Mutual information-based hierarchies on Warsaw's stock exchange

Paweł Fiedor

Uniwersytet Ekonomiczny w Krakowie
s801dok@wizard.uek.krakow.pl

Lublin, 14.05.2014

Motivation

Network analysis of financial markets assumes their being complex systems, yet Pearson's correlation coefficient is used.

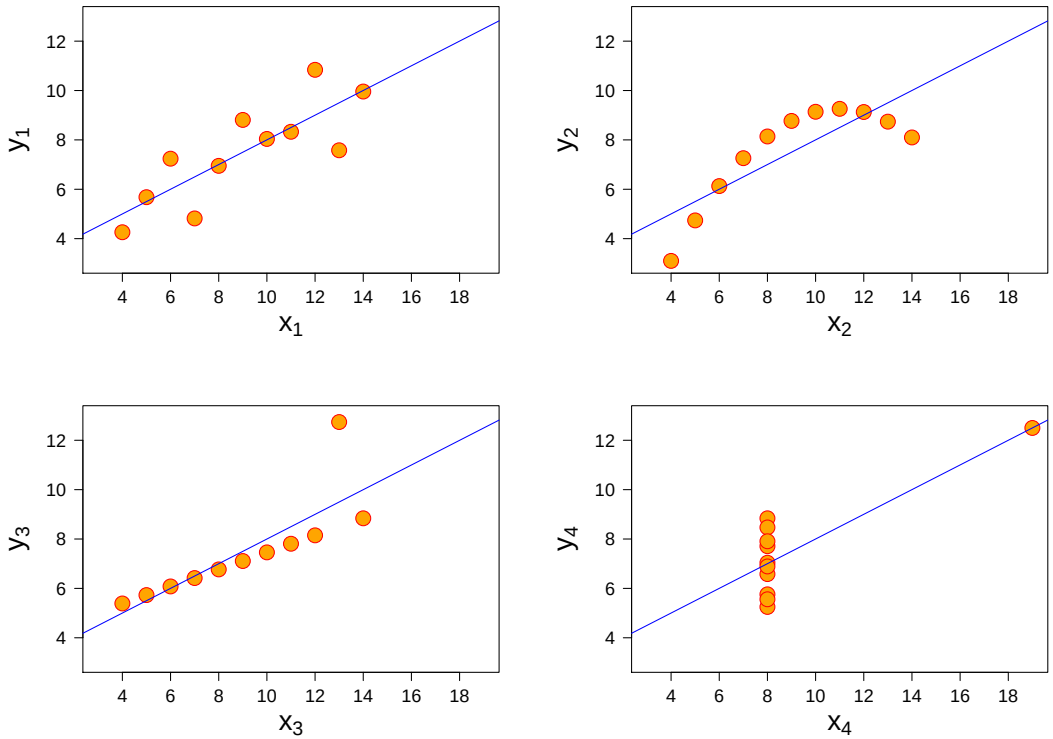


Figure : Anscombe, Francis J. (1973) Graphs in statistical analysis. American Statistician, 27, 17–21

Motivation

Simple validation for threshold based networks. MI for 2 independent random variables (estimated with frequencies) has a distribution close to Γ with $a = (|X| - 1)(|Y| - 1)/2$ and $b = 1/(N \ln 2)$.
Thus a test of significance for α:

$$MI(x_i, y_j) \geq \Gamma_{1-\alpha}(\frac{1}{2}(|X| - 1)(|Y| - 1), \frac{1}{N \ln 2}),$$

where $\Gamma_{1-\alpha}$ is a quantile of Gamma distribution for $1 - \alpha$, instead of computationally expensive permutation tests.

Distance for networks – ρ

Log returns:

$$r_i(t) = ln(P_i(t)/P_i(t-\tau))$$

Pearson's correlation coefficient:

$$\rho_{ij} = \frac{< r_i r_j > - < r_i >< r_j >}{\sqrt{< r_i^2 - < r_i >^2 >< r_j^2 - < r_j >^2 >}}$$

$$d(i,j) = \sqrt{2(1-\rho_{i,j})}$$

Distance for networks – I_S

Mutual information:

$$I_S(X,Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} = H(X)+H(Y)-H(X,Y),$$

$$\hat{H}(X) = -\sum_{x\in X}\frac{\Lambda(x)}{n}\log\frac{\Lambda(x)}{n},$$

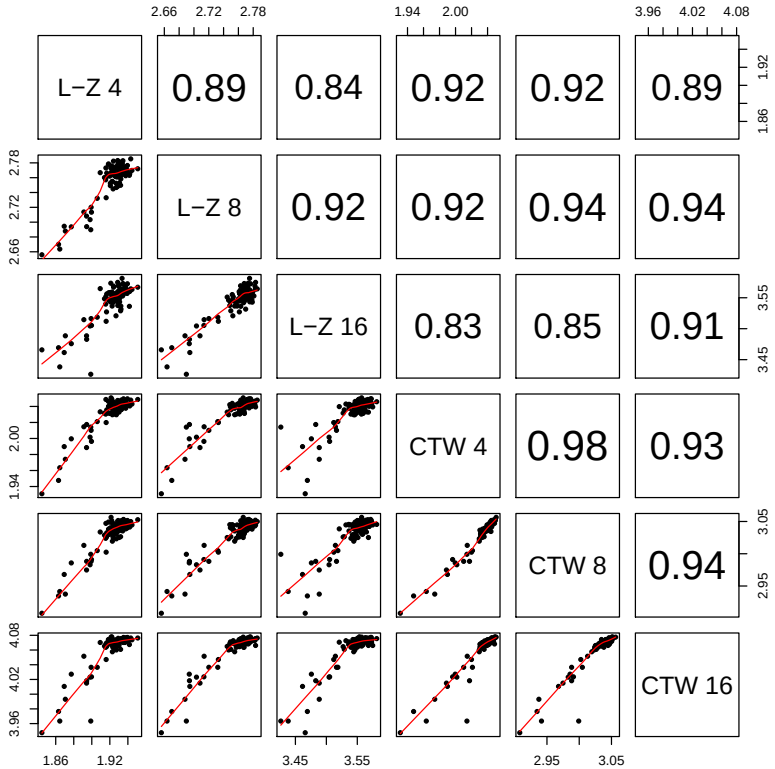
$\Lambda(x)$ is the number of observations equal x, and n is the sample size. Downward biased estimator, others may be used.

$$d(X,Y) = H(X|Y) + H(Y|X) = H(X,Y) - I_S(X,Y),$$

$$d(X,Y) = H(X) + H(Y) - 2I_S(X,Y)$$

Discretization

Alphabet should not be too large. Cardinality of 2 excludes volatility thus 4, 8, or 16 should be used.



Differences between approaches 1

Table : Correlations for computed graphs

	$\rho \sim I_S$		MST∼PMFG	
Year	MST	PMFG	ρ	I_S
2000	0.495	0.559	0.793	0.796
2001	0.754	0.873	0.888	0.891
2002	0.680	0.704	0.825	0.878
2003	0.534	0.718	0.808	0.818
2004	0.523	0.692	0.855	0.853
2005	0.472	0.625	0.817	0.763
2006	0.555	0.727	0.886	0.816
2007	0.773	0.640	0.888	0.913
2008	0.606	0.760	0.883	0.925
2009	0.715	0.696	0.895	0.863
2010	0.666	0.732	0.867	0.886
2011	0.422	0.490	0.901	0.819
2012	0.615	0.669	0.823	0.799
2013	0.382	0.433	0.811	0.718
Average	0.585	0.666	0.853	0.838

Differences between approaches 2

Table : Network comparison

Network	Tree ratio	Graph ratio	Clustering
ρ	62.22%	49.06%	35.20%
I_S	66.67%	55.81%	41.60%
Reference	11.28%	11.28%	100.00%

Sector analysis 1

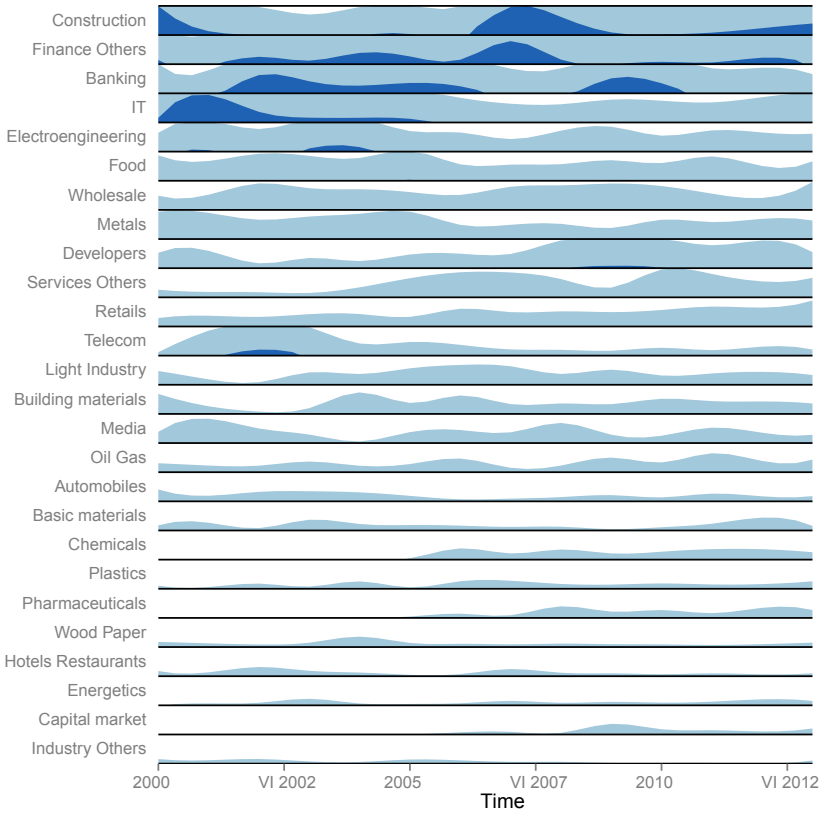


Figure : Aggregate Markov centrality by sector (I_S) (MST)

Sector analysis 2

The most important sectors:

Conclusions

- MI-based approach is more general;
- MI-based networks are significantly different;
- Sector analysis can lead to simple yet useful analytics.

Closing

s801dok@wizard.uek.krakow.pl
pawel@fiedor.eu
Phys. Rev. E **89**, 052801
http://arxiv.org/a/fiedor_p_1
http://ideas.repec.org/f/pfi237.html