Computational Intelligence Meets Financial Forecasting

Edward Tsang et al
Forecasting Research Team
Computing Intelligence Meets Forecasting

What if opportunities are scarce?

Repository Method

Is the market predictable?

Predictions, in the form of:
- prices
- opportunities
- threats

Forecasting System

EDDIE:
Constraint-directed search
For trading precision with recall

EDDIE for Investment and arbitrage opportunities

How to measure success?

How to invest?

Needs motivate new algorithms

22 September 2009
Forecasting

- Will the price go up or down? By how much?


- What is the risk of crashing?

- Are Option and Future prices aligned? (i.e. are there arbitrary opportunities?)
Efficient Market Hypothesis

- Financial assets (e.g. shares) pricing:
  - All available information is fully reflected in current prices
- If EMH holds, forecasting is futile
  - Random walk hypothesis
- Assumptions:
  - Efficient markets (one can buy/sell quickly)
  - Perfect information flow
  - Rational traders
Is the market really efficient?

- Market may be efficient in the long term
- “Fat Tail” observation:
  - big changes today often followed by big changes tomorrow (either up or down)
- How fast can one respond to new information?
  - Faster machines certainly help
  - So should faster algorithms (CIDER)
- Credit crunch: did investors price their risks properly?
Do fundamental values matter?

♦ In boom, markets are liquid but often not driven by fundamentals only (bubbles)
♦ In bust, markets may be driven by fundamentals only, but are not liquid
♦ In neither boom nor bust are markets efficient
  – Willem Buiter (LSE)
Our Research agenda

- What would a reasonable agenda be?
- Predicting the price in 10 days would be good
- But it may be sufficient if I could turn a 50-50 game into a 60-40 game in my favour
- Question asked:
  “Will the price go up (or down) by at least r% within the next n days?”
How can computational intelligence help?
A taste of user input

<table>
<thead>
<tr>
<th>Daily closing</th>
<th>50 days m.a.</th>
<th>Volatility</th>
<th>Define target: 4% in 21 days?</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>80</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>99</td>
<td>82</td>
<td>52</td>
<td>0</td>
</tr>
<tr>
<td>87</td>
<td>83</td>
<td>53</td>
<td>1</td>
</tr>
<tr>
<td>82</td>
<td>82</td>
<td>51</td>
<td>1</td>
</tr>
<tr>
<td>.....</td>
<td>.....</td>
<td>.....</td>
<td>.....</td>
</tr>
</tbody>
</table>
GP: Example Tree

Human users:
- Define grammar
- Assess trees rationality

EDDIE:
- Find interactions
- Discover thresholds

Functions
- If-then-else
- <
- Buy
- >
- Sell
- If-then-else
- Buy

Terminals
- P/E ratio
- 6.4
- 50 days MA
- Current Price

Buy

<

Sell

22 September 2009
All Rights Reserved, Edward Tsang
Syntax of GDTs in EDDIE-2

\[
\text{<Tree>} ::= \text{"If-then-else" \<Condition> \<Tree> \<Tree>} \mid \text{Decision}
\]
\[
\text{<Condition>} ::= \text{<Condition> "And" \<Condition>} \mid
\text{<Condition> "Or" \<Condition>} \mid
\text{"Not" \<Condition>} \mid
\text{Variable \<RelationOperation> \text{Threshold}}
\]
\[
\text{<RelationOperation>} ::= \text{">" \mid "<" \mid "}\text{=}\text{"
}\]

**Terminals:**

- **Variable** is an indicator / feature
- **Decision** is an integer, “Positive” or “Negative” implemented
- **Threshold** is a real number

♦ Richer language \(\Rightarrow\) larger search space
Machine learning basics

What could one learn?
Hypothetical observations
How to summarize success/failure?
Performance measures
Hypothetical Situation

♦ Suppose you’ve discovered a good indicator $R$
  – How can you make use of it?
♦ Suppose it is a fact that whenever
  – $R$ has a value less than 1.4 or greater than 2.7,
  – the volatility of the share prices is above 2.5, and
  – yield is above 5.7%

prices will rise by $\geq 6\%$ within the next 21 days
♦ How can you find this rule
### Hypothetical observations

<table>
<thead>
<tr>
<th>Instance</th>
<th>$R$</th>
<th>Volatility</th>
<th>Yield</th>
<th>Target</th>
<th>Classified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2</td>
<td>3.1</td>
<td>4.8</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>2</td>
<td>1.3</td>
<td>3.0</td>
<td>6.6</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>3</td>
<td>2.8</td>
<td>2.9</td>
<td>5.9</td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>4</td>
<td>2.5</td>
<td>1.7</td>
<td>7.0</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>5</td>
<td>2.4</td>
<td>3.5</td>
<td>6.9</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>6</td>
<td>2.0</td>
<td>2.9</td>
<td>5.6</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>7</td>
<td>3.1</td>
<td>3.3</td>
<td>5.8</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>8</td>
<td>3.1</td>
<td>3.0</td>
<td>5.5</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>9</td>
<td>2.8</td>
<td>2.4</td>
<td>5.0</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>10</td>
<td>2.6</td>
<td>2.5</td>
<td>5.2</td>
<td>False</td>
<td>False</td>
</tr>
</tbody>
</table>
## Confusion Matrix

<table>
<thead>
<tr>
<th>Reality</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>False</td>
<td>5 2 7</td>
</tr>
<tr>
<td>True</td>
<td>1 2 3</td>
</tr>
<tr>
<td>6 4 10</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reality</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>False</td>
<td></td>
</tr>
<tr>
<td>True</td>
<td></td>
</tr>
<tr>
<td>6 4 10</td>
<td></td>
</tr>
</tbody>
</table>

22 September 2009 All Rights Reserved, Edward Tsang
### Performance Measures

<table>
<thead>
<tr>
<th>Reality</th>
<th>Ideal Predictions</th>
<th>Actual Predictions, Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>- 7, 0, 7</td>
<td>- 5, 2, 7</td>
</tr>
<tr>
<td>+</td>
<td>0, 3, 3</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td></td>
<td>7, 3, 10</td>
<td>6, 4, 10</td>
</tr>
</tbody>
</table>

\[
RC = \frac{(5+2)}{10} = 70\%
\]

\[
\text{Precision} = \frac{2}{4} = 50\%
\]

\[
\text{Recall} = \frac{2}{3} = 67\%
\]
Genetic programming in forecasting

EDDIE
Eddie Technical Overview

Tree Representation
- Grammar defined by user

Each tree is a Boolean function
- E.g. Will the price go up by 4% within the next 21 days?

Crossover, Mutation

Constraint-directed fitness
- To improve precision

Fitness eval (RC, RF, RMC)
- Using historic data

Tournament Selection

Update population

Our experience All Rights Reserved, Edward Tsang
Our EDDIE/FGP Experience

- Patterns exist
  - Would they repeat themselves in the future? (EMH debated for decades)

- EDDIE has found patterns
  - Not in every series (we don’t need to invest in every index / share)

- EDDIE extending user’s capability
  - and give its user an edge over investors of the same caliber
Incentive to Improve Precision

### Actual Predictions, Example

<table>
<thead>
<tr>
<th>Reality</th>
<th>-</th>
<th>75</th>
<th>70</th>
<th>5</th>
<th>10</th>
<th>80</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td></td>
<td></td>
<td>75</td>
<td>70</td>
<td>5</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>+</td>
<td>9</td>
<td>5</td>
<td>11</td>
<td>15</td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>75</td>
<td>16</td>
<td>25</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- False positive costs real money
- We cannot change reality
- But we have control over predictions
- Hope: reduced more false positives than true positive

**RC** = \( \frac{70+15}{100} = 85\% \)

**Precision** = \( \frac{15}{25} = 60\% \)

**Recall** = \( \frac{15}{20} = 75\% \)

**RC** = \( \frac{75+11}{100} = 86\% \)

**Precision** = \( \frac{11}{16} = 69\% \)

**Recall** = \( \frac{11}{20} = 55\% \)
FGP: Constrained Fitness

- Constraints can help guiding the search
- Fitness = $w_{rc} \times RC' - w_{rmc} \times RMC - w_{rf} \times RF$
- $RC' = RC$ if $P+ \in [\text{Min, Max}]$
  
  0 otherwise

- One can adjust Min and Max to reflect market expectation (possibly from training), or risk preference

22 September 2009

All Rights Reserved, Edward Tsang
Effect of constraints in FGP-2

Observation: RMC can be traded for RF without significantly affecting RC
EDDIE for arbitrage prediction
Arbitrage Opportunities

- Futures are obligations to buy or sell at certain prices
- Options are rights to buy at a certain price
- If they are not aligned, one can make risk-free profits
  - Such opportunities should not exist
  - But they do in London

A simplified scenario:

- Option price: £0.5
- Option right to buy: £10
- Future selling price: £11
Experience in EDDIE on Arbitrage

- Arbitrage opportunities exist in London
- Naïve approach:
  - Monitor arbitrage opportunities, act when they arise; problem: speed
- Misalignments don’t happen instantaneously
  - Do patterns exist? If so, can we recognize them?
- EDDIE-ARB can find some opportunities
  - With high confidence (precision >75%)
- Commercialisation of EDDIE-ARB
  - Need to harvest more opportunities; Need capital
- Research only made possible by close collaboration between computer scientists and economists
Facing scarce opportunities

Chance Discovery
Problem with scarce opportunities

<table>
<thead>
<tr>
<th>Reality</th>
<th>Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>9,900 0 99%</td>
</tr>
<tr>
<td>+</td>
<td>0 100 1%</td>
</tr>
<tr>
<td></td>
<td>99% 1%</td>
</tr>
</tbody>
</table>

Ideal prediction
Accuracy = Precision = Recall = 100%

<table>
<thead>
<tr>
<th>Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
</tr>
<tr>
<td>#</td>
</tr>
</tbody>
</table>

Moves from - to +
Random move from 99% to 0%
Accuracy = Precision = Recall = 10%
Precision = Recall = 1%
(Accuracy dropped from 99%)

22 September 2009

All Rights Reserved, Edward Tsang
Repository Method

In order to mine the knowledge acquired by the evolutionary process, Repository Method performs the following steps:

1- Rule extraction
   Evolve a GP to create a population of decision trees

2- Rule simplification
   \[
   \begin{align*}
   R_1 & \quad \text{The rule } R_k \text{ is selected by precision;} \\
   R_2 & \quad \text{R}_k \text{ is simplified to } R'_k \\
   \ldots & \\
   R_n & \\
   \end{align*}
   \]

3- New rule detection
   \[ R'_k \text{ is compared to the rules in the repository by similarity (genotype) } \]

4- Add rule to the repository
   If \( R'_k \) is a new rule, \( R'_k \) is added to the rule repository
Where does it go from here?

- Computational finance > CI + Finance
  - Research agenda beyond CI and finance experts
- Finance drives computational intelligence
  - We need more techniques for chance discovery
- Being able to forecast alone is not sufficient
  - If opportunity is predicted, do we invest 100%?
- Financial forecasting is growing rapidly
  - Conferences, IEEE Technical Committee, etc
FAQ in forecasting

- **Is the market predictable?**
  - It doesn’t have to be
  - But if you believe it is, you should code your own expertise
  - Market is not efficient anyway, herding has patterns

- **How can you predict exceptional events?**
  - No, we can’t
  - Neither can human traders

- **How can you be sure that your program works?**
  - No, we can’t
  - Neither were we sure about Nick Leeson at Barrings
  - Codes are more auditable than humans
  - If you can improve your odds from 50-50 to 60-40 in your favour, you should be happy
Reference

The Forecasting Research Team

Edward Tsang
EDDIE / GP

James Butler
EDDIE

Jin Li
FGP

Alma Garcia
Chance Discovery

Sheri Markose
Red Queen

Hakan Er
Arbitrage

Serafin Martinez
EDDIE Red Q

Michael Kampouridis
EDDIE 8

Wang Pu
EDDIE101

Olsen & Associates

Ionic Share-scope

All Rights Reserved, Edward Tsang