Predicting investment fluxes from implicit lead-lag investor networks

D. Challet\textsuperscript{1,2}  M. Lallouache \textsuperscript{1}  R. Chicheportiche \textsuperscript{3,4}

\textsuperscript{1} CentraleSupélec  \textsuperscript{2} Encelade Capital SA  \textsuperscript{3} CFM  \textsuperscript{4} previously at Swissquote Bank SA

September 8, 2015
Broker internal order matching

versus

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Internal order matching

BUY -> 2$ -> BROKER

SELL -> 2$ -> BROKER

?
Broker’s gain

\[ G_T = \sum_{t=1}^{T} I_t r_t = \sum_{t=1}^{T} (I_{t-1} + \delta I_t) r_t \]

Classic optimization problem

1. Fix \( T \) (1 day)
2. Assume random \( \delta I_t \) and \( r_t \)
3. Constraints
4. Minimize cost function
Broker’s gain

\[ G_T = \sum_{t=1}^{T} I_t r_t = \sum_{t=1}^{T} \left( I_{t-1} + \delta I_t \right) r_t \]

Prediction problem

1. predict flux
2. predict price return
3. PROFIT!
**Prediction problem** = Science + Engineering

### Science: flux

1. cluster clients
2. lead-lag networks
3. machine learning
4. ENJOY!

### Engineering

1. Predict price returns
2. Inventory constraints
3. PROFIT!
Investor classification

Supervised
- Individual vs institutional (Odean, Barber, etc.)
- ...

Unsupervised
- Similarity measure
  → categories
Supervised classification

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Predicting investment fluxes from implicit lead-lag investor networks
Unsupervised learning

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Lillo et al. (2008)

- Daily inventory change $V_i(t)$
- Correlation matrix $C \sim E(V_iV_j)$
- Principal Component Analysis + Random Matrix theory
Unsupervised classification I
Equities, Spanish brokers

Lillo et al. (2008)
Zhou et al. (2012)
Unsupervised classification III: FX, individual investors, daily

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Unsupervised classification IV: FX, individual investors, hourly

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Unsupervised classification
Statistically validated networks (SVNs)

Tumminello et al. (2011a)

Agent $i$, $state_i(t) \in \{1, \cdots, S\}$

$$s_i(t) = \text{sign} \frac{Buy(t) - Sell(t)}{Buy(t) + Sell(t)} \in \{-1, 0, 1, 2 = 0\}$$

Agents $i$ and $j$: measure frequency

$$s_i(t) = s \& \& s_j(t) = s'$$

Compute p-values

$O(N^2)$ pairs

$\rightarrow$ multiple hypothesis correction

$\rightarrow$ NETWORK
Origin of trade synchronicity

1. Explicit communication
2. Implicit communication
   1. same news
   2. same strategy: MA(CD)
   3. same parameters
   4. Master in Finance
Trader-trader communication: explicit

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Trader-trader communication: explicit
Do traders look at prices?
Implicit AND explicit communication

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Implicit communication: same price analysis

Predicting investment fluxes from implicit lead-lag investor networks
Implicit communication: same price analysis

Predicting investment fluxes from implicit lead-lag investor networks
Implicit communication: same price analysis

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Example: MACD (1970)

- Parameters 12, 26, and 9 days
- Trading week then: 6 days
- Now: 5 days
Investor SVN

Tumminello et al. (2011b): daily Finish investment fluxes
Datasets

1. Swissquote (SQ): individual traders
2. Large Bank: institutional traders

Number of transactions

D. Challet, M. Lallouache, R. Chicheportiche  Predicting investment fluxes from implicit lead-lag investor networks
Step 1: Implicit network + community, daily

Challet et al. (2015)  EUR.USD SQ
Implicit network + community, hourly

Challet et al. (2015) EUR.USD SQ
Challet et al. (2015)

GBP.USD

EUR.USD
Cluster stability

movie GBP.USD 2014
movie EUR.USD 2014
Step 2: lead-lag SVN

- Determine groups
- Group lead-lag?

SNVs: p-values of

\[ s_g(t) \rightarrow s_{g'}(t + 1) \]
Compute p-value of \( \{s_g(t), s_g'(t + 1)\} \)
Predicting investment fluxes from implicit lead-lag investor networks
Origin of lead-lag

1. Explicit communication
2. Delayed reaction
   1. faster news
   2. same strategy, faster parameters
      \[ \text{MA(CD)} \rightarrow 12, 26, 5 \rightarrow 10, 25, 4: \]
   3. Master in Finance + ...
Lead-lag: number of links vs time

D. Challet, M. Lallouache, R. Chicheportiche
Predicting investment fluxes from implicit lead-lag investor networks
Lead-lag between groups: inter-temporal networks

Goal: predict global flux sign \((B - S)\) out of sample

1. Sliding in-out sample periods
2. Determine groups in-sample
3. Calibrate machine learning in-sample
4. Predict next SIGN

D. Challet, M. Lallouache, R. Chicheportiche
Predicting investment fluxes from implicit lead-lag investor networks
<table>
<thead>
<tr>
<th>$F_{1,t}$</th>
<th>$F_{2,t}$</th>
<th>$F_{1,t-1}$</th>
<th>$F_{2,t-1}$</th>
<th>hour</th>
<th>sign $(B - S)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-2</td>
<td>NA</td>
<td>0</td>
<td>1</td>
<td>+1</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>3</td>
<td>-2</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>3</td>
<td>????</td>
</tr>
</tbody>
</table>
Machine learning: random forest

Random forest = collection of random trees

- Classification tree: $P$ predictors, $T$ data points each
- Each tree: bootstrap of data points
- Each node: cut according to criterion e.g. $\text{predictor}_i > 2$
Prediction results

- 12 weeks in-sample
- 24 hours out-of-sample
- Predict sign of flux
- Predictors: group actions \{-1,0,1,NA\}
- Random Forests

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
12 weeks in-sample
24 hours out-of-sample
Predict sign of flux
Predictors: group actions \{-1,0,1,NA\}
Random Forests
Prediction results: SQ

- 12 weeks in-sample
- 24 hours out-of-sample
- Predict sign of flux
- Predictors: group actions {-1,0,1,NA}
- Random Forests

![Graph showing cumulative sum of flux over time](image)
Out-of-sample performance by hour

- EUR/USD
- EUR/GBP
- USD/JPY

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Origins of performance peaks

D. Challet, M. Lallouache, R. Chicheportiche

Predicting investment fluxes from implicit lead-lag investor networks
Summary

Method

Clustering → communities → lead-lag → machine learning → prediction

Todo

- Find algorithmic strategy of groups
- Other fields
- Fund ownership
- Recommendations systems
- ...