LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
CENTER FOR QUANTITATIVE RISK ANALYSIS
CEQURA

Stefan Mittnik, Nikolay Robinzonov and Martin Spindler

Boosting the Anatomy of Volatility

Working Paper Number 06, 2012
Center for Quantitative Risk Analysis (CEQURA)
Department of Statistics
University of Munich

http://www.cequra.uni-muenchen.de

Boosting the Anatomy of Volatility

Stefan Mittnik[§,*] Nikolay Robinzonov[§] Martin Spindler[§,‡]

November 13, 2012

[§] Institute of Statistics and Center for Quantitative Risk Analysis
Ludwig–Maximilians–Universität München
Akademiestr. 1/I, 80799 Munich, Germany

[‡] Max Planck Institute for Social Law and Social Policy, Munich, Germany

Abstract

Financial risk, commonly measured in terms of asset–price volatility, plays a fundamental role in investment decisions and financial–market regulation. In this paper, we investigate a new modeling strategy to better understand the forces driving market risk. We use componentwise gradient–boosting techniques to identify financial and macroeconomic factors influencing volatility and to assess the specific nature of that influence. Componentwise boosting is a sequential learning method, which is capable of handling a large number of predictors and—in contrast to other machine learning techniques—which gives rise to straightforwardly interpretable estimates.

Considering a range of potential risk drivers, we employ componentwise boosting to derive monthly volatility predictions for stock, bond, commodity, and foreign exchange indices. Comparisons with a common benchmark model show that the approach improves out–of–sample volatility forecasts, especially for medium– and long–run horizons. Moreover, we find that various risk drivers affect future volatility in a nonlinear fashion.

Keywords: componentwise boosting, financial risk, forecasting, GARCH, variable selection.

*Corresponding author. *Email:* finmetrics@stat.uni-muenchen.de, *Tel.:* +49 89 2180 3224, *Fax:* +49 89 2180 5044, *Address:* Ludwig–Maximilians–Universität, Institute of Statistics, Chair of Financial Econometrics, Akademiestr. 1/I, 80799 Munich, Germany.

1 Introduction

The importance of understanding and reliably modeling financial risk has—again—become evident during the market turbulences in recent years. Accurate volatility predictions for asset–prices are crucial when projecting risk measures, such as Value at Risk (VaR) or Expected Shortfall, commonly used in risk assessment and the design of risk–mitigation strategies. Although there has been a long tradition of attempting to predict asset prices (cf. Goyal and Welch, 2003, Welch and Goyal, 2008, Cochrane and Piazzesi, 2005, Lustig et al., 2011), the intense interest in volatility modeling took off only after the seminal works of Engle (1982) and Bollerslev (1986), and has since been extensively researched in financial econometrics.

Despite this tremendous interest, the vast majority of studies on volatility prediction is confined to using only past return histories as consitional information.[1] Only relatively few studies have analyzed to what extent the information contained in other financial or macroeconomic variables helps to improve volatility predictions. Employing autoregressive models, Schwert (1989) analyzes the relation of stock volatility and macroeconomic factors, such as GDP fluctuations, economic activity, and financial leverage. Engle et al. (2008) use inflation and industrial production by combining a daily GARCH process with a mixed–data sampling polynomial applied to monthly, quarterly, and bi-annual macroeconomic variables. Paye (2012) and, especially, Christiansen et al. (2012) consider extended sets of macroeconomic factors and a broader range of asset classes. Both use conventional linear approaches to model log–transformed realized volatility and include lagged volatility as well as financial and macroeconomic factors as regressors. In view of the limited number of studies and their varying approaches, there is little or no consensus concerning the usefulness of financial and macroeconomic variables for volatility prediction.

The question of whether and how macro factors influence the volatility of asset prices is the focus of this paper. To address this question and to gain deeper insight into the nature of volatility processes, we employ so–called boosting techniques, a special machine learning method. As will be shown, boosting techniques enable us not only to identify factors driving market volatility but also to assess the specific nature of their impact. Employing a broad set of potential macroeconomic and financial variables, we specify a flexible model, which is capable of capturing the factor's—linear and nonlinear—influences on volatility. Contrasting most of the existing literature, which focuses on stock market volatility, we are in line with Christiansen et al. (2012) and analyze four diverse asset classes, namely, stocks, bonds, commodities, and foreign exchange.

Although boosting has been proven to be a useful approach in many empirical applications, it has more or less been ignored in empirical economics and finance. Among the very few exceptions are Bai and Ng (2009), who use it for predictor selection in factor-augmented autoregressions, and Audrino and Bühlmann (2009), who apply it to model the daily volatility of stock indices. Our model differs from Audrino and Bühlmann (2009) in several respects, two of which we regard as particularly relevant. First, we go beyond the

[1] A comparison of alternative VaR forecasting strategies that follow this line is given in Kuester et al. (2006).

GARCH(1,1) specification by considering longer histories and allowing for exogenous factors to enter the model. The latter, as it turns out, clearly helps our understanding the nature of volatility processes. Second, we employ componentwise predictor selection instead of the componentwise knot selection in tensor–spline estimation. This leads to genuinely different models and has the attractive feature that certain subjective specification decisions can be avoided. Moreover, given our goal to better understand the impact of macro factors on volatility, we conduct our analyses at a monthly rather than daily frequency.

Altogether, the paper contributes to the existing literature on volatility modeling in several ways. We analyze the volatility of a range of relevant asset classes; we consider a broad set of possible macro drivers; and, by employing boosting techniques, gain deeper insight into the nature of the forces driving asset price volatility.

This paper is organized as follows. Section 2 details and illustrates the specific boosting algorithm we adopt. Section 3 describes the volatility measure and predictor variables employed and the way multi–step forecasting comparisons are conducted. Empirical results of an application to indices of four asset classes are presented in Section 4. Section 5 concludes.

2 Boosting Volatility

Boosting, as proposed in Freund and Schapire (1996), was originally designed to solve two–class classification problems by maximizing the confidence of a binary classifier. To do so and to achieve an arbitrarily high accuracy, it suffices that the classifier, also called *base learner*, performs only slightly better than random guessing, (Kearns and Valiant, 1994, Schapire et al., 1998). Friedman (2001) placed boosting into the regression framework, viewing it as a functional gradient descent technique.Boosting is especially suitable in applications where there is a large number of—possibly "similar"—predictors, as it curbs multicollinearity problems by shrinking their influence towards zero. Volatility modeling via gradient boosting was first proposed by Audrino and Bühlmann (2003), who adopted a GARCH-type framework, assuming a stationary return process of the form $y_t = \sigma_t \varepsilon_t$, $\varepsilon_t \overset{iid}{\sim} N(0,1)$ and a rather general dependence structure between σ_t and past returns. Their approach is, however, mainly suited for prediction rather than economic analysis, as the resulting model lacks interpretability. A similar model, with neural networks as base learners, was proposed by Matías et al. (2010).

Below, we use so-called componentwise gradient boosting (see **?**), which is designed to simultaneously select relevant factors and to model the specific nature of their impact. In the following, we detail our volatility boosting strategy by presenting the underlying model specification and the particular boosting algorithm and, subsequently, using a small simulation study, illustrate our approach.

2.1 Gradient Boosting

Our modeling framework corresponds to the exponential ARCH specification put forth in Nelson (1991) but allows us, in a rather flexible way, to include a large number of risk drivers that potentially affect volatility. The total number of predictors can be very large and may even exceed the sample size. The specific form of our model is given by

$$\begin{aligned} y_t &= \exp({}^{\eta_t}/_2)\varepsilon_t \\ \eta_t &= \beta_0 + f_{\text{time}}(t) + f_{\text{yr}}(n_t) + f_{\text{month}}(m_t) + \sum_{j=1}^{s} f_j(y_{t-j}) + \sum_{k=1}^{q}\sum_{j=1}^{p} f_{j,k}(x_{t-j,k}) =: \eta(\mathbf{z}_t), \end{aligned} \tag{1}$$

where $y_t = \log(P_t/P_{t-1})$ denotes the logarithmic return, P_t is the asset prices at time t, and $\varepsilon_t \overset{iid}{\sim} N(0,1)$. The r-dimensional vector $\mathbf{z}_t = (1, t, n_t, m_t, y_{t-1}, \ldots, y_{t-s}, x_{t-1,1}, \ldots, x_{t-p,1}, \ldots, x_{t-1,q}, \ldots, x_{t-p,q})^\top$, with $r = s + qp + 4$, contains the predictor realizations available at or prior to time $t-1$. Functions $f_{\text{time}}(\cdot), f_{\text{yr}}(\cdot)$ and $f_{\text{month}}(\cdot)$ capture possible deterministic trend and seasonal components in volatility; $f_j(y_{t-j}), j = 1, \ldots, s$, capture the influence of past returns; and $f_{j,k}(x_{t-j,k}), j = 1, \ldots, q$, are functions of lagged predictors.

We specify all $f_{\cdot}(\cdot)$ functions in (1) as regression trees. Regression trees are a nonparametric technique that can handle complex and abruptly varying forms of dependence by recursively partitioning the predictor domain into regions with similar response values and assigning a constant value to the response within each subdomain.[2] Specifically, we use conditional inference trees (Hothorn et al., 2006). In general terms, the regression tree approach can be interpreted as a regime-dependent volatility–response model, which partitions the predictor space according to the magnitude with which conditional volatility responds. For partitioning decisions, we use the permutation test developed in Strasser and Weber (1999), though other split criteria might be conceivable. Both linear estimation and non-parametric smooth estimation of as well as a combination of the two can be employed to estimate $f_{\cdot}(\cdot)$.[3]

We estimate (1) via componentwise gradient boosting, which derives the final model by sequentially combining a series of individual predictor components. To avoid overfitting in the first step, we control the bias–variance tradeoff by using a low-variance/high-bias model. In subsequent steps, this bias will be iteratively reduced, with the variance increasing at a slower rate (Bühlmann and Yu, 2003). Our estimation minimizes the expectation of some (with respect to η differentiable) loss function, L, such that $\hat{\eta} = \arg\min_\eta \mathrm{E}L(y_t, \eta(\mathbf{z}_t))$. To obtain a solution in the data rather than function space, we parameterize η by

$$\hat{\eta} = \arg\min_{\eta} \frac{1}{T}\sum_{t=1}^{T} L(y_t, \eta(\mathbf{z}_t; \boldsymbol{\beta})). \tag{2}$$

The solution to (2) is derived by reducing the empirical loss in successive steps. The final $\boldsymbol{\beta}$ estimate is given by the sum of the estimates obtained in each step. By doing so, our estimation strategy preserves interpretation—a property that does not necessarily hold for other parallel learning techniques, such as bagging or random forests.

To estimate the desired characteristic of the conditional distribution, the loss function, L, needs to be appropriately specified. Under the assumption $y_t|\mathbf{z}_t \sim N(0, e^{\eta_t})$, the negative

[2]For a detailed discussion of the algorithms behind regression trees, see Breiman et al. (1984).

[3]See Hothorn et al. (2011) for a software implementation.

conditional log-likelihood loss function and the negative gradient, respectively, become

$$L_t = \frac{1}{2}\left[\eta_t + \frac{y_t^2}{e^{\eta_t}}\right] \qquad \text{and} \qquad g_t = -\frac{\partial L_t}{\partial \eta_t} = \frac{1}{2}\left[\frac{y_t^2}{e^{\eta_t}} - 1\right]. \tag{3}$$

Boosting favors the direction given by the largest reduction in the empirical loss, i.e., the negative gradient. This means that we seek the solution by fitting the covariates against the negative gradient. Instead of jointly fitting all covariates, they are fitted individually against the gradient through base learners. These are typically—though not necessarily—well–known statistical models (such as linear regressions, generalized additive models or regression trees), which specify the connection between the response and the covariates. At each boosting step, only one covariate is included, namely the one which correlates most strongly with the negative gradient, i.e., the steepest direction to the loss minimum.

We can think of the base learners as isolated "sub-solutions" to the original optimization problem. Hence, we fit the individual covariates against the negative gradient via r individual models—in our case, conditional inference regression trees (Hothorn et al., 2006) with two nodes, also called "stumps." Modeling the dependence between the response and the covariate in terms of two constants assigned to disjoint groups is inflexible and does, in general, not fit the complete signal in a single step.[4] Further iterations will, however, reduce this bias. Furthermore, we shrink the coefficient towards zero, as proposed by Friedman (2001). Shrinkage helps to dampen the "greediness" of the gradient technique, which may otherwise be prone to neglect correlated predictor candidates, and "cures" the typical instability of forward selection methods (Breiman, 1996). The "right" amount of shrinkage is determined empirically and can safely vary between 1% and 10%. The specific choice affects mainly the computational time. Fitting the base learner will modify the gradient evaluation in the next step, so that with each step the covariates and gradients become more and more orthogonal.

Without stopping, boosting with stumps will inevitably overfit and ultimately lead to a perfect fit, making the model useless—especially for prediction. Therefore, an appropriate stopping rule is essential. The optimal number of boosting steps can be determined by bootstrapping, where we sample (with replacement) from the data with probability $1/T$ as if they originated from a multinomial distribution. Doing so, each sample makes use of roughly 64% of the original data for training, with the remaining, unselected data points being used for evaluation. We repeat this twenty–five times for a large number of boosting steps and choose the step number that produces the lowest average out–of–sample loss.

To summarize, the boosting algorithm we employ consists of the following steps:

1. Initialize function estimate $\hat{\eta}_t^{[0]} = \log\left(\frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})^2\right)$, $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$, $t = 1, \ldots, T$.

2. Specify regression trees as base learners: $f_i(z_{t,i}) = \sum_{j=1}^{J} \gamma_j I_{R_j}(z_{t,i})$, $\forall z_{t,i} \in \mathbf{z}_t$. We use stumps, so for each tree $J = 2$. Denote the number of base learners by r and set $m = 0$.

[4] Note that we can choose any statistical model as base learner. In our applications, a specification via stumps turned out to be a better choice than, for example, smooth P-splines or a simple linear model. This seems largely due to the abrupt changes we observe in volatility levels.

3. Increase m by one.

4. (a) Compute the negative gradient in (3) and evaluate $\hat{\eta}^{[m-1]}(\mathbf{z}_t)$, $t = 1, \ldots, T$.
   (b) Estimate the negative gradient, using the stumps specified in Step 2. This yields r vectors, where each vector is an estimate of the gradient.
   (c) Select the base learner, $\hat{f}^{[m]}_{\hat{s}_m}, \hat{s}_m \in \{1, 2, \ldots, r\}$, that correlates most with the gradient according to the residual–sum–of–squares criterion.
   (d) Update the current estimate by setting $\hat{\eta}^{[m]} = \hat{\eta}^{[m-1]} + \nu \hat{f}^{[m]}_{\hat{s}_m}$, where ν is regarded as a shrinkage parameter or as a step size.

5. Repeat Steps 3 and 4 until the stopping condition applies.

2.2 An Illustration

To illustrate our boosting approach, we run a small simulation with data generating process

$$\begin{aligned} y_t &= \exp({}^{\eta_t}/_2)\varepsilon_t \\ \eta_t &= 0.1 + 2x_{t,1} + 2\mathrm{I}_{[0.1,0.5]}(x_{t,2})x_{t,2} - 0.6\mathrm{I}_{[-0.5,-0.2]}(x_{t,3}) + 0x_{t,4} + 0x_{t,5} + 0x_{t,6}, \end{aligned} \tag{4}$$

with $\varepsilon_t \overset{iid}{\sim} N(0,1)$; $x_{t,i}$ being the t-th observation of $X_i \sim U[-0.5, 0.5]$, $i = 1, \ldots, 6$, $t = 1, \ldots, T$, $T = 400$; and $\mathrm{I}_A(\cdot)$ denotes the indicator function, i.e., $\mathrm{I}_A(x) = 1$, if $x \in A \subset \mathbb{R}$, and $\mathrm{I}_A(x) = 0$, otherwise. Note that only the first three covariates affect volatility—the first linearly, the second linearly only for $X_2 \in [0.1, 0.5]$, and the third as a step function. The last three covariates, X_4 through X_6, do not contribute, and are included to check for robustness against false detection. We choose linear base learners for all but the second and third predictors, which are fitted with a regression-tree base learner, so that we fit

$$\begin{aligned} y_t &= \exp({}^{\eta_t}/_2)\varepsilon_t \\ \eta_t &= \beta_0 + \beta_1 x_{t,1} + \sum_{j=1}^{J_1} \gamma_j^{(1)} \mathrm{I}_{R_j^{(1)}}(x_{t,2}) + \sum_{j=1}^{J_2} \gamma_j^{(2)} \mathrm{I}_{R_j^{(2)}}(x_{t,3}) + \beta_4 x_{t,4} + \beta_5 x_{t,5} + \beta_6 x_{t,6}, \end{aligned} \tag{5}$$

where $R_j^{(1)}$ and $R_j^{(2)}$ denote the estimated partitions in the domain of X_2 and X_3. Splitting decisions are made by using the permutation test (Strasser and Weber, 1999), which reflects the level of dependence between the gradient and the corresponding covariate. The test statistic is maximized among all possible split positions (see also Hothorn et al., 2006).

Ideally, the algorithm will recover the β and $\gamma^{(2)}$ parameter values specified in (5). This means that X_4, X_5 and X_6 should not be selected at all, i.e., $\beta_4 = \beta_5 = \beta_6 = 0$, and that the domain of X_3 should be partitioned, such that $X_3 \in [-0.5, -0.2]$ affects volatility. Regarding X_2, although having linear impact for $X_2 \in [0.1, 0.5]$ and none otherwise, we intentionally

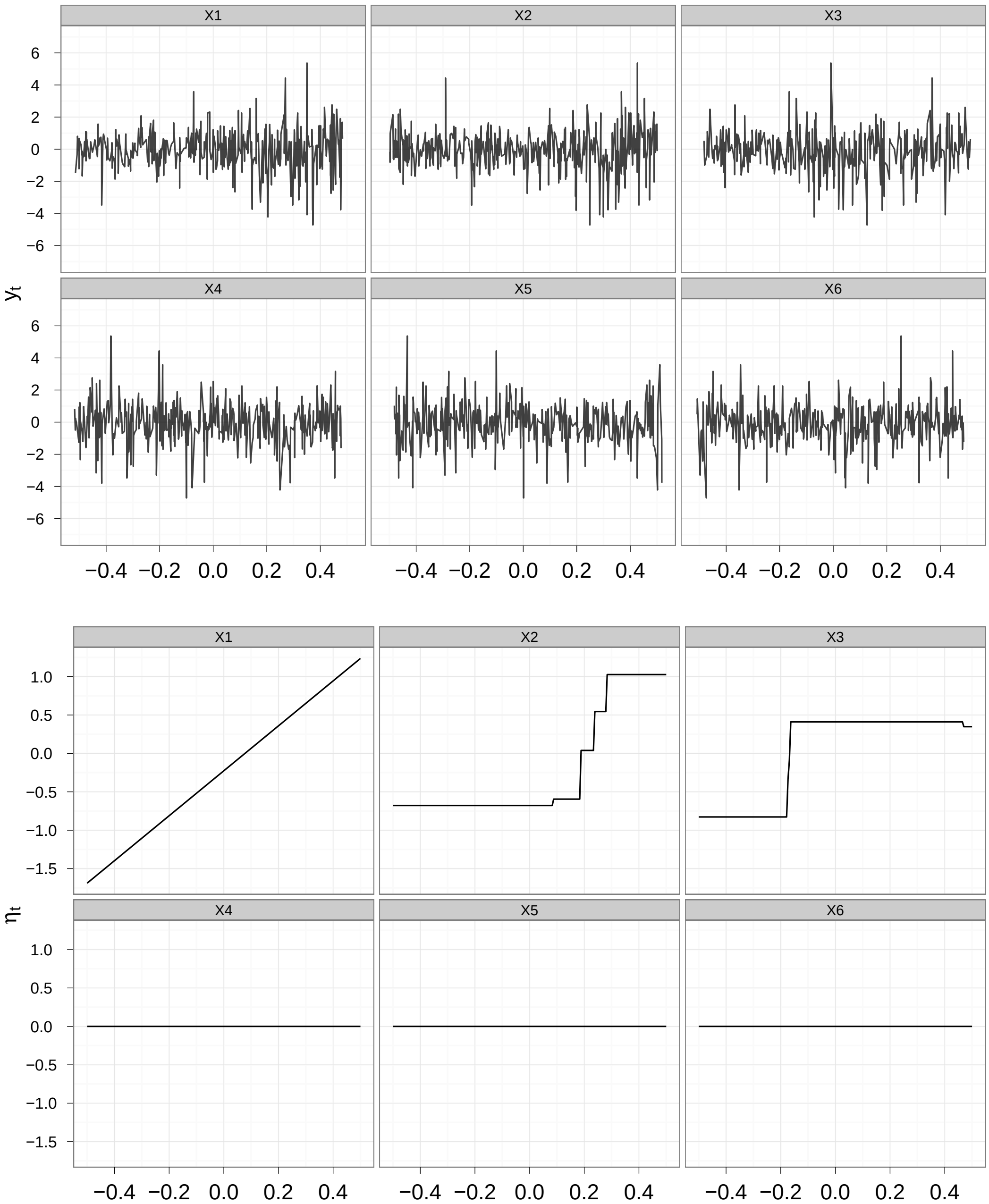

Figure 1: Partial return components (upper panel) simulated from (4), indicating how drivers X_1 through X_6 affect returns, and estimated partial log-volatility (lower panel).

chose an "incorrect" base learner, namely a step function, to see whether the influences can still be adequately approximated.

Figure 1 shows simulated, driver–specific return components (upper panel) and the estimated partial impacts on volatility (lower panel, log scale). The influence of the underlying volatility drivers appears to be captured reasonably well. The parameter estimate $\hat{\beta}_1 = 1.463$

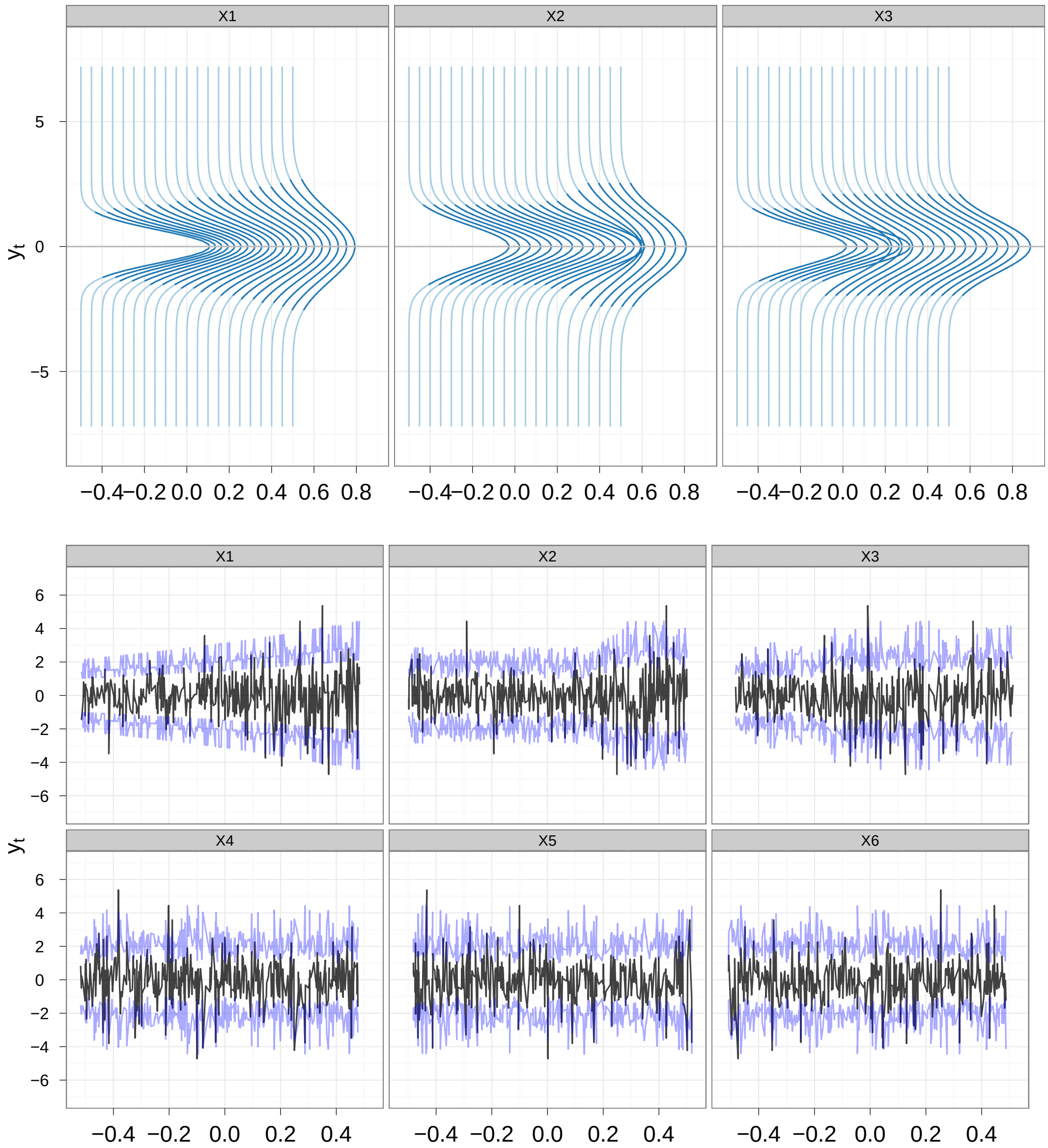

Figure 2: Partial conditional density estimates (upper panel) associated with X_1 through X_3 in (4). Dark line segments indicate estimated 95% interquantile ranges, the lighter ones show the estimated tails. Simulated return components (lower panel, black lines) associated with these partial conditional densities; the lighter lines represent 95% interquantile ranges.

is low due to parameter regularization via early stopping. This is typical for shrinkage methods, where the parameter estimates usually have smaller magnitudes than unregularized solutions and the bias vanishes as the sample size increases. The advantage of early stopping is that no redundant predictors are selected, i.e., $\hat{\beta}_4 = \hat{\beta}_5 = \hat{\beta}_6 = 0$. Furthermore, X_3 has the largest jumps near the right boundary of the interval $[-0.5, -0.2]$, and the linear structure of $X_2 \in [0.2, 0.5]$ is also captured reasonably well, despite the moderate sample size chosen.

The results shown in Figure 1 are typical in the sense that the variations in several hundred repetitions were small. Translating the log scale in Figure 1 back to standard deviations gives the estimate of the conditional density. Figure 2 (upper panel) shows the estimated partial densities for X_1, X_2 and X_3, with the central 95% interquantile ranges represented by the darker segments, and simulated return components associated with these conditional densities (lower panel). Visual inspection reveals that variations in volatility are closely captured, a finding that is supported by the fact that the estimates produce a coverage rate of 95.75% for the 95% interquantile range. The partial contribution of each covariate is readily obtained in an interpretable way: an increase in X_1 causes the variance to increase proportionally; X_2 has an increased impact on the variance for $X_2 \in [0.2, 0.5]$; the variance contribution markedly decreases for $X_3 \in [-0.5, -0.2]$; and, with $\hat{\beta}_4 = \hat{\beta}_5 = \hat{\beta}_6 = 0$, the conditional density of y_t remains, indeed, invariant with respect to X_4, X_5 or X_6.

With more detailed and readily interpretable insights into the role of particular risk drivers, the boosting strategy proposed here helps us to better understand the nature of volatility processes. To what extent this translates into better risk predictions will be addressed next.

3 An Empirical Application to Four Asset Classes

We present an application of our boosting approach to volatility prediction, considering four diverse asset classes. First, we briefly describe the data employed, i.e., the data for the assets to be modeled as well as the financial and macroeconomic factors entertained as potential volatility drivers. Then, we discuss the procedure used to evaluate the predictive performance.

3.1 The Data

We investigate the predictability of volatility of four asset types, namely, stocks, bonds, commodities, and foreign exchange, for each of which we select a representative index. The equity market is represented by S&P 500 futures contracts traded on the Chicago Mercantile Exchange; for the bond market, we use 10-year treasury note futures contracts traded on the Chicago Board of Trade; the commodity market is represented by Standard & Poor's GSCI commodity index; and we use a trade-weighted currency portfolio provided by the Federal Reserve Bank of St. Louis to proxy foreign currency investments. The latter is a weighted average of the foreign exchange value of the U.S. dollar against a broad set of currencies that circulate widely outside their countries of issue, including the euro zone, Canada, Japan, the United Kingdom, Switzerland, Australia, and Sweden. The data cover the period February 1983 to September 2010, amounting to 332 months in total. Summary statistics for the four return series and the logarithmic realized volatility series are given in Tables 1 and 2, respectively.

As potential volatility drivers we consider the 26 financial and macroeconomic factors summarized in Table 3. They include the financial variables Welch and Goyal (2008) use

Table 1: Descriptive statistics of the return series.

	Mean	Std Dev	Skewness	Kurtosis	AR1
Stock	0.0062	0.0457	-1.0116	5.8971	0.0661
Commodity	0.0057	0.0578	-0.6082	6.6808	0.1958
Bond	0.0016	0.0204	0.0504	3.8869	0.0501
FX	-0.0015	0.0213	0.0734	3.5620	0.0743

Table 2: Descriptive statistics of the log realized volatility series.

	Mean	Std Dev	Skewness	Kurtosis	AR1
Stock	-6.3278	0.9254	0.8149	2.0479	0.6669
Commodity	-6.1654	0.9327	0.3341	0.0438	0.7797
Bond	-8.0330	0.7128	-0.0012	0.0987	0.5807
FX	-8.0280	0.6943	0.0503	0.4122	0.5613

to predict stock returns, namely, book–to–market ratio, net equity expansion, term spread, relative T-Bill rate, relative bond rate, long–term bond return, and default spread. In addition, we include the three Fama–French factors: the U.S. market excess returns, the size factor, and the value factor. The set of financial predictors also contains the Pastor and Stambaugh (2003) liquidity factor, the return on the MSCI world stock index, the TED spread (i.e., the difference between the three-month LIBOR rate and the T-Bill rate), the Cochrane and Piazzesi (2005) bond factor, the return on the CRB spot index, the carry trade factor, as in Lustig et al. (2011), the return on dollar risk factor introduced by Lustig et al. (2011), and the FX average bid–ask spread (Menkhoff et al., 2011). Finally, we also include the International Monetary Fund's Financial Stress Index (FSI) for the U.S. (Cardarelli et al., 2009). As macroeconomic factors we include M1 growth, the purchasing manager index, housing starts, inflation, U.S. industrial production growth, and new orders of consumer goods and materials.

3.2 Analyzing the Predictive Performance

Volatility is inherently unobservable, so that measuring volatility is nontrivial. Here, we follow French et al. (1987) and Schwert (1989) and use monthly *realized volatility*, calculated from daily returns, as proxy for market volatility,[5] defined by

$$RV_{i,t} = \log \sum_{\tau=1}^{M_t} r_{i,t,\tau}^2, \quad t = 1, \ldots, T, \tag{6}$$

where $r_{i,t,\tau}$ denotes the τth daily return of asset i in month t and M_t the number of trading days in month t. Figure 3 shows the realized–volatility series for the four indices.

[5] For an in–depth review of the realized–volatility concept, we refer to Andersen et al. (2006).

Table 3: Description of the predictor variables employed.

Variable	Abbrev.	Source	Description
U.S. market excess return	MKTRF	Homepage FF	Fama–French market factor: U.S. stock return minus one-month T-Bill rate
Size factor	SMB	Homepage FF	Fama–French SMB factor: return small minus return on big stocks
Value factor	HML	Homepage FF	Fama–French HML factor: return on value minus return on growth stocks
U.S. IP growth	IPGR	Datastream	Log of annual growth rate of U.S. industrial production
New orders consumer growth	ORDc	Datastream	New orders of consumer goods and materials; annualized growth rate
U.S. M1 growth	M1c	Datastream	Log of annual growth rate of U.S. M1
Consumer sentiment index	SENTc	Datastream	Monthly change in consumer sentiment
Purchasing manager index	PMIc	Datastream	Monthly change in purchasing manager index
Housing starts	HSc	Datastream	Monthly change in housing starts
TED spread	TED	Datastream	Measure of illiquidity: LIBOR minus T-Bill rate
MSCI world stock index	MSCIc	Datastream	Return on the MSCI world stock market index
CRB spot index, annual ret	CRBc	Datastream	Growth in commodity prices: annualized return on CRB spot index
Book–to market–ratio	b/m	Goyal-Welch	Ratio of book value to market value for the Dow Jones Industrial Average
Net equity expansion	ntis	Goyal-Welch	Ratio of 12-month moving sums of net issues by NYSE listed stocks divided by total NYSE market capitalization (end of year)
Inflation	infl	Goyal-Welch	Log of annual growth rate of the U.S. consumer price index
Long–term rate of return	ltr	Goyal-Welch	Rate of return on long term government bonds
Default spread	DEF	Goyal-Welch	Measure of default risk of corporate bonds: BAA minus AAA bond yields
Relative T-Bill rate	RTB	Goyal-Welch	T-Bill rate minus its 12 month moving average
Relative bond rate	RBR	Goyal-Welch	Long-term bond yield minus its 12 month moving average
Term spread	TS	Goyal-Welch	Long-term bond yield minus three-month T-Bill rate
Pastor-Stambaugh liquidity factor	liq_fac	Pastor-Stambaugh	Measure of stock market liquidity based on price reversals
FX average bid–ask spread	BAS	Menkhoff et al.	Bid–ask spreads as measure of illiquidity in foreign exchange market
Return on dollar risk factor	RX	Lustig et al.	FX risk premium measure: average premium for bearing FX risk
Carry trade factor	HML	Lustig et al.	Return on high interest rate currencies minus return on low interest currencies
Cochrance–Piazessi factor	CP		Measure of bond risk premia, recursively estimated based on Fama–Bliss file according to Cochrane and Piazzesi (2005)
Financial stress index	FSI	IMF	U.S. FSI, an aggregate of seven variables capturing financial market stress

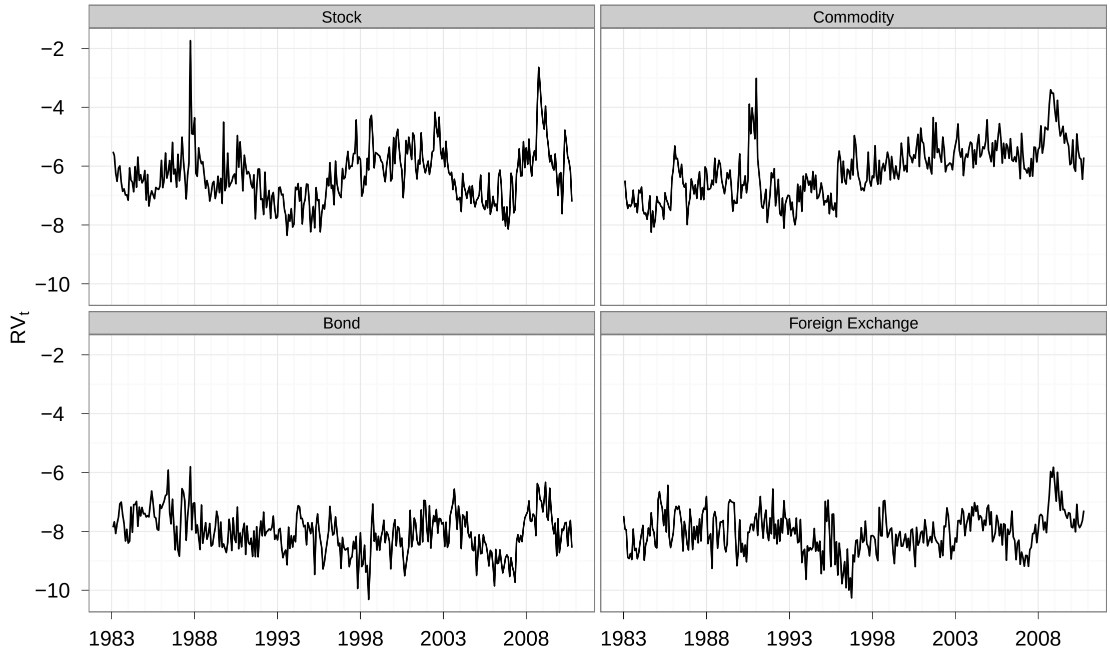

Figure 3: Time series of monthly realized volatility (in logarithms) as defined by (6).

The predictive performance is examined via rolling–window forecasting for the period June 2002 to September 2010. Starting with a history of 230 months, we move the fixed–length window forward month by month, re-estimate, and generate a sequence of one-step-ahead forecasts for 100 months.[6] Applying a direct forecasting approach,[7] we also produce multi–period forecasts for horizons of up to six months by adapting (1) accordingly, i.e.,

$$\begin{aligned} y_{t+h} &= \exp(\eta_{t+h}/2)\varepsilon_{t+h}, \quad h = 1, \ldots, 6, \\ \eta_{t+h} &= \beta_0 + f_{\mathrm{time}}(t+h) + f_{\mathrm{yr}}(n_{t+h}) + f_{\mathrm{month}}(m_{t+h}) + \sum_{j=0}^{s-1} f_j(y_{t-j}) + \sum_{k=1}^{q}\sum_{j=0}^{p-1} f_{k,j}(x_{t-j,k}). \end{aligned} \tag{7}$$

We include the first and second lag of all 26 factors as potential predictors, so that, in (1), $q = 26$ and $p = 2$. In addition, we include lags one and two of realized volatility ($s = 2$), to capture the state dependence and temporal dependence in volatility. Allowing also for seasonal components, we have a total of $r = 58$ predictors.

As volatility is latent, it is common to use the squared returns y_t^2 as a proxy. However, as this surrogate is very noisy, we follow another approach. We evaluate the forecasting performance in terms of the mean squared error between the "true" (realized) volatility, as

[6]Two observations are "lost" due to lagging variables twice.

[7]For direct forecasting via boosting in a nonlinear time series context, see Robinzonov et al. (2012).

defined in (6), and our forecasts for η_{t+h}.[8] Then,

$$\mathrm{ERR}_{i,t+h} = (RV_{i,t+h} - \eta_{i,t+h})^2 \tag{8}$$

represents the h-step squared prediction error for asset i.

4 Empirical Results

In discussing the empirical results we focus on the question: which factors drive realized volatility and in what way do they do so. Comparing the forecasting performance based on linear base learners to that derived from (nonlinear) regression trees, we find that linear base learners produce a lower forecasting accuracy, suggesting that driving factors exert a nonlinear influence on volatility.

To assess the predictive performance, we compare multi–step, out–of–sample boosting forecasts to those from a GARCH(1,1) benchmark model.[9] Clearly, there are many potential alternatives that could serve as benchmarks.[10] However, in the spirit of Lunde and Hansen (2005)—who ask "Does anything beat a GARCH(1,1)?"—the GARCH(1,1) model can be regarded as a natural and challenging benchmark in our context.

In the following subsections, we evaluate the predictive performance of our boosting approach and discuss, in some detail, driving factors in each of the four market.

4.1 Predictive Performance

To evaluate the predictive performance, we compute Theil's U and out–of–sample R^2 statistics for horizons ranging from one to six months. Theil's U is defined as the ratio of the root mean squared error (RMSE) of our model and to that of the GARCH benchmark model. A value below unity indicates that our model outperforms the benchmark. The out–of–sample R^2, proposed by Campbell and Thompson (2008), has an interpretation that is similar to that of Theil's U. Letting in market i, $\eta^M_{i,t+1}$ and $\eta^B_{i,t+1}$ denote the forecasts from our model and those of the benchmark, respectively, the out–of–sample R^2 is defined by

$$R^2_{OOS} = 1 - \frac{\sum_{t=t_0}^{T-1} \left(RV_{i,t+1} - \eta^M_{i,t+1}\right)^2}{\sum_{t=t_0}^{T-1} \left(RV_{i,t+1} - \eta^B_{i,t+1}\right)^2}, \tag{9}$$

where t_0 reflects the initialization period. Positive (negative) values of R^2_{OOS} indicate that boosting provides a superior (inferior) forecasting accuracy relative to the benchmark.

[8] For a detailed discussion on alternative forecast–evaluation criteria for realized volatility see Patton (2011).

[9] We compute multi–step GARCH forecasts recursively. By fitting GARCH models, using data sampled at each frequency, h, we also computed nonrecursive h–step forecasts. These turned out to be rather poor, however.

[10] Christiansen et al. (2012) use an autoregressive model for realized volatility as benchmark.

Table 4: Out-of-sample forecast evaluation.

	Theil's U				Out-of-sample R^2			
Horizon	Stock	Commod.	Bond	FX	Stock	Commod.	Bond	FX
1	0.99	1.07	1.04	1.03	0.00	-0.15	-0.09	-0.06
2	0.98	0.86	1.01	1.01	0.02	0.25	-0.03	-0.02
3	0.89	0.92	1.00	0.91	0.20	0.14	0.00	0.15
4	0.85	0.90	0.98	0.93	0.27	0.18	0.03	0.13
5	0.87	0.79	0.94	0.96	0.24	0.37	0.11	0.06
6	0.83	0.71	0.88	0.95	0.30	0.49	0.21	0.08

Table 5: Modified Diebold–Mariano test results.

Horizon	Stock	Commod.	Bond	FX
1	0.490	0.790	0.858	0.630
2	0.441	0.011 **	0.633	0.555
3	0.104	0.114	0.501	0.215
4	0.052 *	0.071 *	0.357	0.288
5	0.050 **	0.016 **	0.154	0.370
6	0.011 **	0.009 ***	0.030 **	0.355

The prediction results are shown in Table 4. In case of the stock index, boosting outperforms the benchmark over all horizons. For the other markets, the benchmark produces better one–step and in case of bonds and foreign exchange also better two–step predictions. In all other cases, especially for predictions beyond two months, the boosting approach dominates. For commodities and stocks, and to a lesser extend, for bonds, the medium-term performance is considerably better, whereas for FX volatility the differences seem negligible.

We also apply the Diebold–Mariano test (Diebold and Mariano, 1995) in the modified version of Harvey et al. (1997), to assess forecasting accuracy. The null hypothesis of the test presumes that benchmark forecasts are more accurate than those of the proposed model, so that rejection of the null favors our approach. The p-values of the modified Diebold–Mariano test, reported in Table 5, are in line with the results for the Theil's U and the out–of–sample R^2 statistics. Thus, inclusion of exogenous factors and nonlinear boosting help to improve medium– and long–term volatility forecasting—especially for commodity and stock indices.[11]

Overall, the forecasting comparisons suggest that, in the short–term, the forecasting performance of boosting matches that of the GARCH(1,1), but is superior in the medium– and long–term, i.e., beyond two months. An important observation is that the GARCH forecasts have wider MSE ranges, as indicated by the interquartile ranges (IQRs) shown in the MSE–boxplots in Figure 4. This suggests that boosting predictions are more robust relative to the GARCH benchmark.

Summarizing the results of the forecasting comparison for the stock index (upper panel in Figure 4), we find that they are in line with the forecasting statistics reported above.

[11] It should be noted that, when reverting the hypothesis (i.e., the null presumes boosting performs better than the benchmark), we obtain insignificant results in all cases.

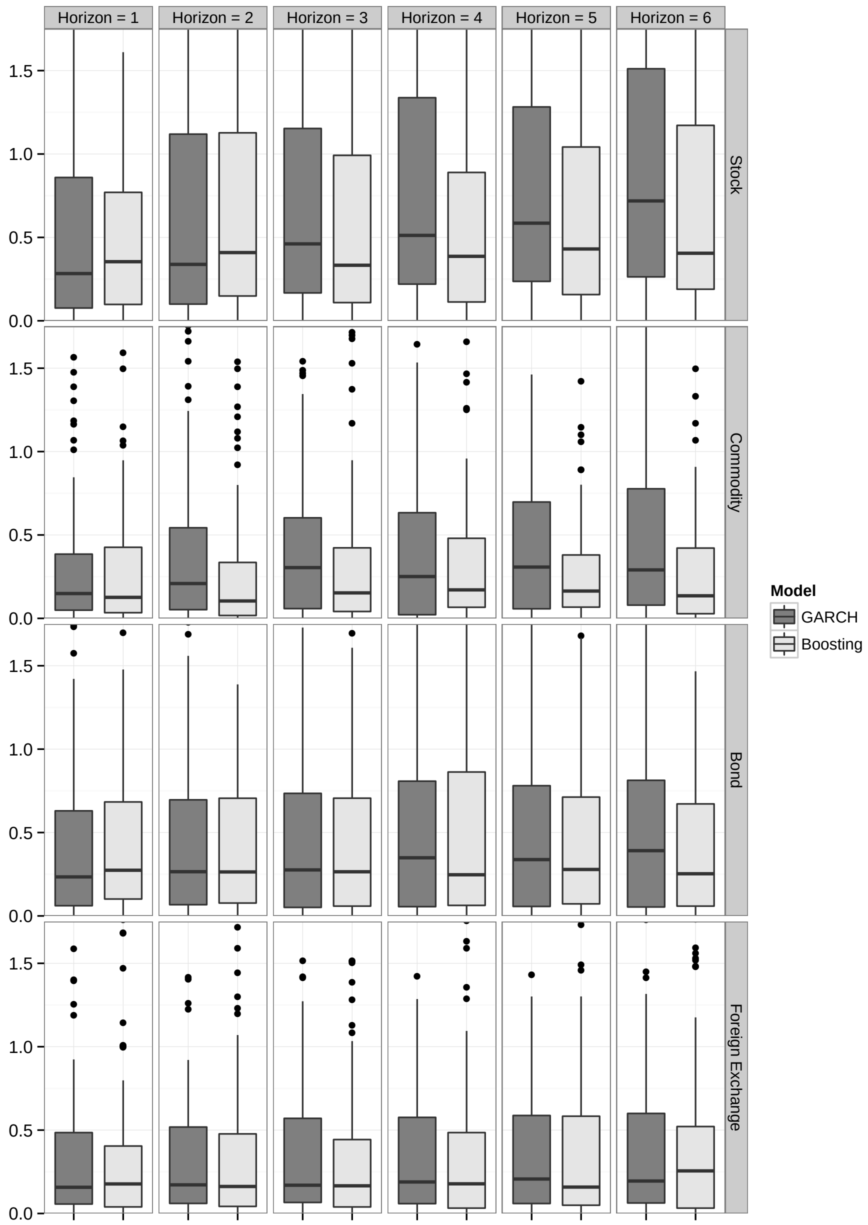

Figure 4: Comparison of predictive MSEs from GARCH(1,1) and boosting.

For the one– and two–period ahead forecasts, the GARCH model produces a lower median MSE; for horizons three to six, boosting produces lower median MSEs. However, for all horizons, the GARCH forecasts have a higher dispersion, as is indicated by the boxplots' IQRs. Thus, not only does the boosting approach provide better medium– and long–term volatility predictions for the S&P 500, its predictions are more robust across all horizons, leading to less extreme MSEs than the GARCH benchmark.

The boxplots for the commoditiy index (second from top in Figure 4) show that, on average, boosting forecasts outperform those of the GARCH model for all horizons. The Theil's U statistics corroborate this: for all horizons, except the first, Theil's U is below 1. For six-month-ahead forecasts, Theil's U decreases to 0.714. This, as well as the modified Diebold and Mariano (1995) tests indicate that the boosting strongly outperforms the GARCH model, especially so for medium and long horizons.

With respect to the short–term prediction MSEs for the bond market (third panel in Figure 4), we have a close race with boosting delivering, on average, better longer–term predictions. This also follows from the Theil's U statistics being below unity. However, modified Diebold-Mariano test statistics indicate that only for the six–month horizon does boosting significantly outperform GARCH.

Finally, in line with the literature (cf., Jorion, 1995, Nowak and Treepongkaruna, 2008), we also find that it is difficult to derive superior model for predicting FX volatility. Here, boosting and GARCH forecasts are about on the same level (bottom panel in Figure 4). For horizons of three to six months, Theil's U is below 1, but none of the modified Diebold-Mariano tests are significant. However, for all six horizons, the boosting produces lower 75% MSE–quantiles, suggesting more robustness also for FX–volatility predictions.

4.2 The Driving Factors

From both a theoretical and practical viewpoint it is of interest to identify the factors that drive market risk and to assess the specific manner in which they affect volatility. A better knowledge about the driving forces could, for example, be used for designing early–warnings mechanisms for market instabilities and for developing stabilization strategies. Clearly, the insights the proposed boosting approach offers make it more suitable for such modeling objectives than the GARCH framework with its black–box nature.

In the following, we discuss the role of risk drivers for each of the indices on the basis of the one–period–ahead model, defined in (1), and on the complete data set. For each of the four indices, we discuss the role of the three most relevant risk drivers in some detail and graphically illustrate their impact.

4.2.1 Stock Market

For the S&P 500, we identify six drivers: the U.S. financial stress index (FSI), the relative bond rate (RBR), (lagged) volatility, S&P 500 returns, U.S. market excess returns, and the CRB spot index. Our built–in variable (and lag) selection process excluded all other drivers considered. Figure 5 shows the impact of the first and second lags of the three most relevant factors. As is to be expected, not all the lags of the relevant variables do necessarily exert influence. For example, S&P 500 returns, the U.S. market excess return and the CRB spot index enter only with their first lag, whereas the FSI and (lagged) realized volatility display a longer–lived impact and enter also with their second lag.

Figure 5 (upper panel) clearly shows regime-dependence of the FSI's impact on volatility. FSI–values above 7.5 increase next month's stock market volatility by about 0.3 (in log scale),

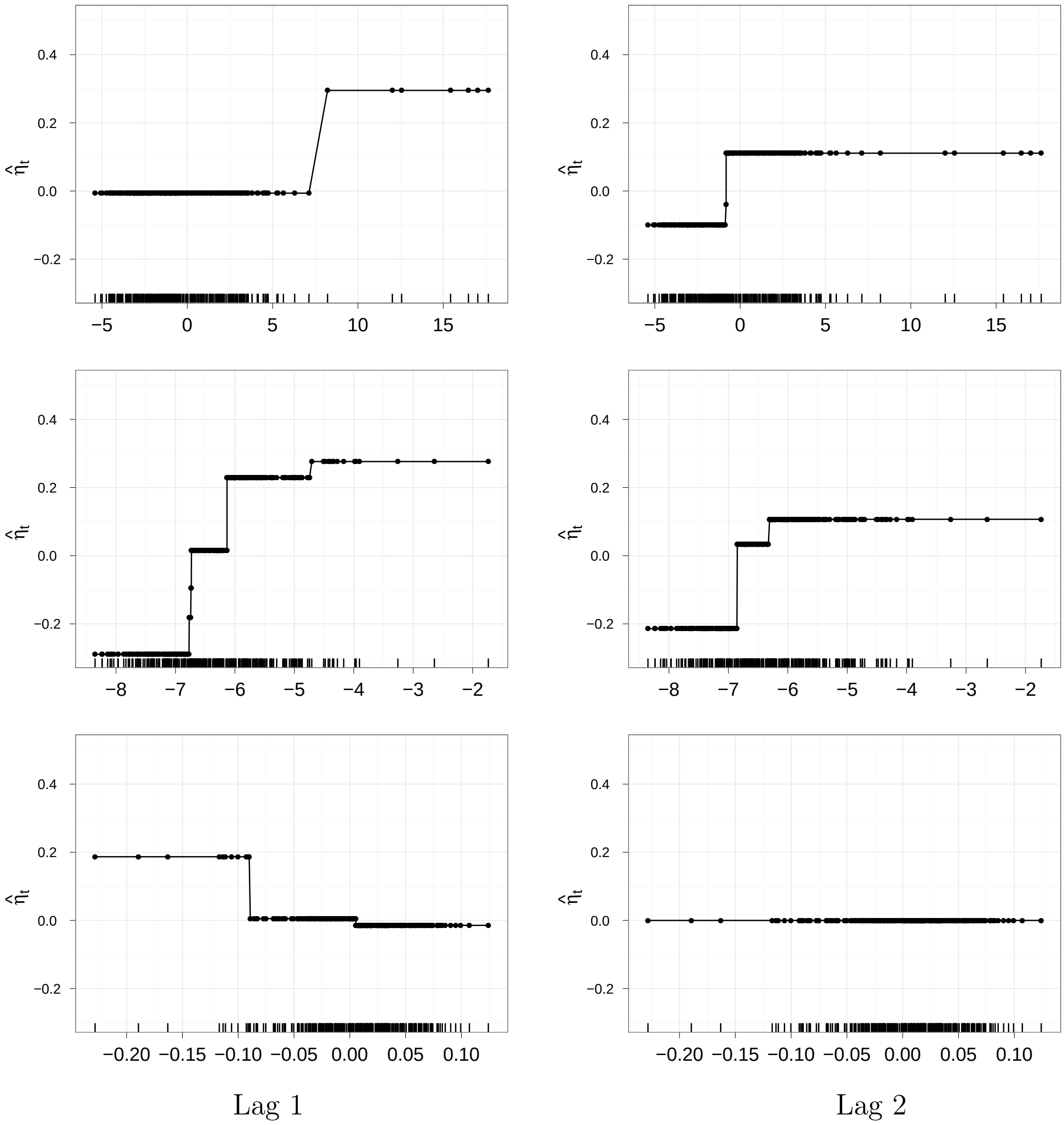

Figure 5: The three most relevant volatility drivers for the S&P 500. Each row shows the estimated partial volatility components for the first and second lag of FSI, RV and S&P 500 returns, respectively.

which corresponds to a volatility increase of about 16%. FSI-values below 7.5 do not affect next month's volatility. For lag two, we find that positive (negative) FSI–values moderately increase (decrease) two–month–ahead volatility in a more or less symmetric fashion.

Another result is that (log) realized volatility depends nonlinearly on past realized volatility. As shown in Figure 5 (middle panel), small values of realized volatility, i.e., $\mathrm{RV} < -7$ or $\exp(\mathrm{RV}) < 0.03$, cause a decrease in next month's volatility. From about $\mathrm{RV} > -6$ onward, the influence becomes positive, i.e., volatility tends to increase in a highly nonlinear fashion.

Also the two-month impact is also nonlinear: values of RV < -7 will reduce volatility by about 10% and values of RV > -6.3 result in an increase of about 5%.

Furthermore, we find that positive changes in the S&P 500 index slightly decrease volatility, whereas small negative changes (between -10% and 0%) moderately increase volatility (Figure 5, bottom panel). However, large negative returns (below -10%) increase volatility by about 10%. In addition to these three drivers, we find that a positive relative bond rate (RBR) induces a considerable increase in volatility, namely, 28% when it increases above 1%. Finally, positive U.S. market excess returns have a moderate calming effect on the stock market, whereas values below -2.5% increase the volatility by 2%.

4.2.2 Commodity Market

Commodity–market volatility is influenced by past realized volatility itself, the net equity expansion, the Cochrane–Piazessi factor, and the U.S. market excess returns. The Cochrane–Piazessi factor impacts via both the first and the second lag, whereas the net equity expansion influences only through the second lag.

Figure 6 (upper panel) reveals that realized volatility depends in a highly nonlinear fashion on its first lag. Highly negative values of lagged realized volatility (below -6.5) dampen volatility by roughly 0.2 (log scale), which roughly translates a 10%–drop in volatility; values above -6.5 lead to a volatility increase, which, for values above -4, jumps to about 60%. Net equity expansion (Figure 6, center panel) has an increasing effect on volatility, if it is below -3%; otherwise it slightly decreases volatility. U.S. market excess returns above -2% dampen and values below increase volatility (Figure 6, bottom panel). The pattern is similar for the Cochrane–Piazessi factor, except that, there, the threshold is at 2%.

4.2.3 Bond Market

For the bond market, we find that volatility is driven by the default spread, the changes in M1, net equity expansion, the changes in the purchasing manager index, the relative bond rate, the change in consumer sentiment, and the book to market ratio.[12]

The influence of the default spread (Figure 7, top panel) exhibits two clearly distinct regimes: a default spread above 1.1% tends to increase volatility by 7% in the following month; and values below that threshold reduce volatility by roughly 4%. The relative bond rate affects volatility only if it exceeds 1%, inducing an increase by 10%. A change in consumer sentiment or the book–to–market ratio produces a similar pattern: below a certain threshold—5% for consumer sentiment and 0.72 for the book–to–market ratio—they have no influence on volatility; only if they exceed these thresholds volatility increases. Sizable increases in M1 (above 5%) let volatility grow by approximately 10%. Smaller expansions or reductions in M1 decrease volatility by 6.8% (Figure 7, center panel).

[12] Bond return volatility has not been extensively studied in the literature. Two exceptions are Huang et al. (2011) and Viceira (2012).

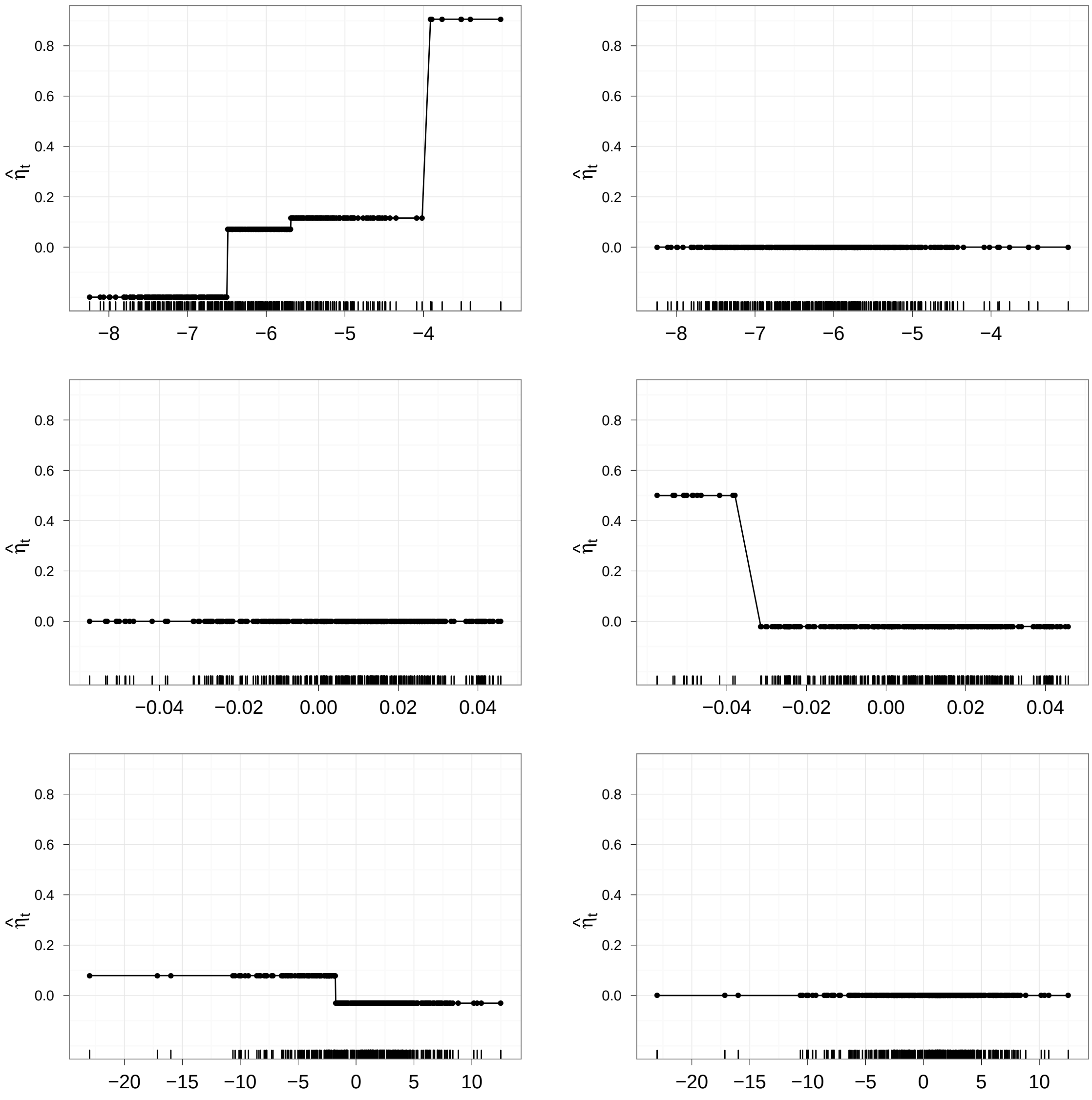

Figure 6: The three most highly relevant volatility drivers in the commodity market. Each row shows the partial impact of the first and second lag of RV, net equity expansion, and U.S. market excess returns, respectively.

4.2.4 Foreign Exchange Market

A large number of factors seem to affect FX volatility. These include the FSI, the default spread, realized volatility, the TED spread, the U.S. market excess return, the long–term rate of return, and changes in M1. Periods of high financial stress, with the FSI assuming values above five, drive up volatility by 12%, whereas low financial stress reduces it, though, by a much smaller amount, namely less than 1% (Figure 8, top panel). Similar to the other markets, once-lagged realized volatility below -7 (log scale), or $\hat{\sigma}^2 < 3\%$, lowers volatility

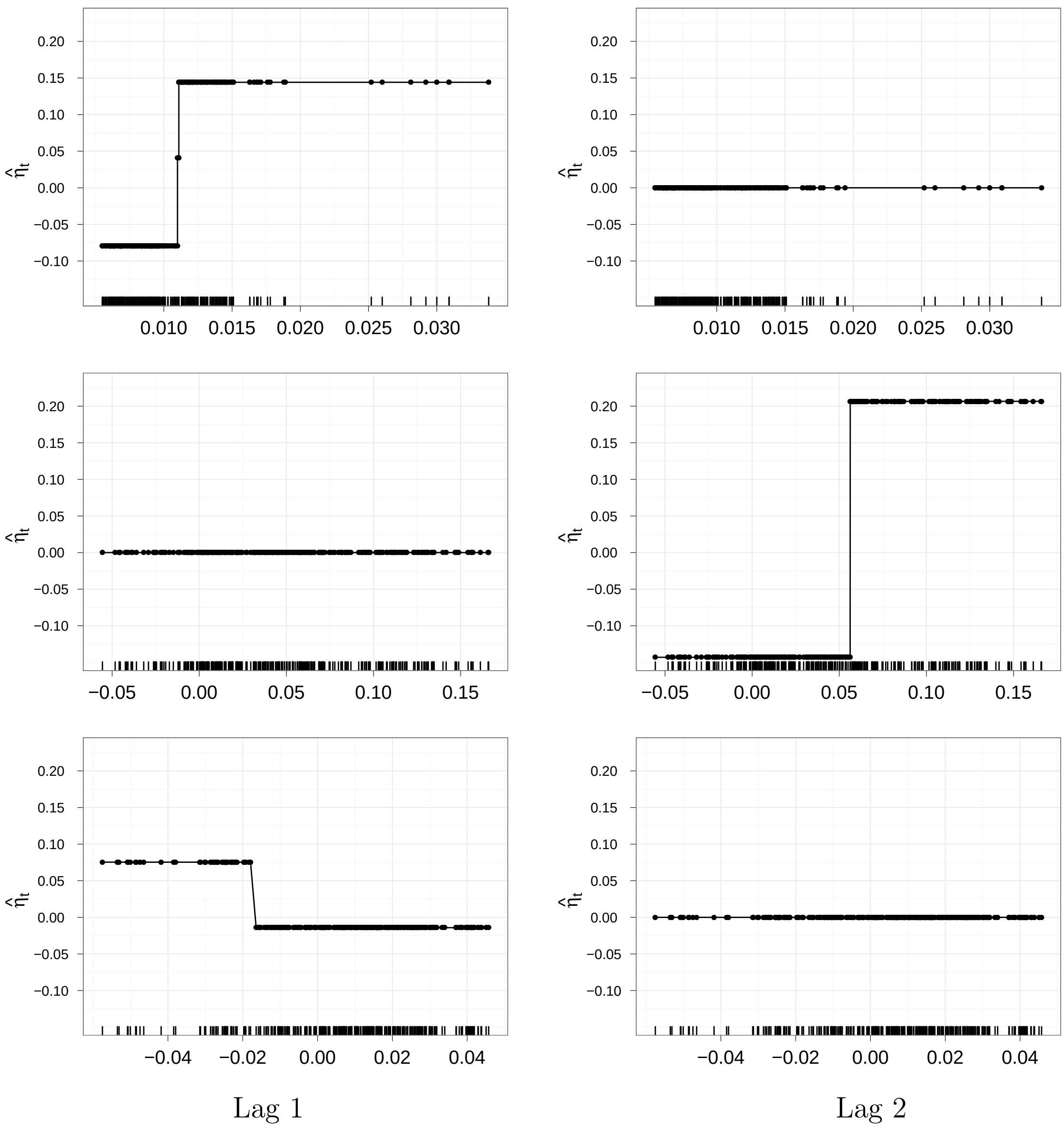

Figure 7: The three most relevant volatility drivers in the bond market. Each row shows the partial impact of first and second lag of the default spread, changes of M1, and the net equity expansion, respectively.

marginally (Figure 8, center panel). Values above this cutoff boost volatility by 15%. U.S. market returns seem to influence volatility only if they are below −10%, in which case they increase the volatility.

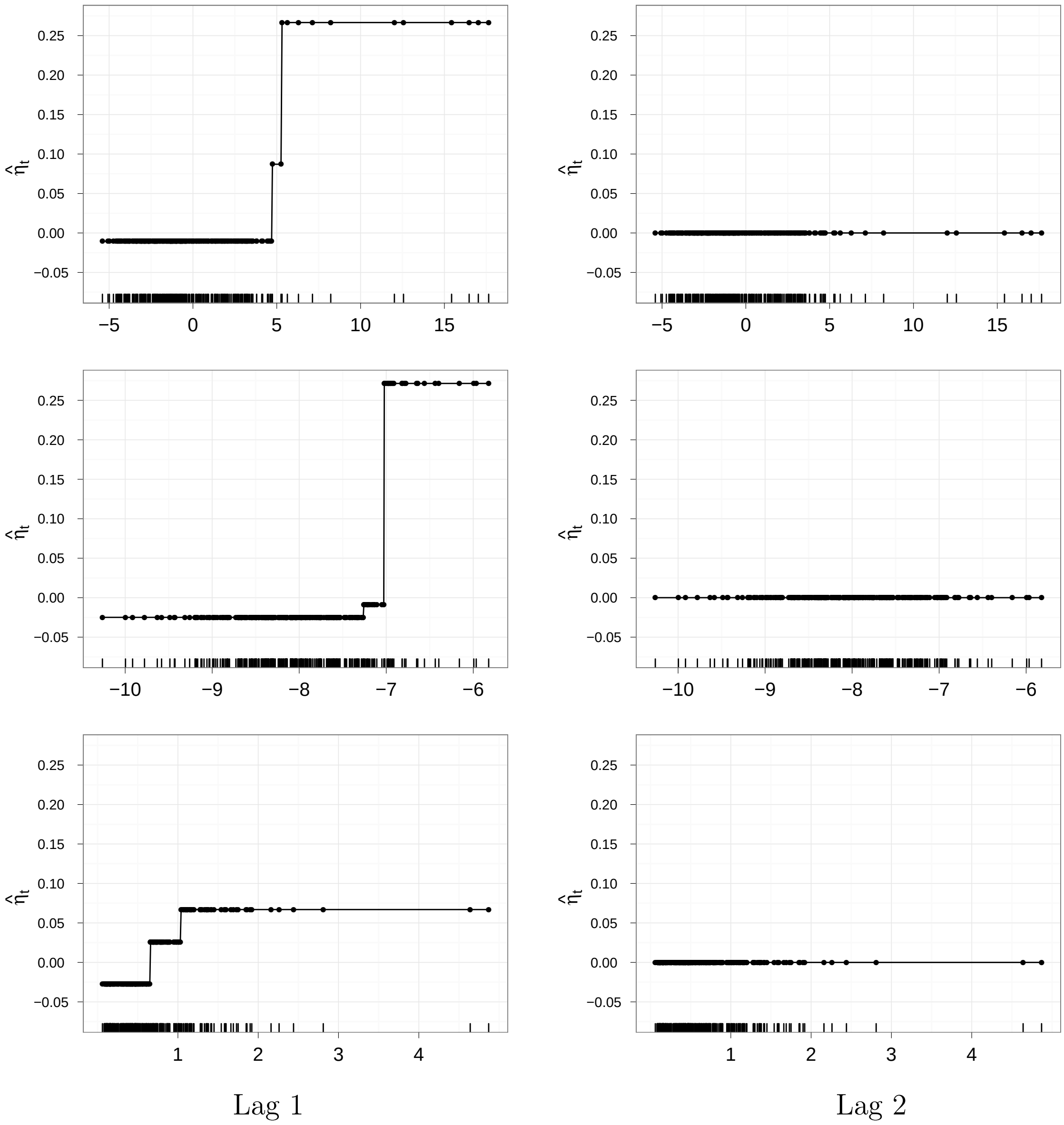

Figure 8: The three most relevant volatility drivers in the foreign exchange market. Each row shows the partial impact of first and second lag of FSI, RV, and the TED spread, respectively.

5 Conclusions

Employing boosting techniques, we have analyzed the determinants of monthly volatility for the four broad asset classes stocks, commodities, bonds, and foreign exchange, considering a wide range of potential financial and macroeconomic drivers. Our boosting approach specifies regression trees as base learners, allowing us to identify influential volatility drivers as well as the functional form of their impact. Specifically, we used componentwise boosting, which is tailor–made for sorting out irrelevant predictors.

Our empirical results give insight into the "anatomy" of volatility by identifying groups of relevant drivers for each market and by estimating driver–specific thresholds, which partition their domains into areas with similar impact on volatility. By doing so, nonlinear dependencies can be identified. We do, indeed, find highly nonlinear influences of financial drivers on volatility. This contrasts the existing literature, which almost exclusively concentrates on linear volatility dynamics.

Out-of-sample forecasting exercises, using realized volatility as a proxy for unobserved volatility, suggests that our boosting approach performs very favorable for stocks and commodities relative to the commonly–used GARCH(1,1) benchmark. The advantages are particularly convincing for longer forecasting horizons. For the bond and foreign exchange markets, boosting produces a comparable short–term and a marginally better medium– to long–term predictive accuracy. However, in all cases does boosting lead to more robust, i.e., less outlier–prone, prediction errors than the GARCH benchmark.

Our findings demonstrate that boosting is well suited for a unified framework to predictor selection and estimation in volatility modeling. This is especially the case in the presence of many potential (and possibly highly dependent) risk drivers. An advantage of the approach is that it can cope with "wide" data situations (Hastie et al., 2009), i.e., applications with a large number of predictors, possibly exceeding the number of observations. Also, models obtained via boosting can be a promising starting point for more detailed nonlinear model specifications.

Acknowledgments

We would like to thank Torsten Hothorn and participants of the 18th International Conference Computing in Economics and Finance in Prague (2012), the SMU-ESSEC Symposium on Empirical Finance and Financial Econometrics in Singapore (2012) and the Annual Meeting of the German Statistical Society in Vienna (2012) for valuable comments.

References

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006), "Volatility and Correlation Forecasting," in *Handbook of Economic Forecasting*, eds. Elliott, G., Granger, C., and Timmermann, A., Elsevier, vol. 1, chap. 15, pp. 777–878.

Audrino, F. and Bühlmann, P. (2003), "Volatility Estimation with Functional Gradient Descent for Very Highdimensional Financial Time Series," *Journal of Computational Finance*, 6, 65–89.

— (2009), "Splines for Financial Volatility," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 71, 655–670.

Bai, J. and Ng, S. (2009), "Boosting Diffusion Indices," *Journal of Applied Econometrics*, 24, 607–629.

Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.

Breiman, L. (1996), “Bagging Predictors,” *Machine Learning*, 24, 123–140.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984), “Classification and Regression Trees,” *Wadsworth, Belmont, CA, USA*.

Bühlmann, P. and Yu, B. (2003), “Boosting with the L_2 Loss: Regression and Classification,” *Journal of the American Statistical Association*, 98, 324–339.

Campbell, J. Y. and Thompson, S. B. (2008), “Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?” *Review of Financial Studies*, 21, 1509–1531.

Cardarelli, R., Elekdag, S., and Lall, S. (2009), “Financial Stress, Downturns, and Recoveries,” IMF Working Paper WP/09/100, International Monetary Fund.

Christiansen, C., Schmeling, M., and Schrimpf, A. (2012), “A Comprehensive Look at Financial Volatility Prediction by Economic Variables,” *Journal of Applied Econometrics*, 27, 956–977.

Cochrane, J. H. and Piazzesi, M. (2005), “Bond Risk Premia,” *American Economic Review*, 95, 138–160.

Diebold, F. X. and Mariano, R. S. (1995), “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13, 253–263.

Engle, R. F. (1982), “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, 50, 987–1007.

Engle, R. F., Ghysels, E., and Sohn, B. (2008), “On the Economic Sources of Stock Market Volatility,” Working paper series, AFA 2008 New Orleans Meetings Paper.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987), “Expected Stock Returns and Volatility,” *Journal of Financial Economics*, 19, 3–29.

Freund, Y. and Schapire, R. (1996), “Experiments With a New Boosting Algorithm,” in *Proceedings of the Thirteenth International Conference on Machine Learning Theory*, San Francisco: Morgan Kaufmann, pp. 148–156.

Friedman, J. H. (2001), “Greedy Function Approximation: A Gradient Boosting Machine,” *The Annals of Statistics*, 29, 1189–1232.

Goyal, A. and Welch, I. (2003), “Predicting the Equity Premium with Dividend Ratios,” *Management Science*, 49, 639–654.

Harvey, D., Leybourne, S., and Newbold, P. (1997), “Testing the Equality of Prediction Mean Squared Errors,” *International Journal of Forecasting*, 13, 281–291.

Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, Berlin, 2nd ed.

Hothorn, T., Buehlmann, P., Kneib, T., Schmid, M., and Hofner, B. (2011), *Model-Based Boosting*, R package version 2.1-1.

Hothorn, T., Hornik, K., and Zeileis, A. (2006), "Unbiased Recursive Partitioning: A Conditional Inference Framework," *Journal of Computational and Graphical Statistics*, 15, 651–674.

Huang, J., Lu, L., and Wu, B. (2011), "Macro Factors and Term Structure of Treasury Bond Volatility," Working Paper WP/09/100.

Jorion, P. (1995), "Predicting Volatility in the Foreign Exchange Market," *Journal of Finance*, 50, 507–28.

Kearns, M. and Valiant, L. (1994), "Cryptographic Limitations on Learning Boolean Formulae and Finite Automata," *Journal of the Association for Computing Machinery*, 41, 67–95.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006), "Value-at-Risk Prediction: A Comparison of Alternative Strategies," *Journal of Financial Econometrics*, 4, 53–89.

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

Lustig, H., Roussanov, N., and Verdelhan, A. (2011), "Common Risk Factors in Currency Markets," *Review of Financial Studies*, 24, 3731–3777.

Matías, J. M., Febrero-Bande, M., González-Manteiga, W., and Reboredo, J. C. (2010), "Boosting GARCH and Neural Networks for the Prediction of Heteroskedastic Time Series," *Mathematical and Computer Modelling*, 51, 256–271.

Menkhoff, L., Sarno, L., Schmeling, M., and Schrimpf, A. (2011), "Carry Trades and Global Foreign Exchange Volatility," CEPR Discussion Papers 8291, C.E.P.R. Discussion Papers.

Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Nowak, S. and Treepongkaruna, S. (2008), "Modeling and Forecasting Volatility in Foreign Exchange Markets," Working Paper Series in Finance FINM0042WP, School of Finance and Applied Statistics, Australian National University.

Pastor, L. and Stambaugh, R. F. (2003), "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 111, 642–685.

Patton, A. J. (2011), "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Journal of Econometrics*, 160, 246–256.

Paye, B. S. (2012), "'Déjà Vol': Predictive Regressions for Aggregate Stock Market Volatility Using Macroeconomic Variables," *Journal of Financial Economics*, in Press.

Robinzonov, N., Tutz, G., and Hothorn, T. (2012), "Boosting Techniques for Nonlinear Time Series Models," *AStA Advances in Statistical Analysis*, 96, 99–122.

Schapire, R., Freund, Y., Bartlett, P., and Lee, W. (1998), "Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods," *The Annals of Statistics*, 26, 1651–1686.

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

Strasser, H. and Weber, C. (1999), "On the Asymptotic Theory of Permutation Statistics," *Mathematical Methods of Statistics*, 8, 220–250.

Viceira, L. M. (2012), "Bond Risk, Bond Return Volatility, and the Term Structure of Interest Rates," *International Journal of Forecasting*, 28, 97–117.

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