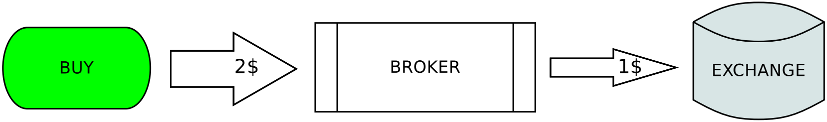
Predicting investment fluxes from implicit lead-lag investor networks

D. Challet[1,2] M. Lallouache [1] R. Chicheportiche [3,4]

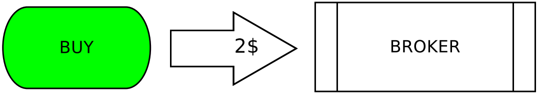
[1] CentraleSupélec [2] Encelade Capital SA [3]CFM [4] previously at Swissquote Bank SA

September 8, 2015

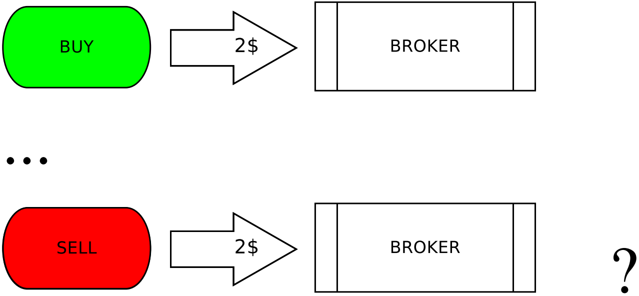
Broker internal order matching



versus



Internal order matching



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 δI_t and r_t
3. Constraints
4. Minimize cost function

Broker's gain

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

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.)
- ...

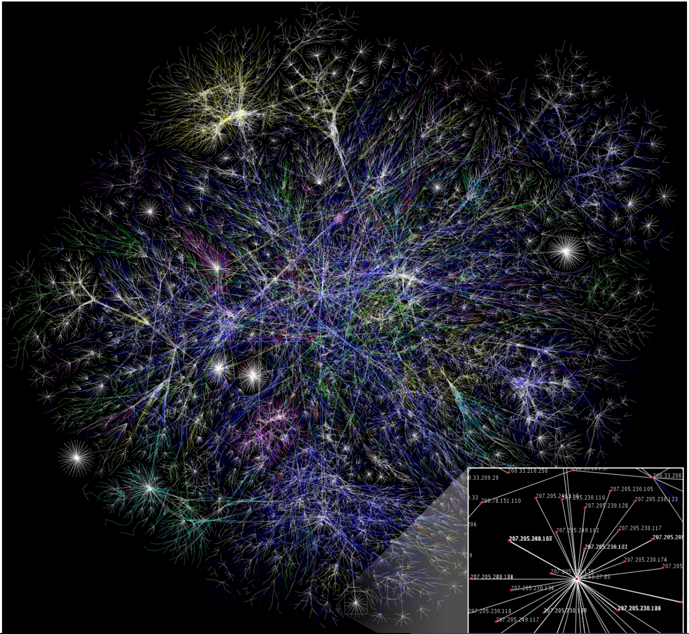
Unsupervised

- Similarity measure

  $\rightarrow$ categories

Supervised classification

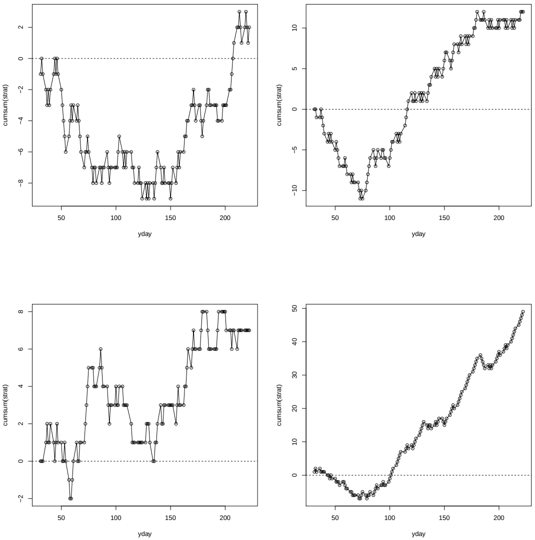
Real networks



Unsupervised learning

Unsupervised learning

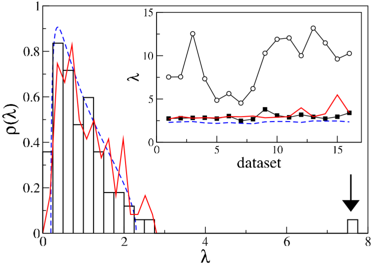
FX traders



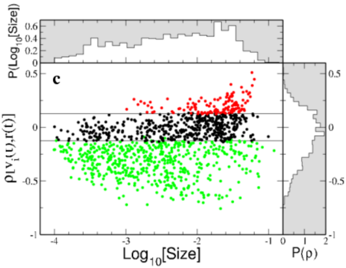
Unsupervised classification I
Spanish brokers

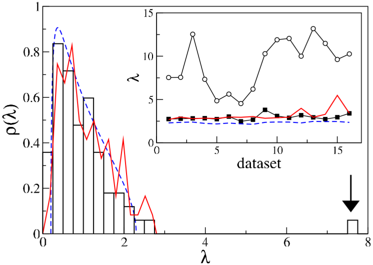
Lillo et al. (2008)

- Daily inventory change $V_i(t)$
- Correlation matrix $C \sim E(V_i V_j)$
- Principal Component Analysis + Random Matrix theory



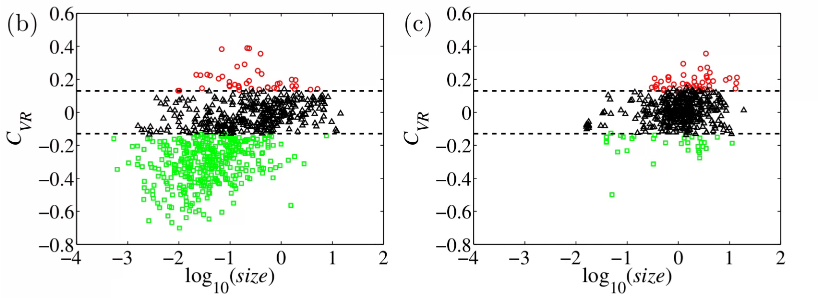
Lillo et al. (2008)



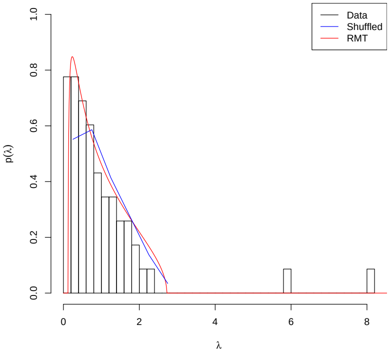


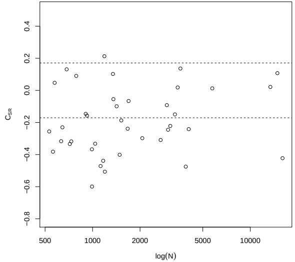
Unsupervised classification II
Equities, Chinese investors

Zhou et al. (2012)

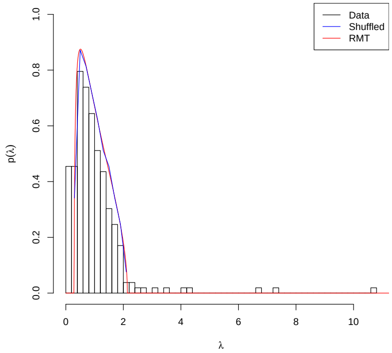


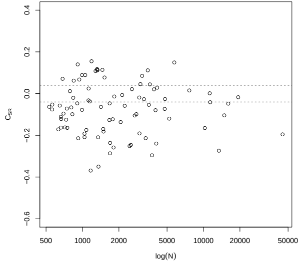
Unsupervised classification III: FX, individual investors, daily





Unsupervised classification IV: FX, individual investors, hourly





Unsupervised classification
Statistically validated networks (SVNs)

Tumminello et al. (2011a)

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

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

- 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

Trader-trader communication: explicit

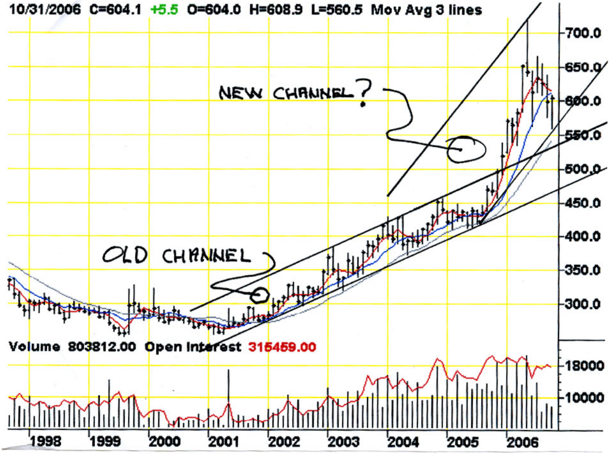
Trader-trader communication: explicit

Do traders look at prices?

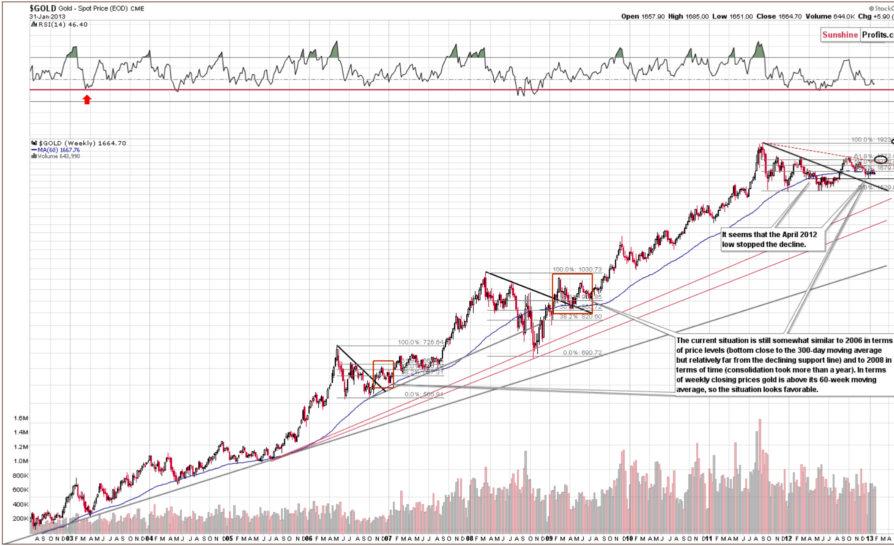
Implicit AND explicit communication

Implicit communication: same price analysis

Implicit communication: same price analysis



Implicit communication: same price analysis

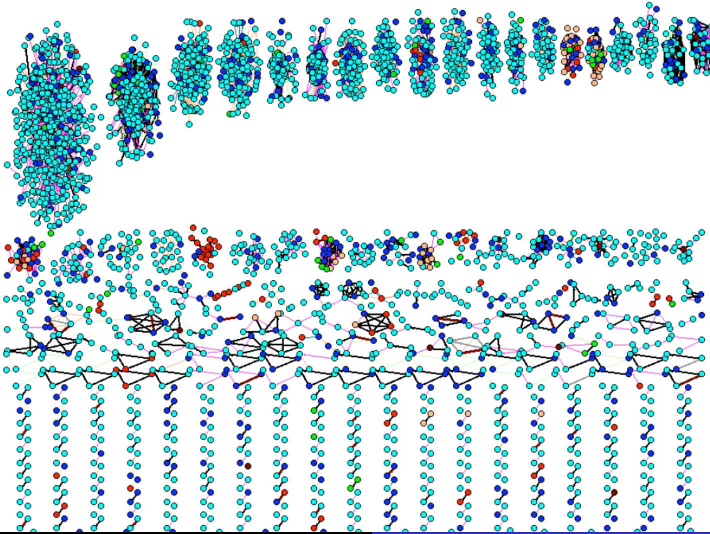


Example: MACD (1970)

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

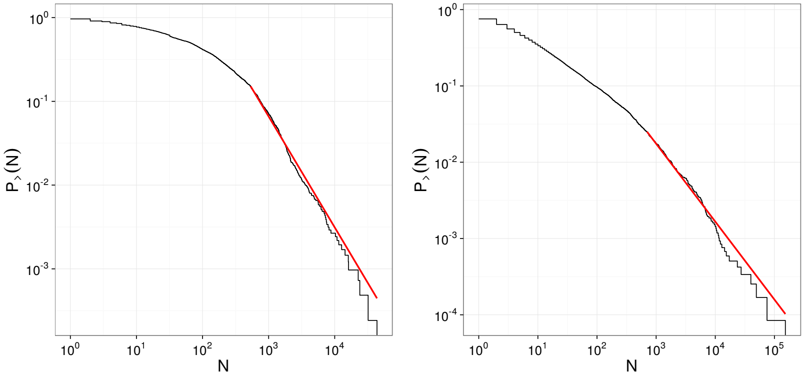
Investor SVNs
group=disconnected subgraphs

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



Datasets

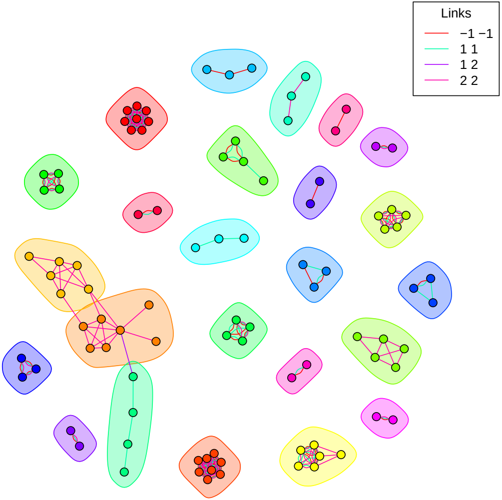
1. Swissquote (SQ): individual traders
2. Large Bank: institutional traders



Number of transactions

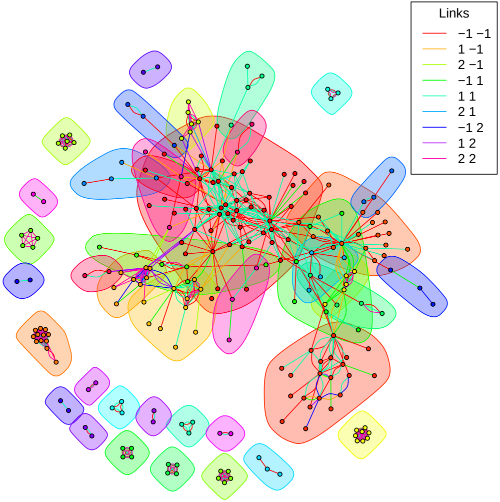
Step 1: Implicit network + community, daily

Challet et al. (2015) EUR.USD SQ



Implicit network + community, hourly

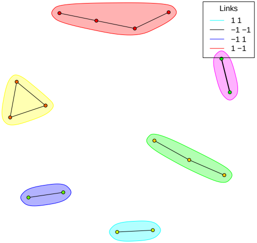
Challet et al. (2015) EUR.USD SQ



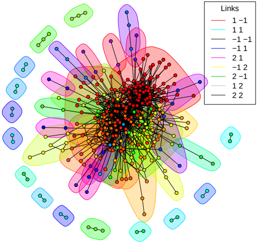
Implicit network + community detection, LB, hourly

Challet et al. (2015)

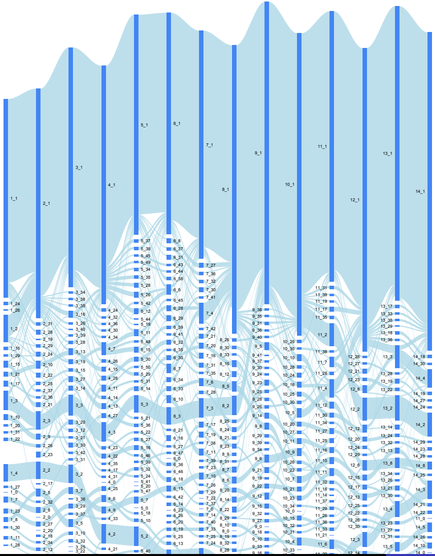
GBP.USD



EUR.USD



Cluster stability



movie GBP.USD 2014

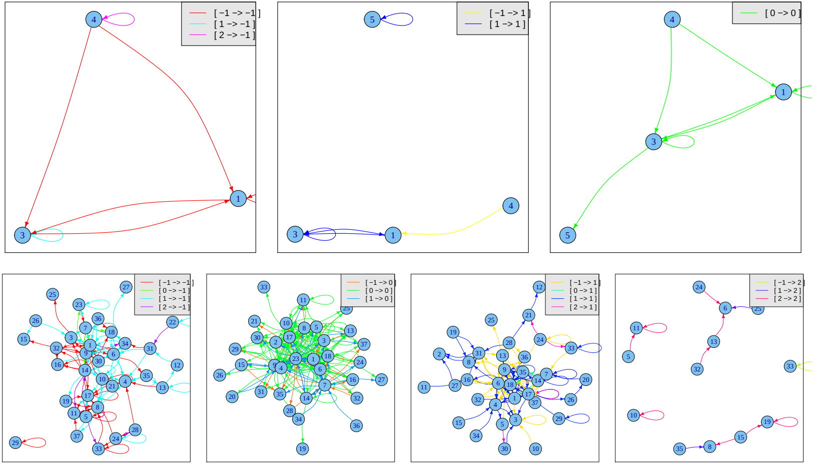
movie EUR.USD 2014

Step 2: lead-lag SVNs

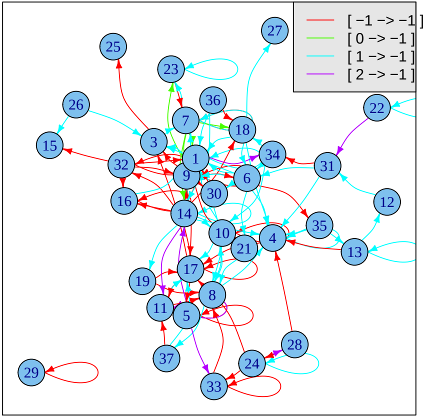
- determine groups
- group lead-lag?
- SNVs: p-values of $s_g(t) \rightarrow s_{g'}(t+1)$

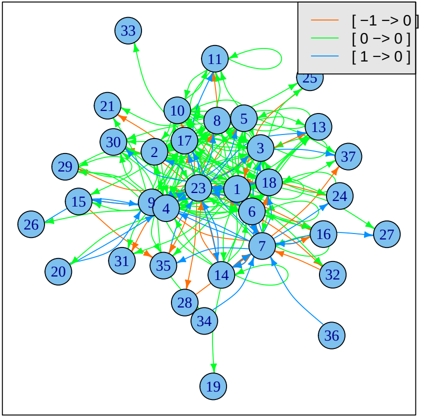
Intertemporal interaction networks (FX LB)

Compute p-value of $\{s_g(t), s_{g'}(t+1)\}$



EUR.USD (LB)



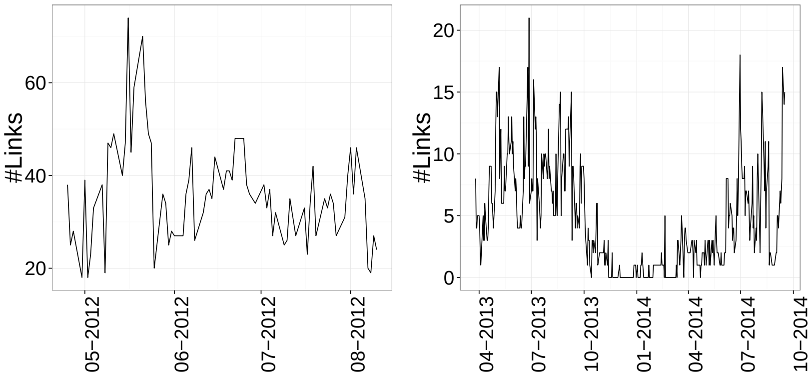


Origin of lead-lag

1. Explicit communication
2. Delayed reaction
    1. faster news
    2. same strategy, faster parameters

       MA(CD) $12, 26, 5 \to 10, 25, 4$:
    3. Master in Finance + ...

Lead-lag: number of links vs time



SQ LB

EUR.USD

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

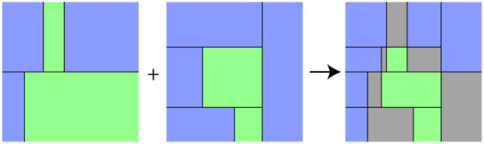
Predictors+predictee

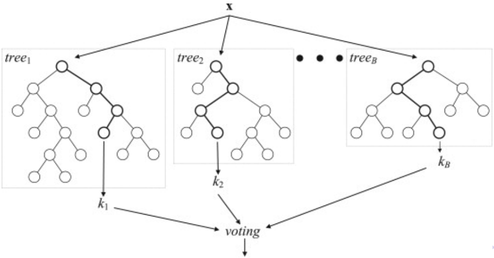
$F_{1,t}$	$F_{2,t}$	$F_{1,t-1}$	$F_{2,t-1}$	hour	sign $(B-S)$
3	-2	NA	0	1	+1
NA	NA	3	-2	2	-1
1	1	NA	NA	3	????

Machine learning: random forest

Random forest = collection of random trees

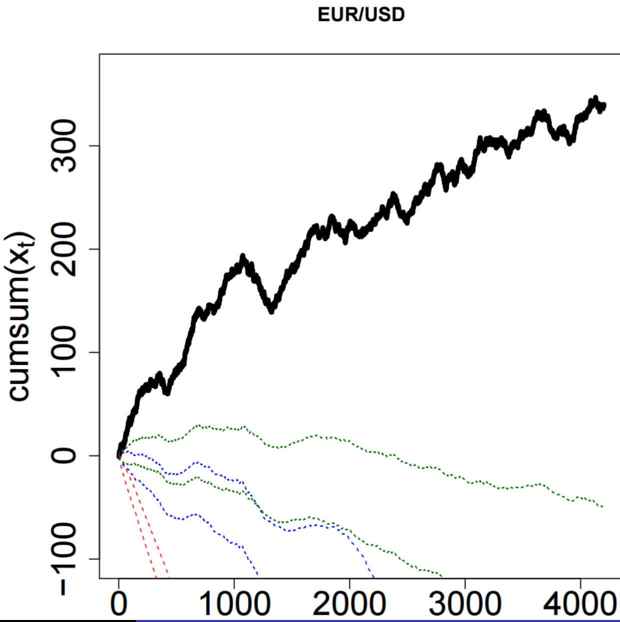
- Classification tree: P predictors, T data points each
- Each tree: bootstrap of data points
- Each node: cut according to criterion e.g. $predictor_i > 2$





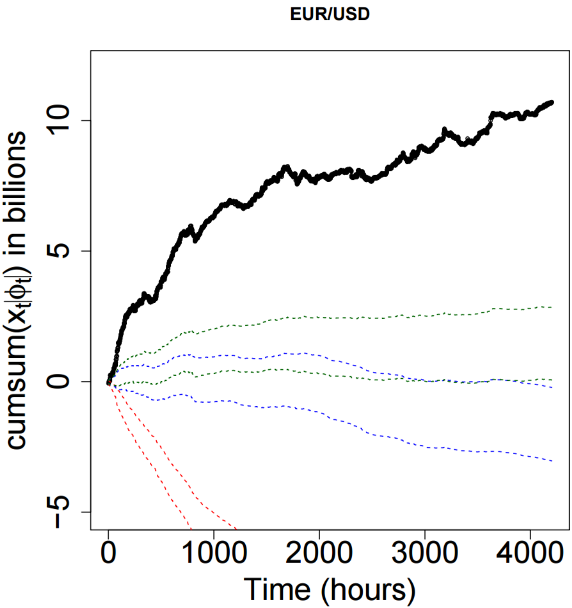
Prediction results

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



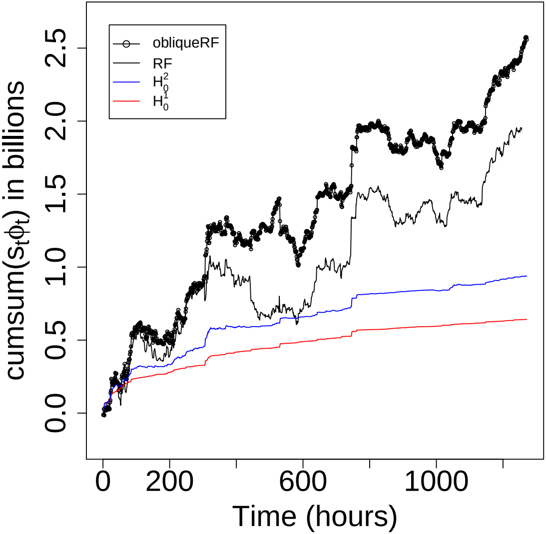
Prediction results

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

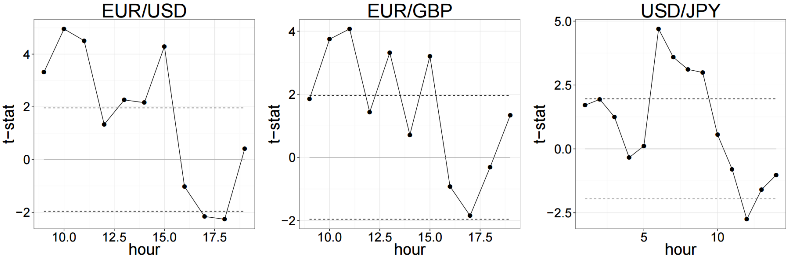


Prediction results: SQ

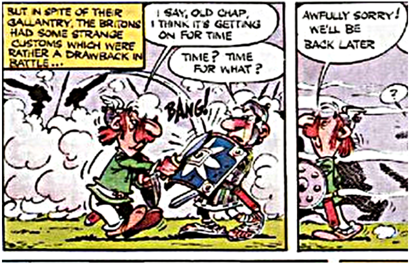
- 12 weeks in-sample
- 24 hours out-of-sample
- Predict sign of flux
- Predictors: group actions {-1,0,1,NA}
- Random Forests



Out-of-sample performance by hour



Origins of performance peaks



Summary

Method

Clustering $\rightarrow$communities $\rightarrow$ lead-lag $\rightarrow$ machine learning $\rightarrow$ prediction

Todo

- Find algorithmic strategy of groups
- Other fields
- fund ownership
- recommendations systems
- ...