The Weak Efficient Market Hypothesis in Light of Statistical Learning

L. Fiévet & D. Sornette
Working Paper
Chair of Entrepreneurial Risks, ETH Zürich
Outline

• Introduction to the Efficient Market Hypothesis
• Concept of Statistical Learning
• Empirical results
The Efficient Market Hypothesis
S&P 500 Index

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Index level</td>
<td>1,527</td>
<td>1,565</td>
</tr>
<tr>
<td>P/E ratio (fwd.)</td>
<td>25.6x</td>
<td>15.2x</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>1.1%</td>
<td>1.8%</td>
</tr>
<tr>
<td>10-yr. Treasury</td>
<td>6.2%</td>
<td>4.7%</td>
</tr>
</tbody>
</table>

Source: Standard & Poor's, First Call, Compustat, FactSet, J.P. Morgan Asset Management.
The chart shows the performance of the S&P 500 Index from 1996 to 2014. Key characteristics include:

- **Index level**:
  - Mar-2000: 1,527
  - Oct-2007: 1,565
  - Dec-2014: 2,059

- **P/E ratio (fwd.)**:
  - Mar-2000: 25.6x
  - Oct-2007: 15.2x
  - Dec-2014: 16.2x

- **Dividend yield**
  - Mar-2000: 1.1%
  - Oct-2007: 1.8%
  - Dec-2014: 1.9%

- **10-yr. Treasury**
  - Mar-2000: 6.2%
  - Oct-2007: 4.7%
  - Dec-2014: 2.2%

The chart highlights significant percentage changes at various points:

- **Dec. 31, 1996**
  - P/E (fwd.) = 16.0x
  - 741

- **Mar. 24, 2000**
  - P/E (fwd.) = 25.6x
  - 1,527

- **Oct. 9, 2002**
  - P/E (fwd.) = 14.1x
  - 777

- **Mar. 9, 2009**
  - P/E (fwd.) = 10.3x
  - 677

- **Oct. 9, 2007**
  - P/E (fwd.) = 15.2x
  - 1,565

- **Dec. 31, 2014**
  - P/E (fwd.) = 16.2x
  - 2,059

The chart also notes **+106%**, **+101%**, **-49%**, **-57%**, and **+204%** at various points over the period.

Source: Standard & Poor's, First Call, Compustat, FactSet, J.P. Morgan Asset Management.
The Efficient Market Hypothesis from 1970

\[ E(r_{t+1} | r_1, ..., r_t) = 0 \]

So that the sequence \( \{ r_t \} \) is a “fair game” (martingale) with respect to the information sequence \( \{ \Phi_t \} \). Fama (1970).

In other words, financial price series are fundamentally unpredictable.
Assume $\epsilon_t \in \mathcal{N}(0, 1)$.

1. Random walk (prices): $p_{t+1} = p_t + \epsilon_t$

2. Random walk (returns): $r_{t+1} = r_t + \epsilon_t$

3. A linear process (returns): $r_{t+1} = \sum_{i=1}^{n} \psi_i \epsilon_{t-i}$
Consider a sequence of returns \( \{ r_1, r_2, \ldots, r_t \} \), mapped to a binary sequence by \( r \rightarrow \text{sign}(r) \).

Niederhoffer & Osborne (1966) noticed that

\[
P(+|++ > P(−|++).
\]

Not profitable as argued by Fama (1970).
The Efficient Market Hypothesis from 1991

Prices fully reflect information

Information & trading costs

Real life prices

Can This Information be Arbitraged?

Assume that the price at time \( t \) can go up or down by \( \Delta p \) with probabilities

\[
|P(p_{t+1} = p_t + \Delta p) - P(p_{t+1} = p_t - \Delta p)| = \delta_t.
\]

The cost of obtaining these probabilities is \( \alpha \).

The transaction cost is \( \beta \) percent.

The invested amount is \( I \).

When would you pay the amount \( \alpha \) to obtain the probabilities?
The Fama-French (1993) Regression Model

• The Fama-French model captures several anomalies:

\[ r_t - r_t^f = f_a + f_b(r_{Mt} - r_{ft}) + f_sSMB_t + f_sHML_t + f_sMOM_t + e_t \]

• Market correlation
• Small Minus Big market capitalization
• High Minus Low book-to-market ratio
• Momentum
Support for the EMH

• No known portfolio profitable in excess of the Fama-French Model.

• Mutual funds do not show excess profits. Fama & French (2009).

• Crashes correct for exuberant stock market increases.
Significance of a Strategy?

Exhibit 2
200 Randomly Generated Trading Strategies

Source: AHL Research.
Distinguishing False Positives and True Negatives?

**EXHIBIT 3**
False Trading Strategies, True Trading Strategies

Panel A: Too Many False Strategies

- **False Trading Strategies**
- **True Trading Strategies**
- **Missed Discoveries**
- **False Discoveries**

**Annual Excess Returns**

- 0%
- 2%
- 4%
- 6%
- 8%
- 10%
- 12%
- 14%
An Operational EMH Definition

- Information set $\Omega_t$
- Models $M_t$
- Search technology $S_t$
- Impossible to make economic profits (=EMH)

Are the Profits Significant?

• Romano & Wolf (2005, 2016) formalized a step-down method to compute p-values adjusted for multiple hypothesis testing.

• Sullivan et al. (2000) showed that in a universe of 7846 trading rules none are statistically significant after 90 years, when adjusting for data-snooping.
Figure 10. Economic and statistical performance of the best model chosen from the full universe according to the mean return criterion: S&P 500 Futures (1984–1996). For a given trading rule, \( n \), indexed on the x-axis, the scattered points plot the mean annualized returns experienced during the sample period. The thin line measures the best mean annualized return among the set of trading rules \( i = 1, \ldots, n \), and the thick line measures the associated data-snooping adjusted p-value.
Asset returns: \( \{r_1, \ldots, r_t\} \)

Strategy 1 signals: \( \{s^1_1, \ldots, s^1_t\} \)

Strategy 2 signals: \( \{s^2_1, \ldots, s^2_t\} \)

\[ \ldots \]

How to randomize the strategies?
Statistical Learning in Finance
Fundamental Concept

- Inputs $X$.
- Outputs $Y$.
- Error term $\epsilon$.
- A training set and a test set.

$Y = f(X) + \epsilon$
Data in the Weak EMH

• All past returns:

\[ \Omega_t = \{r_1, \ldots, r_t\} \]

• Most general training set could be:

\[ D_t = \{(\Omega_1, \bar{\Omega}_1), \ldots, (\Omega_{t-1}, \bar{\Omega}_{t-1})\} \]
Challenges

Returns are sequential and not stationary
No sample independence
Give up cross-validation
No test set at a given time
Binary Returns

• Consider binary returns $r_b = sign(r)$

• Consider only the $m$ last returns.

• Predict only the next step return.

• A sensible requirement is: $\#D > 20 \cdot 2^{m+1}$
Decision Trees Example

\[ \hat{r} = \{r_1, r_2, r_3\} \]

\[ r_1 \leq 0 \]

\[ r_2 \leq 0 \]
\[ r_3 \leq 0 \]
\[ 0 \]
\[ r_2 > 0 \]
\[ r_3 > 0 \]
\[ 1 \]

\[ r_1 > 0 \]

\[ r_2 \leq 0 \]
\[ r_3 \leq 0 \]
\[ 0 \]
\[ r_2 > 0 \]
\[ r_3 > 0 \]
\[ 1 \]
### Variety of Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearest Neighbors</td>
<td>Finds the $k$ nearest neighbors for a new data point and computes their average output.</td>
</tr>
<tr>
<td>Linear Discriminant</td>
<td>Creates classes based on the optimal linear boundaries for a $p$ dimensional Gaussian random variable.</td>
</tr>
<tr>
<td>Quadratic Discriminant</td>
<td>Creates classes based on the optimal quadratic boundaries for a $p$ dimensional Gaussian random variable.</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>Computes the logistic regression, which naturally limits outputs to a lower and upper boundary.</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>Creates classes based on hyperplane separation with maximum margin.</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>Creates classes by recursively finding optimal splits along a single feature.</td>
</tr>
<tr>
<td>Random Forest</td>
<td>Averages the prediction of many decision trees trained on subsets of the training data to avoid overfitting.</td>
</tr>
<tr>
<td>Gradient Boosting</td>
<td>Averages the prediction of decision trees recursively created to reduce the error of the previous set of trees.</td>
</tr>
</tbody>
</table>
Decision Boundaries
Consider $y = a \cdot x_1 + b \cdot x_2$
In-sample & Out-of-sample Length

\[ \text{In-sample length} \quad \text{Out-of-sample length} \]

\( m \)
Empirical Results

<table>
<thead>
<tr>
<th>Family</th>
<th>S&amp;P 500</th>
<th>FTSE</th>
<th>CSI 300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{X_2}^{\text{min}}$</td>
<td>$P_W^{\text{min}}$</td>
<td>$P_{SR}^{\text{min}}$</td>
</tr>
<tr>
<td>GARCH</td>
<td>$6.4 \times 10^{-3}$</td>
<td>0.24</td>
<td>0.19</td>
</tr>
<tr>
<td>NN-R</td>
<td>0.84</td>
<td>0.25</td>
<td>0.10</td>
</tr>
<tr>
<td>SVM-R</td>
<td>0.11</td>
<td>0.25</td>
<td>2.2</td>
</tr>
<tr>
<td>DT-R</td>
<td>0.78</td>
<td>0.48</td>
<td>3.9</td>
</tr>
<tr>
<td>RF-R</td>
<td>0.47</td>
<td>0.55</td>
<td>0.67</td>
</tr>
<tr>
<td>GB-R</td>
<td>0.22</td>
<td>0.20</td>
<td>0.33</td>
</tr>
<tr>
<td>LDA-C</td>
<td>0.14</td>
<td>0.26</td>
<td>0.23</td>
</tr>
<tr>
<td>QDA-C</td>
<td>0.41</td>
<td>0.66</td>
<td>0.65</td>
</tr>
<tr>
<td>LR-C</td>
<td>4.4</td>
<td>10^{-3}</td>
<td>0.37</td>
</tr>
<tr>
<td>NN-C</td>
<td>0.78</td>
<td>0.98</td>
<td>0.84</td>
</tr>
<tr>
<td>SVM-C</td>
<td>8.7</td>
<td>10^{-4}</td>
<td>0.04</td>
</tr>
<tr>
<td>DT-C</td>
<td>0.80</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>RF-C</td>
<td>0.80</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td>GB-C</td>
<td>0.73</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>DT-BR</td>
<td>2.9</td>
<td>10^{-4}</td>
<td>3.6</td>
</tr>
<tr>
<td>DT-BC</td>
<td>7.0</td>
<td>10^{-6}</td>
<td>6.4</td>
</tr>
</tbody>
</table>

Table IV: Summary statistics by model family: out-of-sample S&P 500, FTSE and CSI 300. This table provides for each model family the three best p-values for the directional accuracy ($P_{X_2}$), the compounded wealth ($P_W$), and the Sharpe ratio ($P_{SR}$). The Sharpe ratio p-value is only indicated if $P_{SR} \neq P_W$. The model families are Nearest Neighbors (NN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Logistic Regression (LR). The family names are suffixed as follows: "-R" for regressors; "-C" for classifiers; and "-B" for binary. The p-values are adjusted for multiple testing within each family according to the algorithm in Section II.D. Each family has $49 \times 4 = 196$ models, parametrized by 49 in-sample lengths $l_i \in [20, 500]$ and 4 lags $m \in [2, 3, 4, 5]$. The value $\#_W^{95}$ indicates the number of models significant at the 95% level in compounded wealth.
S&P 500
### Correlations Between Methods

<table>
<thead>
<tr>
<th></th>
<th>GARCH</th>
<th>NN-R</th>
<th>SVM-R</th>
<th>DT-R</th>
<th>RF-R</th>
<th>GB-R</th>
<th>LDA-C</th>
<th>QDA-C</th>
<th>LR-C</th>
<th>NN-C</th>
<th>SVM-C</th>
<th>DT-C</th>
<th>RF-C</th>
<th>GB-C</th>
<th>DT-BR</th>
<th>DT-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>1</td>
<td>0.12</td>
<td>0.28</td>
<td>0.077</td>
<td>0.097</td>
<td>0.13</td>
<td>0.3</td>
<td>0.22</td>
<td>0.65</td>
<td>0.13</td>
<td>0.56</td>
<td>0.06</td>
<td>0.12</td>
<td>0.993</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>NN-R</td>
<td>0.12</td>
<td>1</td>
<td>0.16</td>
<td>0.21</td>
<td>0.3</td>
<td>0.31</td>
<td>0.2</td>
<td>0.21</td>
<td>0.1</td>
<td>0.63</td>
<td>0.11</td>
<td>0.19</td>
<td>0.3</td>
<td>0.25</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td>SVM-R</td>
<td>0.28</td>
<td>0.16</td>
<td>1</td>
<td>0.11</td>
<td>0.14</td>
<td>0.19</td>
<td>0.24</td>
<td>0.34</td>
<td>0.43</td>
<td>0.21</td>
<td>0.46</td>
<td>0.11</td>
<td>0.2</td>
<td>0.18</td>
<td>0.3</td>
<td>0.36</td>
</tr>
<tr>
<td>DT-R</td>
<td>0.077</td>
<td>0.11</td>
<td>0.11</td>
<td>1</td>
<td>0.3</td>
<td>0.35</td>
<td>0.14</td>
<td>0.16</td>
<td>0.066</td>
<td>0.23</td>
<td>0.07</td>
<td>0.28</td>
<td>0.33</td>
<td>0.33</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>RF-R</td>
<td>0.097</td>
<td>0.3</td>
<td>0.14</td>
<td>0.3</td>
<td>1</td>
<td>0.42</td>
<td>0.17</td>
<td>0.17</td>
<td>0.075</td>
<td>0.25</td>
<td>0.085</td>
<td>0.24</td>
<td>0.34</td>
<td>0.32</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>GB-R</td>
<td>0.13</td>
<td>0.31</td>
<td>0.19</td>
<td>0.35</td>
<td>0.42</td>
<td>1</td>
<td>0.23</td>
<td>0.23</td>
<td>0.1</td>
<td>0.27</td>
<td>0.12</td>
<td>0.29</td>
<td>0.37</td>
<td>0.47</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td>LDA-C</td>
<td>0.3</td>
<td>0.2</td>
<td>0.74</td>
<td>0.16</td>
<td>0.17</td>
<td>0.23</td>
<td>1</td>
<td>0.43</td>
<td>0.44</td>
<td>0.25</td>
<td>0.48</td>
<td>0.13</td>
<td>0.24</td>
<td>0.22</td>
<td>0.36</td>
<td>0.43</td>
</tr>
<tr>
<td>QDA-C</td>
<td>0.22</td>
<td>0.21</td>
<td>0.34</td>
<td>0.16</td>
<td>0.17</td>
<td>0.23</td>
<td>0.43</td>
<td>1</td>
<td>0.25</td>
<td>0.27</td>
<td>0.15</td>
<td>0.24</td>
<td>0.23</td>
<td>0.24</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>LR-C</td>
<td>0.65</td>
<td>0.1</td>
<td>0.43</td>
<td>0.066</td>
<td>0.075</td>
<td>0.1</td>
<td>0.44</td>
<td>0.25</td>
<td>0.16</td>
<td>0.04</td>
<td>0.069</td>
<td>0.14</td>
<td>0.11</td>
<td>0.19</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>NN-C</td>
<td>0.13</td>
<td>0.43</td>
<td>0.21</td>
<td>0.23</td>
<td>0.25</td>
<td>0.27</td>
<td>0.25</td>
<td>0.25</td>
<td>0.16</td>
<td>1</td>
<td>0.17</td>
<td>0.24</td>
<td>0.37</td>
<td>0.3</td>
<td>0.2</td>
<td>0.24</td>
</tr>
<tr>
<td>SVM-C</td>
<td>0.56</td>
<td>0.11</td>
<td>0.46</td>
<td>0.07</td>
<td>0.085</td>
<td>0.12</td>
<td>0.48</td>
<td>0.27</td>
<td>0.84</td>
<td>0.17</td>
<td>0.075</td>
<td>0.15</td>
<td>0.13</td>
<td>0.22</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>DT-C</td>
<td>0.06</td>
<td>0.19</td>
<td>0.11</td>
<td>0.28</td>
<td>0.24</td>
<td>0.29</td>
<td>0.13</td>
<td>0.15</td>
<td>0.069</td>
<td>0.24</td>
<td>0.075</td>
<td>0.1</td>
<td>0.38</td>
<td>0.39</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>RF-C</td>
<td>0.12</td>
<td>0.3</td>
<td>0.2</td>
<td>0.33</td>
<td>0.34</td>
<td>0.37</td>
<td>0.24</td>
<td>0.24</td>
<td>0.14</td>
<td>0.37</td>
<td>0.15</td>
<td>0.38</td>
<td>1</td>
<td>0.47</td>
<td>0.19</td>
<td>0.2</td>
</tr>
<tr>
<td>GB-C</td>
<td>0.093</td>
<td>0.25</td>
<td>0.18</td>
<td>0.33</td>
<td>0.32</td>
<td>0.47</td>
<td>0.22</td>
<td>0.23</td>
<td>0.11</td>
<td>0.3</td>
<td>0.13</td>
<td>0.19</td>
<td>0.47</td>
<td>1</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>DT-BR</td>
<td>0.18</td>
<td>0.22</td>
<td>0.3</td>
<td>0.11</td>
<td>0.16</td>
<td>0.22</td>
<td>0.36</td>
<td>0.24</td>
<td>0.19</td>
<td>0.2</td>
<td>0.22</td>
<td>0.12</td>
<td>0.19</td>
<td>0.17</td>
<td>1</td>
<td>0.68</td>
</tr>
<tr>
<td>DT-B</td>
<td>0.19</td>
<td>0.19</td>
<td>0.36</td>
<td>0.12</td>
<td>0.15</td>
<td>0.19</td>
<td>0.43</td>
<td>0.28</td>
<td>0.27</td>
<td>0.24</td>
<td>0.31</td>
<td>0.12</td>
<td>0.2</td>
<td>0.18</td>
<td>0.68</td>
<td>1</td>
</tr>
</tbody>
</table>
Observation & interpretation

• Predictability is high during volatile periods/crashes.

• The stock market as a whole “synchronizes”. Behaves like a single agent.

• EMH is rejected up to 10bps per round trip for the S&P 500.
HOW DEEPMIND CONQUERED GO

Disclaimer: I was not part of this research project, I am merely providing commentary on this work.
Next steps

• Reproduce the results using a deep learning approach. Neural networks with LSTM are Turing complete.

• Create mixed in-sample and memory length models.

• Create multi-agent models: Structural forests.
Summary

• Efficient market hypothesis?

• Advantages/disadvantages of statistical learning?

• Predictability on equity indices?
Thank You For Your Attention

Questions?
References