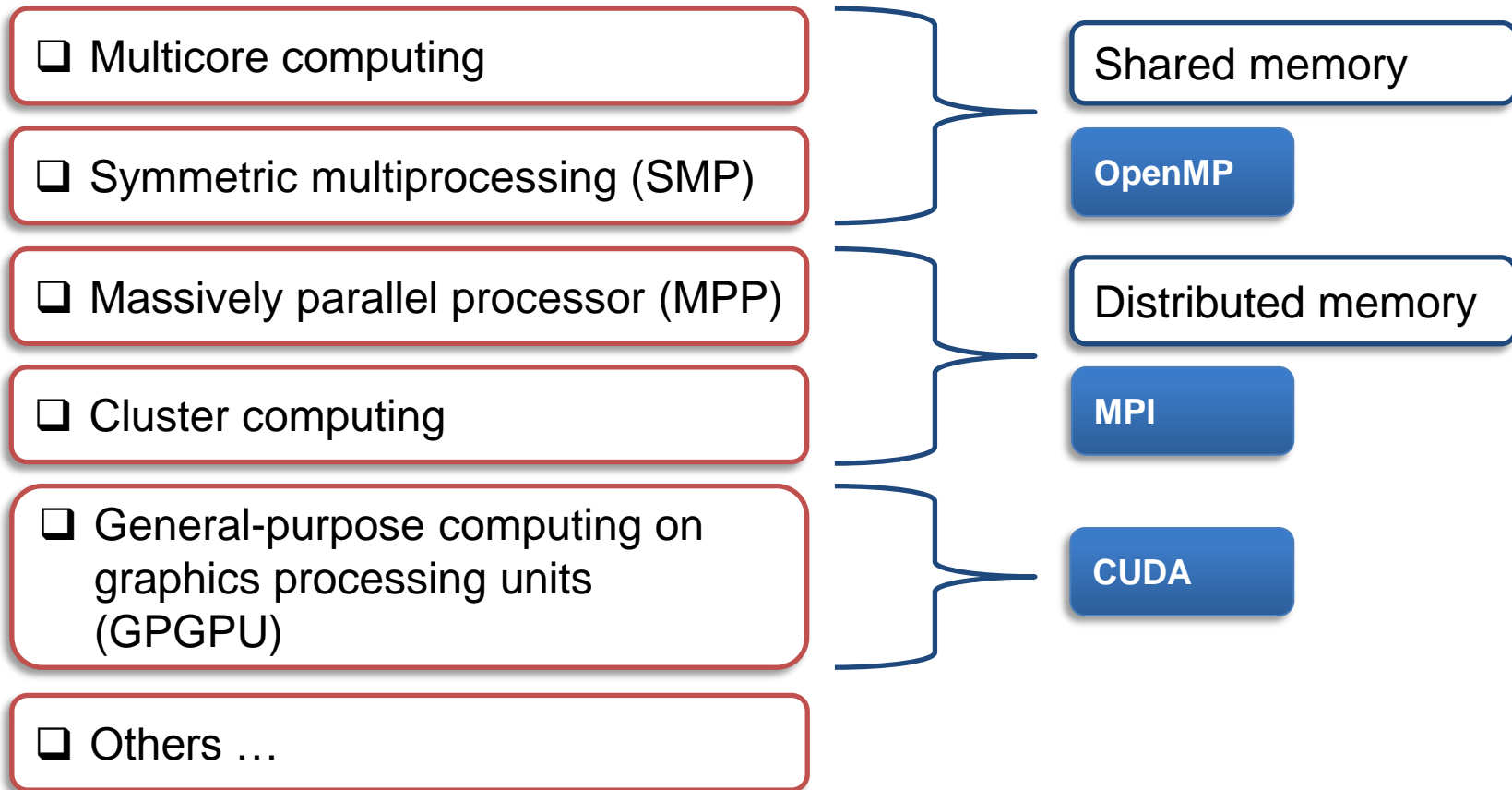


Parallel computations in financial markets research

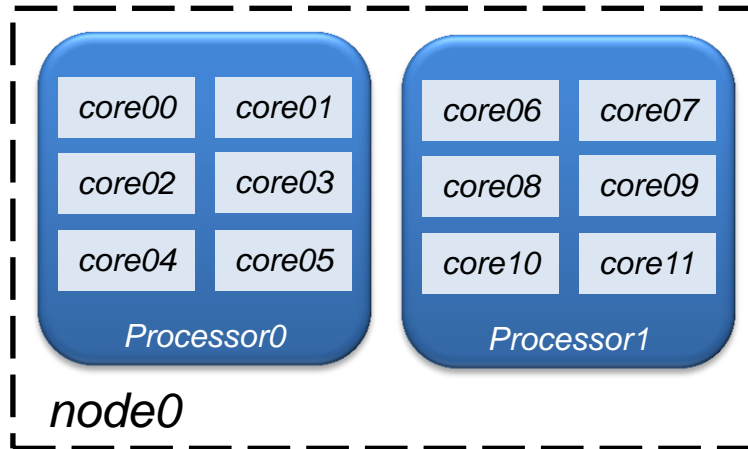
Viacheslav Arbuzov (PSNRU, Prognoz Risk Lab)

Konstantin Gavrilov (PSNRU, Prognoz)



Prognoz Risk Lab has access to the supercomputers:

- MMP Cluster
- Tesla-PGU Cluster



- ✓ Number of nodes with GPU: 3
- ✓ Number of GPU per node: 4
- ✓ GPU type: Nvidia Tesla Fermi
- ✓ RAM per node: 12 Gb



Technical info:

- ✓ Supercomputer type: Cluster
- ✓ Number of nodes: 20
- ✓ Number of Cores per node: 12
- ✓ CPU type: Intel Xeon 5670 (2.93 GHz)
- ✓ RAM per node: 48 Gb
- ✓ OS: Linux CentOS 5
- ✓ Theoretical peak performance: 8992 GFlops
- ✓ Maximal LINPACK performance achieved: 4883 Gflops
- ✓ Communication network: QDR Infiniband
- ✓ Transport network: Gigabit Ethernet
- ✓ Service network: Gigabit Ethernet

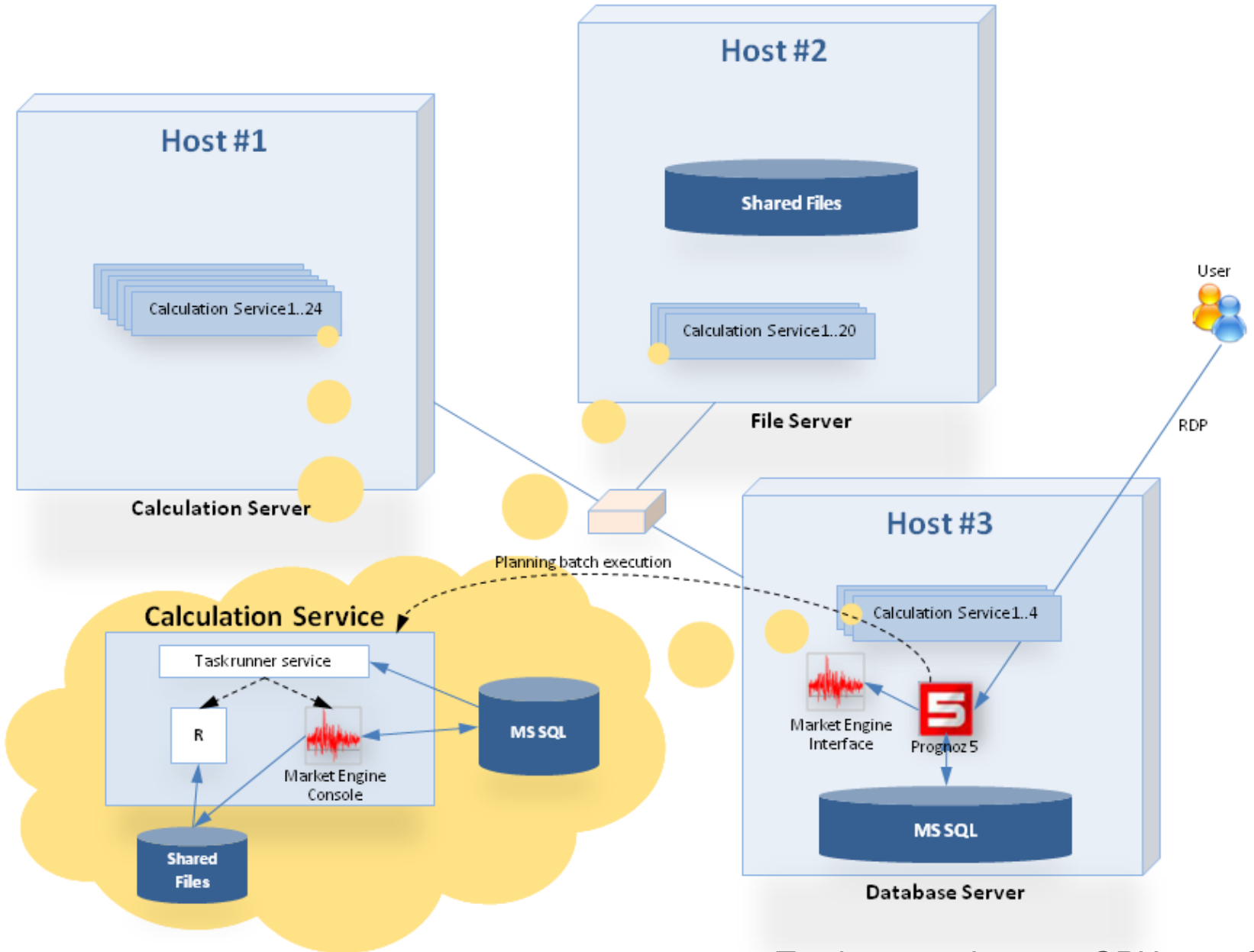
Computing cluster for reverse engineering, agent-based simulation and prediction of microstructure and liquidity of the financial market

Technical info:

- ✓ Installation Site: Perm state university
- ✓ Supercomputer type: Cluster
- ✓ Number of nodes: 3
- ✓ Number of Cores per node: 12
- ✓ CPU type: Intel Xeon 5650 (2.66 GHz)
- ✓ RAM per node: 64 Gb
- ✓ OS: Windows Server 2003



Architecture of cluster



Total: 48 services, 72 CPU, 228 Gb RAM

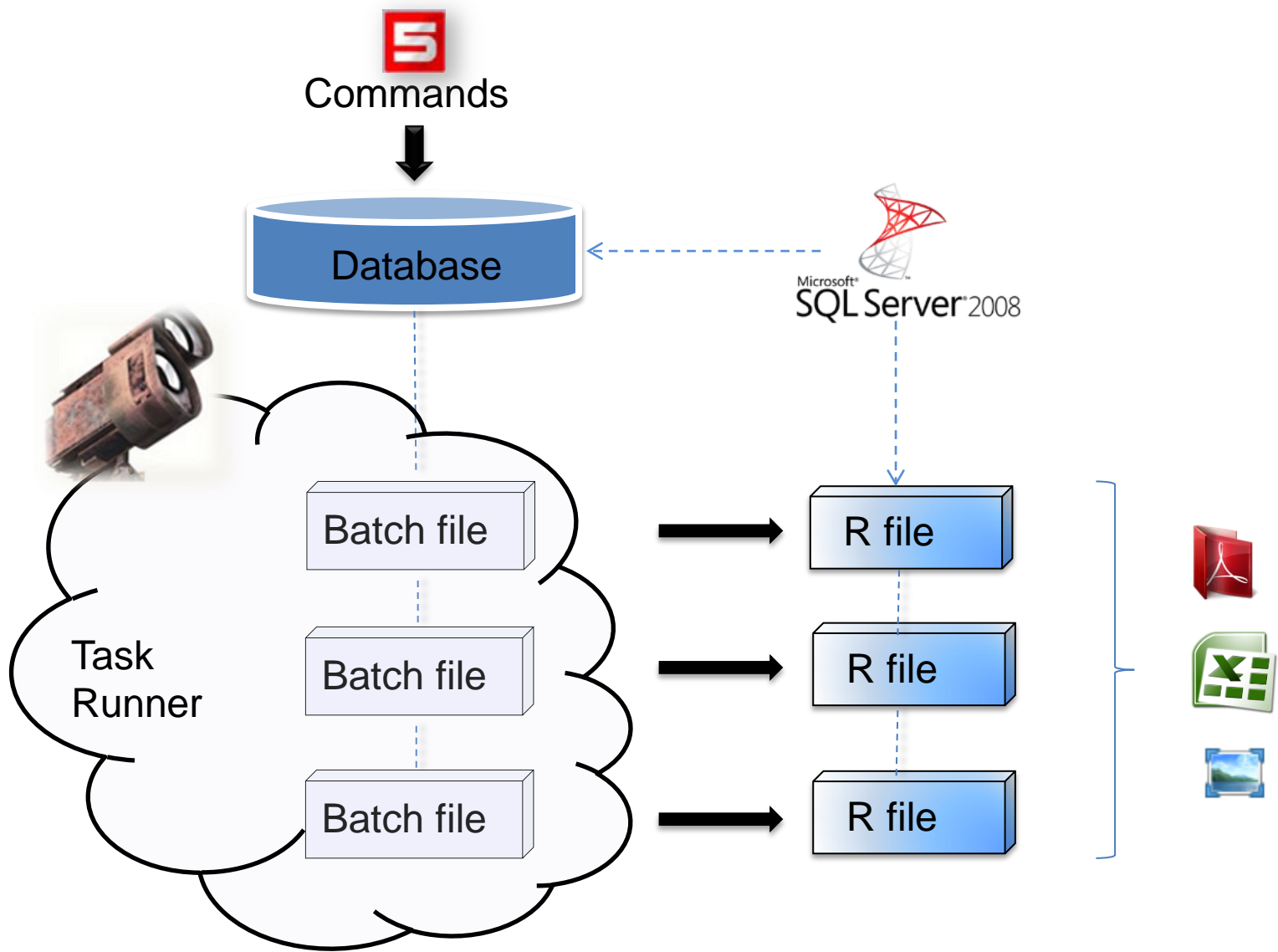


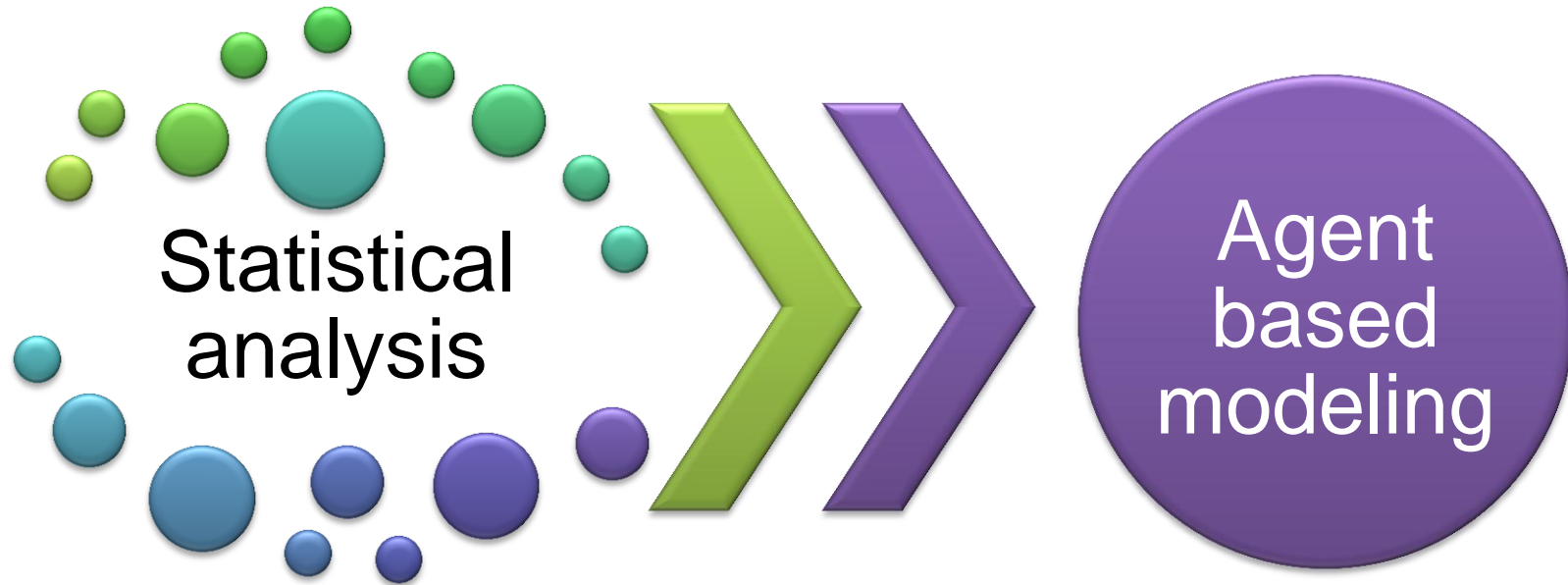
- ❑ R is statistical and graphical programming environment
- ❑ Appeared in 1993 and designed by Ross Ihaka and Robert Gentleman
- ❑ R is a GNU project
- ❑ R – a free implementation of the S language
- ❑ It runs on a variety of platforms including Windows, Unix and MacOS
- ❑ It contains advanced statistical routines not yet available in other packages

There is more than 4300 packages in which allow use specialized statistical techniques, graphical devices, import/export capabilities, reporting tools, etc.

Useful Links:

- ✓ www.r-project.org
- ✓ www.statmethods.net
- ✓ www.r-bloggers.com
- ✓ addictedtor.free.fr/graphiques
- ✓ www.use-r.org
- ✓ www.r-analytics.blogspot.com





- *Data Sample*
- *Long memory (ACF, PACF) and CCDF*
- *Variables PDF fitting*
- *Volatility estimation*
- *Diagonal effect*
- *Leverage effect*
- *Liquidity analysis*
- *Shocks analysis*
- *Other....*

- *Calculation of characteristics*
- *Agents cluster analysis*
- *Herding behavior*
- *Traders activity visualization*
- *CDA Monte-Carlo simulation*
- *Backtesting*

Gaussian (normal)

Log-normal

Student's

Gamma

Exponential

Power-law (Pareto)

Cauchy

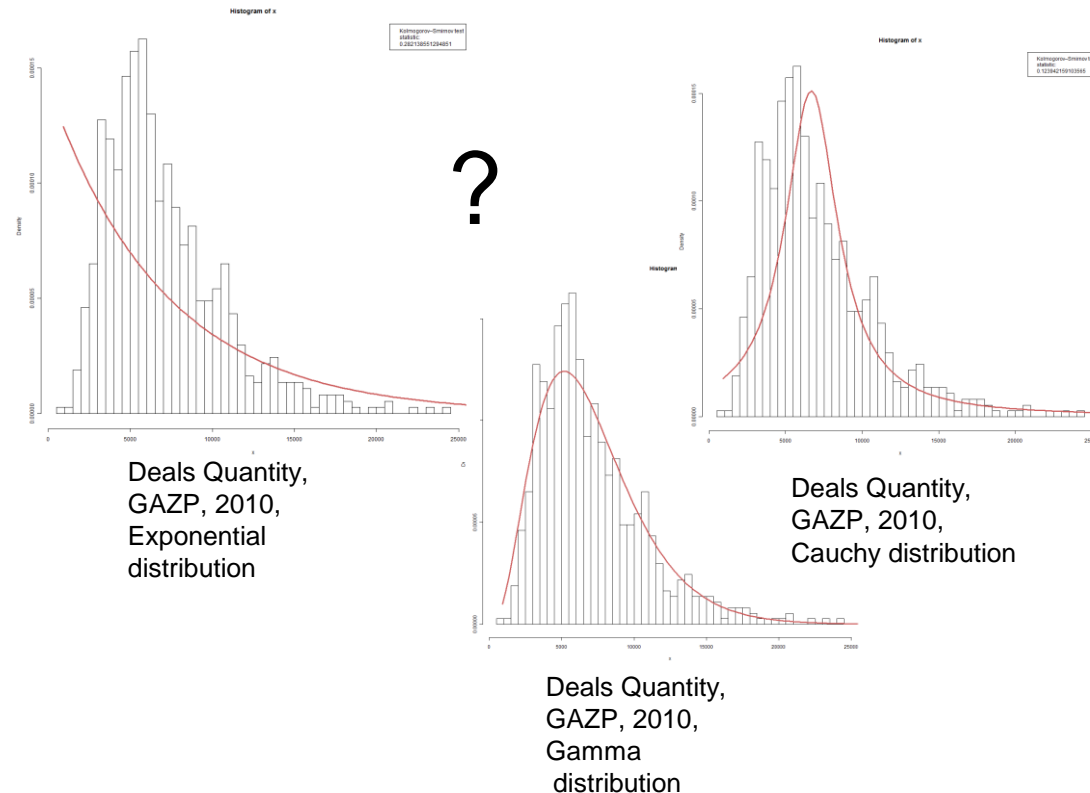
Weibull

Estimated parameters:

- Bid-Ask Spread
- Deals Quantity
- Bid
- Asc
- Price
- Volume-weighted average price

Parameters estimated with MLE, using:

- *MASS* package
- *igraph* package
- *fGarch* package

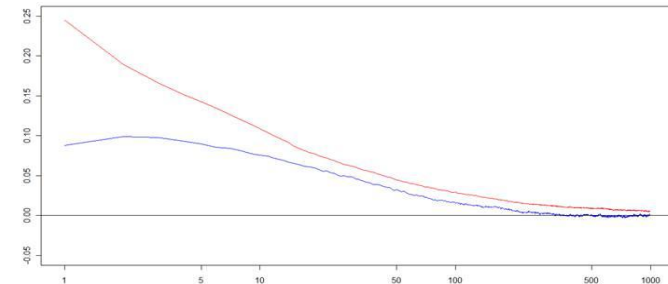
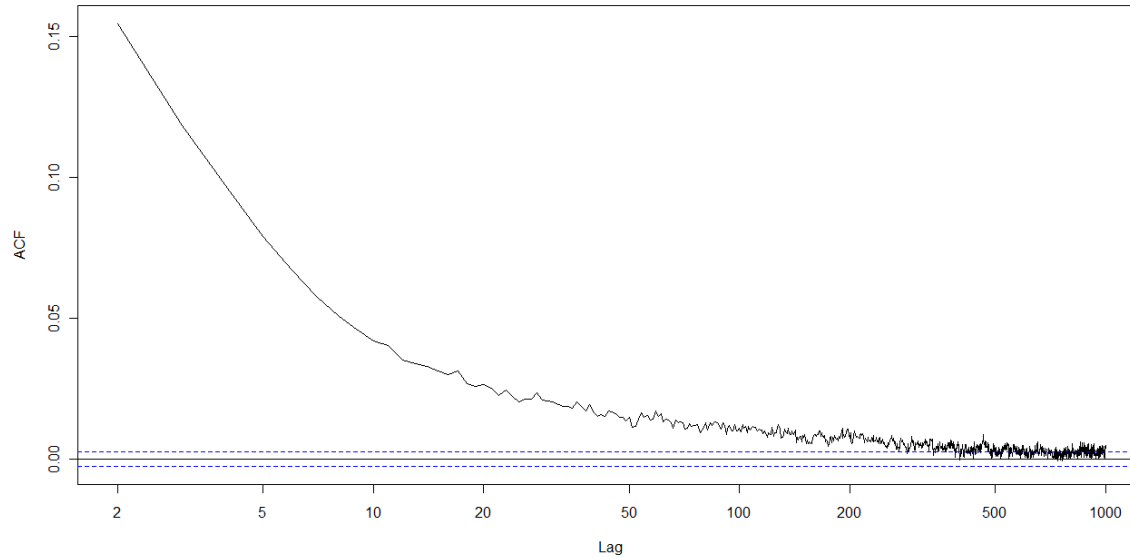


Autocorrelation function of order flow shows that it has long memory. More over some instruments has long memory caused by herding behaviour.

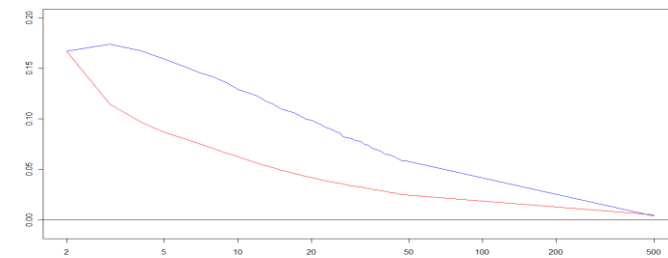
Autocorrelation function

ACF splitting

ACF herding



Autocorrelation function on brokers level



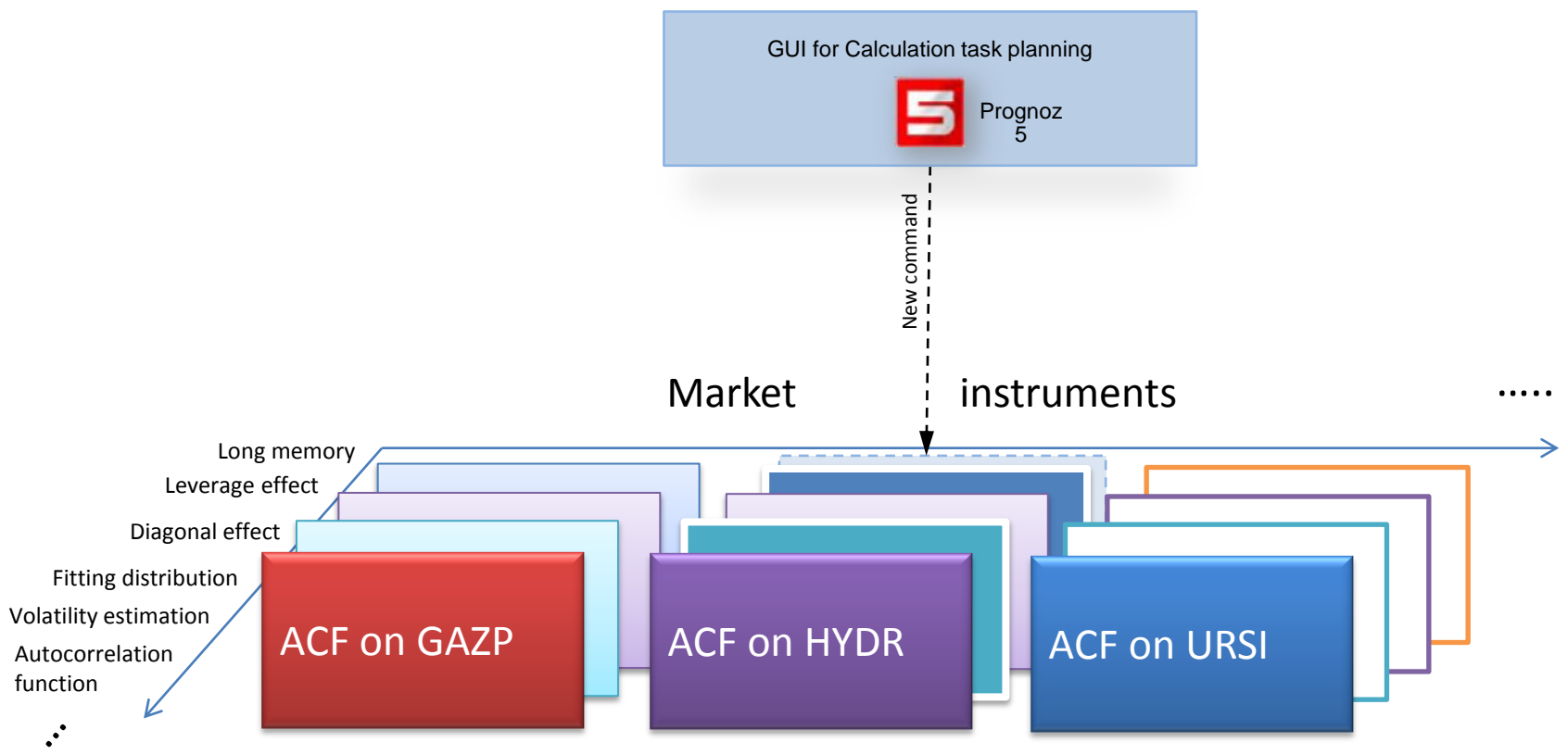
Autocorrelation function on agents level

$$C_{split}(\tau) = \sum_i \left(P^{ii}(\tau) \left[\frac{1}{N^{ii}(\tau)} \sum_t \epsilon_t^i \epsilon_{t+\tau}^i \right] - \left[P^i \frac{1}{N^i} \sum_t \epsilon_t^i \right]^2 \right) *$$

$$C_{herd}(\tau) = \sum_{i \neq j} \left(P^{ij}(\tau) \left[\frac{1}{N^{ij}(\tau)} \sum_t \epsilon_t^i \epsilon_{t+\tau}^j \right] - P^i P^j \left[\frac{1}{N^i} \sum_t \epsilon_t^i \right] \left[\frac{1}{N^j} \sum_t \epsilon_t^j \right] \right) *$$

Red line – splitting in order flow
Blue line – herding in order flow

* Fabrizio Lillo. Order flow in financial markets: Origin of persistence and impact of metaorders, 2011



How to create agent based model (ABM) ?

Grouping agents to homogeneity clusters

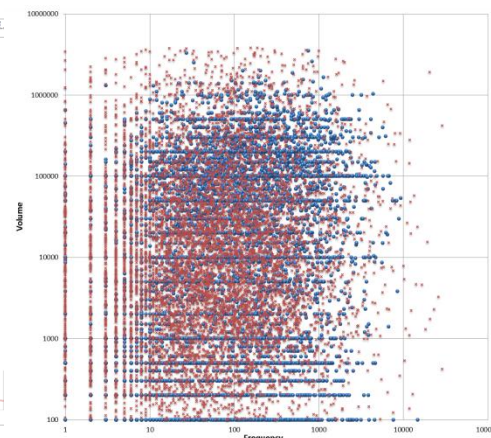
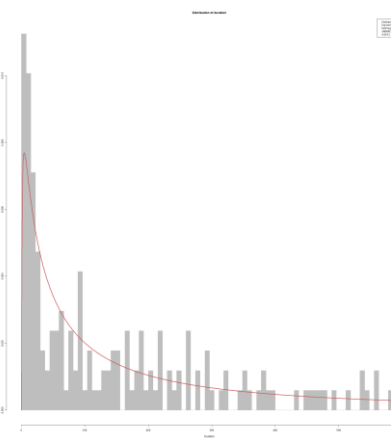
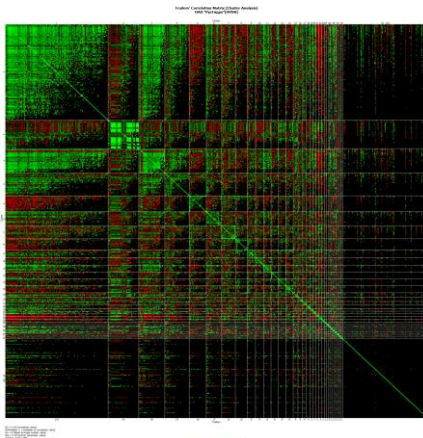
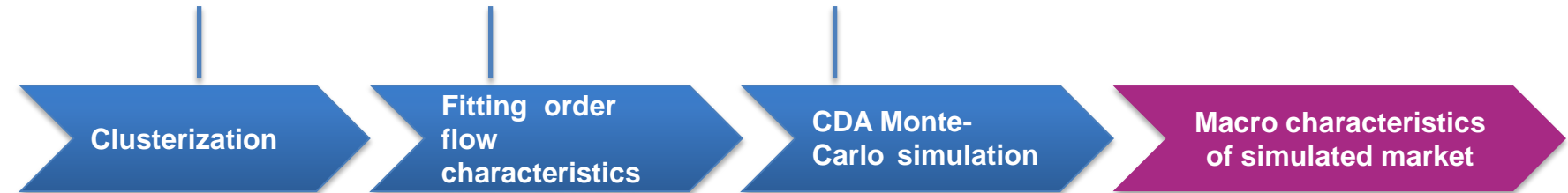
Search and fitting of order flow characteristics

Generation scenarios of order flow based on copula (**CDA mechanism sim.**)

SVN* and Infomap**

Marginal distributions

Copula



* Michele Tumminello, Salvatore Micciche, Fabrizio Lillo, Jyrki Piilo, and Rosario N. Mantegna, Statistically validated networks in bipartite complex systems, (2010)

** Martin Rosvall and Carl T. Bergstrom, Maps of random walks on complex networks reveal community structure, (2008)



Questions:

How to distinguish flows?

What characteristics should we select?

Frequency

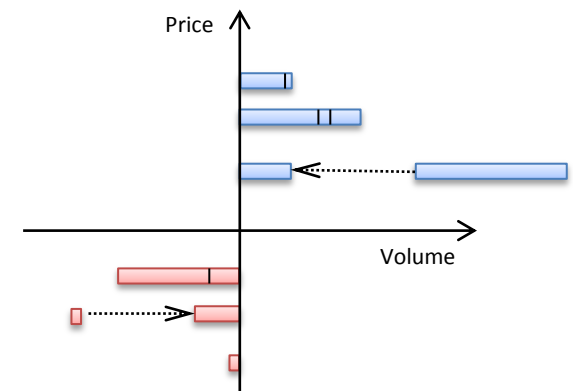
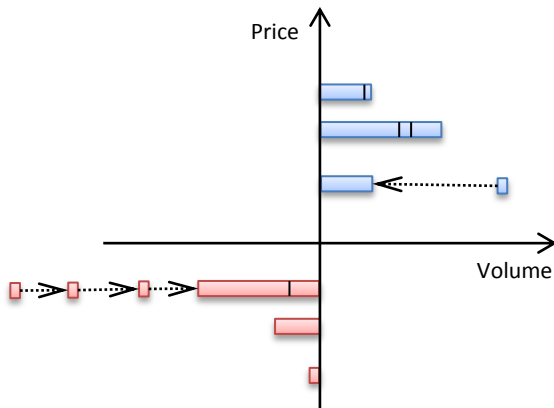
- How often orders are coming?

Width

- In what part of order book orders are coming?

Volume

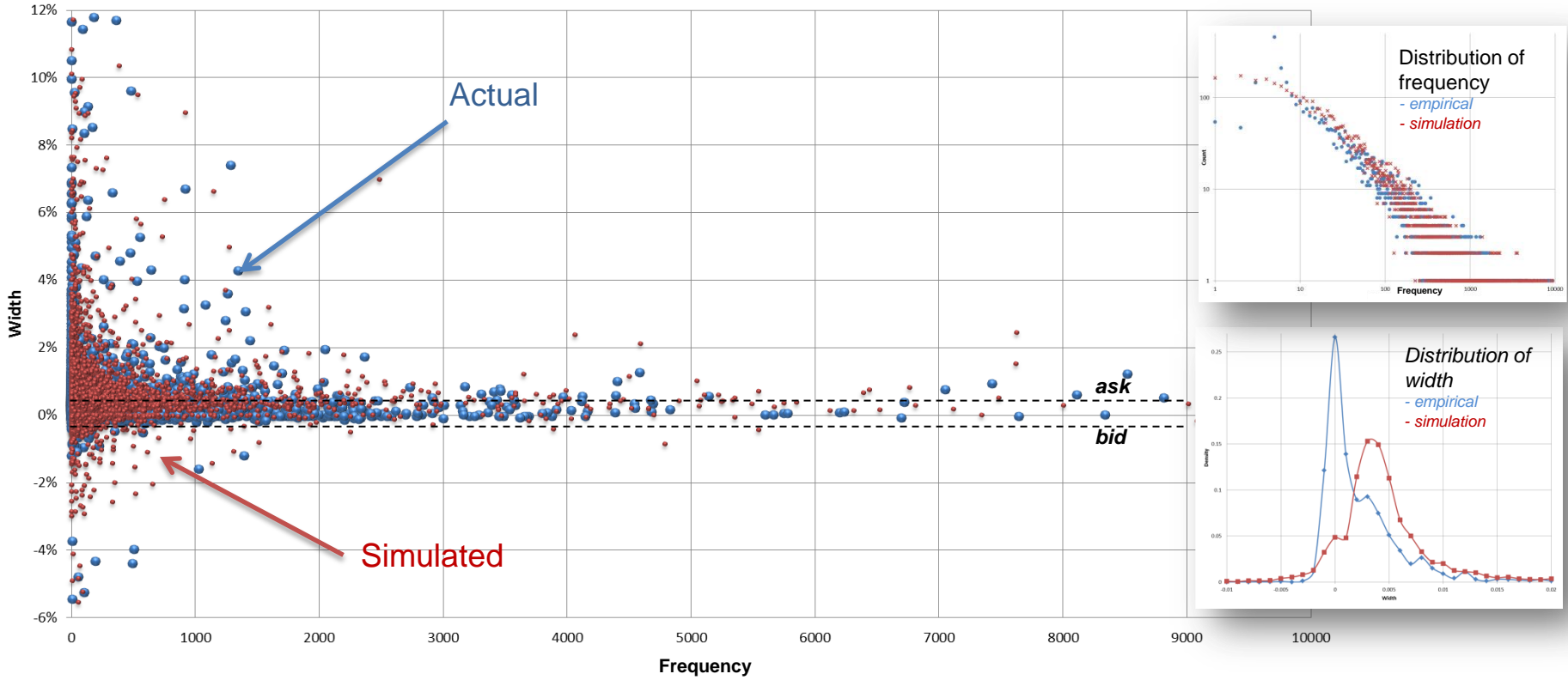
- What is the volume of coming orders?

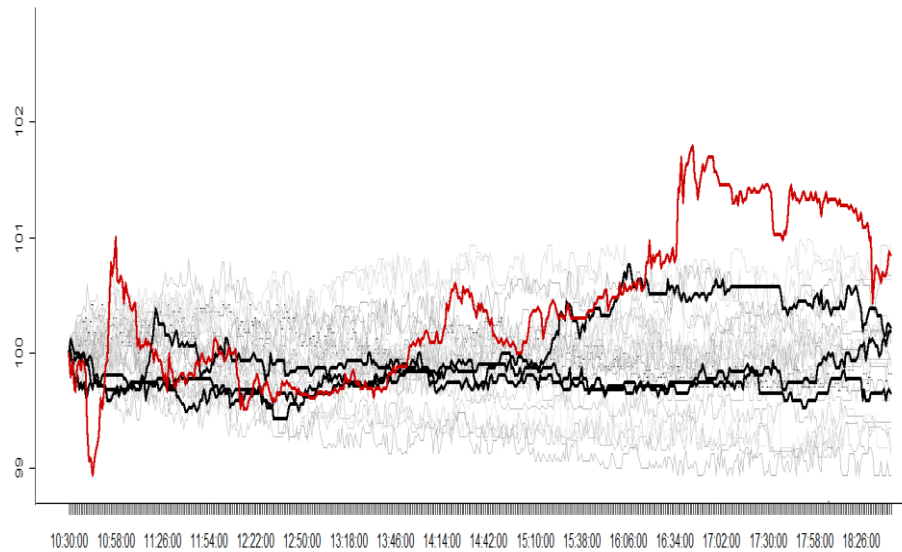


Simulation based on copula

To save the dependences between characteristics of order flow, we use copula mechanism in generation scenarios.

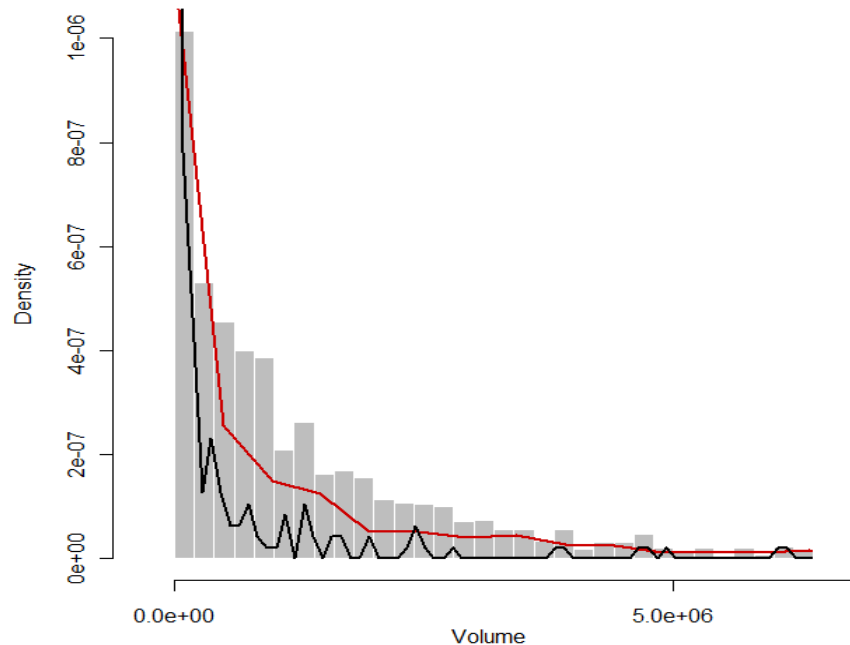
- Copula generation:
- Allows to describe the dependence between random variables.
 - There are many parametric copula families available:
 - Gaussian copula
 - Gumbel
 - Frank
 - others





Macro characteristics for backtesting of model:

- Price volatility
- Distribution of returns
- Distribution of volume
- Quantity of deals
- Presence of stylized facts in price changes
- Presence of stylized facts in order flow



Potential problems:

- Heterogeneity of clusters
- Empirical distributions do not converge to marginal distributions
- 'Zero intelligence' agents
- System does not take into account external shocks

Master class

Parallel computations on **“MMP” cluster**

and

Parallel computations on **“Tesla – PGU” cluster**