Hybrid Statistical Agent Based Model for Financial Market

Pankaj Kumar

Perm State University, Russian Federation

(kumar@econ.psu.ru)

February 5, 2013

Overview

Why fancy name "Statistical" added?

- Financial markets are viewed as (macroscopic) complex systems with an internal (microscopic) structure consisting of many agents interacting so as to generate the systemic property.
- This complexity represents big challenges for researchers to decipher the market microstructure.
- The traditional analytical model posses serious difficulties in analysing market.
- The emerging fields of econophysics (Stanley, (1999), Bouchaud and Potters, (2004), Chakraborti et al., (2011a)), agent based model(Chakraborti et al., (2011b)) and computational finance, provides means to tackle some of the limitations of the analytical models in finance.

Why fancy name "Statistical" added?

- Financial markets are viewed as (macroscopic) complex systems with an internal (microscopic) structure consisting of many agents interacting so as to generate the systemic property.
- This complexity represents big challenges for researchers to decipher the market microstructure.
- The traditional analytical model posses serious difficulties in analysing market.
- The emerging fields of econophysics (Stanley, (1999), Bouchaud and Potters, (2004), Chakraborti et al., (2011a)), agent based model(Chakraborti et al., (2011b)) and computational finance, provides means to tackle some of the limitations of the analytical models in finance.
- However, agent based models also comes with some criticisms about problem of validation, calibration, parameters estimation and simulation complexity.

Objective

Modeling High-Frequency Trading Activity

Objective

Modeling High-Frequency Trading Activity

- Given the proprietary nature of high-frequency trading firms, there is relatively little known in the public domain about how they operate.
- At the same time, their importance in financial markets has grown very quickly in the past decade.
- As a result, the public understanding of financial markets has fallen somewhat behind the times.

Objective

Modeling High-Frequency Trading Activity

- Given the proprietary nature of high-frequency trading firms, there is relatively little known in the public domain about how they operate.
- At the same time, their importance in financial markets has grown very quickly in the past decade.
- As a result, the public understanding of financial markets has fallen somewhat behind the times.
- Establish a frame of reference to begin improving our understanding of the new dynamics of financial markets, of which high-frequency trading is a central contributor

Building Blocks: High Frequency Trading

1. Market Participants and the Trading Environment

Building Blocks: High Frequency Trading

1. Market Participants and the Trading Environment
2. Risks and High-Frequency Trading

Building Blocks: High Frequency Trading

1. Market Participants and the Trading Environment
2. Risks and High-Frequency Trading
3. High-Frequency Statistical Techniques

Building Blocks: High Frequency Trading

1. Market Participants and the Trading Environment
2. Risks and High-Frequency Trading
3. High-Frequency Statistical Techniques
4. Limit Order Books

Building Blocks: High Frequency Trading

1. Market Participants and the Trading Environment
2. Risks and High-Frequency Trading
3. High-Frequency Statistical Techniques
4. Limit Order Books
5. Market Impact, Market Models, and Trading Algorithms

Agent Ecology: Machine Learning Method

Empirically Validated Agent

- The dynamics of stock prices in complex financial market is influenced by market microstructure, diverse cogitative structure of traders, their attitude towards risk and collective activity.

Empirically Validated Agent

- The dynamics of stock prices in complex financial market is influenced by market microstructure, diverse cogitative structure of traders, their attitude towards risk and collective activity.
- The traders are classified into stylized class namely, fundamentalists and chartists (Frankel and Froot, 1990); contrarian and momentum (Chan et al., 1996); informed and uninformed (Grossman and Stiglitz, 1976); noisy zero intelligence (Gode and Sunder, 1993a) etc.

Empirically Validated Agent

- The dynamics of stock prices in complex financial market is influenced by market microstructure, diverse cogitative structure of traders, their attitude towards risk and collective activity.
- The traders are classified into stylized class namely, fundamentalists and chartists (Frankel and Froot, 1990); contrarian and momentum (Chan et al., 1996); informed and uninformed (Grossman and Stiglitz, 1976); noisy zero intelligence (Gode and Sunder, 1993a) etc.
- The classification assumptions are motivated by theoretical considerations, results of surveys and direct investigations of the trading profile (Tumminello et al.,2012).
- Although existing models take into account of agents heterogeneity, analytical tractability or strategics, it need to be verified empirically to validate underlying assumption.

Existing Approach

- Biclustering and Plaid models (Mankad et al., 2011)

Existing Approach

- Biclustering and Plaid models (Mankad et al., 2011)
- Markov Decision Process (MDP) and Inverse Reinforcement Learning(IRL)based approach (Kirilenko et al., 2011)

Existing Approach

- Biclustering and Plaid models (Mankad et al., 2011)
- Markov Decision Process (MDP) and Inverse Reinforcement Learning(IRL)based approach (Kirilenko et al., 2011)
- Network analysis(Tumminello et al., 2012) investigated special database to identify clusters of traders by using Statistically Validated Networks (SVN).

Data

- Moscow Interbank Currency Exchange (MICEX) Stock Exchange agent resolved data provides the identification of the trades, submission , cancellation, or modification of an order.
- National Stock Exchange (NSE) of India for the period January 2011 to June 2012. This database have flags instead of agent ID in first look of the sample.

Plaid Clustering Technique

The plaid clustering technique - a regressionbased method to describe empirical regularities in cross-sectional data was previously used only for a single, static data matrix. Our method extends the plaid model by making use of a time series of data matrices. The extension, which we refer to as the smooth plaid model (Mankad et al, 2011), is able to identify categories of traders and trading outcomes that persist through time

Biclustering

- The term biclustering was first used by Cheng and Church (2000) to refer to grouping procedures appropriate when both the samples and variables are of scientific interest.

Biclustering

- The term biclustering was first used by Cheng and Church (2000) to refer to grouping procedures appropriate when both the samples and variables are of scientific interest.
- In contrast, clustering methods belong to a closely related topic in machine learning and are concerned with discovering the structure of samples only.
- Biclustering methods extract groups of samples (rows) and variables (columns) to find homogeneous submatrices in static data matrix.

Biclustering

- The term biclustering was first used by Cheng and Church (2000) to refer to grouping procedures appropriate when both the samples and variables are of scientific interest.
- In contrast, clustering methods belong to a closely related topic in machine learning and are concerned with discovering the structure of samples only.
- Biclustering methods extract groups of samples (rows) and variables (columns) to find homogeneous submatrices in static data matrix.
- Here, samples are individual traders and the variables are measures of trading activity for each trader: trading volume, net position, change in inventory, trades per second, and median intertrade duration.
- A bicluster is then a group of traders and measurements of their trading activity that are similar

Plaid Model

- It introduces an additive "layer".

Plaid Model

- It introduces an additive "layer".
- A layer is a canonical matrix matching the dimensions of the given data matrix, with zeros everywhere except the biclustered elements.

Plaid Model

- It introduces an additive "layer".
- A layer is a canonical matrix matching the dimensions of the given data matrix, with zeros everywhere except the biclustered elements.
- In our application setting, the background "layer" accounts for market trends that affect trading behavior of all traders, such as for example, a major liquidity event (for example).
- One could also estimate a parametric model that that defies supply and demand rule.

Plaid Model

- It introduces an additive "layer".
- A layer is a canonical matrix matching the dimensions of the given data matrix, with zeros everywhere except the biclustered elements.
- In our application setting, the background "layer" accounts for market trends that affect trading behavior of all traders, such as for example, a major liquidity event (for example).
- One could also estimate a parametric model that that defies supply and demand rule.
- Subsequent layers can also be incorporated to repersent addition effects corresponding to specific traders and variables that exhibit a stong pattern not explained by background layers.

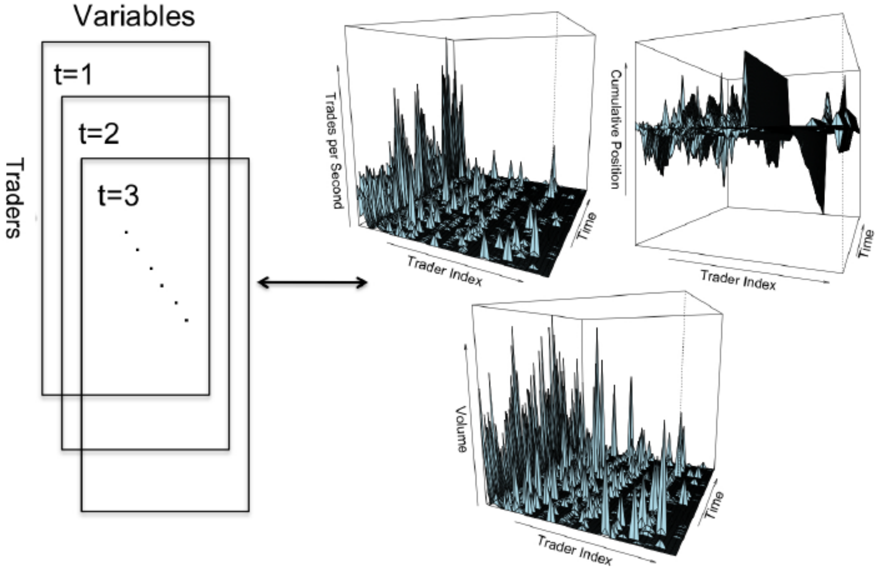
Mathematical Model

Let us consider a time series of matrices $\{X_{ij}^{(t)}\}_{t=1}^{T}$, where t is time, $i = 1 \ldots n$ and $j = 1 \ldots m$. Each row of matrix corresponds to a trader and the columns represent trading strategies at time t. For each n traders, at time t, if we observe similar variables measuring trading strategies, then, data matrix $X^{(t)}$ at time t can be represented by

$$\{X_{ij}^{(t)}\}_{t=1}^{T} = \mu_0^{(t)} + \sum_{k=1}^{K^{(t)}} \theta_{ijk}^{(t)} \alpha_{ik}^{(t)} \beta_{jk}^{(t)} \tag{1}$$

where μ_0 describe the background layer, and θ_{ijk} defines the bicluster effect; k is a layer index running to the number of biclusters K. The parameters α_{ik} and β_{jk} are indicator variables denoting bicluster membership for, respectively, the traders and variables.

Visual Representation



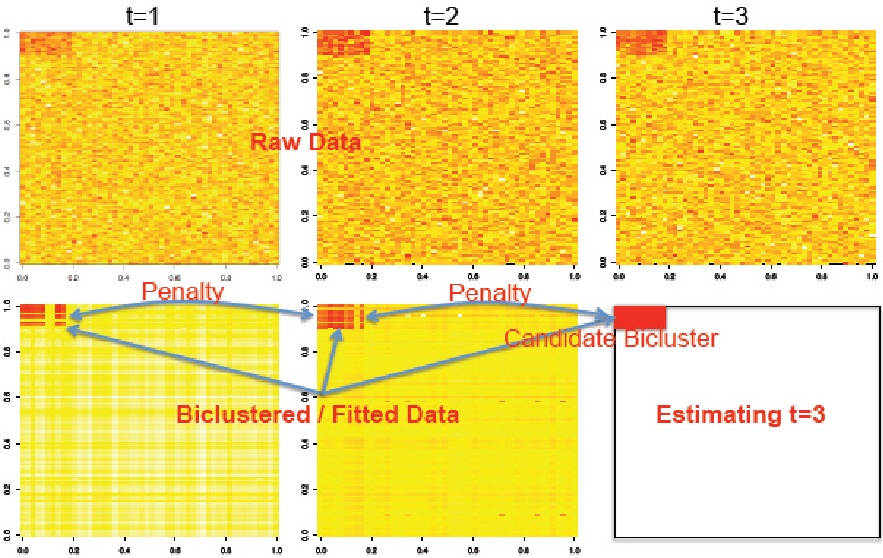
Mathematical Model contd...

Direct analysis of equation above would aggregate problems of losing information, transient pattern or structure irregularity. To mitigate the above problem, we employ constrained optimization to search for search for a K^{th} layer.

$$\begin{aligned} \min_{\{\theta^{T-W},\ldots,\theta^{T}\}} \quad & \sum_{T-W}^{T} \sum_{i-1}^{n} \sum_{j-1}^{m} (\hat{Z}_{ij}^{(t)} - \theta_{ijK}^{(t)} \hat{\alpha}_{iK}^{(t)} \hat{\beta}_{jK}^{(t)})^2 \\ + \quad & \lambda \sum_{T-W+1}^{T} \sum_{i-1}^{n} \sum_{j-1}^{m} (\theta_{ijK}^{(t)} \hat{\alpha}_{iK}^{(t)} \hat{\beta}_{jK}^{(t)} - \theta_{ijK}^{(t-1)} \hat{\alpha}_{iK}^{(t-1)} \hat{\beta}_{jK}^{(t-1)}) \end{aligned} \quad (2)$$

where, $\hat{Z}_{ij}^{(t)}$ is the residual data matrix at time t, λ is a tuning parameter and W is a parameter determining the number of previous time periods to consider.

Algorithm: Toy Example



Algorithms : Main Ideas

- to use the results from previous time steps to form candidate biclusters for the current time period, then apply the penalization framework to filter out transient patterns.
- After candidate biclusters from previous times have been exhausted, a final search is done for other patterns that were not captured previously using the static model.
- Permutation test used in the stopping criterion for the smooth plaid algorithm, and selection of the parameters λ and W.

Algorithms : Main Ideas

- to use the results from previous time steps to form candidate biclusters for the current time period, then apply the penalization framework to filter out transient patterns.
- After candidate biclusters from previous times have been exhausted, a final search is done for other patterns that were not captured previously using the static model.
- Permutation test used in the stopping criterion for the smooth plaid algorithm, and selection of the parameters λ and W.
- Matrix observations at different times should be permuted separately, so that global time effects are maintained.
- For detail description, please refer (Mankad et. al., 2011)

Illustration and Simulation

- We illustrate and validate our methodology with simulated data. We bootstrapped one day data into seven days to illustrate.
- Time series of data matrices, where each matrix has 50 rows and columns, with embedded biclusters that evolve through time.
- We compare the proposed smooth plaid model with scoring rule about trading.
- We considered the effect of background layer, and bicluster mean is considered constant.
- We make λ too large to have persistence smoothing effect.

Illustration and Simulation

- We illustrate and validate our methodology with simulated data. We bootstrapped one day data into seven days to illustrate.
- Time series of data matrices, where each matrix has 50 rows and columns, with embedded biclusters that evolve through time.
- We compare the proposed smooth plaid model with scoring rule about trading.
- We considered the effect of background layer, and bicluster mean is considered constant.
- We make λ too large to have persistence smoothing effect.
- The smooth plaid procedures perform favorably in this synthetic setting by discovering the true, underlying biclustering structure and evolution

Emperical Validation

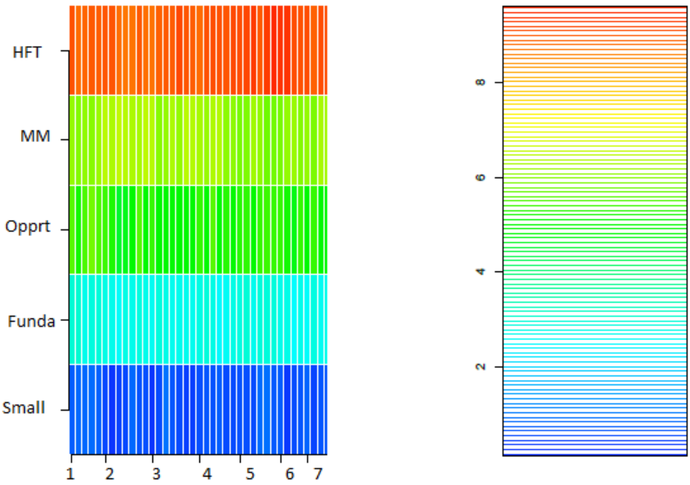
- We use the following variables to cluster traders into groups: trades per second, trading volume (total number of contracts traded), strategic runs (Harsbouch, 2012)and median duration for each trader.
- Each trading variable measures different aspects of how much, in which direction, and how quickly each trader transacts.
- Once organized, the data contains 2549 rows (traders), 4 columns, and 186 time periods.
- After we apply our algorithm to the data, we use additional filtering on the fitted values to separate traders into five broad groups.

Agents Ecology

Table : Agent Clusters

Traders Type	Smooth Plaid	Strategic Scoring
HFT	4	2
Market Maker	46	72
Opportunistic	598	624
Fundamental Buyer/Sellers	670	512
Small	1231	1339

Heatmap: Estimated log average strategic runs



Thank You!