

ОЦЕНКА НЕОПРЕДЕЛЕННОСТИ ПРОЦЕДУР ИДЕНТИФИКАЦИИ РЫНОЧНЫХ ГРАФОВ

В. В. Цыганков, П. А. Колданов

Национальный исследовательский университет «Высшая школа экономики»

Нижний Новгород, Россия

Лаборатория Алгоритмов и Технологий

Анализа Сетевых Структур, Нижний Новгород, Россия,

E-mail: vvtsygankov@edu.hse.ru, pkoldanov@hse.ru

В работе анализируются новые показатели неопределенности процедур идентификации рыночных графов на основе анализа данных 12 фондовых рынков различных стран. Подчеркивается, что преимуществом анализируемых показателей является возможность их расчета только на основе данных наблюдений.

ESTIMATING THE UNCERTAINTY OF MARKET GRAPH IDENTIFICATION

V. V. Tsygankov, P. A. Koldanov

The study analyzes new measures of the uncertainty in procedures for identifying market graphs, using data from 12 stock markets across different countries. It is emphasized that a key advantage of the measures under review is that they can be computed solely from observational data.

Введение. Одним из популярных инструментов для анализа сетевых моделей рынка является рыночный (отсеченный) граф, который иллюстрирует силу и топологию корреляций на фондовом рынке. Большое количество работ посвящено построению такого графа с целью анализа корреляций на конкретных фондовых рынках, см. [2][3][4][5][6][14][13][15]. Однако неопределенность полученных выводов, связанная со случайным характером наблюдений изучена недостаточно. Неопределенности выводов возникают из-за отсутствия информации об истинной корреляции и использования оценки корреляционной связи для получения выводов. Для проведения такого исследования введены показатели неопределенности см. [11], основанные на недавно предложенном методе для разделения выводов о корреляциях на два набора: статистически значимый набор выводов и статистически незначимый набор выводов или зона неопределенности. При таком подходе мера неопределенности естественным образом связана с характеристиками этих двух наборов выводов. Преимуществом предлагаемого подхода является возможность построения и нахождения характеристик этих двух наборов только на основе наблюдаемых данных, в то время как классическое определение неопределенности, основанное на функции риска, требует предположение об истинном отсеченном графе. Целью настоящей работы является проведение сравнительного анализа неопределенности выводов о связях на 12 фондовых рынках, используя подход описанный в [11].

Новизна и обоснование выбранных методов решения задач. Анализ корреляционных связей (γ_{ij}) между акциями, является ключевым инструментом для построения сетевых моделей фондовых рынков. Традиционные методы построения, таких сетевых структур, как максимальное остовное дерево или пороговый граф, сталкиваются с проблемой неопределенности. Классические подходы анализа неопределенности, основанные на функции риска (см. [17][8]) требуют знания истинной сетевой структуры, что делает их неприменимыми в реальных условиях, где такая информация отсутствует.

В данной работе для решения этой проблемы используется разделение выводов о корреляциях на статистически значимые и незначимые (см. [9] [11]). Основная идея заключается в разделении всех возможных пар акций на три категории:

- Статистически значимые сильные связи.
- Статистически значимые слабые связи.
- Статистически незначимые связи заданной силы.

Для формализации этой идеи используются совместные верхние и нижние доверительные границы для множества рёбер отсеченного графа, что позволяет оценить неопределенность без знания истинной структуры этого графа.

Отсеченным графом сетевой модели рынка называется граф, множество вершин которого совпадает с множеством акций фондового рынка, ребро между двумя вершинами добавляется в граф, тогда и только тогда, когда его вес больше заданного порога. Под весом ребра понимается значение коэффициента корреляции между доходностями акций. Более подробно см. [7].

Обозначения. Пусть мы рассматриваем рынок состоящий из N акций. Тогда общее число пар акций вычисляется следующим образом:

$M = |J| = \frac{N(N-1)}{2}$, где $J = \{(i, j): i < j; i, j = 1, \dots, N\}$ и $|J|$ это количество элементов множества J .

Также введем $x = (x_i(t)), i = 1, \dots, N; t = 1, \dots, n$, который содержит выборку из n наблюдений за доходностями акций в количестве N .

За $\gamma_{i,j}$ тут всегда принимается сила связи пары (i, j) из множества J .

Метод построения верхних и нижних границ. *Определение.* Множество $U_e(x, \gamma_0, P^*)$ называется верхней доверительной границей при заданном пороге γ_0 и заданной вероятностью P^* при которой множество $U_e(x, \gamma_0, P^*)$ включает в себя множество ребер истинного рыночного графа (см. [11], стр. 6). Методология построения $U_e(x, \gamma_0, P^*)$ базируется на тесте $\phi_{i,j}^e(x)$, проверки гипотезы (см. [9]): $h_{i,j}^e: \gamma_{i,j} > \gamma_0$ против $k_{i,j}^e: \gamma_{i,j} \leq \gamma_0, i, j = 1, \dots, N(1)$.

Определение. Множество $L_e(x, \gamma_0, P^*)$ называется нижней доверительной границей при заданном пороге γ_0 и заданной вероятностью P^* при которой множество $L_e(x, \gamma_0, P^*)$ содержится в множестве ребер истинного рыночного графа $U_e(x, \gamma_0, P^*)$ (см. [11], стр. 6). Методология построения $L_e(x, \gamma_0, P^*)$ базируется на тесте $\phi_{i,j}^n(x)$, проверки гипотезы (см. [9]): $h_{i,j}^n: \gamma_{i,j} \leq \gamma_0$ против $k_{i,j}^n: \gamma_{i,j} > \gamma_0, i, j = 1, \dots, N(2)$

Тесты $\phi_{i,j}^e(x)$ и $\phi_{i,j}^n(x)$ имеют следующий вид:

$$\begin{aligned}\phi_{i,j}^e &= \{1, T_{i,j}(x) < c_{i,j}^e, 0, T_{i,j}(x) \geq c_{i,j}^e\} \\ \phi_{i,j}^n &= \{1, T_{i,j}(x) > c_{i,j}^n, 0, T_{i,j}(x) \leq c_{i,j}^n\}\end{aligned}$$

$$, \text{ где } T_{i,j} = T_{i,j}^P = \frac{\sqrt{n-2}}{2} \left(\ln \ln \left(\frac{1+r_{i,j}}{1-r_{i,j}} \right) - \ln \ln \left(\frac{1+\gamma_0^P}{1-\gamma_0^P} \right) \right) r_{i,j} = \frac{\sum_{t=1}^n (x_i(t)-\underline{x}_i)(x_j(t)-\underline{x}_j)}{\sqrt{\sum_{t=1}^n (x_i(t)-\underline{x}_i)^2 \sum_{t=1}^n (x_j(t)-\underline{x}_j)^2}}$$

в случае коэффициента корреляции Пирсона (см. [1]),

$$\begin{aligned}T_{i,j} &= T_{i,j}^{Kd} = \frac{\gamma_{i,j}^{Kd \wedge} - \gamma_0}{4\sqrt{P_{cc} \wedge - P_c^2 \wedge}} \\ \gamma_{i,j}^{Kd \wedge} &= \frac{1}{n(n-1)} \sum_{s=1}^n \sum_{t=1, t \neq s}^n I_{i,j}^{Kd}(t, s) \\ I_{i,j}^{Kd}(t, s) &= \{1, (x_i(t) - x_i(s))(x_j(t) - x_j(s)) \geq 0, 0, (x_i(t) - x_i(s))(x_j(t) - x_j(s)) < 0\}\end{aligned}$$

в случае коэффициента корреляции Кендалла (см. [12]).

Пары (i, j) для которых гипотеза (1) принимается, включаются в $U_e(x, \gamma_0, P^*)$;

Пары (i, j) для которых гипотеза (2) отклоняется, включаются в $L_e(x, \gamma_0, P^*)$.

Критические значения $c_{i,j}^e$ и $c_{i,j}^n$ определяются из:

$$P_{\gamma_0}(T_{i,j}(x) > c_{i,j}^n) = P_{\gamma_0}(T_{i,j}(x) < c_{i,j}^e) = \frac{\alpha}{M}$$

где $\alpha = 1 - P^*$ (см. [9]).

Этот метод при фиксированной силе связи γ_0 , позволяет нам разделить все выводы о корреляциях между доходностями акций на фондовом рынке на три типа:

- Статистически значимое множество выводов об отсутствии корреляций уровня γ_0 :

$$L_n(x, \gamma_0, P^*) = J \setminus U_e(x, \gamma_0, P^*)$$

- Статистически значимое множество выводов о наличии корреляций уровня γ_0 :

$$L_e(x, \gamma_0, P^*)$$

- Статистически незначимое множество выводов уровня γ_0 :

$$G(x, \gamma_0, P^*) = U_e(x, \gamma_0, P^*) \setminus L_e(x, \gamma_0, P^*)$$

Предлагаемое решение: новые меры неопределённости. Рассмотрим несколько мер, которые позволяют количественно оценить неопределенность:

- Отношение числа незначимых выводов к общему числу пар акций:

$$K_1(x, \gamma_0, P^*) = \frac{|G(x, \gamma_0, P^*)|}{M}$$

Эта мера показывает долю неоднозначных случаев и позволяет сравнивать рынки разного размера.

- Отношение числа незначимых выводов к числу значимых выводов:

$$K_2(x, \gamma_0, P^*) = \frac{|G(x, \gamma_0, P^*)|}{|L_e(x, \gamma_0, P^*) \cup L_n(x, \gamma_0, P^*)|}$$

Данная мера отражает отношение размера статистически незначимого множества выводов к размеру статистически значимого множества.

- Агрегированные индикаторы неопределенности:

Для характеристики неопределенности i -го рынка предлагаются агрегированные показатели (Более подробно см. [11]):

$$K_1^i(x, P^*) = \sum_{s=1}^K c_s [K_1^i(x, \gamma_s, P^*)]$$

$$K_2^i(x, P^*) = \sum_{s=1}^K c_s [K_2^i(x, \gamma_s, P^*)]$$

$$c_s \geq 0, \sum_{s=1}^K c_s = 1.$$

Результаты анализа данных. Применение предложенных мер к данным 12 фондовых рынков (включая США, Китай, Японию, Великобританию и другие) позволило получить следующие результаты: Анализ табл. 1 показывает, что коэффициент $K_2^i(x, P^*)$ более изменчив, чем коэффициент $K_1^i(x, P^*)$, а именно $K_2^i(x, P^*) \in (0.32, 0.96)$ для корреляции Кендалл, $K_1^i(x, P^*) \in (0.23, 1.29)$ для корреляции Пирсона. Однако можно выделить однородные группы. Для выделения однородных групп, примем во внимание, что K_1^i имеет асимптотически нормальное распределение см. [11] (стр. 15), для выделения однородных групп по K_1^i используется 2σ интервал относительно выборочного среднего. Для выделения однородных групп по K_2^i используется 3σ интервал относительно выборочного среднего. Для корреляции Кендалла, такие группы представлены в табл. 2 (Аргентина, Австралия, Бразилия, Италия, Япония, Испания, Великобритания), для корреляции Пирсона такие группы представлены в табл. 3 (Австралия, Бразилия, Япония, Великобритания). Однородная группа, выбранная обоими корреляциями, совпадает с однородной группой, выбранной с использованием корреляции Пирсона.

Markets	Kendall		Pearson	
	K_1^i	K_2^i	K_1^i	K_2^i
Australia	0.15	0.46	0.2	0.74
China	0.18	0.34	0.19	0.35
India	0.15	0.96	0.21	1.19
Japan	0.14	0.44	0.2	0.7
USA	0.16	0.77	0.22	1.29
Canada	0.14	0.33	0.19	0.45
Argentina	0.22	0.44	0.15	0.23
Brazil	0.14	0.47	0.2	0.76
UK	0.15	0.51	0.2	0.77
Italy	0.17	0.53	0.18	0.58
Spain	0.15	0.48	0.19	0.46
Sweden	0.14	0.32	0.2	0.42

Таблица 1

Stock	K_2^i (Kendall)
Argentina	0.44
Australia	0.46
Brazil	0.47
Italy	0.53
Japan	0.44
Spain	0.48
UK	0.51

Таблица 2

Stock	K_2^i (Pearson)
Australia	0.74
Brazil	0.76
Japan	0.7
UK	0.77

Таблица 3

Преимущества предложенного подхода. Выделим три основных преимущества:

- Не требует знания истинной структуры сети: Все меры вычисляются на основе наблюдаемых данных, что делает метод универсальным.
- Гибкость анализа: Возможность изучать неопределенность для разных порогов и сравнивать рынки между собой.
- Практическая применимость: Результаты могут быть использованы для оценки неопределенности инвестиционных стратегий.

Заключение. Важная проблема оценки неопределенности выводов о различных характеристиках и структурах сетевых моделей получила недостаточное внимание. Лишь немногие работы посвящены этому вопросу, включая исследования [16], [7] и [8]. В этих работах рассматривались методы оценки неопределенности максимальных остовных деревьев и других структур с использованием бутстреп-техник и классической теории рисков. Однако все эти подходы требуют предположения об известности истинных сетевых структур для вычисления неопределенности.

В данной статье анализируется принципиально новый подход к измерению неопределенности, инициированный в [9] и [10]. Этот метод основан на оценки неопределенности средним количеством незначимых выводов. В настоящей работе исследованы два новых показателя неопределенности: доля незначимых выводов от общего числа выводов и отношение числа незначимых выводов к числу

значимых. Основное преимущество предложенных мер заключается в возможности вычисления неопределенности исключительно на основе наблюдаемых данных, без предположения об истинной сетевой структуре.

Для анализа этих мер неопределенности использованы данные 12 фондовых рынков. Показано, что эти меры позволяют выявить группы рынков с похожей неопределенностью и исследовать особенности различных рынков.

СПИСОК ЛИТЕРАТУРЫ

1. *Anderson T. W.* An introduction to multivariate statistical analysis (3rd ed.) // John Wiley & Sons. 2003.
2. *Boginski V., Butenko S., Pardalos P.* Statistical analysis of financial networks // Computational Statistics & Data Analysis. 2005. Vol. 48 (2). P. 431-443.
3. *Boginski V., Butenko S., Pardalos P.* Mining market data: A network approach // Computers & Operations Research. 2006. Vol. 33 (11). P. 3171-3184.
4. *Garas F., Argyrakis P.* Correlation study of the Athens stock exchange // Physica. 2007 Vol. 380. P. 399-410.
5. *Gunawardena A. D. A., Meyer R. R., Dougan W. L., Monaghan P. E., ChotonBasu P. E. M.* Optimal selection of an independent set of cliques in a market graph // International Proceedings of Economics Development and Research. 2012. Vol. 29. P. 281-285.
6. *Huang W. Q., Zhuang X. T., Yao S. A.* A network analysis of the Chinese stock market // Physica. 2009. Vol. 388. P. 2956-2964.
7. *Kalyagin V. A., Koldanov A. P., Koldanov P. A., Pardalos P. M.* Measures of uncertainty in market network analysis // Physica. 2014. Vol. 413. P. 59-70.
8. *Kalyagin V. A., Koldanov A. P., Koldanov P. A.* Reliability of maximum spanning tree identification in correlation-based market networks // Physica A: Statistical Mechanics and its Applications. 2022. P. 127-482.
9. *Koldanov P. A., Koldanov A. P., Semenov D. P.* Confidence bounds for threshold similarity graph in random variable network. // Statistical Analysis and Data Mining: ASA Data Science Journal. 2023. Vol. 16. P. 583-595.
10. *Колданов П. А., Колданов А. П., Семенов Д. П.* Сравнительный анализ неопределенности выводов о связях на фондовых рынках // Журнал Новой экономической ассоциации, 2025. № 1 (66). С. 54-74.
11. *Koldanov P. A., Koldanov A. P., Tsygankov V. V.* Uncertainty of market graph identification 2025 [Preprint]. SSRN. [Electronic resource]. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5268780 (access date: 25.09.2024).
12. *Kruskal W. H.* Ordinal measures of association // Journal of the American Statistical Association. 1958. Vol. 53 (284). P. 814-861.
13. *Namaki A., Shirazi A. H., Raei R., Jafari G. R.* Network analysis of a financial market based on genuine correlation and threshold method // Physica. 2011. Vol. 390. P. 3835-3841.
14. *Shirokikh J., Pastukhov G., Boginski V., Butenko S.* Computational study of the US stock market evolution: A rank correlation-based network model // Computational Management Science. 2013. Vol. 10 (2-3). P. 81-103.
15. *Tse C. K., Liu J., Lau F. C. M.* A network perspective of the stock market // Journal of Empirical Finance. 2010. Vol. 17. P. 659-667.
16. *Tumminello M., Coronello C., Lillo F., Miccichè S.* Trees and bootstrap reliability estimation in correlation based-networks // International Journal of Bifurcation and Chaos. 2006. Vol. 17. P. 2319-2329.
17. *Wald A.* Statistical decision functions // Wiley. 1950.