risky assets. When the house is purchased, the risky assets are sold, reducing dramatically the proportion of risky assets in the household's portfolio (see Faig and Shum (2002)). In addition, we recognize that changes in the economic conditions may vary across regions, so we include the four PSID geographical regions. As a final control variable we include the change in wealth $\Delta_k w_{i,t}$, because habit persistence models imply that changes in wealth affect risk-aversion, and hence investors' portfolio allocations.

Our estimation equation can then be written as

$$\Delta_k \phi_{i,t} = \alpha + \beta \ \Delta_k \widehat{AVG\_S}h_{t-(k-1):t} + \gamma \ \Delta_k w_{i,t} + \delta \ q_{i,t-k} + \theta \Delta_k h_{i,t} + \epsilon_{i,t} \qquad (1.32)$$

We report our results in Table 4 Panel A. Column 1 reports results for Eq. 1.32 while column 2 introduces a number of additional control variables related to asset composition: the labor income/liquid wealth ratio interacted with age, the business wealth/liquid wealth ratio, and the housing wealth/liquid wealth ratio. For both specifications, the estimates of the β coefficients show that changes in the average conditional Sharpe ratio

have a significantly positive effect on the proportion of risky assets held by the investors. The magnitude of the coefficient is 0.392 for the baseline specification, implying an economically significant effect: a 10% increase in the difference between the average conditional Sharpe ratio and its lag value implies a 3.9% increase in the proportion of risky assets held by investors. Furthermore, with a t-statistic equal to 2.72, the coefficient is significant at conventional statistical levels.

We also report the results for the γ coefficient, i.e. the one for the changes in wealth. As highlighted by Brunnermeier and Nagel (2008), this is a particularly important coefficient because habit persistence models imply a positive relation between changes in wealth and the proportion of wealth invested in risky assets. In line with the findings reported by Brunnermeier and Nagel (2008), the coefficient has a sign opposite to the expected one (negative instead of positive), providing strong evidence against the ability of habit persistence models to explain micro-data.

As a robustness check, in Panel B of Table 4 we present alternative estimates of Eq. 1.32 by replacing $\Delta_k \widehat{AVG\_Sh}_{t-(k-1):t}$ with $\Delta_k \widehat{Sh}_t$, i.e. by including the change in the investment opportunity set rather than the average change in the investment opportunity set. The estimates reported in Panel B show that the change in the conditional Sharpe ratio between two survey dates is a statistically significant predictor of the changes in investors' portfolio allocations. The coefficients are economically smaller compared to the ones for the changes in the average Sharpe ratio, but this is mainly due to the effect of averaging in Panel A. They maintain a strong statistical significance though.

Overall, the estimates reported in Panels A and B of Table 4 are consistent with our model of inattention that features time-varying investment opportunities, but not with those frameworks of inattention that imply a constant conditional Sharpe ratio. Hence, our results show that allowing for a price of risk that varies over time is important because it generates a model that explains more precisely investors' behavior.

1.4.3 Empirical Tests Based on Macro-Data

We present next the empirical tests for the macroeconomic implications of our model. We first evaluate empirically the time-varying covariance between stock returns and aggregate consumption using a high-frequency proxy for the latter, i.e. the ADS index. We then assess empirically whether aggregate time-varying trading, proxied by volumes on the S&P 500, is a function of the conditional Sharpe ratios.

Time-varying covariance between aggregate consumption and stock returns.

As in the previous section, we use Boosted Regression Trees and employ monthly observations to compute conditional Sharpe ratios at the one-year horizon. To test whether changes in the conditional Sharpe ratio affect the covariance between consumption and stock returns, we adopt an empirical strategy similar to the one introduced by Rossi and Timmermann (2010), which entails proxying high-frequency consumption using the ADS index. The ADS index is designed to track the business conditions of the US economy at the daily frequency, by incorporating high- and low-frequency as well as stock and flow data. It is obtained by estimating via the Kalman Filter a dynamic factor model that incorporates 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. In unreported results we show that there exists a close relation between the ADS index and total, durable and non-durable consumption with uniformly positive correlations that 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.

The advantage of using the ADS-index as opposed to consumption data is that it allows for the construction of realized co-variances between monthly stock returns and consumption using daily observations, allowing us to test very precisely how consumption and stock returns co-vary as a function of the investment opportunity set.

We compute monthly "realized covariances" between stock returns and scaled changes in the ADS index, $\widehat{cov}_t$, from daily observations as follows:

$$\widehat{cov}_t = \sum_{i=1}^{N_t} \Delta ADS_{i,t} \times r_{i,t}, \qquad (1.33)$$

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

Once the realized covariance measure is obtained, we construct estimates of the conditional covariance

$$\widehat{cov}_{t+1:t+k|t} = f_{cov}(x_t|\hat{\theta}_{cov}), \qquad (1.34)$$

where x_t represents the same set of publicly available predictor variables we employ to construct conditional expected returns and volatilities, $\hat{\theta}_{cov}$ are estimates of the parame-

ters obtained via Boosted Regression Trees and k is a month identifier that ranges from 1 to 12.[16]

In the final step, we characterize parametrically and non-parametrically the relation between the conditional covariance and the conditional Sharpe ratio by estimating the following models

$$\widehat{cov}_{t+1:t+k|t} = \alpha + \beta \ \widehat{Sh}_{t+1:t+12|t} + \epsilon_{t+1:t+k} \tag{1.35}$$

$$\widehat{cov}_{t+1:t+k|t} = f(\widehat{Sh}_{t+1:t+12|t}) + \epsilon_{t:t+k} \tag{1.36}$$

The interesting aspect of working with $\widehat{Sh}_{t+1:t+12|t}$ is twofold. First, it is the same conditional Sharpe ratio measure employed in our PSID panel data estimates. Second, the conditional Sharpe ratio estimates are virtually never below -0.7, threshold below which our model would predict high attention as shown in Figure 1.2. It follows that, empirically, we should expect a monotonic relation between the conditional Sharpe ratio and the conditional covariance between consumption and stock returns.

We report the results for estimates of Eq. (1.35) in Table 5. The β coefficient is positive and significant up to the sixth month, indicating that at short horizons the correlation between stock returns and consumption is time-varying and a function of the conditional Sharpe ratio as implied by our model. The fact that the correlation is not significantly different from zero at longer horizons is natural in our framework given that the longer the time interval employed, the greater the percentage of the population that adjusts its consumption to new levels, even during periods of long inattention.

To corroborate the results reported in Table 5, we present non-parametric estimates of the relation at hand using Boosted Regression Trees. In Figure 1.7 we report the results at 1, 2, 3, 6, 9 and 12 months horizons. From the plots it is clear that the relation is increasing and monotonic until the sixth month while it is effectively flat and non-monotonic at longer horizons.

Overall, the results reported in this section show that an increase in the conditional Sharpe ratio is associated with an increase in the conditional covariance between stock returns and consumption, as implied by our model. Conversely, they are inconsistent with traditional models of asset pricing or models of inattention that do not allow for time-varying investment opportunities, because they both imply a constant covariance between consumption and returns. In the next section we present our final empirical

---

[16]Refer to Rossi and Timmermann (2010) for a more detailed analysis of the conditional covariance series.

exercise testing whether economy-wide trading activity is a function of the conditional Sharpe ratio.

Conditional Sharpe ratio and trading volumes on the S&P 500

Our model predicts that in periods characterized by large (absolute) conditional Sharpe ratios, individuals' opportunity cost of storing wealth in the transactions account is high. It follows that agents should trade rather frequently and, consequently, aggregate volumes on the S&P 500 should be high. The opposite is true for periods where the conditional Sharpe ratio is small in absolute terms. These periods are characterized by a rather small opportunity cost of holding wealth in the transactions account, low attention to the stock market and a feeble aggregate trading activity. We test this implication of our model using aggregate trading volumes on the S&P 500.

Trading volumes have increased dramatically from 1950 through 2008. It follows that the raw volumes series shows clear signs of non-stationarity. In order to study the effect of the conditional Sharpe ratio on trading volumes, we take the natural log transformation of the latter and fit a quadratic trend. The general increase in the trading volumes is commonly attributed to the increase in wealth and financial sophistication of economic agents as well as the technological progress of financial institutions and financial markets. Our model is silent about these, so we limit ourselves to explaining deviations of aggregate volumes from their long-run trend. We define the difference between the observed volumes and their long-run trend as the de-trended volumes series "VOL" and make it the object of our analysis. We estimate

$$VOL_{t+1} = \alpha + \beta \ \widehat{Sh}_{t+1:t+12|t} + \gamma \ \widehat{Sh}^2_{t+1:t+12|t} + \delta \ \widehat{Sh}^3_{t+1:t+12|t} + \epsilon_{t+1}, \qquad (1.37)$$

where the third-order polynomial is employed to capture the non-linearities implied by the model. We also provide an alternative specification whereby we estimate

$$VOL_{t+1} = f(\widehat{Sh}_{t+1:t+12|t}; \theta) + \epsilon_{t+1}, \qquad (1.38)$$

via Boosted Regression Trees. The rationale for Eq.(s) 1.37 and 3 is that after controlling for technological advancements and increased stock market participation, the conditional Sharpe ratio should have an impact on agents' attention to the stock market and, consequently, their trading activity. In Figure 1.8, we report the fitted values for the two specifications proposed. The dashed red line plots the fitted values for the parametric

model, the solid black line the ones for the BRT model. Consistent with our theoretical model, there is a clear non-monotonic relation between aggregate trading volumes and the conditional Sharpe ratio: periods characterized by large Sharpe Ratios in absolute value are also characterized by large trading volumes. This holds irrespective of whether we employ a parametric (ordinary least squares) or non-parametric (Boosted Regression Trees) model to estimate the relation at hand.

Time-varying trading volumes are not consistent with a model of inattention that does not incorporate time-varying investment opportunities. It is also not consistent with a model that does not incorporate information costs, because the latter would entail that agents trade continuously over time. There are, however, a number of potential problems associated with testing our time-varying model of inattention using aggregate S&P 500 volumes. A common objection that could be raised is that only a small fraction of the trading volumes on the S&P 500 is generated by individuals' trades. This objection is based on the misconception that the bulk of equity investments is held by institutional investors. In fact, as reported by Barber and Odean (2000), in 1996 47% of equity investments in the United States were held directly by households, 23% by pension funds and 14% by mutual funds. Furthermore, even if institutional investors generated larger trading volumes than individual investors, it is unlikely that their time-varying trading is a function of the conditional Sharpe ratio. It follows that institutional investors' trading should only introduce noise in our results and hence reduce the precision of our estimates, but not affect the overall identification of our empirical strategy.

1.5   Conclusions

This paper presents a theoretical model of time-varying inattention to the stock market by introducing information costs into a continuous-time model of asset allocation with time-varying investment opportunities. Our model explains why individuals' attention to their investment portfolio is high during recessions as well as during stock market booms and crashes. It also rationalizes why agents do not modify their portfolio allocations gradually with the arrival of new information, but rather alternate extended periods of inertia with brief moments of action where asset allocations are updated according to the current state the economy. By implying a weak and time-varying co-variation between consumption growth and equity returns, our model also contributes to the literature that provides a theoretical explanation of the equity premium puzzle.

Empirical tests based on Odean (1999)'s dataset and estimates conducted on the PSID support the microeconomic predictions of our model against others that have been proposed so far in the literature. Furthermore, CCAPM-type estimates based on high and low frequency economic and financial time-series support the macroeconomic predictions of our model of a time-varying covariance between consumption and stock returns.

While this paper focuses on individuals' portfolio allocations, the potential implications of our theoretical framework are much broader. By combining information costs and time-varying investment opportunities, the multi-agent extension of our model implies that different individuals hold different conditioning information sets at any given point in time. As a consequence, our framework has the potential to serve as the micro-foundation for the reduced form models developed by Mankiw, Reis, and Wolfers (2003) and Carroll (2003) that study heterogeneities in macroeconomic expectations. These implications are currently being explored using survey data on inflation expectations.

Table 1.1: This table presents summary statistics of the knowledge, experience and investment objective variables for the investors populating the brokerage account data provided by Odean (1999). Panel A presents the frequencies for the investment objectives. Panels B and C present the frequencies for the self-reported level of knowledge and experience.

**Summary Statistics for Investors' Knowledge and
Objectives in Odean (1999)'s Dataset**

Panel A. Investment Objectives and associated Frequencies

Investment Objective(s)	Frequency
Growth	55.75%
Income, Growth	11.94%
Conservative, Growth	6.99%
Income	4.57%
Conservative, Income, Growth	5.86%
Conservative	5.00%
Conservative, Income	3.56%
Growth, Speculation	2.74%
Speculation	1.32%
Conservative, Income, Growth, Speculation	1.17%
Income, Growth, Speculation	0.54%
Conservative, Growth, Speculation	0.25%
Conservative, Speculation	0.15%
Income, Speculation	0.13%
Conservative, Income, Speculation	0.02%

Panel B. Knowledge and associated Frequencies

Knowledge	Frequency
None	8.85%
Limited	33.20%
Good	45.82%
Extensive	12.13%

Panel C. Experience and associated Frequencies

Experience	Frequency
None	3.74%
Limited	35.60%
Good	46.66%
Extensive	13.99%

Table 1.2: This table reports the results of a Cox semi-parametric duration model estimating the number of days between investors' transactions as a function of time-varying as well as constant co-variates. The time-varying covariate is the conditional Sharpe ratio estimated at the weekly frequency using observable economic and financial time-series. The constant covariates are the investors' financial sophistication (knowledge), investment objectives, tax rate, income and net worth. For each covariate we report the hazard ratio and z-statistic. Panel A interacts the conditional Sharpe ratio with a recent trade indicator variable. Panel B does not.

Cox Semi-Parametric Proportional Duration Model With and Without Recent Trade Indicator Variable

	A. Specification With the Recent Trade Indicator				**B. Specification Without the Recent Trade Indicator**			
	Time-Varying Covariates							
Account type	**Affluent**	**General**	**Active**	**Total**	**Affluent**	**General**	**Active**	**Total**
Sharpe Ratio	1.465	1.744	1.228	1.628	0.983	0.935	1.097	1.000
	(33.16)**	(65.06)**	(25.03)**	(94.33)**	(-1.48)	(-7.85)**	(11.21)**	(0.00)
	Constant Covariates							
Account type	**Affluent**	**General**	**Active**	**Total**	**Affluent**	**General**	**Active**	**Total**
Investors' Knowledge								
Limited	0.966	0.93	0.835	0.766	0.966	0.928	0.835	0.765
	(-2.91)**	(-10.47)**	(-27.87)**	(-61.31)**	(-2.94)**	(-10.65)**	(-27.85)**	(-61.67)**
Good	1.143	1.146	0.888	0.963	1.144	1.149	0.888	0.965
	(11.79)**	(20.09)**	(-20.50)**	(-9.35)**	(11.92)**	(20.55)**	(-20.41)**	(-8.85)**
Extensive	1.268	1.407	1.002	1.253	1.272	1.417	1.003	1.259
	(19.23)**	(42.85)**	(0.30)	(49.79)**	(19.44)**	(43.72)**	(0.45)	(50.88)**
Objective (Large Groups)								
G	1.064	1.024	0.889	0.993	1.065	1.025	0.888	0.993
	(5.15)**	(2.61)**	(-13.87)**	(-1.34)	(5.28)**	(2.69)**	(-13.92)**	(-1.23)
I; G	1.104	1.017	0.914	1.085	1.104	1.019	0.914	1.086
	(7.67)**	(-1.68)	(-9.77)**	(13.50)**	(7.68)**	(1.86)	(-9.78)**	(13.75)**
C; G	1.129	1.101	0.899	1.075	1.131	1.104	0.899	1.077
	(8.90)**	(8.66)**	(-10.70)**	(11.11)**	(8.99)**	(8.91)**	(-10.73)**	(11.43)**
I	0.928	0.969	0.981	1.027	0.928	0.97	0.981	1.028
	(-4.46)**	(-2.41)*	(-1.68)	(3.47)**	(-4.47)**	(-2.38)*	(-1.67)	(3.60)**
C; I; G	1.197	1.135	0.907	1.217	1.199	1.14	0.907	1.221
	(13.52)**	(10.23)**	(-10.21)**	(30.52)**	(13.63)**	(10.56)**	(-10.20)**	(31.08)**
C; I	0.971	0.917	0.911	1.024	0.97	0.916	0.911	1.024
	(-1.91)	(-5.41)**	(-8.02)**	(3.02)**	(-1.99)*	(-5.53)**	(-8.06)**	(3.02)**
G; S	1.249	1.265	1.117	1.449	1.252	1.27	1.117	1.454
	(11.93)**	(18.64)**	(10.70)**	(51.54)**	(12.04)**	(18.93)**	(10.71)**	(52.08)**
S	1.191	1.128	1.155	1.207	1.192	1.130	1.154	1.208
	(5.85)**	(7.55)**	(9.52)**	(18.41)**	(5.88)**	(7.66)**	(9.47)**	(18.47)**
C; I; G; S	1.060	1.396	1.101	1.528	1.062	1.402	1.101	1.535
	(2.64)**	(18.59)**	(8.33)**	(49.59)**	(2.70)**	(18.84)**	(8.34)**	(50.13)**
Objective (Small Groups)								
I; G; S	1.352	1.314	1.041	1.698	1.357	1.324	1.040	1.707
	(8.57)**	(11.34)**	(2.80)**	(47.24)**	(8.67)**	(11.65)**	(2.74)**	(47.71)**
C; G; S	0.926	0.987	1.277	1.969	0.925	0.994	1.277	1.978
	(-1.28)	(-0.31)	(14.72)**	(47.38)**	(-1.31)	(-0.15)	(14.72)**	(47.69)**
C; S	1.670	0.941	1.399	1.601	1.680	0.942	1.397	1.604
	(7.42)**	(-1.36)	(13.00)**	(22.36)**	(7.51)**	(-1.34)	(12.94)**	(22.44)**
I; S	1.068	0.935	0.802	1.227	1.07	0.938	0.801	1.233
	(-1.09)	(-1.06)	(-6.30)**	(7.54)**	(1.13)	(-1.02)	(-6.35)**	(7.71)**
C; I; S	0.817	0.354	0.303	0.508	0.812	0.354	0.301	0.507
	(-1.10)	(-3.74)**	(-7.63)**	(-6.19)**	(-1.14)	(-3.75)**	(-7.68)**	(-6.23)**
Tax	1.001	0.999	0.997	0.998	1.001	0.999	0.997	0.998
	(2.16)*	(-2.31)*	(-17.38)**	(-14.41)**	(2.22)*	(-2.37)*	(-17.38)**	(-14.47)**
Income	0.977	1.000	0.996	1.011	0.977	1.000	0.995	1.011
	(-5.67)**	(-0.040)	(-1.92)	(6.82)**	(-5.65)**	(-0.07)	(-2.02)*	(6.86)**
Net Worth	1.054	1.018	1.077	1.156	1.054	1.018	1.077	1.157
	(21.83)**	(9.10)**	(51.06)**	(141.49)**	(21.87)**	(9.30)**	(51.14)**	(142.67)**
Observations	1,568,733	3,363,061	1,261,279	6,193,073	1,568,733	3,363,061	1,261,279	6,193,073

Table 1.3: This table reports in Panel A the results of a parametric exponential duration model estimating the number of days between investors' transactions as a function of time-varying as well as constant covariates. The time-varying covariate is the conditional Sharpe ratio estimated at the weekly frequency using observable economic and financial time-series. The constant covariates, whose coefficient estimates are not reported, are the investors' financial sophistication (knowledge), investment objectives, tax rate, income and net worth. For each covariate we report the hazard ratio and z-statistic. In Panel B and C we present the results of two logistic regression specifications. The first employs the same covariates used in Panel A, while the second omits the constant covariates and introduces fixed effects at the household account level instead. In all panels we present the results for a specification that interacts the conditional Sharpe ratio with a recent trade indicator variable and one that does not.

Trading Probabilities and the Investment Opportunity Set

Panel A. Parametric Exponential Duration Model

	With Recent Trade Indicator				Without Recent Trade Indicator			
	Affluent	General	Active	Total	Affluent	General	Active	Total
Hazard Ratio	1.578	1.883	1.265	1.673	1.118	1.048	1.170	1.114
Z-statistic	(43.28)	(79.16)**	(34.66)**	(110.33)**	(10.69)**	(6.01)**	(23.23)**	(23.39)**

Panel B. Logistic Model for Trading Probabilities With Constant Covariates

	With Recent Trade Indicator				Without Recent Trade Indicator			
	Affluent	General	Active	Total	Affluent	General	Active	Total
Coefficient	0.451	0.639	0.261	0.539	0.080	0.024	0.150	0.073
t-statistic	(37.33)**	(70.83)**	(26.25)**	(94.19)**	(6.77)**	(2.75)**	(15.22)**	(12.95)**

Panel C. Logistic Model for Trading Probabilities With Fixed Effects

	With Recent Trade Indicator				Without Recent Trade Indicator			
	Affluent	General	Active	Total	Affluent	General	Active	Total
Coefficient	0.256	0.321	0.218	0.272	0.095	0.038	0.166	0.093
t-statistic	(21.27)**	(35.94)**	(21.17)**	(46.19)**	(8.03)**	(4.34)**	(16.12)**	(15.93)**

Table 1.4: This table reports in Panel A the results of a linear regression model estimating the relation between changes in households' risky assets positions over the course of five years as a function of average changes in the conditional Sharpe ratio and households' wealth over the same time-span. A number of additional covariates that control shifts in preferences, life-cycle and regional effects are included, but their coefficients are not reported. Specification 2 differs from specification 1 in that it includes variables that control for households' asset composition. Panel B repeats the analysis substituting the average changes in the conditional Sharpe ratio with changes in the conditional Sharpe ratio. Coefficients and standard errors are reported for the regressors of interest. Coefficients significant at 5% level are bold-faced.

Evidence of time-varying portfolio allocation from PSID

**Panel A. Portfolio Allocations and Changes
in the Average Conditional Sharpe ratio.**

	Spec 1	Spec 2
β	**0.392**	**0.412**
S.E. β	0.144	0.146
γ	**-0.014**	-0.010
S.E. γ	0.006	0.009
Asset control Composition	No	Yes
Preference Shifters	Yes	Yes
Life-Cycle controls	Yes	Yes
Regions Fixed Effects	Yes	Yes
R^2	3.9%	4.2%
N	1234	1234

**Panel B. Portfolio Allocations and
Changes in the Conditional Sharpe ratio.**

	Spec 1	Spec 2
β	**0.035**	**0.037**
S.E. β	0.014	0.015
γ	**-0.014**	-0.010
S.E. γ	0.006	0.009
Asset control Composition	No	Yes
Preference Shifters	Yes	Yes
Life-Cycle controls	Yes	Yes
Regions Fixed Effects	Yes	Yes
R^2	3.8%	4.0%
N	1234	1234

Table 1.5: This table reports α and β coefficients and standard errors for the linear model: $\widehat{cov}_{t+1:t+k|t} = \alpha + \beta \; \widehat{Sh}_{t+1:t+12|t} + \epsilon_{t+1:t+k}$, where $\widehat{cov}_{t+1:t+k|t}$ is the conditional covariance between consumption and stock returns over "k" months and k ranges from 1 to 12. $\widehat{Sh}_{t+1:t+12|t}$ is the one-year conditional Sharpe ratio. Both conditional covariance and Sharpe ratio measures are obtained via Boosted Regression Trees over the sample 1960-2008. Coefficients significant at 5% level are bold-faced.

Covariance Between Consumption and Stock Returns
as a Function of the Conditional Sharpe Ratio

	Horizon (Months)											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
$\boldsymbol{\alpha}$	**-0.04**	0.05	**0.27**	**0.38**	**0.47**	**1.00**	**1.36**	**2.52**	**3.33**	**3.42**	**3.77**	**3.95**
S.E. $\boldsymbol{\alpha}$	0.01	0.05	0.09	0.12	0.16	0.21	0.25	0.29	0.35	0.39	0.43	0.45
$\boldsymbol{\beta}$	**0.18**	**0.38**	**0.50**	**0.67**	**1.04**	**0.78**	0.65	-0.35	-0.72	-0.55	-0.49	-0.28
S.E. $\boldsymbol{\beta}$	0.02	0.07	0.13	0.18	0.24	0.30	0.37	0.43	0.52	0.57	0.63	0.66
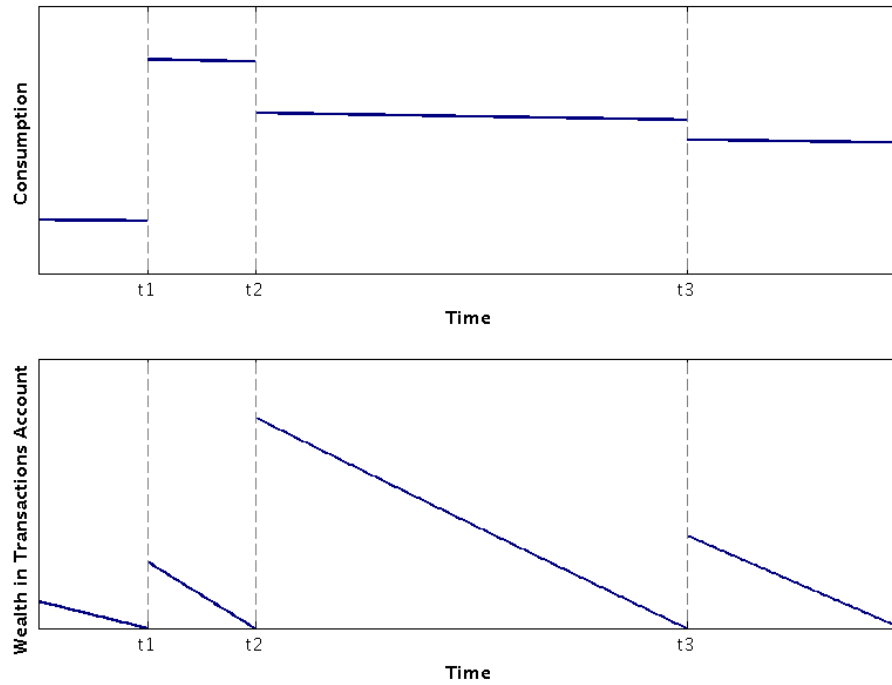
Figure 1.1: This figure plots the dynamics of consumption (Top Panel) as well as the dynamics of the wealth stored in the transactions account (Bottom Panel) as a function of time. Consumption changes in a lumpy fashion at the times of action (represented by vertical dotted lines) and varies very little during the periods of inattention: in fact, if $r^L = \rho$ consumption is constant between the times of action. The behavior of consumption during the periods of optimal inattention is independent from the dynamics of invested wealth and the investment opportunity set. The wealth held in the transactions account is exhausted completely when the period of optimal inattention ends. This is optimal for the agent because the transactions account pays a lower rate of returns compared to the expected returns paid by the investment portfolio.
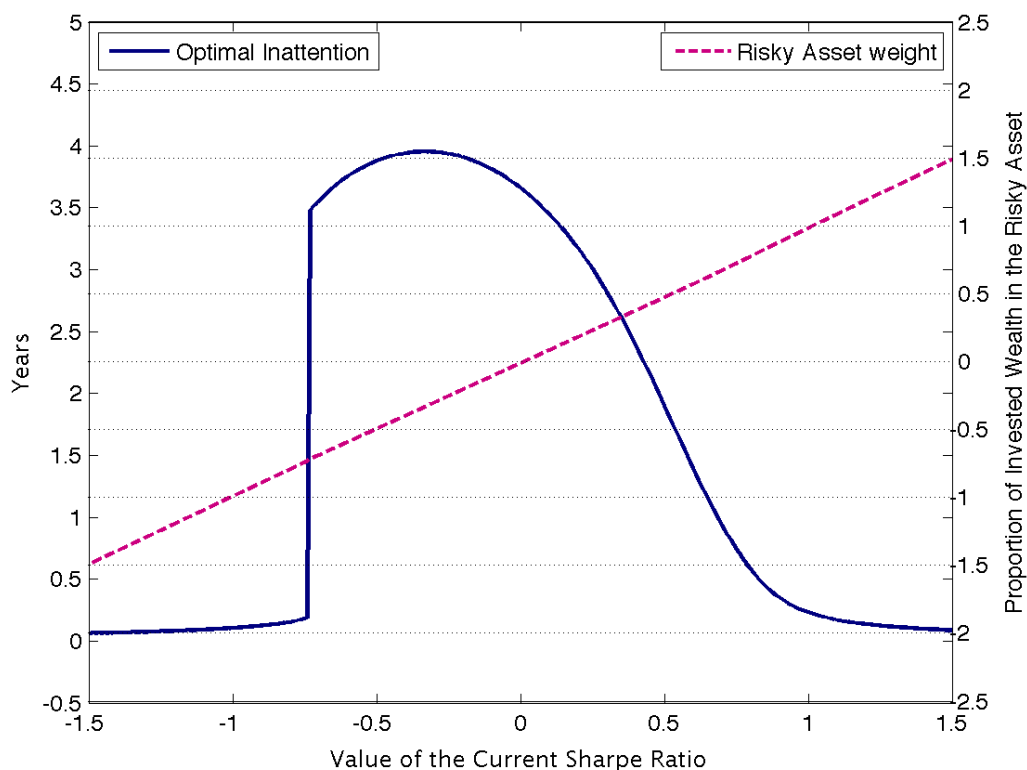
Figure 1.2: This figure plots the optimal inattention span (continuous line) and the optimal risky asset allocation (dashed line) for different values of the Sharpe ratio M. When the expected returns are high, the agent's optimal strategy is to allocate more than 100% of his invested wealth in the risky asset, deposit very little in the transactions account and observe the value of the investment portfolio very frequently. The same holds when the expected Sharpe ratio is very high: the agent shorts the risky asset and reaps the benefits of the negative expected returns. In sluggish economic environments characterized by a Sharpe ratio around zero, the agent's optimal portfolio choice is to invest very little in the risky asset. This translates in a low opportunity cost for the transactions account and a very long period of optimal inattention. The x-axis tracks the variable under study and the y-axis reports the years of inattention on the left and the proportion of invested wealth in the risky asset on the right.

(a) Persistence: λ_M

(b) Volatility: σ_M

(c) Correlation: ρ_{sM}

(d) Information Costs: θ

(e) Risk-Aversion: α

(f) Subjective Discount Factor: $1 - \rho$
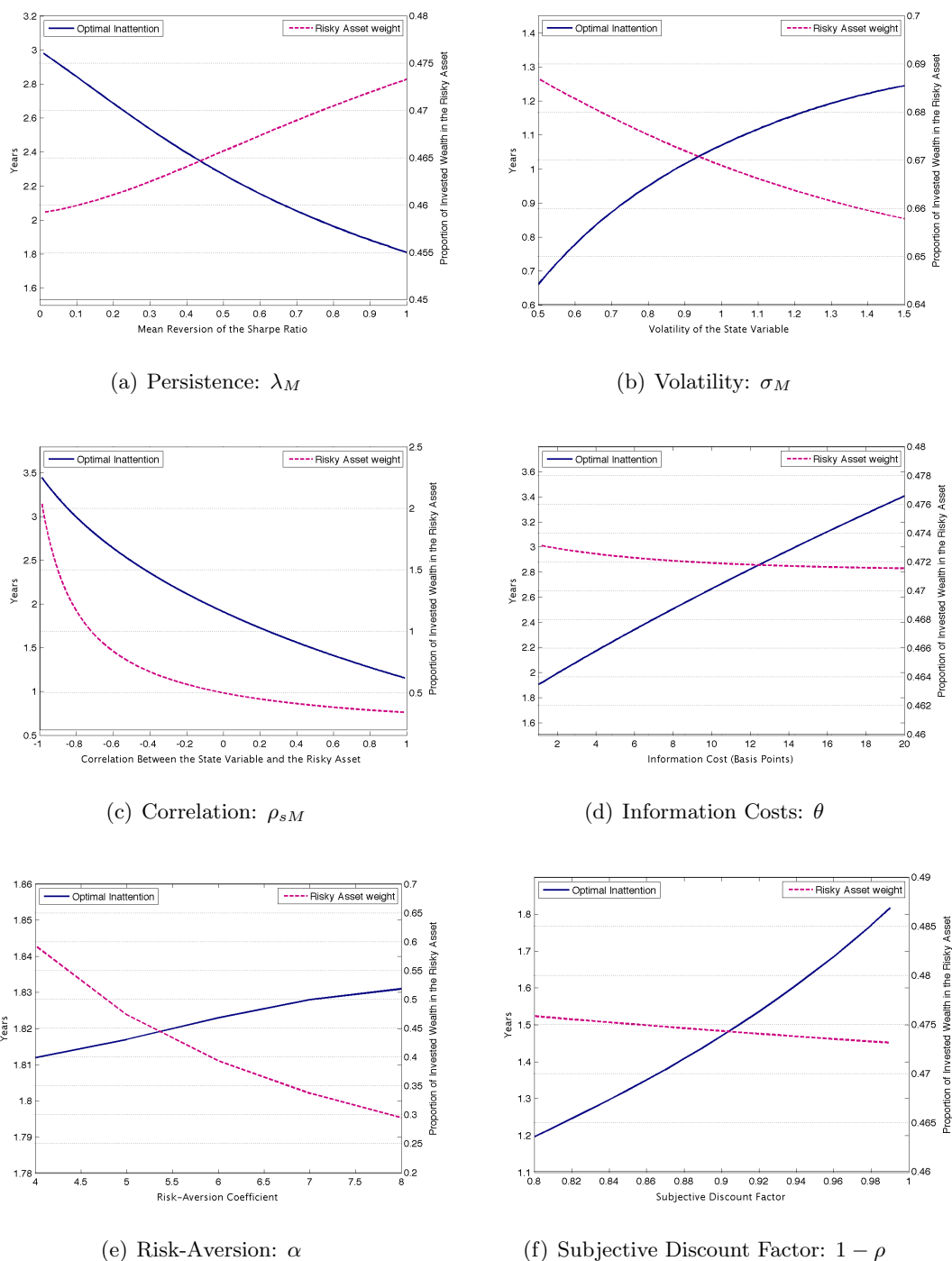
Figure 1.3: This figure plots the optimal inattention interval (continuous line) and the optimal risky asset allocation (dashed line) as a function of the persistence parameter for the Sharpe ratio process λ_M (Panel (a)); its volatility parameter σ_M (Panel (b)); its correlation coefficient with the risky asset ρ_{sM} (Panel (c)); the investor's information costs θ; the investor's risk-aversion coefficient α; and the investor's subjective discount factor $1 - \rho$. In all panels, the x-axis tracks the variable under study and the y-axis reports the years of inattention on the left and the proportion of invested wealth in the risky asset on the right.
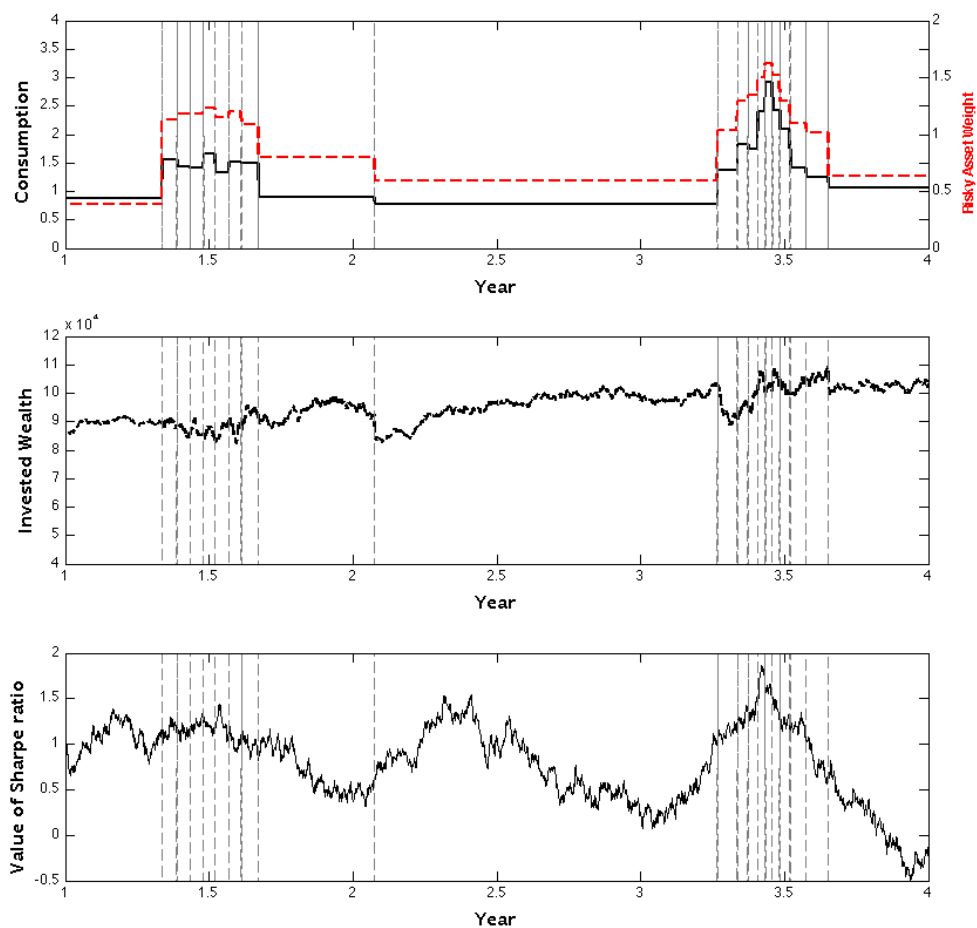
Figure 1.4: This figure plots the portfolio allocation and trading activity for the representative investor. The top panel reports the evolution of the agent's normalized consumption and portfolio allocation. The consumption process is represented by the black continuous line and the portfolio allocation by the red dashed line. The y-axis on the left reports values for the consumption, while the one on the right reports the weight of the risky asset in the investment portfolio. The middle panel presents the process for the agent's invested wealth and the bottom panel plots the simulated process for the Sharpe ratio. The vertical lines in each of the panels represent the agent's trading activity: every period at which the agent changes his portfolio allocation is denoted by a vertical line.
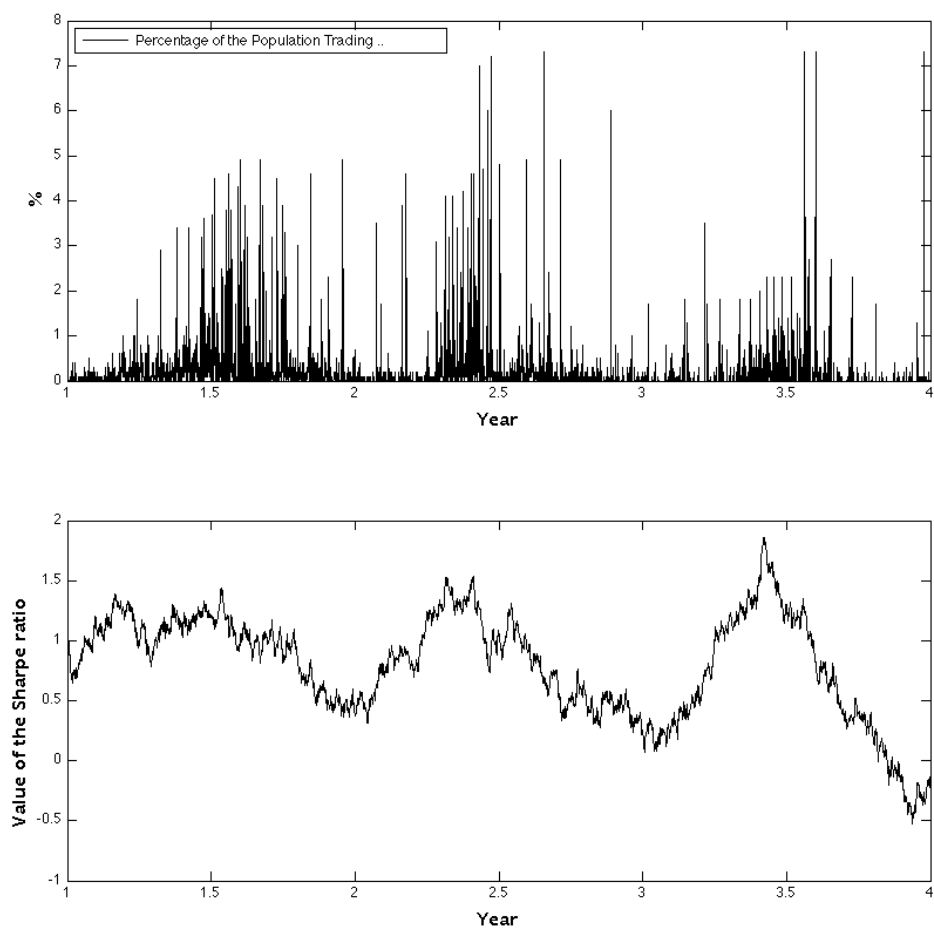
Figure 1.5: This figure plots the trading activity for an economy composed of 1000 agents. The top panel reports the percentage of the population trading at a given point in time. The bottom panel plots the value of the Sharpe ratio over time.
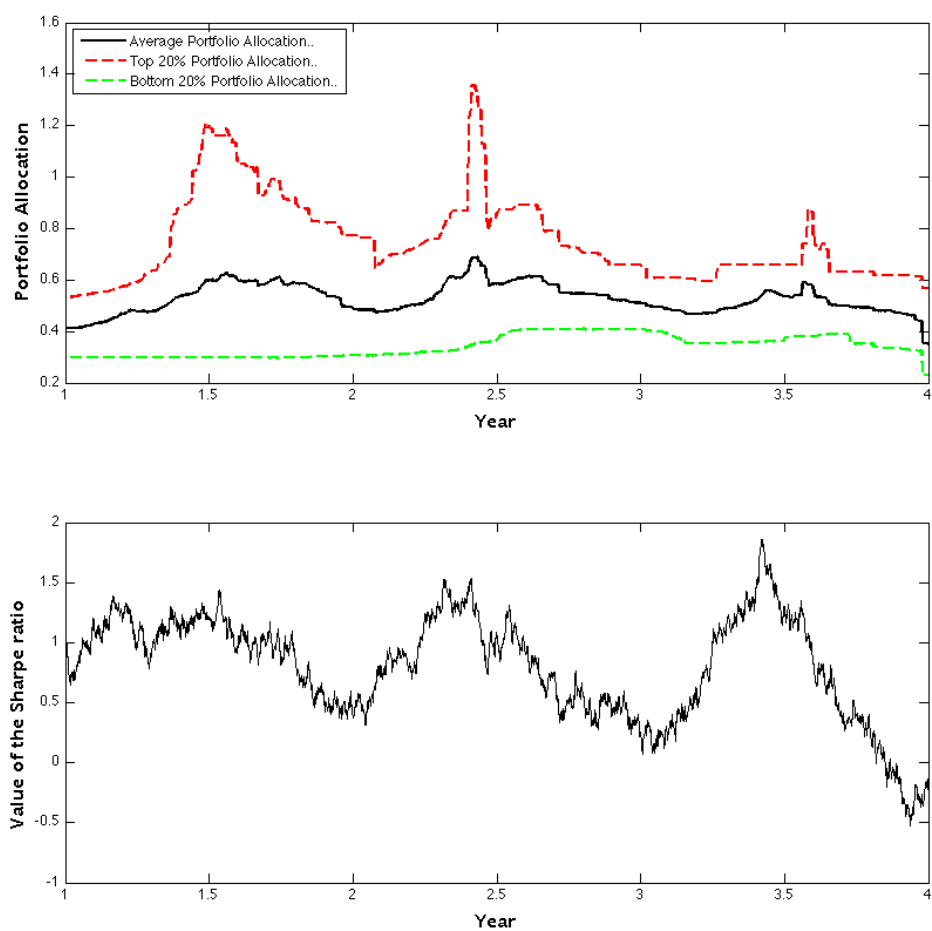
Figure 1.6: This figure plots the dispersion of portfolio allocations across investors for an economy composed of 1000 agents. The top panel reports average portfolio allocation as well as the 20th and 80th percentiles of the portfolio allocations for a given period. The bottom panel plots the value of the Sharpe ratio over time.

(a) Horizon: 1 month   (b) Horizon: 2 months   (c) Horizon: 3 months

(d) Horizon: 6 months   (e) Horizon: 9 months   (f) Horizon: 12 months
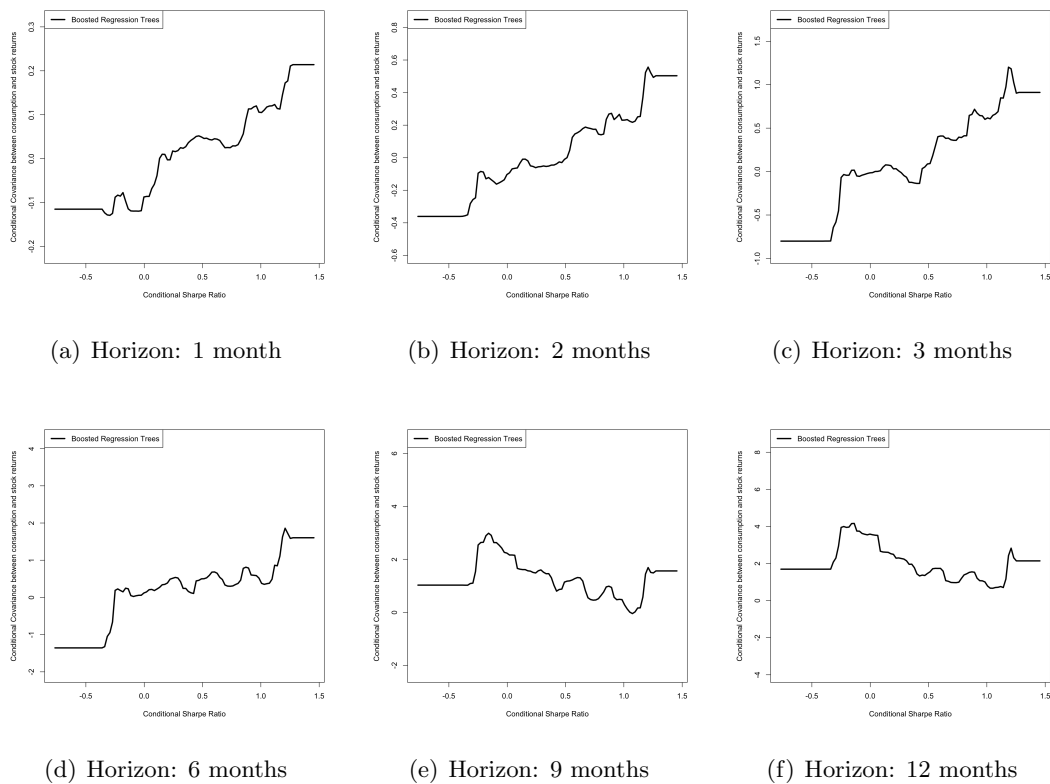
Figure 1.7: These figures represent non-parametric estimates of the relation between the conditional covariance between consumption and stock returns and the conditional Sharpe ratio. The estimates are obtained via Boosted Regression Trees over the sample 1960-2008 and we report results at 1, 2, 3, 6, 9 and 12 months horizons. The horizontal axis covers the sample support of the conditional Sharpe ratio, while the vertical axis tracks the change in the conditional covariance.
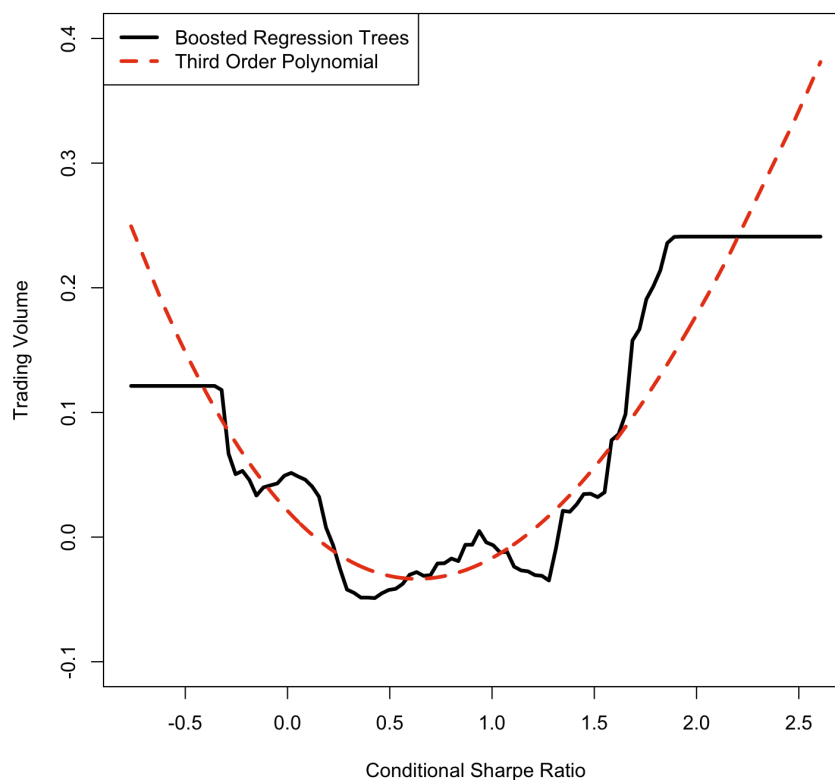
Figure 1.8: This figure plots the fitted values of the aggregate trading volume on the S&P 500 index as a function of the conditional Sharpe ratio. The conditional Sharpe ratio is estimated using Boosted Regression Trees over the sample 1960-2008. The fitted values reported are obtained either by boosted regression trees (black line) or a third-order polynomial model estimated via ordinary least squares.

Appendix

Proof of Proposition 1: For any given amount of wealth X_{t_j} deposited in the transactions account at time t_j and any inattention period of length τ, the agent chooses consumption to solve the following constraint optimization problem.

$$U_{t_j}(\tau) \equiv \max_{\{c_{t+s}\}_{s=0}^{\tau}} \int_0^{\tau} \frac{1}{1-\alpha} c_{t_j+s}^{(1-\alpha)} e^{-\rho s} \, ds \tag{A-1}$$

$$s.t. \quad X_{t_j}(\tau) = \int_0^{\tau} c_{t_j+s} e^{-r^L s} ds, \tag{A-2}$$

where Eq. A-2 is the budget constraint. The budget constraint requires the present value of consumption discounted at rate r^L over the inattention period τ to equal the sum deposited in the transactions account X_{t_j} at t_j. The inter-temporal marginal rate of substitution between consumption at t_j and $t_j + s$ is $\left(\frac{c_{t_j+s}}{c_{t_j}}\right)^{-\alpha} e^{-\rho s}$ and the gross rate of return is $e^{(r^L s)}$. Equating the two we obtain an expression for the evolution of consumption over the interval of inattention:

$$c_{t_j+s} = c_{t_j} \, e^{-\frac{(\rho-r^L)}{\alpha}s}, \quad for \ \ 0 \le s \le \tau, \tag{A-3}$$

Substitute Eq. A-3 into Eq. A-2 to obtain:

$$\begin{aligned}
X_{t_j}(\tau) &= \int_0^{\tau} c_{t_j} e^{-\frac{\rho-r^L}{\alpha}s} e^{-r^L s} ds \\
&= c_{t_j} \int_0^{\tau} exp\left[-\frac{\rho-(1-\alpha)r^L}{\alpha}s\right] ds \\
&= c_{t_j} \int_0^{\tau} e^{-\omega s} ds \\
&= c_{t_j} \left[-\frac{e^{\omega\tau}}{\omega} + \frac{e^{\omega 0}}{\omega}\right] \\
&= c_{t_j} \left[\frac{1-e^{\omega\tau}}{\omega}\right] \\
&= c_{t_j} h(\tau),
\end{aligned}$$

where

$$h(\tau) = \int_0^{\tau} e^{-\omega s} ds = \frac{1-e^{-\omega\tau}}{\omega} \tag{A-4}$$

$$\omega = \frac{\rho-(1-\alpha)r^L}{\alpha}. \tag{A-5}$$

Now substitute Eq. A-3 into Eq. A-1 and use Eq. A-4 to write the maximized utility of consumption as:

$$
\begin{aligned}
U_{t_j}(\tau) &= \int_0^\tau \frac{1}{1-\alpha} c^{*(1-\alpha)}_{t_j+s}\, e^{-\rho s}\, ds \\
&= \int_0^\tau \frac{1}{1-\alpha} \left(c_{t_j}\, e^{[-\frac{(\rho-r^L)}{\alpha} s]} \right)^{(1-\alpha)} e^{-\rho s}\, ds \\
&= \int_0^\tau \frac{1}{1-\alpha} c^{(1-\alpha)}_{t_j}\, e^{-\frac{(\rho-r^L)}{\alpha} s}\, e^{(\rho-r^L)s}\, e^{-\rho s}\, ds \\
&= \int_0^\tau \frac{1}{1-\alpha} c^{(1-\alpha)}_{t_j}\, e^{-\frac{\rho-(1-\alpha)r^L}{\alpha} s}\, e^{\rho s} e^{-\rho s}\, ds \\
&= \frac{1}{1-\alpha} c^{(1-\alpha)}_{t_j} \int_0^\tau e^{-\frac{\rho-(1-\alpha)r^L}{\alpha} s}\, ds \\
&= \frac{1}{1-\alpha} c^{(1-\alpha)}_{t_j} h(\tau) \\
&= \frac{1}{1-\alpha} X^{(1-\alpha)}_{t_j} h(\tau)^\alpha
\end{aligned}
$$

This is the form stated in Proposition 1. ∎

Proof of Proposition 2: We follow Kim and Omberg (1996) and obtain a closed form solution for the optimal expected utility and portfolio allocation. Take τ to be the investment horizon. Denote the wealth in the investment portfolio as W_I, the optimal monetary investment in the risky asset as $y^*(W_I, M, \tau)$ and the optimal expected utility as $J(W_I, M, \tau)$. Since $J(W_I, M, \tau)$ is the investor's expected utility assuming optimal investment over the remaining horizon, it satisfies the dynamic programming condition:

$$
\begin{aligned}
J(W_I, M, \tau) = \max_{y(W_I, M, \tau)} (J - J_\tau dt &+ J_{W_I} E\left\{dW_I\right\} + \frac{1}{2} J_{W_I W_I} E\left\{dW_I^2\right\} \qquad \text{(A-6)} \\
&+ J_M E\left\{dM\right\} + \frac{1}{2} J_{MM} E\left\{dM^2\right\} J_{W_I M} E\left\{dM dW_I\right\}),
\end{aligned}
$$

while the condition for the optimal monetary investment in the risky asset should satisfy:

$$
y^*(W_I, M, \tau) = \left(\frac{J_{W_I}}{-J_{W_I W_I}} \right) \left(\frac{\mu_t - r}{\sigma_t^2} \right) + \left(\frac{J_{W_I M}}{-J_{W_I W_I}} \right) \left(\frac{\rho_{sM} \sigma_M}{\sigma_t} \right) \qquad \text{(A-7)}
$$

The optimal expected utility $J(W_I, M, \tau)$ is the solution to the partial differential equation and horizon-end condition:

$$
-J_\tau + J_{W_I} r W_I + \frac{1}{2} J_{W_I W_I} (y^*)^2 \sigma^2 - J_M \lambda_M (M - \bar{M}) + \frac{1}{2} J_{MM} \sigma_M^2 = 0 \qquad \text{(A-8)}
$$

$$
J(W_I, M, 0) = U(W_I) \qquad \text{(A-9)}
$$

For the myopic case, it can be shown that the solution is additively-separable in W_I and M. For the non-logarithmic case, conjecture trial solutions of the form:

$$J(W_I, M, \tau) = \Phi(M, \tau)\frac{W_I^{1-\alpha}}{1-\alpha}e^{r\tau(1-\alpha)} \qquad \text{(A-10)}$$

$$\Phi(M, \tau) = exp[A(\tau) + B(\tau)M + C(\tau)M^2/2] \qquad \text{(A-11)}$$

$$A(0) = B(0) = C(0) = 0. \qquad \text{(A-12)}$$

Note that the solution is multiplicatively separable in W_I and M, and automatically satisfies the horizon-end condition $J(W_I, M, 0) = U(W_I)$ and the second-order condition for a maximum $J_{W_I W_I}(W_I, M, \tau) < 0$. Substituting the trial solutions in Eq. A-7 and Eq. A-8 yields the following expression for the optimal monetary investment in the risky asset:

$$y^*(W_I, M, \tau) = \frac{W_I}{\alpha}\left(\frac{M}{\sigma} + \frac{\rho_{sM}\sigma_M C(\tau)M + \rho_{sM}\sigma_M B(\tau)}{\sigma}\right) \qquad \text{(A-13)}$$

and the partial differential equation for $\Phi(M, \tau)$

$$-\Phi_\tau + \left(\frac{1-\alpha}{2\alpha}\right)\left(M + \frac{\Phi_M}{\Phi}\rho_{sM}\sigma_M\right)^2\Phi - \lambda_M(M - \bar{M})\Phi_M + \frac{1}{2}\Phi_{MM}\sigma_M^2 = 0 \quad \text{(A-14)}$$

The investment in the risky asset depends only on the functions $B(\tau)$ and $C(\tau)$, while expected utility depends on $A(\tau)$ as well. The functions $C(\tau)$, $B(\tau)$ and $A(\tau)$ should be chosen to satisfy Eqs. $A-11$, $A-12$ and $A-14$. Substitute and collect terms to obtain the quadratic expression in M:

$$\left\{\frac{1}{2}\sigma_M^2 C^2(\tau) - \lambda_M C(\tau) - \frac{1}{2}C'(\tau) + \frac{(1-\alpha)}{2\alpha}[1 + \rho_{sM}^2\sigma_M^2 C^2(\tau) + 2\rho_{sM}\sigma_M C(\tau)]\right\}M^2 \qquad \text{(A-15)}$$

$$+ \left\{\sigma_M^2 B(\tau)C(\tau) - \lambda_M B(\tau) + \lambda_M \bar{M}C(\tau) - B'(\tau) + \frac{(1-\alpha)}{\alpha}\rho_{sM}\sigma_M B(\tau)[1 + \rho_{sM}\sigma_M C(\tau)]\right\}M$$

$$+ \frac{1}{2}\sigma_M^2[B^2(\tau) + C(\tau)] + \lambda_M \bar{M}B(\tau) - A'(\tau) + \frac{(1-\alpha)}{2\alpha}\rho_{sM}^2\sigma_M^2 B^2(\tau) = 0,$$

where the "primes" indicate first derivatives with respect to the horizon τ. In order for the solution to hold for any Sharpe ratio M, the coefficients on M^2, M and 1 must vanish. The result is the following system of first-order non-linear ordinary differential equations:

$$\frac{dC}{d\tau} = c\,C^2(\tau) + b\,C(\tau) + a \qquad \text{(A-16)}$$

$$\frac{dB}{d\tau} = cB(\tau)C(\tau) + \frac{b}{2}B(\tau) + \lambda_M \bar{M}C(\tau) \qquad \text{(A-17)}$$

$$\frac{dA}{d\tau} = \frac{c}{2}B^2(\tau) + \frac{1}{2}\sigma_M^2 C(\tau) + \lambda_M \bar{M}B(\tau) \qquad \text{(A-18)}$$

$$C(0) = B(0) = A(0) = 0,$$

where

$$a = \frac{1-\alpha}{\alpha} \tag{A-19}$$

$$b = 2\left[\left(\frac{1-\alpha}{\alpha}\right)\rho_{sM}\sigma_M - \lambda_M\right] \tag{A-20}$$

$$c = \sigma_M^2\left[1 + \left(\frac{1-\alpha}{\alpha}\right)\rho_{sM}^2\right]. \tag{A-21}$$

The first differential equation involves only $C(\tau)$, the second $C(\tau)$ and $B(\tau)$ and the third $C(\tau)$, $B(\tau)$ and $A(\tau)$. The first is a Riccati equation with constant coefficients and we can re-cast it as the integral equation:

$$\int_0^\tau = \frac{dC}{cC^2 + bC + a} = \tau. \tag{A-22}$$

The left hand side can be determined from standard integral tables, which indicates that the solution can take 4 different forms depending on the parameters a, b and c. The four solutions to the set of three functions $C(\tau)$, $B(\tau)$ and $A(\tau)$ can be categorized using the discriminant

$$q = b^2 - 4ac = 4\lambda_M^2\left[1 - (\gamma - 1)k^*\right] \tag{A-23}$$

of the quadratic equation:

$$\frac{dC}{d\tau} = cC^2(\tau) + bC(\tau) + a = 0. \tag{A-24}$$

The four solutions are:

1. $q > 0$: **normal solution**

2. $q = 0$: **hyperbolic solution** if $b \neq 0$ and **polynomial solution** if $b = 0$

3. $q < 0$: **tangent solution**

The only solution that is interesting to us from an economic point of view is the "normal solution". The analytical expressions for $C(\tau)$, $B(\tau)$ and $A(\tau)$ for the normal solution

case are:

$$C(\tau) = \frac{2\left(\frac{1-\alpha}{\alpha}\right)(1 - e^{-\eta\tau})}{2\eta - (b+\eta)(1 - e^{-\eta\tau})}$$

$$B(\tau) = \frac{4\left(\frac{1-\alpha}{\alpha}\right)\lambda_M \bar{M}(1 - e^{-\eta\tau/2})^2}{\eta[2\eta - (b+\eta)(1 - e^{-\eta\tau})]}$$

$$A(\tau) = \left(\frac{1-\alpha}{\alpha}\right)\left(\frac{2\lambda_M^2 \bar{M}^2}{\eta^2} + \frac{\sigma_M^2}{\eta - b}\right)\tau$$

$$+ \frac{4\left(\frac{1-\alpha}{\alpha}\right)\lambda_M^2 \bar{M}^2 \left[(2b+\eta)e^{-\eta\tau} - 4be^{-\eta\tau/2} + 2b - \eta\right]}{\eta^3 \left[2\eta - (b+\eta)(1 - e^{-\eta\tau})\right]}$$

$$+ \frac{2\left(\frac{1-\alpha}{\alpha}\right)\sigma_M^2}{\eta^2 - b^2}\ln\left|\frac{2\eta - (b+\eta)(1 - e^{-\eta\tau})}{2\eta}\right|,$$

where

$$\eta = \sqrt{4\lambda_M^2\left[1 - \left(\frac{1-\alpha}{\alpha}\right)k^*\right]}$$

$$k^* = \left(\frac{\sigma_M}{\lambda_M}\right)^2 + 2\rho_{sM}\left(\frac{\sigma_M}{\lambda_M}\right) \qquad\blacksquare$$

Proof of Proposition 3: Start from the Ornstein-Uhlenbeck process for the Sharpe ratio dynamics:

$$dM = -\lambda_M(M - \bar{M}) + \sigma_M dZ_M,$$

which has the integrated form, from t_j to $t_j + \tau$, equal to:

$$M(t_j, t_j + \tau) = \bar{M} + \left(M_{t_j} - \bar{M}\right)e^{-\lambda_M\tau} + \sigma_M \int_{t_j}^{t_j+\tau} e^{\lambda_M(\tau-u)}dZ_M(u).$$

Following Karatzas and Shreve (1991), the future risk premium is normally distributed with mean, variance, auto-covariance and covariance with the risky-asset's return equal to:

$$E\{M(t_j, t_j + \tau)\} = \bar{M} + (M_{t_j} - \bar{M})e^{\lambda_M\tau}$$

$$Var\{M(t_j, t_j + \tau)\} = \left(\frac{\sigma_M^2}{2\lambda_M}\right)(1 - e^{-2\lambda_M\tau})$$

$$Cov\{M(u), M(v)\} = \left(\frac{\sigma_M^2}{2\lambda_M}\right)\left(e^{-\lambda_M|u-v|} - e^{-\lambda_M(u+v)}\right)$$

$$Cov\{M(u), dZ(v)\} = \begin{cases} \rho_{sM}\sigma_M e^{-\lambda_M(u+v)} & \text{for } u < v \\ 0 & \text{for } u > v \end{cases}$$

Recall now the expression for the normalized return process:

$$dR_t = M_t \, dt + dZ.$$

Its expected value from t_j to $t_j + \tau$:

$$E\{R(t_j, t_j + \tau)\} = \bar{M} + (M_{t_j} - \bar{M})e^{\lambda_M \tau}$$

while its variance is

$$Var\{R(t_j, t_j + \tau)\} = \tau + \int_{t_j}^{t_j+\tau} \int_{t_j}^{t_j+\tau} Cov(X(u), X(v)) du \, dv$$

$$+ \int_{t_j}^{t_j+\tau} \int_{t_j}^{t_j+\tau} Cov[X(u), dZ(v)] du$$

Using the expressions reported above we can re-write this last expression as:

$$Var\{(R(t_j, t_j + \tau)\} = \tau + \left(\frac{\sigma_M}{\lambda_M}\right)^2 \left(\tau + \frac{2e^{-\lambda_M \tau}}{\lambda_M} - \frac{e^{-2\lambda_M \tau}}{2\lambda_M} - \frac{3}{2\lambda_M}\right)$$

$$+ \left(\frac{2\rho_{sM}\sigma_M}{\lambda_M}\right)\left(\tau + \frac{e^{-\lambda_M \tau}}{\lambda_M} - \frac{1}{\lambda_M}\right) \qquad \blacksquare$$

Proof of Theorem 1: Given the Sharpe ratio value M_{t_j}, wealth W_{t_j} and the investment horizon τ, the value function has the following form:

$$\frac{1}{1-\alpha}\gamma W_{t_j}^{1-\alpha} = \max_{X_{t_j}} \frac{1}{1-\alpha} X_{t_j}^{1-\alpha}[h(\tau)]^{\alpha} \tag{A-25}$$

$$+ \frac{1}{1-\alpha}\gamma (W_{t_j} - X_{t_j})^{1-\alpha}\chi^{\alpha} e^{\left[(r(1-\alpha)-\rho)\tau + A(\tau)+B(\tau)M+C(\tau)\frac{M^2}{2} \mid M=M_{t_j}\right]}$$

where

$$\chi \equiv (1-\theta)^{\frac{1-\alpha}{\alpha}}. \tag{A-26}$$

To obtain the optimal amount of wealth deposited in the transactions account $X_{t_j}^*$, differentiate Eq. A-25 w.r.t X_{t_j} and set it equal to zero

$$X^*(M_{t_j}, \tau) = h(\tau) \, \gamma^{-\frac{1}{\alpha}} \chi^{-1}(W_{t_j} - X_{t_j}^*) \, e^{\left[\frac{1}{\alpha}\left((\rho - r(1-\alpha))\tau - A(\tau) - B(\tau)M - C(\tau)\frac{M^2}{2}\right)\mid M=M_{t_j}\right]}. \tag{A-27}$$

Now define:

$$S(M_{t_j}, \tau) \equiv \left[\frac{1}{\alpha}\left((\rho - r(1-\alpha))\tau - A(\tau) - B(\tau)M - C(\tau)\frac{M^2}{2}\right) \mid M = M_{t_j}\right] \tag{A-28}$$

$$K(M_{t_j}, \tau) \equiv h(\tau) \, \gamma^{-\frac{1}{\alpha}}\chi^{-1}e^{S(M_{t_j},\tau)} \tag{A-29}$$

and use Eq. A-29 to re-write Eq. A-27 as:

$$X^*_{t_j} = \frac{K(M_{t_j},\tau)}{K(M_{t_j},\tau)+1}\, W_{t_j}. \tag{A-30}$$

Plug this last expression into Eq. A-25 together with Eq. A-28 and solve for γ to obtain:

$$\gamma(M_{t_j},\tau) = \left(\frac{K(M_{t_j},\tau)}{K(M_{t_j},\tau)+1}\right)^{1-\alpha} [h(\tau)]^\alpha + \gamma(M_{t_j},\tau)\left(\frac{1}{K(M_{t_j},\tau)+1}\right)^{1-\alpha}\chi^\alpha e^{-\alpha S(M_{t_j},\tau)}. \tag{A-31}$$

Equations (A-31) and (A-29) constitute a system of two equations and two unknowns, conditioning on the value of the current Sharpe ratio M_{t_j}. The solution is found by substituting $K(M_{t_j},\tau)$ into $\gamma(M_{t_j},\tau)$:[17]

$$\gamma = \frac{\left(\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^{1-\alpha} h^\alpha}{\left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^{1-\alpha}} + \frac{\gamma}{\left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^{1-\alpha}}\chi^\alpha e^{-\alpha S}$$

$$\gamma = \frac{\left(\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^{1-\alpha} h^\alpha + \gamma\chi^\alpha e^{-\alpha S}}{\left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^{1-\alpha}}$$

$$\gamma = \frac{\gamma\left[\gamma^{-\frac{1}{\alpha}} h^{1-\alpha}\chi^{\alpha-1}e^S e^{-\alpha S}h^\alpha + \chi^\alpha e^{-\alpha S}\right]}{\left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^{1-\alpha}}$$

$$\gamma = \frac{\gamma e^{-\alpha S}\chi^\alpha\left[\gamma^{-\frac{1}{\alpha}}\chi^{-1}e^S h + 1\right]}{\left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^{1-\alpha}}.$$

Which can be re-written as:

$$\gamma = \frac{\gamma e^{-\alpha S}\chi^\alpha}{\left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^{-\alpha}}$$

$$1 = \frac{e^{-S\alpha}\chi^\alpha}{\left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^{-\alpha}}$$

$$1 = \left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^\alpha e^{-S\alpha}\chi^\alpha$$

$$e^{S\alpha}\chi^{-\alpha} = \left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)^\alpha$$

$$e^S\chi^{-1} = \left(1+\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)$$

$$e^S\chi^{-1} - 1 = \left(\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right)$$

$$\left(\gamma^{-\frac{1}{\alpha}} h\, \chi^{-1}e^S\right) = e^S\chi^{-1} - 1.$$

---

[17]We temporarily suppress the dependence on M_{t_j} and τ to ease the notation

Solve the expression above for γ to obtain:

$$\gamma^{-\frac{1}{\alpha}} = \frac{e^S \chi^{-1} - 1}{h \, \chi^{-1} e^S}$$

$$\gamma = \left(\frac{e^S \chi^{-1} - 1}{h \, \chi^{-1} e^S}\right)^{-\alpha}$$

$$\gamma = \left(\frac{h \, \chi^{-1} e^S}{e^S \chi^{-1} - 1}\right)^{\alpha}$$

$$\gamma = \left(1 - e^{-\omega\tau}\right)^{\alpha} \omega^{-\alpha} \left(\frac{e^S \chi^{-1}}{e^S \chi^{-1} - 1}\right)^{\alpha}$$

$$\gamma = \left(\frac{1 - e^{-\omega\tau}}{1 - \chi e^{-S}}\right)^{\alpha} \omega^{-\alpha}$$

Replace the expression for γ derived above into K to obtain:

$$K = \gamma^{-\frac{1}{\alpha}} h \chi^{-1} e^S$$

$$= \left(\frac{1 - \chi e^{-S}}{1 - e^{-\omega\tau}}\right) \omega \frac{(1 - e^{-\omega\tau})}{\omega} \chi^{-1} e^S$$

$$= \left(1 - \chi e^{-S}\right) \chi^{-1} e^S$$

$$= \chi^{-1} e^S - 1.$$

Hence, the solution to the system of equations is:

$$\gamma(M_{t_j}, \tau) = \left(\frac{1 - e^{-\omega\tau}}{1 - \chi e^{-S(M_{t_j}, \tau)}}\right)^{\alpha} \omega^{-\alpha}$$

$$K(M_{t_j}, \tau) = \chi^{-1} e^{S(M_{t_j}, \tau)} - 1. \qquad \blacksquare$$

Bibliography

ABEL, A., J. EBERLY, AND S. PANAGEAS (2007): "Optimal inattention to the stock market," *American Economic Review*, 97(2), 244–249.

——— (2009): "Optimal Inattention to the Stock Market with Information Costs and Transactions Costs," Working Paper.

AMERIKS, J., AND S. ZELDES (2002): "How do household portfolio shares vary with age," *TIAA-CREF working paper*, pp. 6–120101.

BARBER, B., AND T. ODEAN (2000): "Trading is hazardous to your wealth: The common stock investment performance of individual investors," *The Journal of Finance*, 55(2), 773–806.

BAUMOL, W. (1952): "The transactions demand for cash: An inventory theoretic approach," *The Quarterly Journal of Economics*, 66(4), 545–556.

BOUDOUKH, J., M. RICHARDSON, AND R. WHITELAW (1997): "Nonlinearities in the relation between the equity risk premium and the term structure," *Management Science*, 43(3), 371–385.

BRENNAN, M. (1998): "The role of learning in dynamic portfolio decisions," *Review of Finance*, 1(3), 295.

BRENNAN, M., AND Y. XIA (2000): "Stochastic interest rates and the bond-stock mix," *Review of Finance*, 4(2), 197.

——— (2001): "Assessing asset pricing anomalies," *Review of Financial Studies*, 14(4), 905.

——— (2002): "Dynamic asset allocation under inflation," *The journal of finance*, 57(3), 1201–1238.

BRUNNERMEIER, M., AND S. NAGEL (2008): "Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals' asset allocation," *The American Economic Review*, 98(3), 713–736.

CARROLL, C. (2003): "Macroeconomic Expectations Of Households And Professional Forecasters*," *Quarterly Journal of Economics*, 118(1), 269–298.

COOLEY, T., AND G. HANSEN (1989): "The inflation tax in a real business cycle model," *The American Economic Review*, 79(4), 733–748.

DUFFIE, D., AND T. SUN (1990): "Transactions Costs and Portfolio Choice in a Discrete-Continuous-Time Setting.," *Journal of Economic Dynamics and Control*, 14(1), 35–51.

FAIG, M., AND P. SHUM (2002): "Portfolio choice in the presence of personal illiquid projects," *The Journal of Finance*, 57(1), 303–328.

GABAIX, X., AND D. LAIBSON (2002): "The 6D Bias and the Equity-Premium Puzzle," *NBER macroeconomics annual*, 16, 257–312.

HUBERMAN, G., AND W. JIANG (2006): "Offering versus choice in 401 (k) plans: Equity exposure and number of funds," *The Journal of Finance*, 61(2), 763–801.

JOHNSON, W. (2004): "Predictable investment horizons and wealth transfers among mutual fund shareholders," *The Journal of Finance*, 59(5), 1979–2012.

——— (2010): "Do investors trade uniformly through time?," *Journal of Empirical Finance.*

KANDEL, S., AND R. STAMBAUGH (1990): "Expectations and volatility of consumption and asset returns," *Review of Financial Studies*, pp. 207–232.

KARATZAS, I., AND S. SHREVE (1991): *Brownian motion and stochastic calculus.* Springer.

KIM, T., AND E. OMBERG (1996): "Dynamic nonmyopic portfolio behavior," *Review of Financial Studies*, pp. 141–161.

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

LIU, J. (2007): "Portfolio selection in stochastic environments," *Review of Financial Studies*, 20(1), 1.

LYNCH, A. (1996): "Decision frequency and synchronization across agents: Implications for aggregate consumption and equity return," *The Journal of Finance*, 51(4), 1479–1497.

MANKIW, N., R. REIS, AND J. WOLFERS (2003): "Disagreement about inflation expectations," *NBER Macroeconomics Annual*, 18, 209–248.

MERTON, R. (1969): "Lifetime portfolio selection under uncertainty: The continuous-time case," *The Review of Economics and Statistics*, 51(3), 247–257.

MOSCARINI, G. (2004): "Limited information capacity as a source of inertia," *Journal of Economic Dynamics and Control*, 28(10), 2003–2035.

ODEAN, T. (1998): "Are investors reluctant to realize their losses?," *The Journal of Finance*, 53(5), 1775–1798.

——— (1999): "Do investors trade too much?," *American Economic Review*, 89(5), 1279–1298.

PEREZ-QUIROS, G., AND A. TIMMERMANN (2000): "Firm size and cyclical variations in stock returns," *The Journal of Finance*, 55(3), 1229–1262.

ROSSI, A. (2010): "Financial Markets Prediction. Can Boosted Regression Trees Benefit the Mean-Variance Investor?," *Unpublished Manuscipt, University of California San Diego.*

ROSSI, A., AND A. TIMMERMANN (2010): "What is the Shape of the Risk-Return Relation?," *UCSD Discussion Paper*.

SANGVINATSOS, A., AND J. WACHTER (2005): "Does the Failure of the Expectations Hypothesis Matter for Long-Term Investors?," *Journal of Finance*, pp. 179–230.

SIMS, C. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50(3), 665–690.

TOBIN, J. (1956): "The interest-elasticity of transactions demand for cash," *The Review of Economics and Statistics*, 38(3), 241–247.

VISSING-JORGENSEN, A. (2002): "Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures," *NBER working paper*.

WACHTER, J. (2003): "Risk aversion and allocation to long-term bonds," *Journal of Economic Theory*, 112(2), 325–333.

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

WHITELAW, R. (1997): "Time variation Sharpe ratios and market timing," *Unpublished working paper. New York University*.

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

2 Asset Allocation with High-Dimensional Information Sets

Abstract

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

2.1 Introduction

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

Our paper contributes to the long-standing literature assessing the predictability of stock returns. Over the nineties and the beginning of the twenty-first century the combination of longer time-series and greater statistical sophistication have spurred a large number of attempts to add evidence for or against the predictability of asset returns and volatility. In-sample statistical tests show a high degree of predictability for a number of variables: Rozeff (1984), Fama and French (1988), Campbell and Shiller (1988a,b), Kothari and Shanken (1997) and Pontiff and Schall (1998) find that valuation

ratios predict stock returns, particularly so at long horizons; Fama and Schwert (1977), Keim and Stambaugh (1986), Campbell (1987), Fama and French (1989), Hodrick (1992) show that short and long-term treasury and corporate bonds explain variations in stock returns; Lamont (1998), Baker and Wurgler (2000) show that variables related to aggregate corporate payout and financing activity are useful predictors as well. While these results are generally encouraging, there are a number of doubts regarding their accuracy as most of the regressors considered are very persistent, making statistical inference less than straightforward; see, for example, Nelson and Kim (1993), Stambaugh (1999), Campbell and Yogo (2006) and Lewellen, Nagel, and Shanken (2010). Furthermore, data snooping may be a source of concern if researchers are testing for many different model specifications and report only the statistically significant ones; see, for example, Lo and MacKinlay (1990), Bossaerts and Hillion (1999) and Sullivan, Timmermann, and White (1999). While it is sometimes possible to correct for specific biases, no procedure can offer full resolution of the shortcomings that affect the in-sample estimates.

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

The results arising from the out-of-sample studies are mixed and depend heavily on the model specification and the conditioning variables employed.[1] In particular, many of the studies conducted so far are characterized by one or more of these limitations. First, the forecasts are generally formulated using simple linear regressions. The choice is dictated by simplicity and the implicit belief that common functional relations can be

---

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

approximated reasonably well by linear ones.[2] Most asset pricing theories underlying the empirical tests, however, do not imply linear relationships between the equity premium and the predictor variables, raising the issue whether the mis-specification implied by linear regressions is economically large. Second, linear models overfit the training dataset and generalize poorly out-of-sample as the number of regressors increases, so parsimonious models need to be employed at the risk of discarding valuable conditioning information. Approaching the forecasting exercise by way of standard non-parametric or semi-parametric methods is generally not a viable option because these methods encounter "curse-of-dimensionality" problems rather quickly as the size of the conditioning information set increases. Third, the models tested are generally constant: different model specifications are proposed and their performance is assessed ex-post. Although interesting from an econometric perspective, these findings are of little help for an investor interested in exploiting the conditioning information in real time as he would not know what model to choose ex-ante.[3] Finally, apart from some important exceptions, much of the literature on financial markets prediction focuses on formulating return forecasts and little attention is dedicated to analyzing quantitatively the economic value associated with them for a representative investor.[4]

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

---

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

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

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

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

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

Our analysis answers three questions. The first is whether macroeconomic and financial variables contain information about expected stock returns and volatility that can be exploited in real time by a mean-variance investor. For stock returns we use the major conditioning variables proposed so far in the literature and summarized by Welch and Goyal (2008). We propose two models of volatility forecasts. The first models volatility as a function of monthly macroeconomic and financial time-series as well as past volatility. The second is inspired by the family of MIDAS models proposed by Ghysels, Santa-Clara, and Valkanov (2006) and models monthly volatility as a function of lagged daily squared returns. We call this model "semi-parametric MIDAS" and show that its performance is superior to that of its parametric counterpart. Genuine out-of-sample forecasts require not only that the parameters are estimated recursively, but also that the conditioning information employed is selected in real-time. For this reason, every predictive framework under consideration starts from the large set of predictor variables employed by Welch and Goyal (2008) and selects recursively the model specification. Our estimates show that BRT forecasts outperform the established benchmarks and possess significant market timing in both returns and volatility.

A related question we address is whether the conditioning information contained

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

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

The rest of the paper is organized as follows. Section 2.2 introduces our empirical framework and describes how stock returns and volatility are predicted. In Section 2.3 we show how we employ boosted regression trees to directly select optimal portfolio allocations. Section 2.4 presents results for the out-of-sample accuracy of the model,

---

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

conducts formal tests of market timing in both returns and volatility and evaluates the performance of empirical trading strategies based on BRT forecasts. Section 2.5 concludes.

2.2 Empirical Framework and Full-Sample Results

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

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

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

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

leading to the optimal portfolio weights

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

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

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

The objects $E_t\{r_{t+1}\} = \mu_{t+1|t}$ and $Var_t\{r_{t+1}\} = \sigma^2_{t+1|t}$ in Eq. 2 represent conditional expectations of returns and variance on the basis of the investor's conditioning information at time t. In this paper we allow these conditional expectations to be non-linear

functions of observable macroeconomic and financial time-series, the idea being that the linearity assumption generally adopted in financial economics may be costly in terms of forecasting accuracy and portfolio allocation profitability.

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

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

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

where x_t represents a set of publicly available predictor variables and β_μ is a vector of parameter estimates obtained via ordinary least squares. The linear specification is generally imposed for simplicity at the expense of being potentially misspecified. The

---

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

sources of misspecification are at least two. The first relates to what information is incorporated in the formulation of the forecasts. Asset pricing models suggest a wide array of economic state variables for both returns and volatility, but linear frameworks are prone to over-fitting if the number of parameters to be estimated is large compared to the number of observations, forcing the agent to exclude a large portion of the conditioning information available. The second relates to how information is incorporated in the forecasts: theoretical frameworks rarely identify linear relations between the variables at hand, so empirical estimates based on ordinary least squares may not be appropriate. Note however that, in our context, misspecification *per se* is not a source of concern as long as it does not translate into lower predictive accuracy, which is ultimately what matters for portfolio allocation.

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

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

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

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

2.2.1   Regression Trees

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

---

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

written as an additive model:

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

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

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

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

that minimize the sum of squared residuals:

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

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

$$\hat{c}_1 = \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_1(p,s)\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_1(p,s)\},$$

$$\hat{c}_2 = \frac{1}{\sum_{t=1}^{T} I\{x_t \in S_2(p,s)\}} \sum_{t=1}^{T} y_{t+1} I\{x_t \in S_2(p,s)\}. \tag{5}$$

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

Regression trees are generally employed in high-dimensional datasets where the relation between predictor and predicted variables is potentially non-linear. This becomes important when modeling stock returns because numerous predictor variables

have been proposed so far in the literature. Furthermore, the theoretical frameworks rarely imply a linear or monotonic relation between predictor and predicted variable. On the other hand, the approach is sequential and successive splits are performed on fewer and fewer observations, increasing the risk of fitting idiosyncratic data patterns. Furthermore, there is no guarantee that the sequential splitting algorithm leads to the globally optimal solution. To deal with these problems, we next consider a method known as boosting.

2.2.2 Boosting

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

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

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

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

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

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

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

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

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

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

2.2.3 Implementation

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

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

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

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

The second refinement is *subsampling* and is inspired by "bootstrap aggregation" (bagging), see Breiman (1996). Bagging is a technique that computes forecasts over bootstrap samples of the data and averages them in a second step, therefore reducing

---

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

the variance of the final predictions. In our context, the procedure is adapted as follows: at each boosting iteration we sample without replacement one half of the training sample and fit the next tree on the sub-sample obtained.

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

2.2.4   Conditional Volatility Estimates

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

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

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

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

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

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

More recently, Ghysels, Santa-Clara, and Valkanov (2005) proposed a new family of models under the heading MIDAS (MIxed DAta Sampling). Their model adopts the

following estimator for the conditional variance of monthly returns:

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

where

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

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

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

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

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

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

2.2.5 MIDAS and Semi-Parametric MIDAS

MIDAS originates from the idea that expanding the information set on which volatility forecasts are conditioned should increase their precision. While realized volatility frameworks generally use arbitrary volatility lags to forecast the future, MIDAS models fit volatility by weighting past observations optimally. For the case of monthly data,

this translates into forecasting a given month's volatility using daily squared returns lagged up to a year: i.e. 250 days. In order to reduce the dimensionality of the problem specific weighting functions are assumed (the Beta weights reported in Eq. 12 is one possible example) and their parameters are obtained via Quasi-Maximum Likelihood. The optimal weights determine how past information is incorporated in the predictive model to generate volatility forecasts.

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

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

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

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

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

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

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

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

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

The fact that the two procedures exploit differently the same set of conditioning information raises the question of which one exploits it in a more effective fashion. We answer this question in Section 2.4 that is dedicated to assessing the out-of-sample performance of the forecasting models at hand.

2.3   Predicting Optimal Portfolio Allocations

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

---

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

optimal portfolio weights represent a natural measure of the expected price of risk.

We construct optimal portfolio allocations based on market returns r_{t+1} and volatility σ_{t+1} as well as the risk-aversion coefficient γ

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

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

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

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

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

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

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

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

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

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

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

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

Rossi and Timmermann (2010) perform a similar analysis for return and volatility prediction. Their results highlight that the three most important predictors for the equity premium are inflation, the log earnings price ratio and the de-trended T-bill rate; the ones for volatility are the lag volatility, the default spread and the lag return on the market

portfolio. They also uncover strong non-linearities in the functional form relating the predictor and predicted variables. These findings imply that linear models are potentially misspecified. In order to assess whether correcting for such misspecification translates into greater predictive accuracy, we assess next the out-of-sample forecasting performance of boosted regression trees and other benchmark models commonly employed in the literature.

2.4   Out-of-Sample Results.

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

2.4.1   Returns

Panel A of Table 1 reports the results for return prediction in terms of mean-squared error (MSE) and directional accuracy. The results are reported for BRT as well as two competing models proposed in the literature: the prevailing mean proposed by Welch and Goyal (2008) and a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2.2. Over the sample 1969-2008 BRT outperforms both benchmarks in terms of L^2 norm and achieves an out-of-sample R^2 of 0.30%. The second best is the prevailing mean, with an R^2 of -0.79%. Finally, the performance of the multivariate linear model is disappointing, with an R^2

of -2.28%.[11]

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

As mentioned above, MSE may not be an appropriate loss function in financial settings as directional accuracy has, in some circumstances, proved to be more closely related to the profitability of financial forecasts. For this reason, we also analyze the directional accuracy of the competing models. Because there is no obvious benchmark of no-predictability in this case, we employ the recursively estimated prevailing direction as a benchmark. Furthermore, because the prevailing direction has been that of positive excess returns over the sample under consideration, our benchmark is equivalent to predicting positive excess returns at all times and has a predictive accuracy of 55.26%. BRT has a hit ratio of 57.35%, 2% greater than the prevailing direction. The linear model has a rather disappointing accuracy of 52.17% and so ranks last among the models considered.

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

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

where τ is the first month we start our calculations from, "pm" stands for prevailing mean and "$model$" is alternatively BRT or the multivariate linear model. This measure is useful at tracking the performance of the forecasting framework on each period because an upward sloping "CumSum" curve over a given month entails that the predictive

---

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

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

model outperforms the prevailing mean, and vice versa. The plot shows that BRT outperforms the multi-variate linear model and the prevailing mean over the sample under consideration. BRT outperforms the prevailing mean from 1970 to 1975 and from 1980 to 2005. It underperforms the prevailing mean over the period 1975-1980 and the last few years characterized by the financial crisis. The multivariate linear model outperforms the prevailing mean over the period 1970-1975 and in 2008. It underperforms consistently the prevailing mean from 1975 until the end of 2007.

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

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

2.4.2   Volatility

Panel B of Table 1 presents forecasting accuracy results for volatility. BRT that uses both macroeconomic and financial variables as conditioning information performs best with an R^2 of 40.84%. The "Semi-Parametric MIDAS" specification we propose

has an R^2 of 34.95%, which is comparable to the 35.69% of MIDAS (beta). Garch (1,1) has the worst performance as its R^2 is only 14.13%.[13]

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

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

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

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

The middle panel of Figure 5 reports the out-of-sample directional accuracy difference between BRT, Semi-Parametric MIDAS, MIDAS (Beta) and GARCH (1,1) that we employ as a benchmark. The vertical axis counts the differential performance between the prediction model and the benchmark in terms of correct volatility directional

---

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

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

forecasts. The plot shows that the BRT model outperforms its competitors on most of the sample only to lose its lead in the last few years. MIDAS (Beta) and Semi-Parametric MIDAS display a virtually identical performance throughout the test period.

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

2.4.3   Weight Prediction

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

The bottom panel of Figure 4 presents estimates of

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

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

Overall, the results reported in this section highlight the superiority of BRT compared to the established benchmarks at predicting the equity premium and volatility. The results hold whether mean squared error or directional accuracy is employed as an evaluation criterion. As far as the optimal portfolio allocation is concerned, our

preliminary results suggest that BRT performs well in terms of mean squared error, but the evidence for directional accuracy is not convincing.

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

2.4.4   Formal Tests of Predictive Accuracy

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

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

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

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

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

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

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

Zarnowitz-type test, we run the regression:

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

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

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

In Panel C of Table 2 we report the market timing tests for the optimal portfolio weights. In accordance with the rather weak directional accuracy results of Table 1, BRT does not display significant directional accuracy according to the HM test. In the tests

---

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

where not only directional accuracy but also magnitudes matter (CM and BH tests) BRT displays instead significant market timing, possibly implying profitable portfolio allocations.

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

2.4.5   Mean-Variance Investor and Portfolio Allocation

We construct portfolio weights for the mean-variance investor described in Section 2.2 who uses boosted regression trees to formulate return and volatility predictions.[16]

We have a free parameter γ that determines the investor's degree of risk-aversion. We set it to 4 as that corresponds to a moderately risk-averse investor. We first present the results for unconstrained portfolio weights and evaluate the performance of two models. In the first, market returns and volatility are forecasted separately and the optimal mean-variance portfolio allocation is computed in a second stage (two-step BRT model). In the second, the optimal portfolio allocation is forecasted directly following the procedure described in Section 2.3 (one-step BRT model).

We compare our results to three benchmark strategies. The first uses stock returns predictions from a multivariate linear regression model selected recursively using

---

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

the BIC on the full set of predictor variables listed in Section 2.2 and volatility predictions from a GARCH (1,1) model. The second model replaces the GARCH (1,1) model with a MIDAS (Beta) model (Ghysels et al (2005)). The third benchmark is a passive investment strategy that allocates 100% of wealth in the market portfolio at all times.

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

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

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

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

The performance is evaluated over the sample 1969-2008. The best model is the BRT that forecasts the optimal portfolio allocation in two steps with a monthly Sharpe ratio of 15.81%. The second best is the one-step BRT model with a Sharpe ratio of 14.81%. The two-step BRT model has a higher and more significant Jensen's *Alpha* measure of market outperformance, but a lower and less significant Treynor-Mazuy (TM) measure of market timing compared to the one-step BRT model. The benchmark models perform considerably worse. The two active benchmark strategies have a negative and insignificant Jensen's *Alpha*: -0.072% for the model that uses a GARCH (1,1) and -0.069% for the model that uses a MIDAS (Beta) specification. The Treynor-Mazuy measure of market timing is positive but insignificant for both models. The Sharpe ratios, 2.70% and 3.18% respectively, are smaller than that of the passive strategy (Panel D) that invests 100% of wealth in the market portfolio (7.71%).

In Panel B of Table 3 we repeat the same exercise imposing short-selling and borrowing constraints. As expected, the constraints reduce the profitability of both BRT-based investment strategies. The Sharpe ratio of the two-step BRT model drops by approximately 1.8% to 14.01%, while the Sharpe ratio of the one-step BRT model

drops by approximately 2.2% to 12.59%. The lower Sharpe ratios are due to both lower mean returns and volatility of the constrained portfolio allocations compared to the unconstrained ones. The magnitudes of both Jensen's *Alpha*'s and TM market timing measures are lower than their unconstrained counterparts, but they maintain a strong level of statistical significance. The opposite holds true for the two active benchmark strategies. The Sharpe ratio for the investment strategy that exploits GARCH (1,1) volatility forecasts increases by 2.8% to 5.52% and the one for the MIDAS (Beta) increases by 3.5% to 6.69%. There are a number of reasons for this, the most important being that there is a fair degree of estimation error in the estimated optimal portfolio allocations, particularly so for the linear regression, GARCH (1,1) and MIDAS (Beta) models that minimize L^2 criterion functions. We show in section 2.4.7 that accounting for estimation uncertainty improves the performance of the investment strategies.

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

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

frictions.

2.4.6 Portfolio Allocation Performance with Transaction Costs

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

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

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

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

2.4.7 Economic Value of the Investment Strategies

Even though very popular and widely used, the performance measures employed until now are affected by a number of shortcomings. As argued by Fleming, Kirby, and Ostdiek (2001) and Marquering and Verbeek (2005), the Sharpe ratio is not a correct measure of risk-adjusted returns in the presence of time-varying volatility and Jensen's *Alpha* is a biased estimator of abnormal returns in the presence of time-varying portfolio weights. Also, the TM measure provides an indication of whether the portfolio exhibits market timing, but it does not quantify the economic value it generates .

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

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

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

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

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

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

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

Another issue that has been ignored until now is that of estimation uncertainty. Expected returns and volatilities are not known but estimated from data. Even though we used the longest possible time-series available, noise may be responsible for making the portfolio weights too volatile over time. A number of approaches have been proposed

to mitigate the effect of estimation uncertainty, see, for example, Ter Horst, De Roon, and Werker (2000) and Maenhout (2004). Here, for simplicity, we follow Marquering and Verbeek (2005) and estimate the optimal portfolio weights using a "pseudo-risk aversion coefficient" that is double the true risk-aversion coefficient of the agent.

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

Finally, we compute the maximum monthly portfolio fees that a representative investor would be willing to pay in order to have his wealth invested in one of the investment strategies we propose as opposed to holding the market portfolio. In Figure 6 we present the results for the one and two-step BRT models that account for estimation uncertainty. The top panel presents the break-even monthly portfolio fees for the unconstrained, constrained and "return predictability only" strategies for the two-step BRT models with zero, low and high transaction costs (left, middle and right plot respectively). The bottom panel repeats the exercise for the one-step BRT model. Each

figure reports the break-even portfolio fees (in percentage) on the y-axis and the risk-aversion coefficient (ranging from 1 to 10) on the x-axis. When transaction costs are not accounted for, the monthly break-even portfolio fees for the two-step BRT models are quite sizable and range from 0.47% to 1.20% for the unconstrained investment strategy, from 0.22% to 0.91% for the constrained investment strategy and from 0.18% to 0.47% for the investment strategy that exploits only the predictability in returns. Similar, but somewhat lower, are the results for the one-step BRT model. The results remain very strong and are almost unaffected by low transaction costs, but they deteriorate for high transaction costs. For example, an investor with a coefficient of risk aversion equal to 2 would rather hold the market portfolio than the unconstrained investment strategy formulated using the one or two-step BRT models. All the other risk-aversion, transaction costs and investment strategy combinations deliver positive and sizable monthly break-even portfolio fees, implying that the investment strategies proposed are economically valuable.

2.5 Conclusions

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

Table 2.1: This table reports the out-of-sample R^2 and directional accuracy for a boosted regression tree model with 10,000 boosting iterations used to forecast monthly stock returns, realized volatility and the optimal portfolio weight in the risky asset for the mean-variance investor. For return prediction (Panel A) we also present results for the prevailing mean model proposed by Welch and Goyal (2008) and a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2 of the paper. For volatility prediction (Panel B) we report the results for a semi-parametric MIDAS specification described in Section 2 of the paper, a MIDAS (Beta) model (Ghysels et al (2005)) and a GARCH (1,1). For optimal portfolio allocation prediction (Panel C) we also report the results for the prevailing mean. The out-of-sample performance is computed over the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Out-of-Sample Forecasting Accuracy

A. Return Models

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

B. Realized Volatility Models

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

C. Optimal Weight Prediction Models

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

Table 2.2: This table reports forecasting accuracy tests for a boosted regression tree model with 10,000 boosting iterations used to forecast monthly stock returns in Panel A, realized volatility in Panel B and the optimal portfolio weight in the risky asset for the mean-variance investor in Panel C. Each panel presents the results for three tests: the Henriksson and Merton (HM), the Cumbey and Modest (CM) and the Bossaerts and Hillion (BH) test. We report the coefficient of interest, α_1, whose significance determines the directional predictive ability of each forecasting framework. The numbers in brackets are t-test values. For comparison, in Panel A we also present results for a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2 of the paper; in Panel B we report the results for a semi-parametric MIDAS specification described in Section 2 of the paper, a MIDAS (Beta) model (Ghysels et al (2005)) and a GARCH (1,1). The tests are performed for the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Tests of Out-Of-Sample Forecasting Accuracy

A. Return Models

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

B. Volatility Models

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

C. Models for the Optimal Portfolio Weight

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

Table 2.3: This table reports the performance of an investment portfolio based on a boosted regression tree model with 10,000 boosting iterations. The first row of each panel presents the results obtained by forecasting stock market returns and volatility separately and computing the optimal mean-variance portfolio allocation for an agent with risk-aversion coefficient of 4 (two-step BRT model). In the second row we forecast the optimal portfolio allocation directly on the basis of the conditioning variables listed in Section 2 of the paper (one-step BRT model). For comparison, we also present results for investment portfolios constructed using stock return predictions from a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables described in Section 2 of the paper and volatility predictions from alternatively a GARCH (1,1) or a MIDAS (Beta) model (Ghysels et al (2005)). For each investment portfolio we report the average monthly return, standard deviation, Sharpe ratio as well as its Jensen's *Alpha* and Treynor-Mazuy market timing measure. In brackets are t-test values. Panel A reports the results for unconstrained portfolio allocations. Panel B imposes shortselling and borrowing constraints. Panel C exploits only return predictability. Panel D report results for a passive investment strategy that places 100% of wealth in the market portfolio. The performance is computed over the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Portfolio Allocation Performance

A. Unconstrained Weights	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
Two-Step BRT	2.10%	10.19%	15.81%	1.37%	4.484
				(3.09)	(3.93)
One-Step BRT	2.02%	10.38%	14.81%	1.12%	4.939
				(2.77)	(4.79)
Linear Model & GARCH (1,1)	0.62%	4.84%	2.70%	-0.07%	0.397
				(-0.35)	(0.81)
Linear Model & MIDAS (Beta)	0.69%	6.35%	3.18%	-0.07%	0.213
				(-0.31)	(0.34)
B. Short-Selling and Borrowing Constraints	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
Two-Step BRT	0.93%	3.17%	14.01%	0.26%	0.931
				(2.77)	(3.90)
One-Step BRT	0.93%	3.57%	12.59%	0.22%	0.825
				(2.45)	(3.54)
Linear Model & GARCH (1,1)	0.64%	2.86%	5.52%	-0.02%	0.039
				(-0.21)	(0.19)
Linear Model & MIDAS (Beta)	0.69%	3.07%	6.69%	0.01%	0.126
				(0.17)	(0.58)
C. Return Predictability Only	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
Two-Step BRT	0.94%	3.55%	12.83%	0.24%	0.729
				(2.44)	(2.89)
One-Step BRT	0.91%	3.78%	11.18%	0.18%	0.741
				(1.93)	(3.15)
Linear Model	0.80%	3.91%	8.09%	0.05%	0.070
				(0.60)	(0.31)
D. Passive Strategy	Mean	St. Dev.	Sh. Ratio	Jensen's α	TM
100% Market	0.83%	4.45%	7.71%	—	—

Table 2.4: This table reports the performance of an investment portfolio based on a boosted regression tree model with 10,000 boosting iterations accounting for transaction costs. Panel A presents the results obtained by forecasting stock market returns and volatility separately and computing the optimal mean-variance portfolio allocation for an agent with risk-aversion coefficient of 4 (two-step BRT model), while accounting for low and high transaction costs. For each transaction costs level we report the performance of the unconstrained optimal portfolio, the optimal portfolio with shortselling and borrowing constraints as well as an investment strategy that invests 100% of wealth in the risky asset if expected excess returns are greater than zero and 0% of wealth in the risky asset otherwise. In Panel B we conduct the same exercise, but we forecast the optimal portfolio allocation directly on the basis of the conditioning variables listed in Section 2 of the paper (one-step BRT model). The tests are performed on the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Portfolio Performance with Transaction Costs

A. Two-Step BRT

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

B. One-Step BRT

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

Table 2.5: This table reports the realized utility of an investment portfolio based on a boosted regression tree model with 10,000 boosting iterations for different coefficients of risk-aversion. The top rows of Panel A present the results obtained by forecasting stock market returns and volatility separately and computing the optimal mean-variance portfolio allocation in a second step (two-step BRT model). The bottom rows of Panel A present the results obtained by forecasting the optimal portfolio allocation directly on the basis of the conditioning variables listed in Section 2 of the paper (one-step BRT model). In Panel B we conduct the same exercise using a "pseudo risk-aversion coefficient" that doubles the investor's genuine risk-aversion γ to adjust portfolio weights for estimation uncertainty. For comparison, in Panel C we report results for a passive investment strategy that places 100% of wealth in the market portfolio. The analysis is conducted over the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Average Realized Utilities

A. Without Accounting for Estimation Uncertainty

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

B. Accounting for Estimation Uncertainty

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

C. Passive Strategy

	$\gamma = 2$	$\gamma = 4$	$\gamma = 6$	$\gamma = 8$	$\gamma = 10$
100% Market	0.60%	0.37%	0.14%	-0.09%	-0.32%
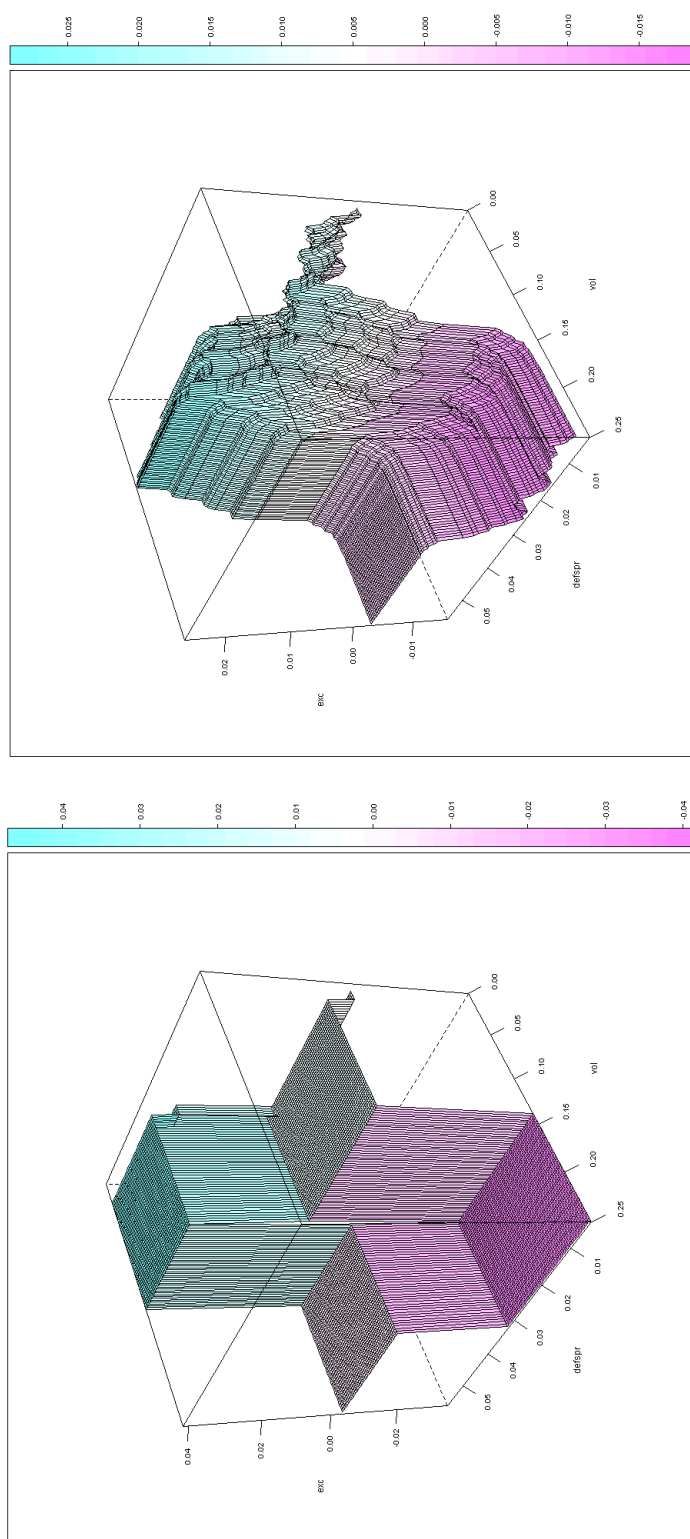
(a) 3 Iterations

(b) 5000 Iterations

Figure 2.1: Fitted values of excess returns (exc) as a function of volatility (vol) and the default spread (defspr). Both plots are based on boosted regression trees with two terminal nodes. The panel on the left uses three boosting iterations, while the right panel uses 5,000 iterations. The scale for volatility has been inverted. The plots are based on monthly data from 1927-2008.
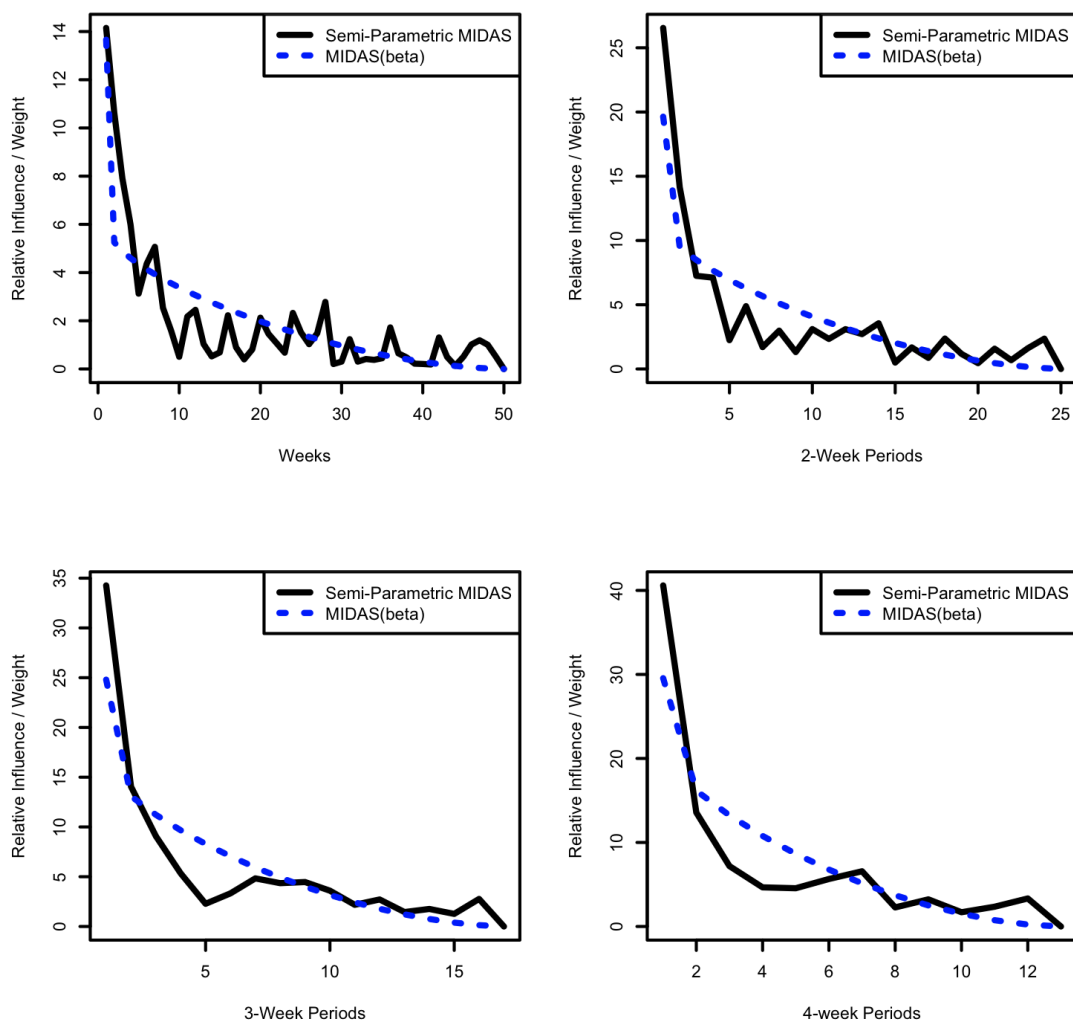
Figure 2.2: This figure presents the relative influence (in percent) of lagged squared returns at predicting monthly volatility for both the MIDAS (beta) model (Ghysels et al (2005)) and the Semi-Parametric MIDAS model. Monthly volatility is predicted using 250 lagged squared daily returns as conditioning information. For ease of interpretation, in each panel we aggregate individual days' relative influence in 1, 2, 3, and 4 week periods.
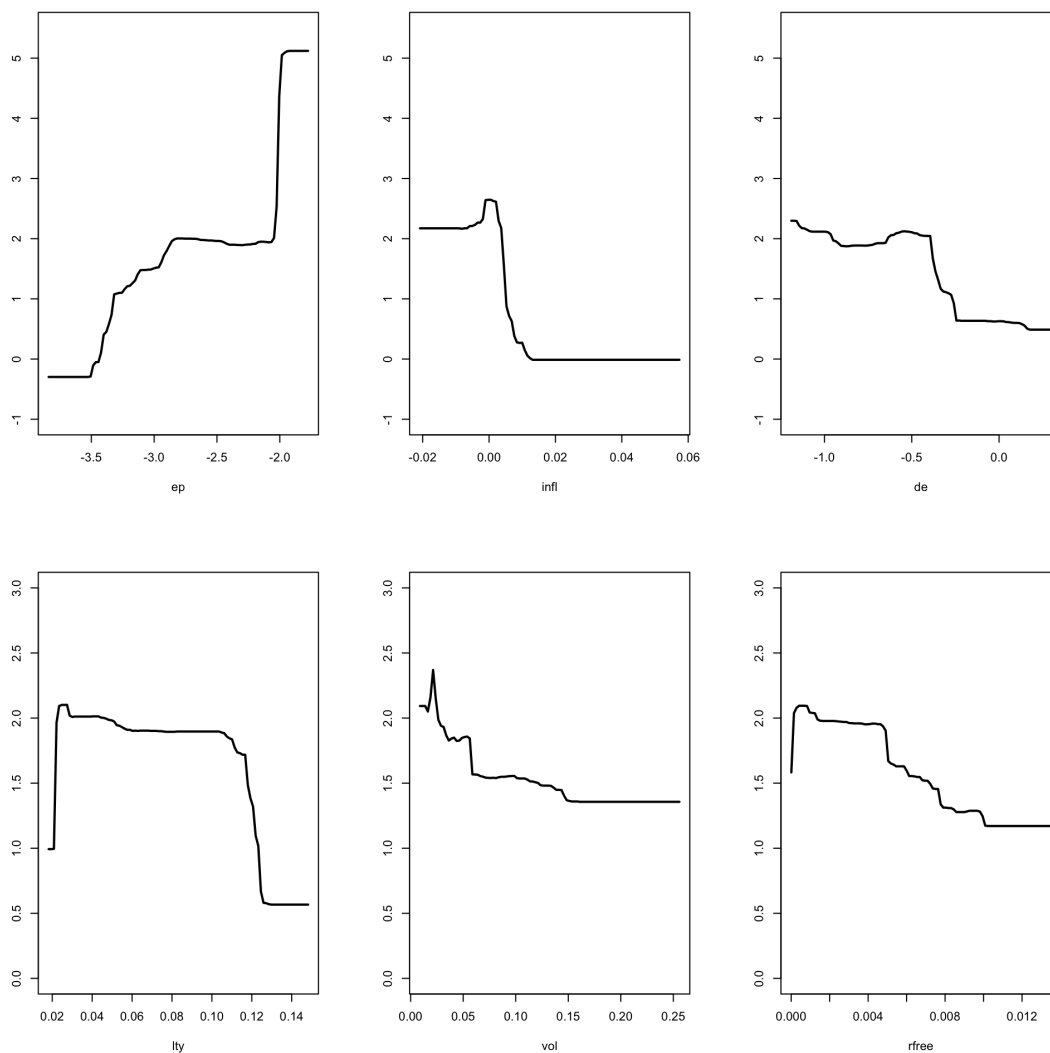
Figure 2.3: This figure presents the partial dependence plots for the optimal allocation to risky assets as a function of the six predictor variables with the highest relative influence over the sample 1927-2008. The most influential predictors are the log earnings price ratio (ep), inflation (infl), the log dividend earnings ratio (de), the yield on long term government bonds (lty), stock market volatility (vol) and the three-month T-bill rate (Rfree) that have a relative influence of 16.63%, 15.88%, 13.94%, 9.38%, 7.23% and 6.67%, respectively. The horizontal axis covers the sample support of each predictor variable, while the vertical axis tracks the change in the optimal weight in the market portfolio as a function of the individual predictor variables.
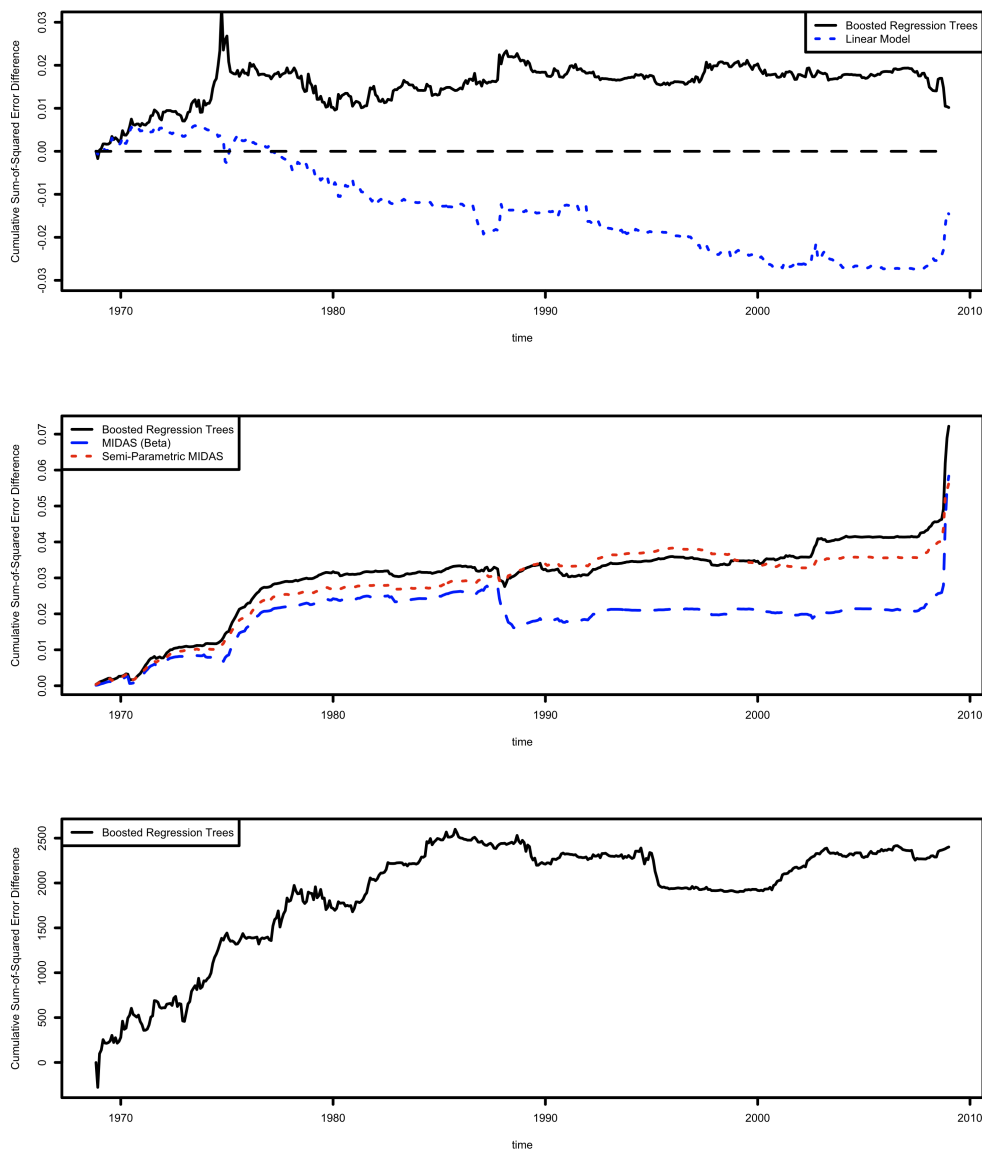
Figure 2.4: This figure presents cumulative out-of-sample sum-of-squared error differences between a boosted regression tree and benchmark forecasting models for returns, volatility and portfolio weight predictions. The top graph performs the analysis for stock returns: it uses the prevailing mean proposed by Goyal and Welch (2008) as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations and a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2 of the paper. The middle graph performs the analysis for volatility: it uses a GARCH (1,1) as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations, a MIDAS (beta) model (Ghysels et al (2005)) and a semi-parametric MIDAS specification described in Section 2 of the paper. The bottom graph performs the analysis for portfolio weight predictions: it uses the prevailing mean as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations.
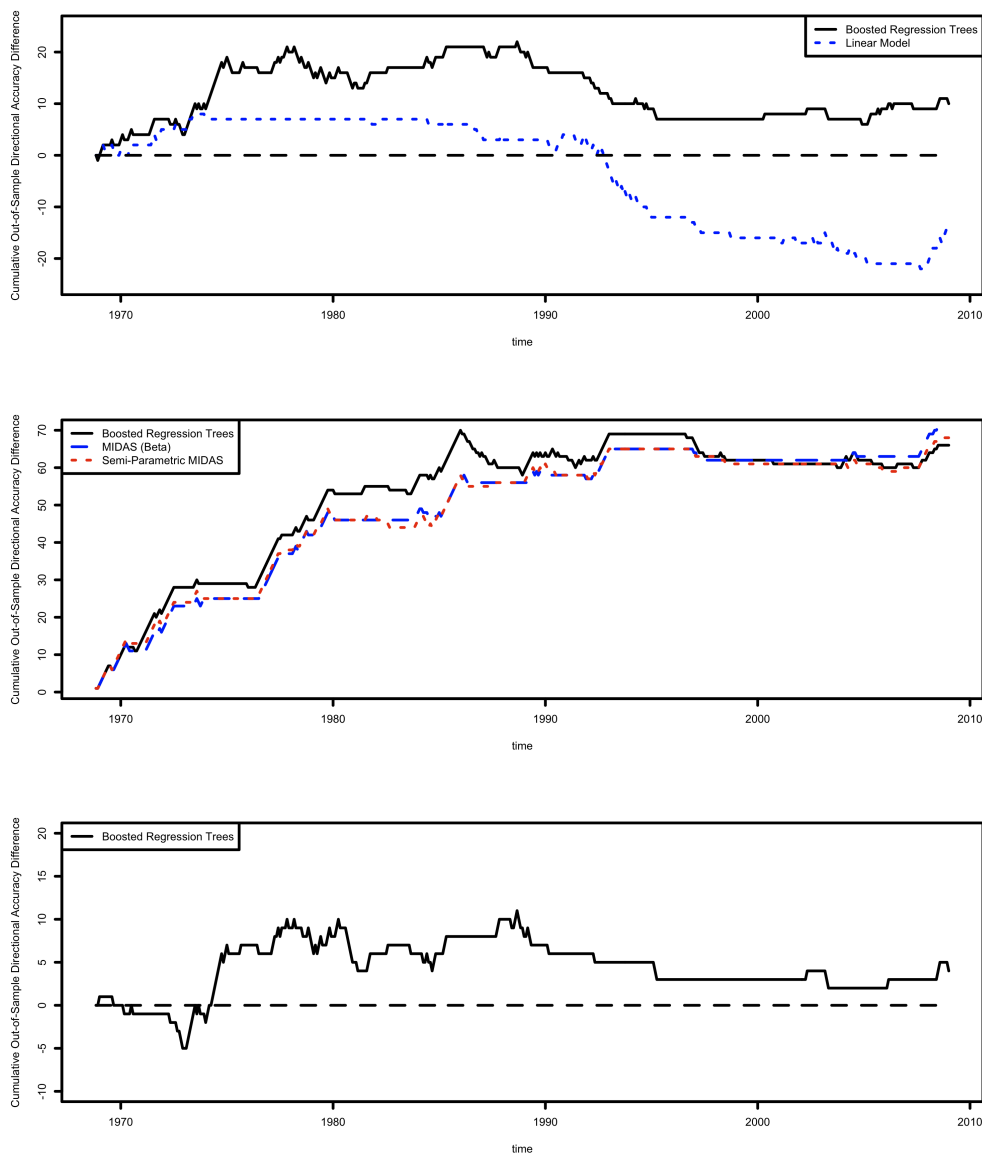
Figure 2.5: This figure presents cumulative out-of-sample directional accuracy differences between a boosted regression tree and benchmark forecasting models for returns, volatility and portfolio weight predictions. The top graph performs the analysis for stock returns: it uses the prevailing direction as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations and a multivariate linear regression model selected recursively using the BIC on the full set of predictor variables listed in Section 2 of the paper. The middle graph performs the analysis for volatility: it uses a GARCH (1,1) as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations, a MIDAS (beta) model (Ghysels et al (2005)) and a semi-parametric MIDAS specification described in Section 2 of the paper. The bottom graph performs the analysis for portfolio weight predictions: it uses the prevailing direction as a benchmark and plots the performance of a boosted regression tree model with 10,000 boosting iterations.
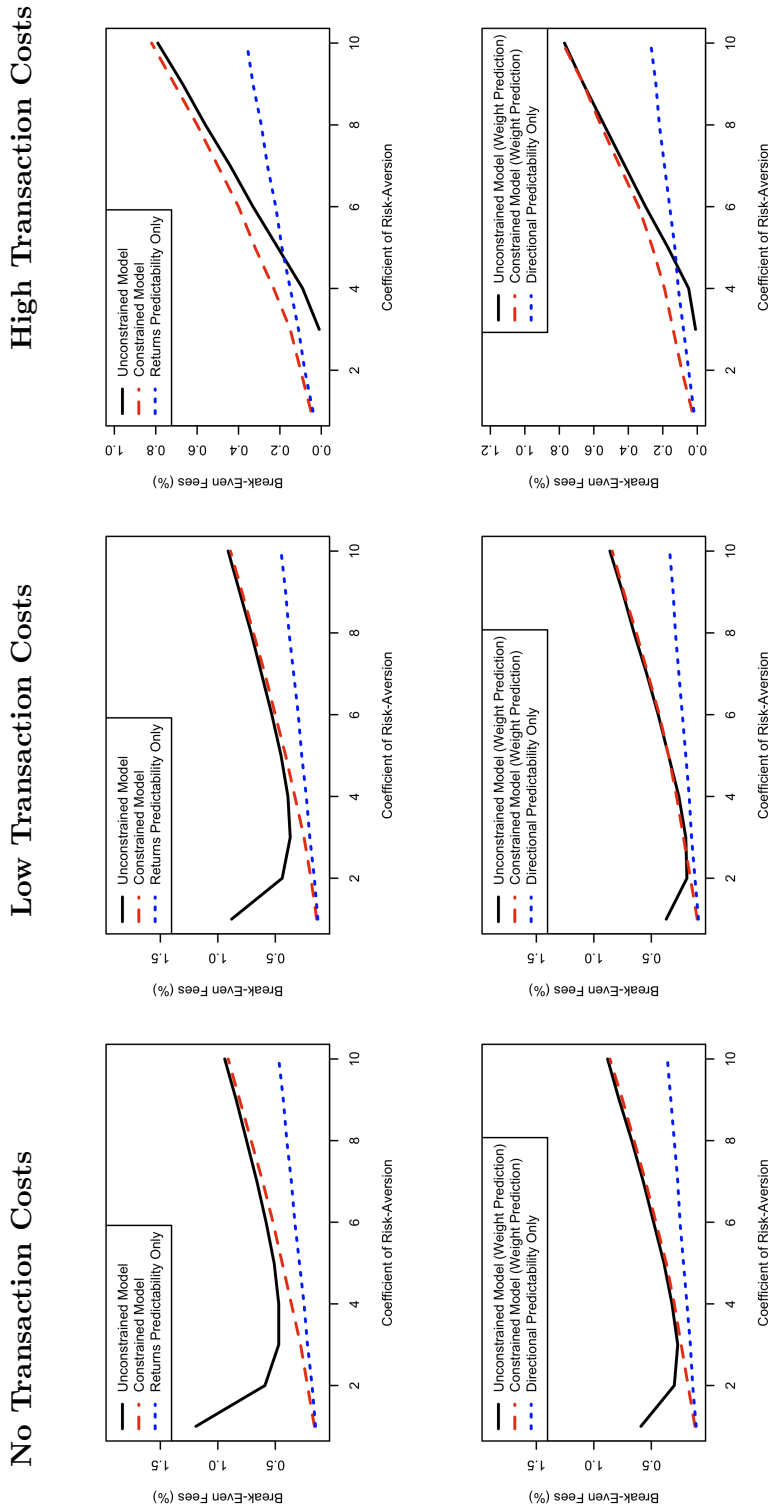
Figure 2.6: This figure plots break-even portfolio fees that a representative investor would be willing to pay in order have his wealth invested in one of the investment strategies we propose as opposed to investing 100% of his wealth in the market portfolio. All investment strategies proposed are based on a boosted regression tree model with 10,000 boosting iterations. The analysis is conducted for different coefficients of risk-aversion. Each plot presents the results for an unconstrained model, a model with short-selling and borrowing constraints and a model where the portfolio allocation is decided on the basis of return predictability only. The top panel presents the results obtained by forecasting stock market returns and volatility separately and computing the optimal mean-variance portfolio allocation in a second step (two-step BRT model). The bottom panel presents the results obtained by forecasting the optimal portfolio allocation directly on the basis of the conditioning variables listed in Section 2 of the paper (one-step BRT model). In each panel, the left, middle and right plots report the results for zero, low and high transaction costs. In both panels, the optimal portfolio allocations are computed using a "pseudo risk-aversion coefficient" that doubles the investor's genuine risk-aversion γ to adjust portfolio weights for estimation uncertainty. The analysis is conducted over the sample 1969-2008. The parameters of the forecasting models are estimated recursively through time.

Bibliography

AIT-SAHALIA, Y., AND M. BRANDT (2001): "Variable Selection for Portfolio Choice," *The Journal of Finance*, 56(4), 1297–1351.

BAI, J., AND S. NG (2009): "Boosting diffusion indices," *Journal of Applied Econometrics*, 24(4), 607–629.

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

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

BREEN, W., L. GLOSTEN, AND R. JAGANNATHAN (1989): "Economic significance of predictable variations in stock index returns," *Journal of Finance*, 44(5), 1177–89.

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

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

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

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

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

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

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

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

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

DANGL, T., AND M. HALLING (2008): "Predictive regressions with time-varying coefficients," *Working Paper, Vienna University of Technology*.

ENGLE, R., E. GHYSELS, AND B. SOHN (2006): "On the Economic Sources of Stock Market Volatility," *Manuscript, New York University*.

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

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

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

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

FLEMING, J., C. KIRBY, AND B. OSTDIEK (2001): "The Economic Value of Volatility Timing," *Journal of Finance*, 56, 329–352.

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

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

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2005): "There is a Risk-Return Tradeoff After All," *Journal of Financial Economics*, 76(3), 509–548.

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2006): "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, 131(1-2), 59–95.

GOYAL, A., AND I. WELCH, I. (2003): "Predicting the Equity Premium with Dividend Ratios," *Management Science.*

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

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.

HENRIKSSON, R., AND R. MERTON (1981): "On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business*, 54(4), 513.

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

JOHANNES, M., A. KORTEWEG, AND N. POLSON (2009): "Sequential learning, predictive regressions, and optimal portfolio returns," *Mimeo, Columbia University.*

KEIM, D., AND R. STAMBAUGH (1986): "Predicting returns in the stock and bond markets," *Journal of Financial Economics*, 17(2), 357–390.

KOTHARI, S., AND J. SHANKEN (1997): "Book-to-market, dividend yield, and expected market returns: A time-series analysis," *Journal of Financial Economics*, 44(2), 169–203.

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

LEITCH, G., AND J. TANNER (1991): "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures," *American Economic Review*, 81(3), 580–590.

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

LEWELLEN, J., S. NAGEL, AND J. SHANKEN (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96(2), 175–194.

LO, A., AND A. MACKINLAY (1990): "Data-snooping biases in tests of financial asset pricing models," *Review of Financial Studies*, 3(3), 431–467.

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

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

MARQUERING, W., AND M. VERBEEK (2005): "The Economic Value of Predicting Stock Index Returns and Volatility," *Journal of financial and Quantitative Analysis.*

NELSON, C., AND M. KIM (1993): "Predictable stock returns: The role of small sample bias," *Journal of Finance*, 48(2), 641–661.

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

PESARAN, H., AND A. TIMMERMANN (2005): "Real-time econometrics," *Econometric Theory*, 21(01), 212–231.

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

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

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

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

ROSSI, A., AND A. TIMMERMANN (2010): "What is the Shape of the Risk-Return Relation?," *UCSD Discussion Paper.*

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

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

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

TER HORST, J., F. DE ROON, AND B. WERKER (2000): "Incorporating Estimation Risk in Portfolio Choice," *CentER Working Papers Series*.

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

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

Abstract

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

3.1 Introduction

The existence of a systematic trade-off between market risk and expected returns is central to modern finance. Yet, despite more than two decades of empirical research, there is little consensus on the basic properties of the relation between the equity premium and conditional stock market volatility. Studies such as Campbell (1987), Breen, Glosten, and Jagannathan (1989), Glosten, Jagannathan, and Runkle (1993), Whitelaw (1994), and Brandt and Kang (2004) find a negative trade-off, while conversely

French, Schwert, and Stambaugh (1987), Bollerslev, Engle, and Wooldridge (1988), Harvey (1989), Harrison and Zhang (1999), Ghysels, Santa-Clara, and Valkanov (2005), Guo and Whitelaw (2006), and Ludvigson and Ng (2007) find a positive trade-off. While these studies use different methodologies and sample periods, it remains a puzzle why empirical results vary so much.

Theoretical asset pricing models do not generally imply a linear or even monotonic, risk-return relation. For example, in the context of a simple endowment economy, Backus and Gregory (1993) show that the shape of the relation between the risk premium and the conditional variance of stock market returns is largely unrestricted with increasing, decreasing, flat, or non-monotonic patterns all possible. Similar conclusions are drawn by studies such as Abel (1988), Gennotte and Marsh (1993), Veronesi (2000), Whitelaw (2000), and David and Veronesi (2009).[1] It follows that the conventional practice of measuring the risk-return trade-off by means of a single slope coefficient in a linear model offers too narrow a perspective and can lead to biased results since it limits the analysis to monotonic or flat relations. Typically, the risk-return trade-off cannot be summarized in this manner without making strong auxiliary modeling assumptions whose validity need to be separately tested. Instead it is necessary to consider the shape of the entire risk-return relation at different levels of risk.

This paper introduces a novel and flexible regression approach that does not impose strong modeling assumptions such as linearity to analyze the shape of the risk-return relation. Our approach uses regression trees to carve out the state space through a sequence of piece-wise constant models that approximate the unknown shape of the risk-return relation. By using additive expansions of simple regression trees—a process known as boosting—we obtain smooth and stable estimates that let us map the shape of the risk return relation as well as empirically test if it is monotonic.[2]

We adopt the boosted regression tree approach to empirically analyze the risk-return relation for US monthly stock returns over the period 1927-2008 and find strong empirical evidence of a non-monotonic risk-return relation. At low and medium levels of conditional volatility, there is a strongly positive relation between the conditional mean

---

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

[2]By combining several models, boosted regression trees have advantages similar to those found in forecast combinations, see, e.g., Rapach, Strauss, and Zhou (2010).

and volatility of stock market returns. Conversely, at high levels of volatility, the relation appears to be flat or inverted, i.e., higher levels of conditional volatility are associated with lower expected returns. Formal statistical tests that account for sampling error soundly reject a monotonically increasing mean-volatility relation.

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

A second reason why our results differ from previous estimates comes from differences in the underlying mean and volatility estimates. Empirical evidence in Glosten, Jagannathan, and Runkle (1993) and Harvey (2001) suggests that inference on the risk-return relationship can be very sensitive to how the expected return and volatility models are specified.[3] Estimates may again be biased if the models for expected returns or conditional volatility are misspecified as a result of using overly restrictive models or including too few predictor variables (Ludvigson and Ng (2007)). Our model specification tests show that linear models used to compute expected returns and conditional volatility are misspecified even when a large set of conditioning variables is used. The same holds for GARCH or MIDAS variance estimates based on past squared returns. These results suggest that it is important to maintain a flexible functional form for the conditional mean and volatility models while also considering a large conditioning information set. We find that the boosted regression tree approach accomplishes this and leads to reliable estimates of expected returns and conditional volatility that pass the model specification tests.

Dynamic asset pricing models provide economic intuition for our findings of a non-monotonic risk-return relation. In consumption asset pricing models, the conditional equity premium reflects the correlation between stock returns and the marginal rate of

---

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

substitution which only depends on growth in next period's consumption. Stock returns reflect changes in expectations of the entire infinite future stream of consumption and asset payoffs. In a model with regime switching in the consumption process, Whitelaw (2000) shows that this difference in horizons can reduce the relation between stock market returns and the marginal rate of substitution. While the traditional positive risk-return relation emerges in "normal" states with little uncertainty, states with high probability of transitioning to a new regime and hence high levels of uncertainty and return volatility can have low expected returns. These effects can lead to a non-monotonic, inverted risk-return relation.

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

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

variations in conditional volatility.

Our analysis generalizes and provides a synthesis of many existing approaches from the literature on the risk-return trade-off. Ludvigson and Ng (2007) argue that most studies consider too few conditioning variables and provide a factor-based approach that parsimoniously summarizes information from a large cross-section of variables. Once the conditioning information set is expanded in this way, they find evidence of a positive risk-return trade-off. Like these authors, we consider a large set of conditioning variables to compute the conditional equity premium and market volatility. Linear mean-variance models such as Bollerslev, Engle, and Wooldridge (1988), and Ghysels, Santa-Clara, and Valkanov (2005) arise as special cases of our setup when the variance estimates are based on past (squared) returns. Guo and Whitelaw (2006) argue that findings of a negative or insignificant mean-variance relation is due to the omission of an intertemporal hedging component leading to a downward bias in the variance coefficient. We consider a model that includes both variance and covariance terms and find that our results are robust to the inclusion of both measures of risk. Following papers such as Harrison and Zhang (1999), we do not impose monotonicity on the risk-return relation, but allow its shape to be freely estimated.

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

The remainder of the paper is organized as follows. Section I introduces our

approach to modeling the risk-return relation. Section II describes the data, analyzes the risk-return relation empirically, compares our approach to existing methods and discusses reasons why our empirical findings differ from previously reported ones. Section III provides economic intuition for the non-monotonic risk-return relation in the context of dynamic asset pricing models. Section IV introduces the new covariance measure of risk, while Section V concludes.

3.2   Empirical Methodology

Dynamic asset pricing models do not generally restrict the relation between conditional market volatility and expected returns to be linear. To quote from Gennotte and Marsh (1993, page 1039), ".. in a general equilibrium framework, the market risk premium is a complicated function of the cash flow uncertainty, implying that the simple regression and time series fits of the relation between equity risk premiums and asset price volatility are likely to be misspecified." To avoid biases that follow from restricting the shape of the risk-return trade-off, it is therefore important to adopt an empirical modeling approach that is flexible, yet as emphasized by Ludvigson and Ng (2007) can simultaneously deal with large sets of predictor variables.

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

Specifically, we first split the sample support into two states and compute the mean of y in each state. We choose the state variable (x_1 or x_2) and the splitting point

to achieve the best fit. Next, one or both of these states is split into two additional states. Boosted regression trees are additive expansions of regression trees, where each additional tree is fitted on the residuals of the previous tree until some stopping criterion is reached. The number of trees used in the summation is known as the number of boosting iterations.

An economic illustration of the approach is provided in Figure 1, which shows boosted regression trees that use the (lagged) log payout ratio (i.e., the dividend-earnings ratio) and the log earnings-price ratio to predict excess returns on the S&P500 index. Each iteration fits a tree with two terminal nodes, so every new tree stub generates two regions. The graph on the left uses only three boosting iterations. The resulting model ends up with one split along the payout ratio axis and two splits along the earnings-price ratio axis. Within each state the predicted value of stock returns is constant. With only three boosting iterations the model is quite coarse and shows that the expected return is at its lowest (highest) when the payout ratio is high (low) and the earnings-price ratio is low (high). The fit improves as more boosting iterations are added. As an illustration, the figure on the right is based on 5,000 boosting iterations. Now the plot is much smoother, but clear similarities between the two graphs remain.

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

3.2.1   Regression Trees

Consider a time-series with T observations on a single dependent variable, y_{t+1}, and P predictor (state) variables, $x_t = (x_{t1}, x_{t2}, ..., x_{tp})$, for $t = 1, 2, ..., T$. As illustrated in Figure 1, implementing a regression tree requires deciding, first, which predictor variables to use to split the sample space and, second, which split points to use. A given split point may lead to J disjoint sub-regions or states, $S_1, S_2, ..., S_J$, and the dependent variable is modeled as a constant, c_j, within each state, S_j. For example, in Figure 1 there are $J = 2$ nodes at each split point. For the resulting states c_j corresponds to the value of the flat spots on the vertical (expected return) axis. The value fitted by a

regression tree, $\mathcal{T}(x_t, \Theta_J)$, with J nodes and parameters $\Theta_J = \{S_j, c_j\}_{j=1}^J$ can thus be written

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

where the indicator variable $I\{x_t \in S_j\}$ equals one if $x_t \in S_j$ and is zero otherwise.

Estimates of S_j and c_j can be obtained as follows. Under the conventional objective of minimizing the sum of squared forecast errors, the estimated constant, $\widehat{c}_j$, is the average of y_{t+1} in state S_j:

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

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

Regression trees are very flexible and can capture local features of the data that linear models overlook. Moreover, they can handle cases with large-dimensional data. This becomes important when modeling stock returns because the identity of the best predictor variables is unknown and so must be determined empirically. On the other hand, the approach is sequential and successive splits are performed on fewer and fewer observations, increasing the risk of overfitting. There is also no guarantee that the sequential splitting algorithm leads to the globally optimal solution. To deal with these problems, we next consider a method known as boosting.

3.2.2   Boosting

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

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

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

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

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

where $e_{t+1,B-1} = y_{t+1} - f_{B-1}(x_t)$ is the forecast error remaining after $B-1$ boosting iterations. The solution is the regression tree that most reduces the average of the squared residuals $\sum_{t=1}^{T} e_{t+1,B-1}^2$ and $\hat{c}_{j,B}$ is the mean of the residuals in the jth state. Figure 1 shows that as the number of boosting iterations increases, the area covered by individual states shrinks and the fit becomes better.

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

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

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

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

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

Each tree is fitted on a randomly drawn subset of the training data, whose length is set at one-half of the full sample, the default value most commonly used. Again this reduces the risk of overfitting. Importantly, since individual trees are fitted on different subsets of the data, the starting point of the algorithm, or more specifically the particular sequence of splits selected by the regression trees, has very little effect on our results.

Finally, the empirical analysis minimizes mean absolute errors. We do this in light of a large literature suggesting that squared-error loss places too much weight on observations with large residuals. This is a particular problem for fat-tailed distributions such as those observed for stock returns and volatility. By minimizing absolute errors, our regression model is more robust to outliers.

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

3.2.3 Measuring the Effect of Individual Variables

In a linear model the importance of a particular state variable can be measured through the magnitude and statistical significance of its slope coefficient. This measure is not applicable to regression trees since these do not impose linearity. As an alternative measure of influence, we instead consider the reduction in the forecast error every time a particular variable, x_p, is used to split the tree. Summing the reductions in forecast errors (or improvements in fit) across the nodes in the tree and across boosting iterations gives a measure of each variable's influence (Breiman (1984)). The more frequently a variable is used for splitting and the bigger its effect on reducing the forecast errors, the

greater its influence. If a variable never gets chosen to conduct the splits, its influence will be zero. Finally, the resulting measure of influence is divided by the summed influence across all variables to get a measure of relative influence. This sums to one and can be compared across predictor variables.

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

3.3 Empirical Risk-Return Estimates

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

3.3.1 Data

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

Specifically, our empirical analysis uses a data set comprising monthly stock returns along with a set of predictor variables previously analyzed in Welch and Goyal (2008) extended to cover the sample 1927-2008.[4] Stock market returns are tracked by the S&P 500 index and include dividends. A short T-bill rate is subtracted to obtain excess returns. For brevity we refer to these simply as the returns.

The predictor variables fall into three broad categories. First, there are valuation

---

[4]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.

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 $t + 1$.

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

Market variance is unobserved, so we follow a large recent literature in proxying it through the realized variance. Specifically, let $r_{i,t}$ be the daily return on day i during month t and let N_t be the number of trading days during that month. Following, e.g., French, Schwert, and Stambaugh (1987) and Schwert (1989) we construct the realized

variance measure

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

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

3.3.2 Estimates of Expected Return and Conditional Volatility

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

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

Here x_t is a set of publicly available predictor variables and $\hat{\theta}_\mu$ and $\hat{\theta}_\sigma$ are estimates of the parameters of the expected return and volatility models, respectively. We provide further details of these estimates below, but first turn to the risk-return model.

3.3.3 Linear Estimates of the Risk-Return Relation

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

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

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

Empirical estimates of this model are shown in Panel A1 of Table 1. For the full sample, 1927-2008, we find evidence of a positive and significant linear relation between the contemporaneous volatility and expected returns with a t-statistic of 2.7.

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

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

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

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

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

As a simple first measure of the nature of the misspecification of the linear risk-return relation, we adopt the threshold regression approach of Hansen (2000) and esti-

mate a piece-wise linear model relating the expected return to the conditional variance:

$$\hat{\mu}_{t+1|t} = \begin{cases} -0.003 \ + \ 25.538 \ \ \hat{\sigma}^2_{t+1|t} \text{ for } \hat{\sigma}^2_{t+1|t} \leq (0.0253)^2 \\ (-1.492) \quad (6.229) \\ \\ \ \ \ 0.009 \ - \ \ 0.542 \ \ \hat{\sigma}^2_{t+1|t} \text{ for } \hat{\sigma}^2_{t+1|t} > (0.0253)^2 \\ (20.730) \quad (-5.219) \end{cases} \tag{10}$$

At low levels of the conditional variance (corresponding to annualized volatility levels below 9%), a strongly positive and significant mean-variance relation emerges. In contrast, for higher values of the conditional variance, the relationship is negative and strongly significant. A linearity test for equal slope of the two segments is rejected with a p-value well below 0.01%.[5] This demonstrates the limitations of linear and monotonic models for the relation between the conditional mean and variance and indicates that the risk-return relation is inverted at high levels of volatility.

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

3.3.4 Flexible Risk-Return Model

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

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

where now f is estimated by means of the BRTs. As noticed in Section I.C, we can no longer measure the importance of the explanatory variables through their slope coeffi-

---

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

cients, so instead Panels A2 and B2 in Table I present estimates of the relative influence of the three variables in this model using either $\hat{\sigma}_{t+1|t}$ or $\hat{\sigma}_{t+1|t}^2$ as the measure of risk.[6] The relative weight on current conditional volatility is 8% for the full sample which is statistically significant at the 5% level.[7] This weight is similar to that of the lagged volatility (9%). The weight on the lagged expected return (83%) is higher, which is unsurprising since the expected return is quite persistent and so its lagged value is likely to be important in this model. This finding is also consistent with the larger coefficient and t-statistic on lagged expected returns in the linear model. Interestingly, in the first subsample, 1927-67, the current conditional volatility obtains a large (and significant) weight of 21%, but the weight declines to 10% in the second subsample, 1968-2008, where it fails to be significant.

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

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

---

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

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

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

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

A second concern is that the BRTs may overfit the data. To address this point, we computed out-of-sample forecasts of stock returns and volatility, using 1970-2008 as the evaluation period and making use only of historically available information as opposed to full-sample estimates. In unreported results we find that, for the return forecasts, the prevailing mean model of Welch and Goyal (2008) and a multivariate linear regression model generate negative out-of-sample R^2-values, whereas the BRT model generates more precise out-of-sample forecasts and a positive R^2-value. For the volatility forecasts, we find that the BRTs produce more precise out-of-sample forecasts than those from a GARCH(1,1) model, a MIDAS model from Ghysels, Santa-Clara, and Valkanov (2005), and an autoregressive 'realized volatility' model. This out-of-sample analysis suggests that the boosted regression tree estimates of the conditional mean and volatility are not overfitting the data.

---

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

3.3.5 Comparison with Existing Approaches

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

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

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

To understand why linear models for the volatility and expected returns are misspecified, we study the BRT estimates of the conditional mean and volatility in more detail. These are unconstrained and so are able to reveal the nature of any deviations from linearity. The top row in Figure 4 presents partial dependence plots for the three most important predictor variables in the BRT model for expected returns, namely inflation, the earnings-price ratio and the relative return. The relation between expected

stock returns and these predictor variables is highly nonlinear. At negative levels of inflation the relation between the rate of inflation and expected returns is either flat or rising. Conversely, at positive levels of inflation, higher consumer prices are associated with lower mean returns. Although the relation between expected stock returns and the log earnings-price ratio is always positive, it is strongest at low or high levels of this ratio, and gets weaker at medium levels.

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

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

3.3.6 Tests of Monotonicity

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

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

with high conditional volatility, mean returns should rise.[9]

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

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

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

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

To test this hypothesis, we use the test statistic of Wolak (1989). The null that the conditional mean increases monotonically in the level of conditional volatility is rejected if there is sufficient evidence against it. Conversely, a failure to reject the null implies that the data is consistent with a monotonically increasing relation between the conditional mean and conditional volatility. The test statistic has a distribution that, under the null, is a weighted sum of chi-squared variables whose critical values can be computed via Monte Carlo simulation.

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

Test results are reported in Panel A of Table III. 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. Panel B shows similar findings for

---

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

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

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

the VIX measure of volatility. These results demonstrate that a monotonically increasing relation between the conditional mean and the conditional volatility is strongly rejected, providing evidence of a nonlinear mean-volatility or mean-variance relation.

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

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

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

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

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

is sufficiently high, this can lead to a non-monotonic shape of the risk-return relation.

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

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

---

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

is a negative relation between the price-earnings ratio and expected returns.[13]

We next provide a more detailed example of these effects using a model that matches the non-monotonic risk-return relation observed empirically.

3.4.1 Illustration from a Simple Model

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

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

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

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

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

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

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

---

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

excess returns are given by:

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

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

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

Following the analysis in Garcia, Meddahi, and Tedongap (2008), we consider a four-state model with parameters for the consumption and dividend growth processes calibrated to match the empirical moments reported by Lettau, Ludvigson, and Wachter (2008). Specifically, the means and volatilities of the consumption and dividend growth processes across the four states (in percent/year) are $\mu_c = \left(\begin{array}{cccc} 0.62 & 0.62 & -0.32 & -0.32 \end{array}\right)$; $\mu_d = \left(\begin{array}{cccc} 2.80 & 2.80 & -1.45 & -1.45 \end{array}\right)$; $\sigma_c = \left(\begin{array}{cccc} 0.75 & 0.40 & 0.75 & 0.40 \end{array}\right)$; $\sigma_d = \left(\begin{array}{cccc} 3.36 & 1.82 & 3.36 & 1.82 \end{array}\right)$. Volatility of consumption and dividend growth is highest in states one and three and lowest in states two and four, while the mean growth rates are highest in states one and two and lowest in states three and four. Consistent with a configuration in Garcia, Meddahi, and Tedongap (2008), preference parameters are set to $\psi = 0.5$, $\gamma = 30$, $\delta = 0.9925$. The only additional parameters in our analysis are the pairwise correlations between dividend and consumption growth which we set to $\rho = \left(\begin{array}{cccc} 3/4 & 1/4 & 1/4 & 1/4 \end{array}\right)$, so that the first state has a higher correlation parameter than the other states. However, the shape of the risk-return relation does not appear to be overly sensitive to the exact choice of correlation parameters.

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

the market portfolio acts as a hedge in state three. Figure 5 shows how this translates into a non-monotonic relation between the conditional volatility and expected returns in a way that matches our empirical findings.

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

$$Cov_t(R^e_{t+1}, \Delta c_{t+1}) = \left(\lambda'_2 \zeta_t\right) \left((\lambda_1 + e)' P \zeta_t\right) \rho' \zeta_t. \tag{19}$$

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

3.5   Conditional Covariance Risk

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

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

Here a_W measures the representative investor's relative risk aversion and b_{Wx} depends on the sensitivity of the investor's indirect utility function with respect to wealth (W)

and state variables (x). Leaving out the conditional covariance term from Eq. (20) could lead to omitted variable bias and so must be addressed (Guo and Whitelaw (2006)).

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

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

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

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

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

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

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

3.5.1 Realized Covariance

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

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

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

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

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

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

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

3.5.2   Empirical Results

We next extend the risk-return model to include estimates of our new conditional covariance measure in addition to the earlier measure of conditional volatility.

We estimate both linear and flexible risk-return specifications that control for dynamic effects:

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

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

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

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

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

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

Figure 7 shows that the partial dependence plots for the joint model in Eq. (26) further corroborate these findings. Expected returns increase monotonically in the conditional covariance, whereas the expected return-conditional volatility relation rises

---

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

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

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

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

at first but then declines at higher levels of volatility. Moreover, expected returns vary by approximately 5% per annum due to variations in the conditional covariance but change by less than 2% per annum due to variations in the conditional variance.

Guo and Whitelaw (2006) also include a covariance estimate of time-varying investment opportunities in their analysis of the risk-return relation. In common with much of the literature, they assume that f is a linear function of both $\widehat{\sigma}^2_{t+1|t}$ and $\widehat{cov}_{t+1|t}$ and they compute $\widehat{\sigma}^2_{t+1|t}$ and $\widehat{cov}_{t+1|t}$ from linear projections on observable state variables. However, the plots in the bottom row in Figure 4 show that the conditional covariance is a highly nonlinear function of the most important state variables such as inflation, the payout rate and long term returns, suggesting that linear models for the covariance are misspecified. Corroborating this, the bottom rows of Panel B in Table II show that linear specifications for the conditional covariance are clearly misspecified. While there is evidence that the BRT covariance estimates are also misspecified, this is driven by a single observation following the default of Lehmann (October 2008) which represents an extreme outlier for the realized covariance measure. Without this single observation, the BRT estimates appear to be correctly specified, while the linear estimates are not.

3.6 Conclusion

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

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

The non-monotonic trade-off between conditional volatility and expected returns

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

Chapter 3 is currently being prepared for submission for publication of the material. Rossi, Alberto; Timmermann, Allan. The dissertation author was the primary investigator and author of this material.

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

Estimates of the risk-return trade-off

A1. Linear volatility model

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

Sub-Samples	$\hat{\sigma}_{t+1|t}$ (t-stat)	$\hat{\sigma}_{t|t-1}$ (t-stat)	$\hat{\mu}_{t|t-1}$ (t-stat)	R^2
1927-2008	0.058 (2.76)	-0.079 (-3.79)	0.659 (26.78)	44.40%
1927-1967	0.056 (2.07)	-0.104 (-3.92)	0.584 (15.41)	41.54%
1968-2008	0.048 (1.42)	-0.025 (-0.72)	0.662 (19.29)	42.93%
Common Factor Model				
1960-2007	0.007 (0.13)	-0.011 (-0.21)	0.607 (17.65)	36.8%
VIX Volatility Proxy				
1986-2008	0.053 (1.75)	-0.042 (-1.33)	0.656 (14.07)	42.35%

A2. Flexible volatility model

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

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

B1. Linear variance model

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

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

B2. Flexible variance model

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

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

Table 3.2: This table presents Ramsey RESET specification tests applied to different models for the risk-return relation (Panel A) or the underlying conditional mean (Panel B.I), variance (Panel B.II) and covariance (Panel B.III). The null is that a model is correctly specified, so a small p-value (i.e., below 0.05) indicates misspecification. The linear and boosted regression tree (BRT) models use the state variables described in Section II.A as predictors, while the MIDAS and GARCH(1,1) models are based on past returns. The baseline results are based on the full sample (1927-2008), while the factor results use a shorter period (1960-2007) for which data on 131 underlying economic variables are available.

Model Specification Tests

A. Risk-Return Relation

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

B. Moment Estimates
B.I. Mean

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

B.II. Variance

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

B.III. Covariance

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

Table 3.3: This table presents the results of a test of whether the relationship between conditional risk and expected returns is monotonic after marginalizing out the effect of lagged risk and lagged expected returns. The test uses pairs of expected return, risk observations that are sorted on the basis of the conditional risk measure (volatility or VIX), yielding groups of observations corresponding to different levels of risk. The number of monthly observations in the small, medium and large groups or "portfolios" are approximately 40, 50 and 65. We then test if the associated mean return is monotonically increasing as we move from low to high risk observations. Each entry reports the p-value of a Wolak test that entertains the null hypothesis of an increasing and monotonic relation between expected returns and conditional volatility against the alternative of a decreasing or non-monotonic relation. Small p-values indicate rejection of a monotonically increasing risk-return relation. In panel A, estimates of the conditional mean and volatility are based on boosted regression trees that use the state variables from Section II.A. The Chicago Board Options Exchange Volatility Index, also known as the VIX, is used as a proxy for the conditional volatility in Panel B.

Tests for a monotonically increasing risk-return relation

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

Table 3.4: This table reports estimates of the effect on the conditional mean return, $\hat{\mu}_{t+1|t}$, of the conditional variance, $\hat{\sigma}^2_{t+1|t}$, the conditional covariance between stock returns and changes to economic activity, $\widehat{cov}_{t+1|t}$, the lagged conditional mean return, $\hat{\mu}_{t|t-1}$, lagged variance, $\hat{\sigma}^2_{t|t-1}$, and lagged covariance, $\widehat{cov}_{t|t-1}$. Panel A reports estimates and t-statistics from a linear model. Panel B shows relative influence estimates (in percent) from the boosted regression tree model. In parentheses we present the significance of the relative influence estimates by way of Monte Carlo p-values. Panel C presents the results of a test of whether the relation between conditional variance or conditional covariance) and expected returns is monotonic after marginalizing out the effect of the other variables in the model. Each entry reports the p-value of a Wolak test that entertains the null hypothesis of an increasing and monotonic relation between expected returns and conditional variance (or conditional covariance). Small p-values indicate rejection of a monotonically increasing relation. The number of observations in the small, medium and large portfolios are approximately 40, 50 and 65.

Estimates and tests of the covariance model

A. Linear model

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

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

B. Flexible model
Relative influence measures

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

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

C. Monotonicity tests

	Group size		
	Small	**Medium**	**Large**
Conditional Variance	0.0000	0.0260	0.2567
	Small	**Medium**	**Large**
Conditional Covariance	0.2355	0.2359	0.8242
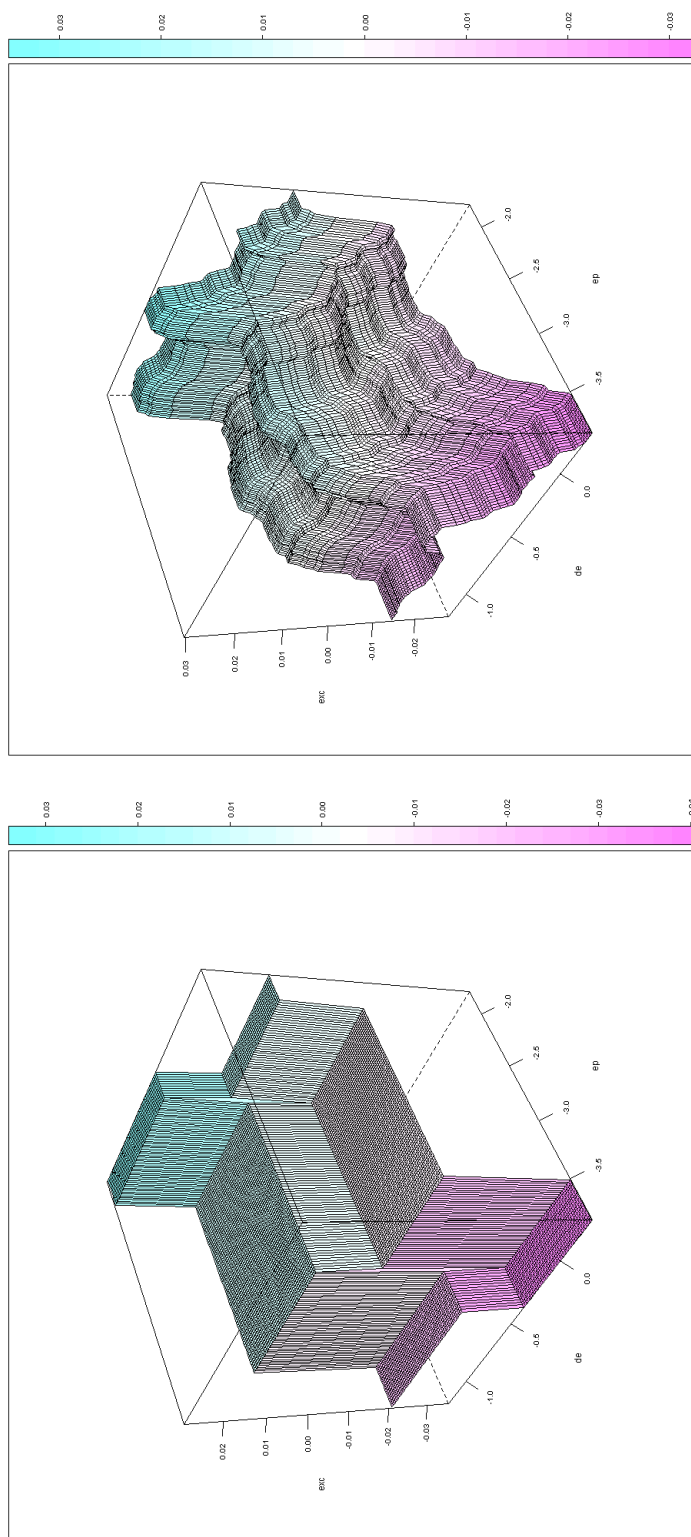
(a) 3 Iterations

(b) 5000 Iterations

Figure 3.1: Fitted values of market excess returns (exc) as a function of the log dividend-earnings ratio (de) and the log earnings-price ratio (ep). Both plots are based on boosted regression trees with two nodes in each split. The panel on the left uses three boosting iterations, while the right panel uses 5,000 iterations. The plots are based on monthly data from 1927 to 2008.
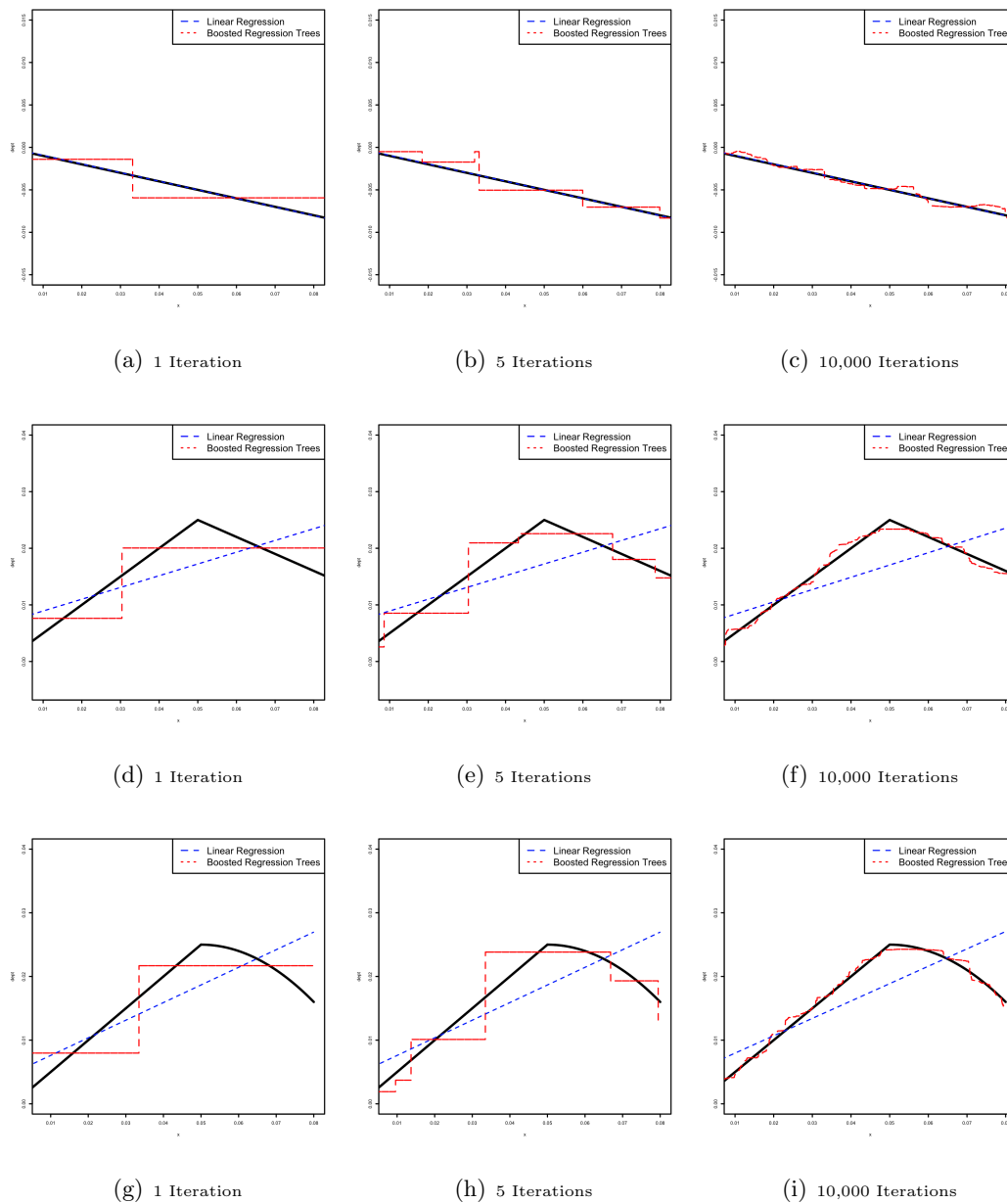
(a) 1 Iteration   (b) 5 Iterations   (c) 10,000 Iterations

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

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

Figure 3.2: **Fitted values from linear regression and boosted regression trees.** The top row assumes the true relation is linear, the middle row assumes an inverted V-shaped relation, while the bottom row assumes the true relation is first linear, then quadratic. The number of boosting iterations is set to one (left column), five (middle column), or 10,000 (right column).
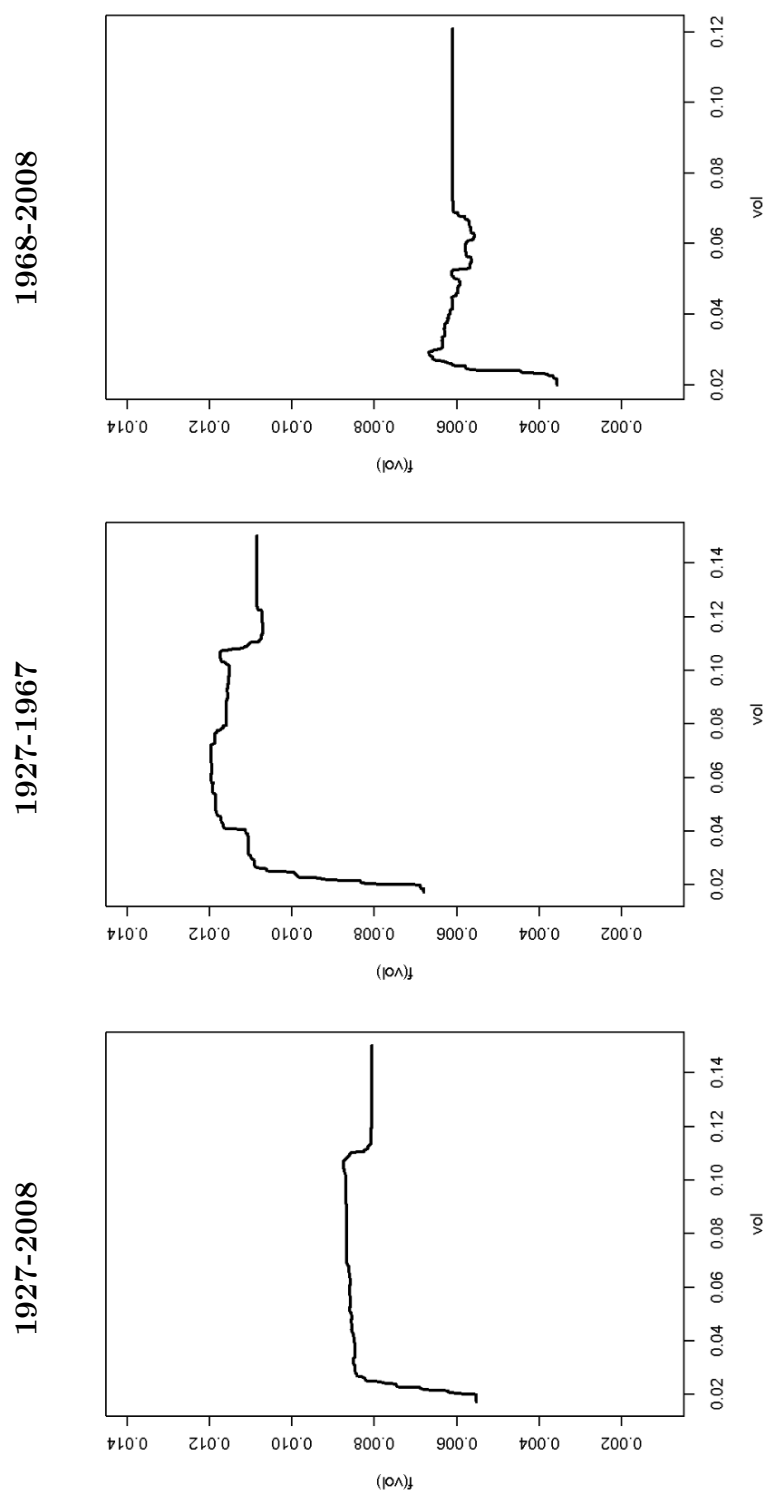
Figure 3.3: **Expected return-conditional volatility trade-off.** The figures show partial dependence plots for the conditional equity premium as a function of the conditional volatility (vol). The plots are based on boosted regression trees, using data for three samples, 1927-2008 (left panel), 1927-1967 (middle panel) and 1968-2008 (right panel). The horizontal axis covers the sample support of the conditional volatility, while the vertical axis tracks the resulting change in the conditional mean as a function of the conditional volatility.

Panel A. Expected Returns
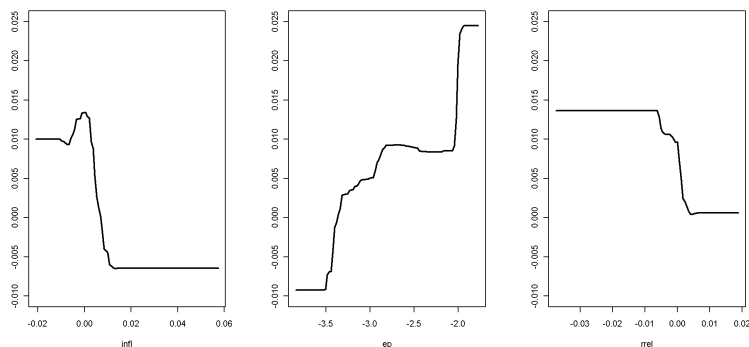
Panel B. Volatility
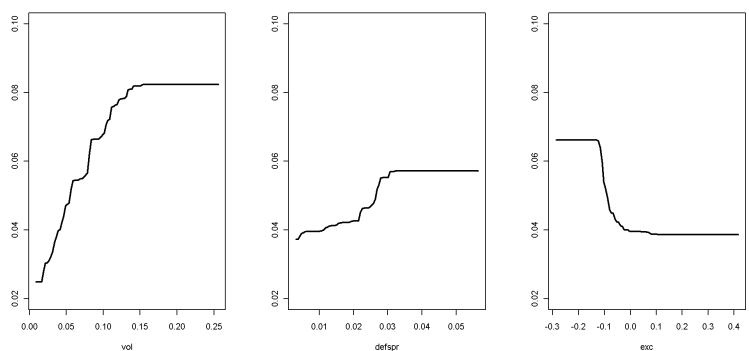
Panel C. Covariance
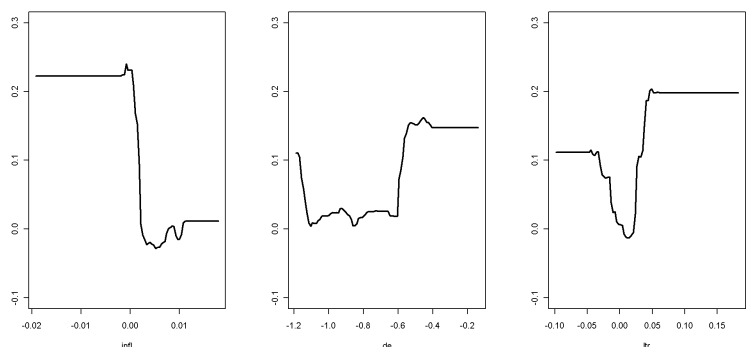
Figure 3.4: **Effect of predictor variables on conditional moments.** The figures present partial dependence plots for the mean excess return (Panel A), the conditional volatility (Panel B) and the conditional covariance (Panel C) based on the three predictor variables with the highest relative influence during 1927-2008, namely inflation (infl), the log earnings price ratio (ep), and the detrended T-Bill rate (rrel) for returns; the lagged stock market volatility (vol), the default spread (defspr) and the excess return on stocks (exc) for volatility; inflation (infl), the log dividend earnings ratio (de), and the long-term rate of return (ltr) for the covariance. The horizontal axis covers the sample support of the individual predictor variables, while the vertical axis tracks the change in the conditional equity premium as a function of the individual predictor variables.
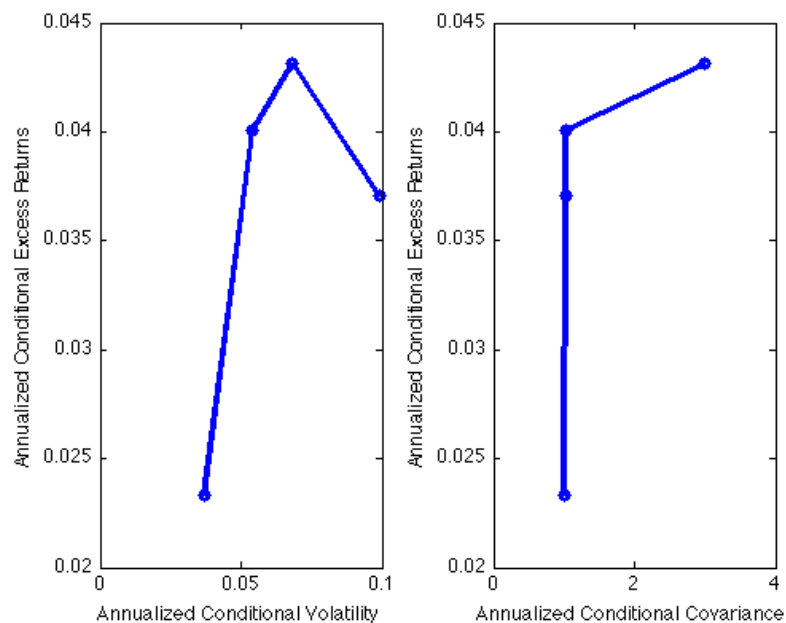
Figure 3.5: **Risk-Return relation in a dynamic asset pricing model.** This figure plots the conditional expected excess returns against the conditional volatility (left graph) and the conditional covariance (right graph) implied by the four-state regime switching model proposed by Garcia et al. (2008). The left plot shows a non-monotonic relation between conditional volatility and expected returns. The state with the highest conditional volatility has a weaker correlation between consumption and dividend growth than the other states. This means that the market portfolio provides a partial hedge against adverse shocks to consumption in this state, resulting in a reduced equity premium. Conversely, the right plot shows that there is a monotonically increasing relation between expected returns and the conditional covariance between consumption growth and stock returns.
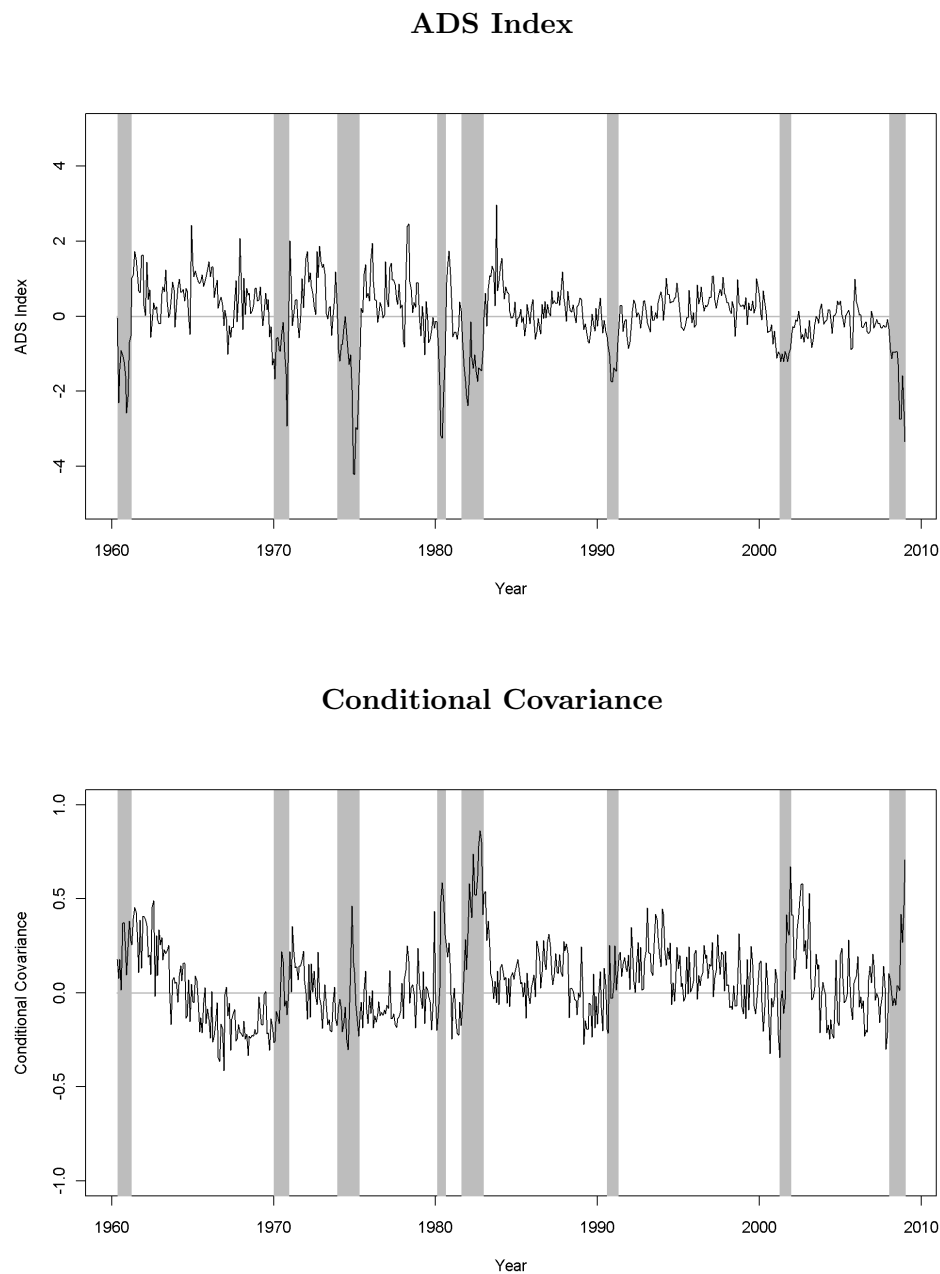
ADS Index

Conditional Covariance

Figure 3.6: **ADS index and the conditional covariance measure.** This figure plots the ADS index at the monthly frequency in the top panel and the scaled conditional covariance between changes in the ADS index and stock returns in the bottom panel. The scaled conditional covariances are obtained as follows. First, monthly realized covariances between changes in the ADS index and stock returns are obtained using observations at the daily frequency. The scaled changes in the ADS Index are obtained by dividing the change to the ADS Index by the standard deviation of returns times the standard deviation of changes to the index. Finally, the conditional covariances are estimated by way of boosted regression trees.

Panel A. Baseline Model
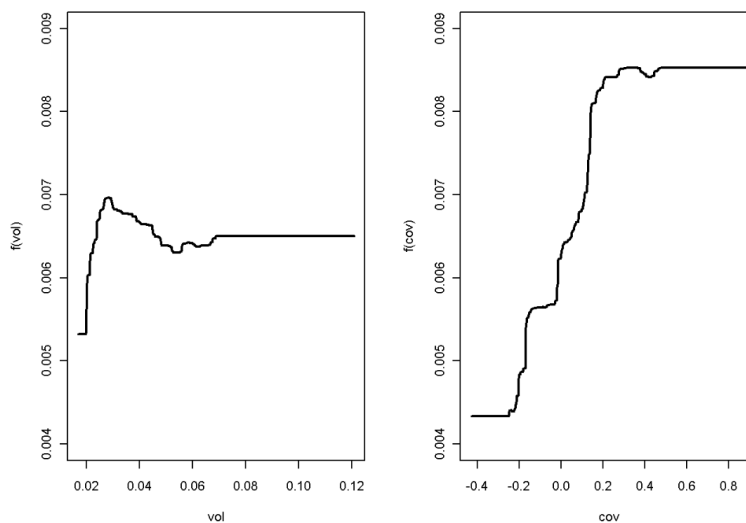
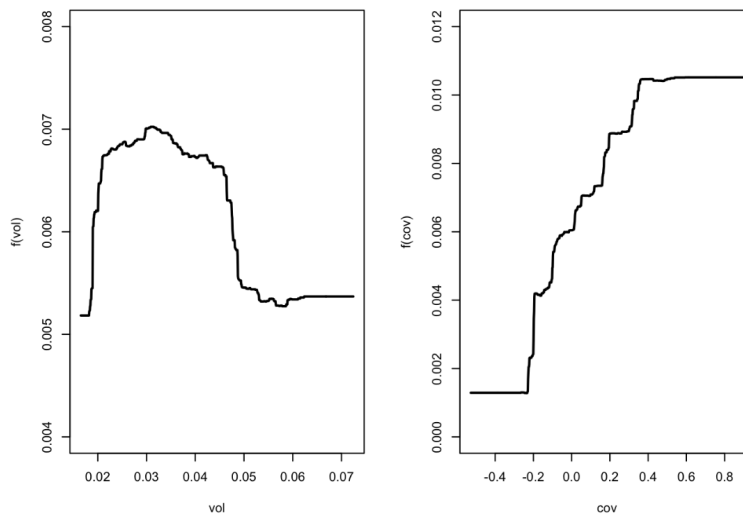**Panel B. Model With Factors**

Figure 3.7: **Risk-return trade-off in the model with the conditional covariance risk measure.** The figure shows partial dependence plots for the conditional mean return as a function of the conditional volatility (vol) and the conditional covariance between stock returns and changes to economic activity (cov). The plot is based on a boosted regression tree, using data over the sample 1960-2008 in Panel A and 1960-2007 in Panel B. The horizontal axis covers the sample support of each predictor variable, while the vertical axis tracks the change in the conditional mean as a function of the individual state variables. In Panel A, the conditioning information is the predictor variables described in Section II.A. In Panel B the conditioning information is the principal components derived from a set of 131 state variables and the three most important variables selected from the predictors in Section II.A.

Bibliography

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

AROUBA, S., F. DIEBOLD, AND C. SCOTTI (2009): "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 27(4).

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

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

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

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

BONOMO, M., R. GARCIA, N. MEDDAHI, AND R. TÉDONGAP (2011): "Generalized Disappointment Aversion, Long-run Volatility Risk, and Asset Prices," *Review of Financial Studies*, 24(1), 82–122.

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

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

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

BREEN, W., L. GLOSTEN, AND R. JAGANNATHAN (1989): "Economic significance of predictable variations in stock index returns," *Journal of Finance*, 44(5), 1177–89.

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

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

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

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

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

FRENCH, K. R., G. W. SCHWERT, AND R. F. STAMBAUGH (1987): "Expected stock returns and volatility," *Journal of Financial Economics*, 19(1), 3–29.

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

GARCIA, R., N. MEDDAHI, AND R. TEDONGAP (2008): "An analytical framework for assessing asset pricing models and predictability," *Université de Montréal Working Paper*.

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

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2005): "There is a Risk-Return Tradeoff After All," *Journal of Financial Economics*, 76(3), 509–548.

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

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

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

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

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

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

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.

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

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

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

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

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

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

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

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

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

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

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

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

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