Boosting the Anatomy of Volatility

Stefan Mittnik (CEQURA, LMU)
Nikolay Robinzonov (CEQURA, LMU)
Martin Spindler (CEQURA, MPI)

18th International Conference
Computing in Economics and Finance
June 27, 2012, Prague
Outline

1 Introduction and Motivation
2 Econometric Model
3 Estimation
4 Illustration
5 Data Set
6 Empirical Results
Research Question

**Explain:** Insights into the “anatomy” of volatility:
1. Identify small groups of influential drivers for different markets.
2. Identify volatility regimes of these drivers.
3. Quantify regime dependence.

**Forecast:** Can we use these factors to forecast volatility?
Motivation

• Importance of understanding, modeling and forecasting volatility:
  + We quantify risk.

• Applications:
  + Risk management (VaR, ES)
  + Portfolio optimization and asset allocation (time-varying betas, conditional Sharpe ratios, time-varying covariances)
  + Option valuation (dynamic volatility)
Introduction and Motivation

Linear model

\[
y_t \mid x_t \sim N(\mu_t, \sigma^2)
\]

\[
\mu_t = x_t^\top \beta, \quad \hat{\sigma}^2 = \frac{\hat{\epsilon}^\top \hat{\epsilon}}{n - p}
\]

The variance does not vary w.r.t the observations \((y_t, x_t)\).
\[ y_t | x_t \sim N(\mu_t, \sigma^2) \]
\[ \mu_t = x_t^\top \beta, \quad \hat{\sigma}^2 = \frac{\hat{\varepsilon}^\top \hat{\varepsilon}}{n - p} \]

The variance \underline{does not vary w.r.t} the observations \((y_t, x_t)\).
Proposed Model

\[ y_t = \exp(\eta_t/2) \cdot N(0, 1) \Leftrightarrow y_t|z_{t-1} \sim N(0, e^{\eta_t}) \]
\[ \eta_t = f(z_{t-1}) = \beta_0 + f_{\text{time}}(t) + f_{\text{year}}(n_t) + f_{\text{month}}(m_t) + \]
\[ + \sum_{j=1}^{s} f_j(y_t-j) + \sum_{k=1}^{q} \sum_{j=1}^{p} f_{k,j}(x_{k,t-j}) \]

\[ z_{t-1} = (1, t, n_t, m_t, y_{t-1}, \ldots, y_{t-s}, x_{1,t-1}, \ldots, x_{1,t-p}, \ldots, x_{q,t-1}, \ldots, x_{q,t-p})^\top \in \mathbb{R}^r, \quad r = s + qp + 4. \]

\begin{align*}
\text{GARCH}(1,1) & \quad y_t = \sigma_t \cdot N(0, 1) \\
& \quad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
\text{SVM} & \quad y_t = \exp(\eta_t/2) \cdot N(0, 1) \\
& \quad \eta_t = \beta_0 + \beta_1 \eta_{t-1} + N(0, \tau^2) \end{align*}
Estimation
Estimation Problem

- Objective: Obtain an estimate $\eta^*$:

$$\eta^* = \arg\min_{\eta} \mathbb{E} [L(y, \eta(z))]$$

(2)

where $L$ is some loss (or risk) function, assumed to be differentiable with respect to $\eta(z)$.

- In practice, we minimize the empirical risk:

$$\eta^* = \arg\min_{\eta} \frac{1}{T} \sum_{t=1}^{T} L(y_t, \eta(z_t)).$$

(3)

- The minimization is performed w.r.t. parameters $\beta$:

$$\eta^* = \eta(z; \hat{\beta}) = \arg\min_{\beta} \frac{1}{T} \sum_{t=1}^{T} L(y_t, \eta(z_t; \beta)).$$

(4)
Steepest Descent

\[ f(Y_1, Y_2) = 1.5 \cdot Y_1^2 + 2 \cdot Y_2^2 \]

\[- \frac{\partial f}{\partial Y_1 \partial Y_2} = g(Y_1, Y_2) = \left( -3 \cdot Y_1, -4 \cdot Y_2 \right)^\top \]
Steepest Descent

\[ f(Y_1, Y_2) = 1.5 \cdot Y_1^2 + 2 \cdot Y_2^2 \]

\[ -\frac{\partial f}{\partial Y_1 \partial Y_2} = g(Y_1, Y_2) = \left(-3 \cdot Y_1, -4 \cdot Y_2\right)^T \]

\[ Y^{[0]} = (-1.5, 2)^T \]

\[ g(Y^{[0]}) = (4.5, -8)^T \]

\[ Y^{[1]} = Y^{[0]} + \frac{1}{10} \cdot g(Y^{[0]}) \]

\[ = (-1.05, 1.2)^T \]
**Steepest Descent**

\[
f(Y_1, Y_2) = 1.5 \cdot Y_1^2 + 2 \cdot Y_2^2
\]

\[
- \frac{\partial f}{\partial Y_1 \partial Y_2} = g(Y_1, Y_2) = \left( -3 \cdot Y_1, -4 \cdot Y_2 \right)^T
\]

\[
Y[0] = (-1.5, 2)^T
\]

\[
g(Y[0]) = (4.5, -8)^T
\]

\[
Y[1] = Y[0] + \frac{1}{10} \cdot g(Y[0])
\]

\[
= (-1.05, 1.2)^T
\]

\[
\vdots
\]

\[
Y[k] = Y[k-1] + \frac{1}{10} \cdot g(Y[k-1])
\]

\[
\vdots
\]
Estimation

Steepest Descent

\[ \eta^* = \eta(z; \hat{\beta}) = \operatorname{arg\,min}_{\beta} \frac{1}{T} \sum_{t=1}^{T} L(y_t, \eta(z_t; \beta)) \]  \hspace{1cm} (4)

1. Given any approximation \( \eta(z_t; \hat{\beta}^{[m-1]}) \) the current negative gradient is

\[ g^{[m]} := -\left[ \frac{\partial}{\partial \eta} L(y_t, \eta) \right]_{\eta = \eta(z_t; \hat{\beta}^{[m-1]})} \]

which gives the steepest-descent direction.

2. Then, we estimate the gradient:
   - Estimate the negative gradient by each predictor separately
     \[ \hat{\gamma}^{[m]}_j = \operatorname{arg\,min}_{\gamma_j} \left[ f_j(z_t; \gamma_j) \rightarrow \hat{g}^{[m]}_j \right] \]
   - \[ \hat{s}_m = \operatorname{arg\,min}_{j \in \{1, \ldots, r\}} \sum_{t=1}^{T} \left( \hat{g}^{[m]}_j - f_j(z_t; \hat{\gamma}^{[m]}_j) \right)^2 \]

3. Update

\[ \eta(z_t; \hat{\beta}^{[m]}) = \eta(z_t; \hat{\beta}^{[m-1]}) + \nu \cdot g^{[m]}, \quad t = 1, \ldots, T. \]

\[ \text{steepest descent} \]
Estimation

Steepest Descent

\[ \eta^* = \eta(z; \hat{\beta}) = \arg \min_{\beta} \frac{1}{T} \sum_{t=1}^{T} L(y_t, \eta(z_t; \beta)) \]  \hspace{1cm} (4)

1. Given any approximation \( \eta(z_t; \hat{\beta}^{[m-1]}) \) the current negative gradient is

\[ g^{[m]} := -\left[ \frac{\partial}{\partial \eta} L(y_t, \eta) \right]_{\eta=\eta(z_t; \hat{\beta}^{[m-1]})} \]

which gives the steepest-descent direction.

2. Then, we estimate the gradient:

   - Estimate the negative gradient by each predictor separately
     \[ \hat{\gamma}^{[m]}_j = \arg \min_{\gamma_j} \left[ f_j(z_t; \gamma_j) \rightarrow \hat{g}^{[m]}_j \right], \]

   - \( \hat{s}_m = \arg \min_{j \in \{1, \ldots, r\}} \sum_{t=1}^{T} \left( \hat{g}^{[m]}_j - f_j(z_t; \hat{\gamma}^{[m]}_j) \right)^2 \)

3. Update

\[ \eta(z_t; \hat{\beta}^{[m]}) = \eta(z_t; \hat{\beta}^{[m-1]}) + \nu \cdot \frac{\hat{s}_m(z_t; \hat{\gamma}_{\hat{s}_m})}{\hat{g}^{[m]}} , \quad t = 1, \ldots, T. \]
Loss Function and Negative Gradient

We assume $y_t | z_{t-1} \sim N(0, \sigma^2_t)$ with log-likelihood function:

$$l(y_t | \sigma^2_t) = \ln \left[ 2\pi \sigma^2_t \right]^{-1/2} - \frac{y_t^2}{2\sigma^2_t} = -\frac{1}{2} \left[ \ln 2\pi + \ln \sigma^2_t + \frac{y_t^2}{\sigma^2_t} \right]$$

$\Rightarrow$ Loss function: $L_t = \frac{1}{2} \left[ \ln \sigma^2_t + \frac{y_t^2}{\sigma^2_t} \right]$

$$\sigma^2_t = e^{\eta_t} \frac{1}{2} \left[ \eta_t + \frac{y_t^2}{e^{\eta_t}} \right]$$

$\Rightarrow$ Negative gradient: $g_t = -\frac{\partial L_t}{\partial \eta_t} = \frac{1}{2} \left[ -1 + \frac{y_t^2}{e^{\eta_t}} \right]$  

(Audrino and Bühlmann, 2009)
Why Boosting?

In contrast to conventional fitting methods, ...
Why Boosting?

In contrast to conventional fitting methods, ...

... boosting is applicable to many different risk functions (absolute loss, quantile regression)

... boosting can be used to carry out variable selection during the fitting process ⇒ No separation of model fitting and variable selection

... boosting is applicable even if $p \gg n$ (wide data)

... boosting addresses multicollinearity problems (by shrinking effect estimates towards zero)

... boosting optimizes prediction accuracy (w.r.t. the risk function).
Illustration
Simulation:

\[ y_t = \exp(\eta_t/2) \cdot N(0, 1) \]

\[ \eta_t = 0.2 + 0.5 \cdot X_{1,t} - 0.4 \cdot X_{2,t} + 1.2 \cdot I_{[1,2]}(X_{3,t}) + 0 \cdot X_{4,t} + 0 \cdot X_{5,t} + 0 \cdot X_{6,t}, \quad X_i \in U[0, 4]. \]

The model:

\[ y_t = \exp(\eta_t/2) \cdot N(0, 1) \]

\[ \eta_t = \beta_0 + \beta_1 \cdot X_{1,t} + \beta_2 \cdot X_{2,t} + \sum_{j=1}^{J} \gamma_j I_{R_j}(X_{3,t}) + \beta_4 \cdot X_{4,t} + \beta_5 \cdot X_{5,t} + \beta_6 \cdot X_{6,t} \]
Partial Effects

\[ y_t \sim N(0, e^{\eta_t}) \]

\[ \eta_t = 0.2 + 0.5 \cdot X_{1,t} - 0.4 \cdot X_{2,t} + 1.2 \cdot I_{[1,2]}(X_{3,t}) \]
Conditional Densities

\[ y_t \sim N(0, e^{\eta_t}) \]

\[ \eta_t = 0.2 + 0.5 \cdot X_{1,t} - 0.4 \cdot X_{2,t} + 1.2 \cdot I_{[1,2]}(X_{3,t}) \]
Data Set
Analysis of four different asset classes:

- Bonds represented by 10-year Treasury note futures contracts traded on the Chicago Board of Trade (CBOT).
- Commodities represented by Standard & Poor’s GSCI commodity index.
- Foreign Exchange (FX) represented by a trade-weighted portfolio provided by the Federal Reserve Bank of St. Louis.

All data are monthly and cover the period February 1983 to September 2010 (332 months).
Figure 1: The logarithm of the monthly realized volatility
### Data Set

## Macroeconomic and Financial Factors

<table>
<thead>
<tr>
<th>1. dividend price ratio</th>
<th>14. book to market ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. cross sectional premium</td>
<td>15. net equity expansion</td>
</tr>
<tr>
<td>3. term spread</td>
<td>16. relative T-bill rate</td>
</tr>
<tr>
<td>4. relative Bond rate</td>
<td>17. investor sentiment</td>
</tr>
<tr>
<td>5. U.S. market excess return</td>
<td>18. size factor</td>
</tr>
<tr>
<td>6. return on the MSCI world index</td>
<td>19. value factor</td>
</tr>
<tr>
<td>7. Cochrane and Piazzesi bond factor</td>
<td>20. carry trade factor</td>
</tr>
<tr>
<td>8. purchasing manager index</td>
<td>21. inflation</td>
</tr>
<tr>
<td>9. return on the CRB spot index</td>
<td>22. orders</td>
</tr>
<tr>
<td>10. industrial production growth</td>
<td>23. housing starts</td>
</tr>
<tr>
<td>11. FX average bid-ask spread (Menkhoff et al., 2011)</td>
<td>24. TED spread</td>
</tr>
<tr>
<td>13. Financial Stress Index (FSI) for advanced economies</td>
<td>26. M1 growth</td>
</tr>
</tbody>
</table>
Forecasting Strategy

- Sample period: February 1983 to September 2010 (332 months)
- Estimation: 26 regressors with first and second lag ($q = 26, p = 2$) and first and second lag of realized volatility ($s = 2$)
  In total: $r = 58$ predictors
- Evaluation period: June 2002 to September 2010 (100 months)
- Rolling window scheme (window size: 230 months)
- Direct forecasting approach; multi-period forecasts for one to six months
- Forecast evaluation by mean squared error. We will compare ex-post (realized volatility) estimations $\sigma^2_{RV}$ with ex-ante predictions $\exp(\eta_t)$. 
Direct Forecasting

\[ y_{t+h} = \exp(\eta_t/2)\varepsilon_{t+h} \]

\[ \eta_t = \beta_0 + f_{\text{time}}(t) + f_{\text{year}}(n_t) + f_{\text{month}}(m_t) + \sum_{j=1}^{s} f_j(y_{t-j}) + \sum_{k=1}^{q} \sum_{j=1}^{p} f_{k,j}(x_{k,t-j}) \]

Accuracy measure:

\[ \text{MSE}_{t+h} = \frac{1}{T} \sum_{t=1}^{T} \left[ \hat{\eta}_t - \ln(\sigma_{\text{RV},t+h}^2) \right]^2 \]
Empirical Results
Figure 2: Financial Stress Index (FSI), lag 1 (left) and lag 2 (right).
Stock Market II

Figure 3: Lagged Realized Volatility, lag 1 (left) and lag 2 (right).
Stock Market III

Driving Factors:
financial Stress Index (FSI, Fig. 2)
lagged realized volatility (RV, Fig. 3)
lagged returns
U.S. market excess return
relative bond rate (RBR, second lag)
CRB spot index

All other factors seem irrelevant.

Other markets
Figure 4: Forecasting comparison of the MSE between GARCH(1,1) and our model for the stock market.
Empirical Results

S&P500

Goldman Sachs Commodity Index

Foreign Exchange

Bond
## Forecast Evaluation

### Table 1: Modified Diebold-Mariano test results

<table>
<thead>
<tr>
<th>Horizon</th>
<th>FX</th>
<th>GSCI</th>
<th>S&amp;P500</th>
<th>TNOTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.630</td>
<td>0.790</td>
<td>0.4901</td>
<td>0.8585</td>
</tr>
<tr>
<td>2</td>
<td>0.555</td>
<td>0.011 **</td>
<td>0.4407</td>
<td>0.6330</td>
</tr>
<tr>
<td>3</td>
<td>0.215</td>
<td>0.114</td>
<td>0.1039</td>
<td>0.5011</td>
</tr>
<tr>
<td>4</td>
<td>0.288</td>
<td>0.071 *</td>
<td>0.0519 *</td>
<td>0.3568</td>
</tr>
<tr>
<td>5</td>
<td>0.370</td>
<td>0.016 **</td>
<td>0.0496 **</td>
<td>0.1542</td>
</tr>
<tr>
<td>6</td>
<td>0.355</td>
<td>&lt;0.01 ***</td>
<td>0.0112 **</td>
<td>0.0297 **</td>
</tr>
</tbody>
</table>

The null hypothesis is that the GARCH forecasting error is smaller than the boosting forecasting error.

[Theil’s U]
Summary and Conclusion

- Boosting with regression trees identifies macro and financial factors in different asset classes.
- Insights into the “anatomy” of volatility:
  1. Identify small groups of influential drivers for each market
  2. Identify volatility regimes of these drivers.
  3. Quantify regime dependence.
- The influence of the factors is highly nonlinear.
- Forecasts for commodities and stocks significantly outperform the GARCH(1,1) model, especially for medium and long horizons.
Thank you for your attention!