

**А.П. Колданов**  
НИУ ВШЭ, Нижний Новгород

**П.А. Колданов**  
НИУ ВШЭ, Нижний Новгород

**Д.П. Семенов**  
НИУ ВШЭ, Нижний Новгород

## **Построение доверительного множества связанных акций фондового рынка<sup>1</sup>**

**Аннотация.** В работе рассматривается задача анализа связей между парами акций фондового рынка по результатам наблюдений за их доходностями. Такая задача возникает при сетевом анализе фондового рынка. Предполагается, что совместное распределение доходностей принадлежит классу эллиптических распределений. В качестве мер связи рассматриваются классический коэффициент корреляции Пирсона, коэффициент корреляции Кендалла и коэффициент корреляции Фехнера. Исследуются способы построения множества пар акций, которое с вероятностью не менее заданной содержит все сильносвязанные пары акций. Для построения таких множеств используются процедуры одновременной проверки многих гипотез о величине рассматриваемых коэффициентов корреляции. Свойства таких процедур изучаются методом статистического моделирования. Приведены результаты, показывающие, что процедуры, основанные на комбинировании индивидуальных тестов Кендалла и Фехнера, гарантируют построение множеств с заданной доверительной вероятностью. Показано, что одновременное применение тестов проверки гипотез о значении классического коэффициента корреляции Пирсона не приводит к доверительным множествам при отклонении распределения от нормального. В случае распределения Стьюдента построенное множество незначительно отличается от доверительного. Предложена и апробирована на реальных данных процедура проверки согласия с эллиптической моделью и исследованы некоторые особенности этой модели.

**Ключевые слова:** *сетевая модель фондового рынка, отсеченный граф, классический коэффициент корреляции Пирсона, корреляция Кендалла, корреляция Фехнера, доверительное множество, процедуры множественной проверки гипотез.*

Классификация JEL: C02.

DOI: 10.31737/2221-2264-2021-50-2-1

### **1. Введение**

Один из ключевых вопросов при исследовании фондового рынка заключается в анализе связей между доходностями акций, обращающихся на рынке. В частности, ковариационная матрица играет существенную роль в теории портфелей Марковица. Анализ связей доходностей активов естественным образом возникает при сетевом анализе фондового рынка. Сетевой анализ заключается в построении сетевой модели или простого полного взвешенного графа, вершины которого соответствуют акциям, а веса ребер задаются степенью взаимосвязи между их доходностями.

<sup>1</sup> Статья подготовлена в результате проведения исследования в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ). Работа А.П. Колданова и П.А. Колданова частично поддержана грантом Российского фонда фундаментальных исследований (проект 18-07-00524). Работа Д.П. Семенова частично поддержана грантом Российского фонда фундаментальных исследований (проект 19-31-90088).

Для выделения ключевой информации из сетевой модели извлекаются различные сетевые структуры. Одной из наиболее популярных сетевых структур является отсеченный граф. Множество вершин отсеченного графа совпадает с множеством вершин сетевой модели. Ребро (без веса) между двумя вершинами проводится в отсеченном графе тогда и только тогда, когда его вес в сетевой модели превышает заданный порог. Таким образом, при анализе фондового рынка отсеченный граф выделяет пары акций, степень взаимосвязи между которыми больше заданного порога.

Отсеченный граф, или граф рынка, использовался в работах (Boginski, Butenko, Pardalos, 2005, 2006; Garas, Argyrakis, 2007; Gunawardena et al., 2012; Huang, Zhuang, Yao, 2009; Namaki et al., 2011; Shirokikh et al., 2013; Tse, Liu, Lau, 2010; Калягин и др., 2017; Визгунов и др., 2012) для анализа фондовых рынков различных стран. Авторы этих работ выделяют общие черты и особенности исследуемых рынков, но достоверность полученных выводов недостаточно подкреплена. Необходимость изучения связана с тем, что с современной точки зрения данные о доходностях акций являются наблюдениями над случайными величинами, что неизбежно приводит к ошибкам при получении выводов по выборке конечного объема.

В настоящей работе исследование вопросов идентификации отсеченного графа по наблюдениям основано на понятии сети случайных величин, введенном в (Kalyagin et al., 2017). Сеть случайных величин представляет пару  $(X, \gamma)$ , где  $X = (X_1, \dots, X_N)$  – случайный вектор,  $\gamma_{i,j} = \gamma(X_i, X_j)$  – мера связи между случайными величинами  $X_i, X_j$ . Опираясь на сеть случайных величин, можно определить истинный отсеченный граф, т.е. невзвешенный граф с ребром между вершинами  $i$  и  $j$ , если  $\gamma_{i,j} \geq \gamma_0$ , где  $\gamma_0$  – заданный порог.

В настоящее время существует несколько подходов к задачам выбора одной из многих гипотез (Hochberg, Tamhane, 1987). Подход к задачам выбора одной из многих гипотез о матрице смежности отсеченного графа с позиций теории статистических решений (Lehmann, 1957) исследовался в работах (Koldanov, 2013; Kalyagin et al., 2017, 2018). Подход позволяет строить статистические процедуры идентификации отсеченного графа, минимизирующие линейную комбинацию средних чисел неправильно включенных ребер и неправильно невключенных ребер в выборочный отсеченный граф. Однако вопросы контроля других характеристик статистических процедур идентификации отсеченного графа не изучались.

Целью настоящей работы является исследование способов построения множества пар акций, которое с вероятностью не менее заданной  $P^*$  содержит все пары акций со связями, превышающими заданный порог  $\gamma_0$ . Такое множество будем называть доверительным множеством для ребер (ДМР) отсеченного графа уровня  $P^*$ . Такие множества представляют интерес для выделения наиболее влиятельных

акций (Hero, Rajaratnam, 2012), построения групп сильносвязанных попарно акций (клик) (Boginski et al., 2005), анализа ковариационных матриц больших размерностей, а также для построения фактор-графов, используемых в анализе графических моделей (Jordan, 2004).

Для построения доверительных множеств в настоящей работе используется подход, известный как выбор подмножества (subset selection (Gupta, Pancharakesan, 2002)). При этом мы опираемся на эквивалентность между процедурами выбора подмножества и процедурами множественной проверки гипотез (multiple hypotheses testing) (Finner, Giani, 1994). Такая эквивалентность позволяет строить как одношаговые, так и многошаговые процедуры построения ДМР( $\gamma_0, P^*$ ). В настоящей работе исследованы одношаговые процедуры построения ДМР( $\gamma_0, P^*$ ) в сетях случайных величин  $X, \gamma$ .

В качестве модели распределения вектора  $X = (X_1, \dots, X_N)$  используется класс эллиптических распределений, широко применяемый в экономике (Vodnar, 2013). Рассмотрение такого широкого класса распределений ставит вопрос выбора меры близости при построении сетевой модели. В настоящей работе исследуются три меры близости: классический коэффициент корреляции Пирсона, коэффициент корреляции Фехнера (или вероятность совпадения знаков) и коэффициент корреляции Кендалла. Для проверки индивидуальных гипотез о значениях рассматриваемых коэффициентов корреляции применяются стандартные тесты.

При исследовании свойств процедур построения ДМР( $\gamma_0, P^*$ ) основное внимание сосредоточено на изучении свойства устойчивости к изменению вероятностной модели. Свойства любой процедуры множественной проверки гипотез определяются свойствами индивидуальных тестов и способом комбинирования этих тестов. Одношаговые процедуры построения ДМР( $\gamma_0, P^*$ ) основаны на неравенстве Бонферрони при комбинировании индивидуальных тестов уровня  $2(1 - P^*) / [N(N - 1)]$ . Известно, что такое неравенство дает оценку сверху для вероятности хотя бы одного ложного отвержения верной индивидуальной гипотезы (FWER). Таким образом, применение процедуры Бонферрони может приводить к построению множеств, которые содержат все пары сильносвязанных акций, с вероятностью значительно большей вероятности  $1 - P^*$ . Однако многие индивидуальные тесты не обладают нужными свойствами. В частности, стандартные тесты проверки гипотез о величине классического коэффициента корреляции и коэффициента корреляции Кендалла неустойчивы при отклонении вероятностной модели от гауссовской. А именно уровень значимости таких индивидуальных тестов может превосходить заданный (Koldanov, 2019a). Неустойчивость индивидуальных тестов может быть компенсирована грубостью процедуры Бонферрони. Таким образом, комбинирование неустойчивых индивидуальных тестов может приводить к построению множеств, которые с вероятностью не менее

заданной  $P^*$  содержат все сильносвязанные пары акций. Насколько нам известно, вопросы подобного рода не рассматривались в научной литературе.

В настоящей работе исследуется устойчивость процедур построения доверительных множеств, содержащих все пары акций, связи между которыми не менее заданного порога  $\gamma_0$ .

Под *устойчивостью* понимается гарантия неравенства  $P \geq P^*$  независимо от модели совместного поведения доходностей акций и объема наблюдений, где  $P$  – вероятность того, что построенное множество содержит все сильносвязанные пары акций.

В работе с помощью метода статистического моделирования изучается зависимость размера построенного множества и среднего числа связей, ошибочно непопавших в это множество, от модели совместного поведения доходностей акций, типа индивидуальных тестов и объема наблюдений.

Результаты статистического моделирования подтверждают теоретически доказанную устойчивость процедур, основанных на комбинации частот совпадения знаков или индивидуальных тестов Фехнера (Kalyagin, 2017). Результаты статистического моделирования показывают устойчивость процедур, основанных на комбинации индивидуальных тестов Кендалла, несмотря на их неустойчивость при отклонении вероятностной модели от гауссовской, показанной в (Koldanov, 2019a). Вместе с тем, грубость процедуры Бонферрони оказывается недостаточной для построения устойчивых процедур, основанных на комбинации индивидуальных тестов проверки гипотез о классическом коэффициенте корреляции. Однако процент сильных связей, ошибочно невключенных в множество, построенное как комбинация индивидуальных тестов о значении классического коэффициента корреляции, достаточно мал.

Оценивая степень различия отсеченных графов, построенных на различных мерах близости, мы покажем, что такое различие может быть велико. Однако это не противоречит гипотезе согласия с эллиптической моделью. Нами предложена процедура проверки гипотезы согласия распределения с эллиптической моделью, основанная на симметрической разности матриц смежности графов рынка, построенных при согласованных порогах в разных сетях случайных величин. Анализ реальных данных с помощью этой процедуры не приводит к отклонению гипотезы согласия с эллиптической моделью.

Статья организована следующим образом: в разд. 2 приведены основные обозначения и постановка задачи. В разд. 3 предложена общая схема построения доверительных множеств. В разд. 4 описаны процедуры построения доверительных множеств в различных сетях случайных величин. В разд. 5 приведены результаты анализа устойчивости доверительных множеств. В разд. 6 обсуждаются различия отсеченных графов, построенных на различных мерах близости, и сфор-

мулированы результаты проверки согласия с эллиптической моделью. В разд. 7 сделаны выводы.

## 2. Основные определения и постановка задачи

### 2.1. Сеть случайных величин и истинный отсеченный граф

Сеть случайных величин – это пара  $(X, \gamma)$ , где  $X = (X_1, \dots, X_N)$  – случайный вектор,  $\gamma$  – мера зависимости между случайными величинами. Такая сеть порождает сетевую модель, т.е. простой полный взвешенный граф  $(V, \Gamma)$ , где  $V = \{1, \dots, N\}$  – множество вершин,  $\Gamma = (\gamma_{i,j})$  – матрица весов ребер  $\gamma_{i,j} = \gamma(X_i, X_j)$ ,  $i \neq j$ ,  $\gamma_{i,i} = 0$  между вершинами  $i, j \in V$ . Одной из характеристик сетевой модели является отсеченный граф с порогом  $\gamma_0$ , который представляет собой невзвешенный граф на  $N$  вершинах. Ребро между вершинами  $i, j$  добавляется в отсеченный граф тогда и только тогда, когда его вес в сетевой модели  $\gamma_{i,j} \geq \gamma_0$ . Такой граф будем называть истинным отсеченным графом.

В зависимости от модели распределения вектора  $X = (X_1, \dots, X_N)$  и выбора меры  $\gamma$  можно рассматривать различные сети случайных величин (ССВ). Простым примером ССВ является сеть Пирсона с нормальным распределением, в которой предполагается, что вектор  $X = (X_1, \dots, X_N)$  имеет многомерное нормальное распределение  $N(\mu, \Sigma)$ , а мера  $\gamma^p$  задается коэффициентом корреляции Пирсона:

$$\gamma^p = \rho_{i,j} = \rho(X_i, X_j) = \frac{E(X_i - \mu_i)(X_j - \mu_j)}{\sqrt{\sigma_{ii}\sigma_{jj}}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}},$$

где  $\mu = (\mu_1, \dots, \mu_N)$  – вектор математических ожиданий,  $\Sigma = (\sigma_{ij})_{N \times N}$  – ковариационная матрица.

В настоящей работе рассматривается одна из популярных моделей распределения доходностей акций – эллиптическая модель. Случайный вектор  $X = (X_1, \dots, X_N)$  имеет эллиптическое распределение  $ECD_N(\mu, \Sigma, g)$  (Anderson, 2003), если его плотность имеет вид

$$f(x; \mu, \Sigma) = |\Sigma|^{-1/2} g\{(x - \mu)' \Sigma^{-1} (x - \mu)\}, \quad (1)$$

где  $\Sigma = (\sigma_{ij})$  – положительно определенная матрица; функция  $g(x) \geq 0$ ,

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y'y) dy_1 \dots dy_p = 1.$$

Этот класс включает, в частности, многомерное нормальное распределение и многомерное распределение Стьюдента. Известно, что если математическое ожидание существует, то  $E(X_i) = \mu_i$ ,  $i = 1, \dots, N$ .

В настоящей работе рассматриваются 3 ССВ:

- сеть Пирсона  $(X, \gamma^p)$  с эллиптическим распределением, в которой вектор  $X = (X_1, \dots, X_N)$  имеет эллиптическое распределение  $ECD_N(\mu, \Sigma, g)$ ;
- знаковая сеть  $(X, \gamma^{sg})$  с эллиптическим распределением, в которой вектор  $X = (X_1, \dots, X_N)$  имеет эллиптическое распределение  $ECD_N(\mu, \Sigma, g)$ , а мера  $\gamma_{i,j}^{sg} = p^{i,j} = P((X_i - \mu_i)(X_j - \mu_j) > 0)$ . Заметим,

что такая мера линейно связана с коэффициентом корреляции Фехнера  $\gamma_{i,j}^{Fh} = 2\gamma_{i,j}^{Sg} - 1$ ;

- эллиптическая сеть корреляций Кендалла  $(X, \gamma^{Kd})$ , в которой вектор  $X = (X_1, \dots, X_N)$  имеет эллиптическое распределение  $ECD_N(\mu, \Sigma, g)$  и используется одна из популярных ранговых мер близости – коэффициент корреляции Кендалла, равный (Kruskal, 1958)

$$\tau_{i,j} = P[(X_i(t) - X_i(s))(X_j(t) - X_j(s)) > 0],$$

где  $\begin{pmatrix} X_i(t) \\ X_j(t) \end{pmatrix}, \begin{pmatrix} X_i(s) \\ X_j(s) \end{pmatrix}$  – независимые копии вектора  $\begin{pmatrix} X_i \\ X_j \end{pmatrix}$ .

В (Kalyagin et al., 2017; Калягин и др., 2017) исследованы знаковые ССВ с эллиптическим распределением  $X - ECD(\mu, \Sigma, g)$ . Показано, что  $\gamma_{i,j}^{Sg} = p^{i,j} = P((X_i - \mu_i)(X_j - \mu_j) > 0) = 0,5 + (1/\pi)\arcsin \rho_{i,j}$ , т.е. неравенство  $\rho_{i,j} \geq \rho_0$  эквивалентно неравенству  $p^{i,j} \geq p^0$ . Следовательно, истинный отсеченный граф, построенный в знаковой сети с эллиптическим распределением с порогом  $p^0$ , совпадает с истинным отсеченным графом, построенным в сети Пирсона с порогом  $\rho_0$ , если  $p^0 = 0,5 + (1/\pi)\arcsin \rho_0$ .

В (Koldanov, 2019b) доказано, что в классе эллиптических распределений  $ECD(\mu, \Sigma, g)$   $\tau_{i,j} = p^{i,j}$ . Поэтому истинный отсеченный граф, построенный в знаковой сети с эллиптическим распределением с порогом  $p^0$ , совпадает с истинным отсеченным графом, построенным в сети Кендалла  $(X, \gamma^{Kd})$ , в которой  $\gamma_{i,j}^{Kd} = \tau_{i,j}$ , с тем же порогом.

Таким образом, в рассматриваемых в настоящей работе сетях случайных величин истинные отсеченные графы при согласованных порогах совпадают.

## 2.2. Постановка задачи построения доверительного множества ребер

Пусть  $(X, \gamma)$  – сеть случайных величин, вектор  $X = (X_1, \dots, X_N)$  имеет распределение  $f(x; \theta)$ ,  $\theta \in \Omega$ ,  $\gamma_{i,j}(\theta) = \gamma(X_i, X_j)$  – мера зависимости между  $X_i, X_j$ . Матрица связей имеет вид

$$\Gamma(\theta) = \begin{pmatrix} 1 & \gamma_{1,2}(\theta) & \dots & \gamma_{1,N}(\theta) \\ \gamma_{2,1}(\theta) & 1 & \dots & \gamma_{2,N}(\theta) \\ \dots & \dots & \dots & \dots \\ \gamma_{N,1}(\theta) & \gamma_{N,2}(\theta) & \dots & 1 \end{pmatrix}.$$

Пусть  $I = \{(i, j) : i \neq j; i, j = 1, \dots, N\}$  – множество пар индексов случайных величин  $X_1, \dots, X_N$ ;

$$J(\theta, \gamma_0) = \{(i, j) : \gamma_{i,j}(\theta) > \gamma_0\}, \quad J(\theta) \subset I \quad (2)$$

множество пар индексов случайных величин  $X_1, \dots, X_N$ , соответствующих вершинам отсеченного графа, между которыми проведено ребро в истинном отсеченном графе при пороге  $\gamma_0$ .

Пусть  $x = (x_1(t), \dots, x_N(t))$ ,  $t = 1, \dots, n$  – множество наблюдений над случайным вектором  $X$ . Предполагается, что множество наблюдений формирует повторную выборку, т.е. вектора  $X(t) = (X_1(t), \dots, X_N(t))$ ,  $t = 1, \dots, n$  независимы.

Пусть

$$\Phi(x) = \begin{pmatrix} 0 & \varphi_{1,2}(x) & \dots & \varphi_{1,N}(x) \\ \varphi_{2,1}(x) & 0 & \dots & \varphi_{2,N}(x) \\ \dots & \dots & \dots & \dots \\ \varphi_{N,1}(x) & \varphi_{N,2}(x) & \dots & 0 \end{pmatrix} \quad (3)$$

представляет собой выборочный отсеченный граф, т.е. граф, построенный каким-либо алгоритмом по имеющимся наблюдениям.

Пусть  $J(x, \gamma_0) = \{(i, j) : \varphi_{i,j}(x) = 0 \ \forall i \neq j\}$  – множество пар индексов случайных величин  $X_1, \dots, X_N$ , соответствующих вершинам графа рынка, между которыми проведено ребро в выборочном отсеченном графе. Задача, рассматриваемая в настоящей работе, заключается в построении процедуры выбора подмножества  $J(x, \gamma_0) \subset I$ , удовлетворяющего условию

$$P_0(J(X, \gamma_0) \supset J(\theta, \gamma_0)) \geq P^* \quad \forall \theta \in \Omega, \quad \forall \gamma_0. \quad (4)$$

Множество  $J(X, \gamma_0)$ , удовлетворяющее условию (4), будем называть доверительным множеством для ребер отсеченного графа уровня  $P^*$  и обозначать ДМР( $\gamma_0, P^*$ ).

### 3. Общая схема построения доверительного множества для ребер

Используя эквивалентность между процедурами выбора подмножества и процедурами множественной проверки гипотез (Finner, Gianì, 1994), задача построения ДМР( $\gamma_0, P^*$ ) может быть сведена к задаче множественной проверки гипотез

$$h_{ij} : \gamma_{i,j} \geq \gamma_0 \text{ против альтернативы } k_{ij} : \gamma_{i,j} < \gamma_0. \quad (5)$$

Заметим, что проверяемые индивидуальные гипотезы удовлетворяют условию свободной комбинации. Поэтому проблем когерентности и консонантности не возникает (Finner, 1994).

Пусть

$$\varphi_{i,j}(x) = \begin{cases} 0, & T_{i,j}(x) \geq c_{i,j}; \\ 1, & T_{i,j}(x) < c_{i,j} \end{cases}$$

тесты проверки индивидуальных гипотез (5). Результаты совместного применения тестов  $\varphi_{i,j}(x)$  можно представить в виде матрицы

$$\Phi(x) = \begin{pmatrix} 0 & \varphi_{1,2}(x) & \dots & \varphi_{1,N}(x) \\ \varphi_{2,1}(x) & 0 & \dots & \varphi_{2,N}(x) \\ \dots & \dots & \dots & \dots \\ \varphi_{N,1}(x) & \varphi_{N,2}(x) & \dots & 0 \end{pmatrix}. \quad (6)$$

При этом если процедура (6), рассматриваемая как процедура множе-

ственной проверки гипотез (5), контролирует  $FWER \leq 1 - P^*$  в сильном смысле (т.е. при любой конфигурации истинных и ложных гипотез), то множество нулей в (6), не лежащих на главной диагонали, представляет собой множество ребер, входящих в ДМР( $\gamma_0, P^*$ ), т.е.  $DMR(\gamma_0, P^*) = \{(i, j) : \varphi_{i,j}(x) = 0, i \neq j\}$ . При этом можно использовать различные процедуры множественной проверки гипотез, контролирующие FWER в сильном смысле.

Построение таких процедур обычно основано на комбинировании индивидуальных тестов, каждый из которых контролирует уровень значимости, согласованный с заданной вероятностью хотя бы одного ложного отвержения. Неожиданным результатом настоящей работы является иллюстрация того, что комбинирование индивидуальных тестов, не контролируемых заданный уровень значимости, также может привести к доверительному множеству с заданным уровнем  $P^*$ , т.е. ДМР( $\gamma_0, P^*$ ).

#### 4. Процедуры построения доверительного множества для ребер в различных сетях случайных величин

Как отмечалось во введении, свойства любой процедуры множественной проверки гипотез определяются двумя моментами: свойствами индивидуальных тестов и способом комбинирования этих тестов. Одношаговые процедуры построения ДМР( $\gamma_0, P^*$ ) основаны на использовании неравенства Бонферрони при комбинировании индивидуальных тестов. Неравенство Бонферрони является грубой оценкой вероятности хотя бы одного ложного отвержения верной индивидуальной гипотезы. Однако многие индивидуальные тесты не обладают нужными свойствами, например нужной устойчивостью, или не контролируют уровень значимости. Такой недостаток индивидуальных тестов может компенсироваться грубостью неравенства Бонферрони. В настоящем разделе приведены стандартные тесты проверки индивидуальных гипотез (5) и известные результаты исследования устойчивости этих тестов.

В сети Пирсона мера  $\gamma$  определяется коэффициентом корреляции Пирсона  $\gamma_{i,j} = \rho_{i,j}$ . Стандартные тесты (Anderson, 2003) проверки индивидуальных гипотез (5) в сети Пирсона с нормальным распределением имеют вид:

$$\varphi_{ij}^P(x) = \begin{cases} 0, & \frac{1}{2\sqrt{n-3}} \left( \ln \frac{1+r_{i,j}}{1-r_{i,j}} - \ln \frac{1+\gamma_0^P}{1-\gamma_0^P} \right) \geq c_{i,j}^P; \\ 1, & \frac{1}{2\sqrt{n-3}} \left( \ln \frac{1+r_{i,j}}{1-r_{i,j}} - \ln \frac{1+\gamma_0^P}{1-\gamma_0^P} \right) < c_{i,j}^P, \end{cases} \quad (7)$$

$$r_{i,j} = \frac{\sum_{t=1}^n (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)}{\sqrt{\sum_{t=1}^n (x_i(t) - \bar{x}_i)^2 \sum_{t=1}^n (x_j(t) - \bar{x}_j)^2}}$$



Если вектор  $X$  имеет многомерное нормальное распределение  $N(\mu, \Sigma)$  и  $c_{i,j}^P - (1 - P^*) / C_N^2$  – квантиль стандартного нормального распределения, процедура (6) с тестами (7) является процедурой Бонферонни множественной проверки гипотез (5) с контролем  $FWER \leq 1 - P^*$  в сильном смысле. Следовательно, доверительное множество для ребер, построенное одношаговой процедурой в сети Пирсона с нормальным распределением, имеет вид

$$\text{ДМР}_{SS}^P(\gamma_0^P, P^*) = \{(i, j) : \varphi_{ij}^P(x) = 0, i \neq j\}.$$

В (Koldanov, 2019a) показано, что тесты (7) неустойчивы при отклонении от нормальности.

В знаковой сети тесты проверки индивидуальных гипотез (5) определяются формулой

$$\varphi_{i,j}^{Sg} = \begin{cases} 0, & \hat{p}^{i,j} \geq c_{i,j}^{Sg}; \\ 1, & \hat{p}^{i,j} < c_{i,j}^{Sg}, \end{cases} \quad (8)$$

где  $\hat{p}^{i,j} = \sum_{t=1}^n I_{ij}(t)$ .

Из (Kalyagin et al., 2017) следует, что если вектор  $X$  имеет эллиптическое распределение  $ECD_N(\mu, \Sigma, g)$  и  $c_{i,j}^{Sg} - (1 - P^*) / C_N^2$  – квантиль биномиального распределения  $b(n, \gamma_0^{Sg})$ , где  $\gamma_0^{Sg} = 0,5 + (1/\pi) \arcsin \gamma_0^P$ , процедура (6) с тестами (8) является процедурой Бонферонни с контролем  $FWER \leq 1 - P^*$  в сильном смысле. Следовательно, доверительное множество для ребер, построенное одношаговой процедурой в знаковой сети с эллиптическим распределением, равно

$$\text{ДМР}_{SS}^{Sg}(\gamma_0^{Sg}, P^*) = \{(i, j) : \varphi_{ij}^P(x) = 0, i \neq j\}.$$

Стандартные тесты проверки индивидуальных гипотез (5) в сети Кендалла  $(X, \gamma^{Kd})$  имеют вид

$$\varphi_{ij}^{Kd} = \begin{cases} 0, & \sqrt{\frac{9n(n-1)}{2(2n+5)}} (T_{ij}^{Kd} - \gamma_0^{Kd}) \geq c_{i,j}^{Kd}; \\ 1, & \sqrt{\frac{9n(n-1)}{2(2n+5)}} (T_{ij}^{Kd} - \gamma_0^{Kd}) < c_{i,j}^{Kd}, \end{cases} \quad (9)$$

где  $\gamma_0^{Kd} = (2/\pi) \arcsin \gamma_0^P$ , а статистика –

$$T_{ij}^{Kd} = \frac{1}{n(n-1)} \sum_{s=1}^n \sum_{t=1}^n \text{sign}((x_i(t) - x_i(s))(x_j(t) - x_j(s))).$$

Из (Daniels, Kendall, 1947; Hoffding, 1947) следует, что если компоненты вектора  $X$  – независимые случайные величины и  $c_{i,j}^{Kd} - (1 - P^*) / C_N^2$  – квантиль стандартного нормального распределения, то при  $\gamma_0 = 0$ ,  $n \rightarrow \infty$  процедура (6) с тестами (9) является процедурой Бонферонни с контролем  $FWER \leq 1 - P^*$  в сильном смысле. Следовательно, в этом случае доверительного множества для ребер, построенное одношаговой процедурой в сети Кендалла, равно  $\text{ДМР}_{SS}^{Kd}(\gamma_0^{Kd}, P^*) = \{(i, j) : \varphi_{ij}^{Kd}(x) = 0, i \neq j\}$ .

Как показано в (Koldanov, 2019a), в общем случае (т.е. при  $\gamma_0 \neq 0$  и при зависимых компонентах вектора  $X$ ) уровень значимости тестов (9) неустойчив при отклонении от нормальности в классе эллиптических распределений. Однако, как показывают результаты статистического моделирования, приведенные ниже, комбинирование тестов (9) уровня  $(1 - P^*) / C_N^2$  с помощью процедуры Бонферонни приводит к построению ДМР<sub>SS</sub><sup>Kd</sup> ( $\gamma_0^{Kd}, P^*$ ).

### 5. Анализ свойств процедур построения множества связанных акций

В этом разделе приведены результаты статистического моделирования процедур построения множества связанных акций в сетях случайных величин  $(X, \gamma^P), (X, \gamma^{Sg}), (X, \gamma^{Kd})$  при размерах вектора  $X$ , равных  $N = 5; 10; 20$ , и объемах наблюдений  $n = 100; 500; 1000$ . Наблюдения моделировались из плотности  $f(x) = \varepsilon f_{Norm}(\mu, \Sigma) + (1 - \varepsilon) f_{St}(\mu, \Sigma)$ , где  $f_{Norm}(\mu, \Sigma)$  – функция плотности многомерного нормального распределения,  $f_{St}(\mu, \Sigma)$  – функция плотности распределения Стьюдента с тремя степенями свободы. В качестве параметров  $\mu, \Sigma$  использовались оценки максимального правдоподобия, вычисленные по данным  $N = 5; 10; 20$  самых ликвидных акций фондового рынка США (тикеры акций: GE, BAC, F, T, CHK, TWTR, ABEV, PBR, PFE, VALE, WFC, BABA, SWN, ITUB, FCX, ORCL, C, VZ, NOK, JCP) за период с 1.01.2018 до 1.01.2019.

При этом изучались следующие вопросы.

1. Зависимость вероятности того, что построенные множества содержат все ребра сетевой модели с весами не менее  $\gamma_0$  от используемой меры близости ( $\gamma^P, \gamma^{Sg}, \gamma^{Kd}$ ), значения параметра смеси  $\varepsilon$ , значения  $\gamma_0$  (которое определяет число ребер в истинном графе, например, при  $\gamma_0 = 0,1; N = 20$  число ребер равно 174) и объема наблюдений.

2. Зависимость среднего числа ребер в построенном множестве от используемой меры близости ( $\gamma^P, \gamma^{Sg}, \gamma^{Kd}$ ), значения параметра смеси  $\varepsilon$ , значения  $\gamma_0$  и объема наблюдений.

3. Зависимость среднего числа ребер истинного графа с порогом  $\gamma_0$ , непопавших в построенное множество, от используемой меры близости ( $\gamma^P, \gamma^{Sg}, \gamma^{Kd}$ ), значения параметра смеси  $\varepsilon$ , значения  $\gamma_0$  и объема наблюдений.

Ниже приведены результаты статистического моделирования, усредненные по 300 повторениям эксперимента.

#### 5.1. Доверительные уровни построенных множеств

Данные в табл. 1 показывают, что в сети Пирсона вероятность  $P^* = 0,9$ , что все ребра  $(i, j): \gamma_{i,j} \geq \gamma_0$  попадут в построенное множество, не контролируется при уменьшении параметра смеси  $\varepsilon$  при размере вектора  $X$  ( $N = 20$ ),  $\gamma_0 = 0,1; 0,3$  и числе наблюдений  $n = 100, 500, 1000$  начиная с  $\varepsilon = 0,8$ . Аналогичные результаты показывают, что вероят-

ность  $P^* = 0,9$  не контролируется начиная с  $\varepsilon = 0,7$  при размере вектора  $X$  ( $N = 10$ ) и числе наблюдений  $n = 100$ ; не контролируется начиная с  $\varepsilon = 0,6$  при размере вектора  $X$  ( $N = 5$ ) и числе наблюдений  $n = 100$ . При этом большая неустойчивость вероятности при увеличении числа наблюдений связана с уменьшением размера построенного множества при уменьшении параметра  $\varepsilon$ . Например, из 174 истинных ребер (при  $\gamma_0 = 0,1$ ) большее число гипотез  $h_{i,j} : \gamma_{i,j} \geq \gamma_0$  будет отвергнуто при генерировании наблюдений из распределения Стьюдента (распределения с тяжелыми хвостами), чем при генерировании наблюдений из нормального распределения.

Таблица 1

Оценка доверительного уровня построенного множества в сети Пирсона (для  $N = 20$ )

$\gamma_0$	$\varepsilon$										
	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
$n = 100$											
0,1 (174)	0,40	0,50	0,54	0,62	0,71	0,71	0,80	0,82	0,89	0,94	0,99
0,3 (72)	0,44	0,52	0,62	0,58	0,72	0,78	0,75	0,86	0,87	0,95	1,00
0,5 (12)	0,85	0,88	0,90	0,88	0,90	0,91	0,93	0,95	0,97	0,98	1,00
0,7 (3)	0,97	0,98	0,97	0,97	0,97	0,97	0,99	0,99	0,99	0,99	1,00
$n = 500$											
0,1 (174)	0,31	0,40	0,44	0,54	0,54	0,68	0,77	0,74	0,86	0,91	1,00
0,3 (72)	0,33	0,33	0,42	0,54	0,61	0,65	0,77	0,80	0,87	0,93	0,99
0,5 (12)	0,77	0,78	0,80	0,81	0,87	0,87	0,91	0,94	0,95	0,95	1,00
0,7 (3)	0,98	0,93	0,95	0,96	0,96	0,96	0,97	0,99	0,99	1,00	1,00
$n = 1000$											
0,1 (174)	0,31	0,32	0,40	0,50	0,60	0,64	0,69	0,82	0,84	0,89	0,99
0,3 (72)	0,30	0,33	0,44	0,49	0,56	0,58	0,72	0,79	0,86	0,91	1,00
0,5 (12)	0,70	0,76	0,78	0,79	0,86	0,88	0,92	0,92	0,95	0,97	1,00
0,7 (3)	0,95	0,94	0,97	0,93	0,96	0,98	0,99	0,99	0,99	0,99	1,00

**Примечание.** В столбце 1 в скобках указано число ребер при заданном  $\gamma_0$ .

Согласно (Kalyagin et al., 2017) естественно ожидать, что вероятность  $P^* = 0,9$  того, что все ребра  $(i, j) : \gamma_{i,j} \geq \gamma_0$  в знаковой сети попадут в построенное множество, контролируется при изменении параметра смеси  $\varepsilon$ . Эксперимент подтверждает это предположение при любом размере  $N = 5; 10; 20$  и любом  $\varepsilon \in [0,1]$  (табл. 2). Эксперименты показали, что такая вероятность практически равна 1. Данный факт можно объяснить грубостью неравенства Бонферрони. Аналогичный результат наблюдается и при других объемах наблюдений.

**Таблица 2**

Оценка доверительного уровня построенного множества в знаковой сети и сети Кендалла,  $N = 20, n = 100$

$\gamma_0$	$\varepsilon$										
	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Знаковая сеть											
0,1 (174)	0,99	0,99	0,98	0,99	0,98	1,00	0,99	0,98	0,99	0,99	0,98
0,3 (72)	0,99	0,99	0,98	0,99	0,98	0,99	0,99	0,99	0,99	1,00	1,00
0,5 (12)	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00	0,99
0,7 (3)	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Сеть Кендалла											
0,1 (174)	0,97	0,98	0,99	0,98	0,97	0,99	0,99	0,97	0,98	0,99	0,98
0,3 (72)	0,99	0,99	0,98	0,99	0,99	0,99	0,98	0,99	0,99	0,99	0,99
0,5 (12)	1,00	0,99	1,00	1,00	1,00	1,00	0,99	1,00	1,00	0,99	0,99
0,7 (3)	1,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00

Свойства теста Кендалла (9) в зависимости от параметра смеси  $\varepsilon$  изучались в (Koldanov, 2019a), где было показано, что уровень значимости теста (9) неустойчив при уменьшении параметра смеси  $\varepsilon$ . Вместе с тем, согласно проведенным экспериментам (см. табл. 2) вероятность того, что построенное множество содержит все пары  $(i, j) : \gamma_{i,j} \geq \gamma_0$ , несильно зависит от параметра смеси  $\varepsilon$ , поэтому условие  $P \geq P^*$  выполняется при  $P^* = 0,9$ . Данное утверждение справедливо в сети Кендалла при любом размере случайного вектора  $X$  ( $N = 5; 10; 20$ ) и любом  $\varepsilon \in [0, 1]$ . Отметим, что как и в случае знаковой сети, вероятность  $P$  значительно больше  $P^* = 0,9$ . При увеличении числа наблюдений  $n = 500; 1000$  вероятность  $P^*$  очень близка к 1 (см. табл. 2).

### 5.2. Среднее число ребер в построенных множествах

В табл. 3 приведены средние размеры (среднее число ребер) построенных множеств при  $n = 100, N = 20$  в зависимости от порога  $\gamma_0$  и параметра смеси  $\varepsilon$ .

В сети Пирсона с ростом  $\varepsilon$  среднее число ребер в построенном множестве возрастает при  $\gamma_0 = 0,1; 0,3; 0,5$  и убывает при  $\gamma_0 = 0,7$ . Это можно объяснить двумя разнонаправленными процессами.

1. При  $\gamma_{i,j} \geq \gamma_0$  имеем  $P_{St}(T_{i,j} > c) \leq P_N(T_{i,j} > c)$ , где  $P_{St}(T_{i,j} > c)$  – вероятность, что тест (7) не отвергнет гипотезу  $h_{i,j} : \gamma_{i,j} \geq \gamma_0$  при генерировании наблюдений из распределения Стьюдента,  $P_N(T_{i,j} > c)$  – вероятность, что тест (7) не отвергнет гипотезу  $h_{i,j} : \gamma_{i,j} \geq \gamma_0$  при генерировании наблюдений из нормального распределения. Поэтому среднее число правильно принятых гипотез при генерировании наблюдений из распределения Стьюдента меньше, чем при генерировании наблюдений из нормального распределения.

Таблица 3

Средний размер построенного множества

$\gamma_0$	$\varepsilon$					
	0	0,2	0,4	0,6	0,8	1
Сеть Пирсона						
0,1	184,40	186,16	187,38	187,82	189,01	189,93
0,3	169,24	173,83	175,93	177,85	181,15	184,26
0,5	117,50	118,55	118,85	119,15	122,01	120,15
0,7	33,22	31,37	29,06	26,01	21,90	19,21
Знаковая сеть						
0,1	189,93	189,95	189,93	189,93	189,96	189,95
0,3	188,52	188,4	188,5	188,57	188,43	188,47
0,5	165,21	164,21	165,08	165,11	164,01	166,05
0,7	76,07	76,24	76,5	77,34	76,72	75,69
Сеть Кендалла						
0,1	189,94	189,93	189,94	189,95	189,96	189,95
0,3	188,37	188,54	188,38	188,39	188,56	188,62
0,5	165,14	164,85	165,46	163,84	164,64	164,1
0,7	77,93	75,68	77,73	77,38	76,19	77,2

2. Если  $\gamma_{i,j} < \gamma_0$ , то  $P_{St}(T_{i,j} > c) \geq P_N(T_{i,j} > c)$ . Поэтому среднее число ложно принятых гипотез при генерировании наблюдений из распределения Стьюдента больше, чем при генерировании наблюдений из нормального распределения.

Из данных в табл. 3 следует, что средний размер построенных множеств в знаковой сети и сети Кендалла больше, чем средний размер построенного множества в сети Пирсона. Это может означать преимущество использования классического коэффициента корреляции. При этом следует учитывать результаты в табл. 1, которые показывают зависимость вероятности  $P$  от распределения наблюдений.

При анализе устойчивых ДМР( $\gamma_0, P^*$ ) в знаковой сети и сети Кендалла следует обратить внимание, как меняются размеры ДМР( $\gamma_0, P^*$ ) в этих сетях в зависимости от числа наблюдений.

В табл. 4 приведены средние размеры ДМР( $\gamma_0, P^*$ ) в знаковой сети, сети Кендалла и сети Пирсона при  $\varepsilon = 1$ , построенные по  $n = 100; 500; 1000$  наблюдениям при  $N = 20$ . Легко заметить, что размеры ДМР( $\gamma_0, P^*$ ) в этих сетях совпадают при различных объемах наблюдений.

В табл. 5 показан средний размер построенного множества в сети Пирсона при  $n = 100; 500; 1000$  наблюдениях.

**Таблица 4**

Изменение среднего размера ДМР( $\gamma_0, P^*$ ) в зависимости от объема наблюдений  $n$

$\gamma_0$	$n$			
	100	500	1000	Число истинных ребер
Знаковая сеть				
0,1	189,94	189,82	189,48	174
0,3	188,5	172,8	154,8	72
0,5	165,2	78,54	47,17	12
0,7	76,5	10,46	7,05	3
Сеть Кендалла				
0,1	189,94	189,81	189,50	174
0,3	188,48	172,84	154,65	72
0,5	165,08	78,57	47,18	12
0,7	76,5	10,50	7,05	3
$\varepsilon = 1$				
0,1	189,9	189,3	188,3	174
0,3	184	148,9	132,8	72
0,5	121,9	39,4	28,98	12
0,7	19,45	6,21	4,63	3

**Таблица 5**

Изменение среднего размера построенного множества в сети Пирсона в зависимости от объема наблюдений  $N = 20$ ,  $\varepsilon = 0$

$\gamma_0$	$N$			
	100	500	1000	Число истинных ребер
0,1	185	183,6	180,9	174
0,3	170,3	141,6	127	72
0,5	117,6	50,98	36,05	12
0,7	35,29	8,75	6,19	3

### 5.3. Сеть Пирсона. Среднее число истинных ребер в построенном множестве

В табл. 6 приведено среднее число истинных ребер в построенном множестве в сети Пирсона при  $n = 100$ ,  $N = 20$  в зависимости от  $\gamma_0$  и  $\varepsilon$ . Полученные результаты показывают, что максимальная средняя ошибка равна 4, т.е. максимальное число истинных ребер, которые не попадают в построенное множество в сети Пирсона, равно 4.

Таблица 6

Среднее число истинных ребер в построенном множестве в сети Пирсона

$\gamma_0$	N			
	100	500	1000	Число истинных ребер
0,1 (174)	169,80	171,27	172,23	172,42
0,3 (72)	69,66	70,59	70,82	71,19
0,5 (12)	11,76	11,85	11,84	11,88
0,7 (3)	2,95	2,95	2,97	2,98

#### 5.4. Отношение среднего числа неправильно невключенных ребер к общему числу истинных ребер в построенном множестве в сети Пирсона

Табл. 7 содержит отношение среднего числа неправильно невключенных ребер к общему числу истинных ребер в построенном множестве в сети Пирсона при  $n = 100$ ,  $N = 20$  в зависимости от  $\gamma_0$  и  $\varepsilon$ . Эти данные свидетельствуют, что отношение среднего числа неправильно невключенных ребер в построенное множество к общему числу истинных ребер остается стабильным и очень маленьким. Эксперименты для  $N = 5; 10$  дали аналогичные результаты.

Таблица 7

Отношение среднего числа неправильно невключенных ребер к общему числу истинных ребер в построенном множестве в сети Пирсона

$\gamma_0$	$\varepsilon$					
	0	0,2	0,4	0,6	0,8	1
0,1	0,02	0,02	0,01	0,01	0	0
0,3	0,03	0,02	0,02	0,01	0,01	0
0,5	0,02	0,01	0,01	0,01	0	0
0,7	0,01	0,01	0,01	0,01	0,01	0

## 6. Проверка согласия с эллиптической моделью

Предыдущие результаты основаны на предположении эллиптической модели доходностей акций. В соответствии с изложенным в п. 2.1, истинные графы рынка в сетевых моделях, порожденных эллиптическими сетями случайных величин  $(X, \gamma^P), (X, \gamma^{Sg}), (X, \gamma^{Kd})$ , при порогах, удовлетворяющих условию

$$\gamma_{i,j}^{Sg} = \gamma_{i,j}^{Kd} = \frac{1}{2} + \frac{1}{\pi} \arcsin \gamma_{i,j}^P, \quad (10)$$

совпадают. Однако при этом выборочные отсеченные графы могут отличаться.

Оценим степень различия выборочных отсеченных графов, заданных (3) с индивидуальными тестами (7), (8), (9), и исследуем, согласуется ли наблюдаемое различие с условием (10).

Заметим, что существует подкласс эллиптических распределений вектора  $X$ , в которых  $\text{cov}(X_i, X_j)$  не существует. Поэтому сеть Пирсона  $(X, \gamma^P)$  не существует. В этом случае последовательность (по  $n$ ) оценок  $(1/n) \sum_{t=1}^n (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)$  расходится (Феллер, 1984). Поэтому применение процедуры (6) с тестами (7) к имеющимся наблюдениям приводит, вообще говоря, к неинтерпретируемым результатам. Такого недостатка лишены сети  $(X, \gamma^{Sg})$ ,  $(X, \gamma^{Kd})$ .

Исследование тестов проверки гипотез согласия с эллиптической моделью распределения началось только в последнее время. В (Chicherortiche, Bouchaud, 2012) было показано, что совместное распределение доходностей акций не является эллиптическим. Такой результат был получен с помощью сравнения парных мер зависимости между парами акций. Авторы (Chicherortiche, Bouchaud, 2012) особо отметили, что применяемая ими методология проверки согласия не является статистической, что не позволяет оценить надежность полученных выводов.

В (Koldanov, Lozgageva, 2016; Semenov, Koldanov, 2018) предложены статистические процедуры проверки свойства симметрии распределения доходностей акций. Построены свободные от распределения тесты проверки симметрии доходностей каждой пары акций. Для проверки симметрии совместного распределения доходностей акций фондовых рынков США, Великобритании, Германии, Франции, Китая, Индии применена процедура Холма (Holm, 1979) множественной проверки гипотез, использующая эти индивидуальные тесты. Показано, что при удалении небольшого числа акций гипотеза симметрии совместного распределения доходностей остальных акций не отвергается.

В (Koldanov, 2019b) доказано, что в классе эллиптических распределений вектора  $X$  для любых  $i, j$  выполняется  $\gamma_{i,j}^{Kd} = \gamma_{i,j}^{Sg}$ ; построены свободные от распределения тесты проверки гипотез  $h_{i,j} : \gamma_{i,j}^{Kd} = \gamma_{i,j}^{Sg}$  против альтернатив  $k_{i,j} : \gamma_{i,j}^{Kd} \neq \gamma_{i,j}^{Sg}$  и процедура множественной проверки гипотез  $h_{i,j}$  совместного распределения доходности акций. Показано, что при удалении небольшого числа акций применение такой процедуры к данным фондовых рынков Китая, США, Великобритании и Германии приводит к неотвержению проверяемой гипотезы совместного распределения доходностей остальных акций.

### 6.1. Сравнение выборочных графов в сетях случайных величин $(X, \gamma^{Sg}), (X, \gamma^{Kd}), (X, \gamma^P)$

Пусть  $X_i(t)$ ,  $i=1, \dots, N$ ;  $t=1, \dots, n$  — множество наблюдений над доходностями акций. Построим три выборочных графа, т.е. три матрицы вида (3), используя индивидуальные тесты (7)–(9). Для сравнения этих матриц будем использовать симметрическую разность:



$$(X, \gamma^P), (X, \gamma^{Sg}): SR_{PS} = \sum_{i < j} (I(\varphi_{i,j}^P = 1, \varphi_{i,j}^{Sg} = 0) + I(\varphi_{i,j}^P = 0, \varphi_{i,j}^{Sg} = 1));$$

$$(X, \gamma^P), (X, \gamma^{Kd}): SR_{PK} = \sum_{i < j} (I(\varphi_{i,j}^P = 1, \varphi_{i,j}^{Kd} = 0) + I(\varphi_{i,j}^P = 0, \varphi_{i,j}^{Kd} = 1));$$

$$(X, \gamma^{Sg}), (X, \gamma^{Kd}): SR_{SK} = \sum_{i < j} (I(\varphi_{i,j}^{Sg} = 1, \varphi_{i,j}^{Kd} = 0) + I(\varphi_{i,j}^{Sg} = 0, \varphi_{i,j}^{Kd} = 1)).$$

Число ребер в выборочных графах рынка в рассматриваемых сетях, построенных по данным разд. 5 при  $N = 10; 20$ , приведено в табл. 8.

**Таблица 8**

Число ребер и число элементов в симметрической разности выборочных графов (по данным фондового рынка США) для двух размеров вектора  $X$

Размер вектора $X$	Сеть	$\gamma_0$				
		0	0,1	0,3	0,5	0,7
N=10	Число ребер					
	В сети Пирсона	45	42	16	1	0
	В знаковой сети	45	42	18	5	1
	В сети Кендалла	45	44	19	3	0
	Число элементов в симметрической разности					
	В сети Пирсона и знаковой сети $sr_{PS}$	0	4	8	4	1
	В знаковой сети и сети Кендалла, $sr_{SK}$	0	2	5	2	1
В сети Пирсона и сети Кендалла, $sr_{PK}$	0	4	9	4	0	
N=20	Число ребер					
	В сети Пирсона	189	174	72	12	3
	В знаковой сети	189	180	87	18	6
	В сети Кендалла	188	179	86	17	4
	Число элементов в симметрической разности					
	В сети Пирсона и сети Кендалла $sr_{PK}$	1	13	26	9	1
	В сети Пирсона и знаковой сети, $sr_{PK}$	0	18	39	10	3
В сети Кендалла и знаковой сети, $sr_{PK}$	1	7	25	7	2	

Из анализа данных в табл. 8 следует, что при  $N = 10$  выборочные графы рынка в различных сетях случайных величин могут сильно отличаться. Например, при пороге  $\gamma_0 = 0,3$ ,  $N = 10$  графы рынка в сети Пирсона и в знаковой сети отличаются в 8 элементах, при том что графы при данном пороге содержат 16, 18 ребер соответственно.

Таким образом, имеется 5 ребер (т.е. примерно 27%), которые присутствуют в одном графе и не присутствуют в другом. Аналогичный вывод справедлив и при  $N = 20$  как для графов рынка в сети Пирсона и сети Кендалла, так и для графов рынка в знаковой сети и сети Кендалла.

Отсюда вытекают два вопроса. Какой выборочный граф предпочтительней для практического применения? Приводят ли полученные результаты к отвержению гипотезы эллиптичности?

Ответ на первый вопрос следует из результатов (Kalyagin et al., 2017), в соответствии с которыми целесообразно отдать предпочтение графу рынка в знаковой сети как более устойчивому. Ответ на второй вопрос приведен в п. 6.2.

### 6.2. Процедура проверки согласия с эллиптической моделью

Для проверки соотношения (10) предложим процедуру проверки гипотезы согласия с эллиптической моделью, имеющую непосредственное отношение к задаче идентификации графа рынка. Процедура основана на числе элементов в симметрической разности матриц (3) с различными индивидуальными тестами. Гипотеза согласия наблюдаемых данных с эллиптической моделью будет отвергаться, если число элементов в симметрической разности матриц (3) с различными индивидуальными тестами больше порога  $ts_{1-\alpha}$ . Порог  $ts_{1-\alpha}$  определяется как квантиль уровня  $1-\alpha$  функции распределения случайной величины  $SR$ , равной числу элементов в симметрической разности двух матриц (3). Функция распределения  $SR$  находится в предположении истинности гипотезы эллиптичности. Таким образом, процедура проверки согласия с эллиптической моделью имеет вид

$$\delta_{gf} = \begin{cases} 0, & SR \leq ts_{1-\alpha}; \\ 1, & SR > ts_{1-\alpha}. \end{cases} \quad (11)$$

В настоящей работе пороги  $ts_{1-\alpha}$  находятся методом статистического моделирования по 100 повторениям эксперимента, описанного в начале разд. 5 при  $n = 250$ , что соответствует выбранному периоду наблюдения.

В табл. 9 приведены квантили  $ts_{0,9}$  уровня 0,9 эмпирического распределения статистики  $SR_{PS}$ ,  $SR_{PK}$ ,  $SR_{SK}$  при  $N = 10$ .

Из анализа соответствующих частей табл. 8 и 9 следует, что гипотеза об эллиптической модели распределения доходностей не отвергается:

- в сети Пирсона и в знаковой сети при любом  $\varepsilon \in [0; 1]$  и  $\gamma_0 \in [0; 0,5]$ , а также при  $\varepsilon \in [0; 0,2]$  и  $\gamma_0 = 0,7$ ;
- в сети Пирсона и в сети Кендалла при любом  $\varepsilon \in [0; 1]$  и  $\gamma_0 = 0$ ,  $\gamma_0 = 0,7$ , при  $\varepsilon \in [0; 0,9]$  и  $\gamma_0 = 0,1$ , при  $\varepsilon \in [0; 0,8]$  и  $\gamma_0 = 0,3$ , при  $\varepsilon \in [0; 0,7]$  и  $\gamma_0 = 0,5$ ;
- в знаковой сети и в сети Кендалла при любом  $\varepsilon \in [0; 1]$  и  $\gamma_0 < 0,7$ .

Таким образом, в основном гипотеза об эллиптической модели распределения доходностей на уровне  $\alpha = 0,1$  не отвергается.

Таблица 9

Доверительные границы для числа элементов в симметрической разности выборочных графов,  $N=10$

$\varepsilon$	Сеть Пирсона и знаковая сеть ( $SR_{PS}$ )					Сеть Пирсона и сеть Кендалла ( $SR_{PK}$ )					Знаковая сеть и сеть Кендалла ( $SR_{SK}$ )				
	$\gamma_0$					$\gamma_0$					$\gamma_0$				
	0	0,1	0,3	0,5	0,7	0	0,1	0,3	0,5	0,7	0	0,1	0,3	0,5	0,7
0	8	11	14	9	1	8	11	14	9	1	3	6	9	4	0
0,1	8	12	15	8	1	8	12	15	8	1	3	6	8	4	0
0,2	7	11	14	7	1	7	11	14	7	1	3	6	9	3	0
0,3	7	10	13	6	0	7	10	13	6	0	3	6	9	3	0
0,4	6	10	13	6	0	6	10	13	6	0	3	6	9	4	0
0,5	6	10	12	5	0	6	10	12	5	0	3	6	9	4	0
0,6	5	8	11	5	0	5	8	11	5	0	3	6	8	4	0
0,7	4	7	10	4	0	4	7	10	4	0	3	6	8	4	0
0,8	3	6	9	3	0	3	6	9	3	0	3	6	9	3	0
0,9	1	4	5	2	0	1	4	5	2	0	3	6	8	4	0
1,0	1	3	4	1	0	1	3	4	1	0	3	6	9	4	0

## 7. Заключение

В работе исследованы способы построения множества пар акций, содержащего с заданной вероятностью  $P^*$  все пары акций, сила связи между которыми не менее заданного порога. Такие процедуры представляют собой комбинирование результатов совместного применения стандартных тестов проверки гипотез о величине классического коэффициента корреляции Пирсона, коэффициентов корреляции Фехнера и Кендалла.

Результаты статистического моделирования применения процедур построения таких множеств в различных сетях случайных величин показывают, что данные процедуры в эллиптической сети корреляций Пирсона не приводят к построению доверительных множеств с заданной доверительной вероятностью  $P^*$ . А именно вероятность того, что построенное множество содержит все ребра, веса которых больше  $\gamma_0$ , не контролируется при изменении параметра смеси многомерного нормального распределения и многомерного распределения Стьюдента. В то же время число правильно включенных ребер в построенное множество в сети Пирсона мало отличается от истинного числа ребер. В частности, отношение среднего числа неправильно невключенных ребер к общему числу истинных ребер остается стабильным и не превышает 2%.

Результаты статистического моделирования сильно зависят от числа истинных ребер. В частности, результаты экспериментов, приведенные в табл. 1, показывают, что в сети Пирсона при  $\gamma_0 = 0,5$ ,  $\varepsilon \geq 0,6$  и при  $\gamma_0 = 0,7$  при любом  $\varepsilon$ , т.е. при малом числе истинных ребер, доверительная вероятность  $P^*$  контролируется. Поэтому при имеющейся априорной информации о малом числе истинных ребер, несильном отклонении ( $\varepsilon \geq 0,6$ ) от нормального распределения, при любом объеме наблюдений для построения ДМР( $\gamma_0, P^*$ ) можно рекомендовать использовать тесты Пирсона (7), которые в среднем приводят к ДМР( $\gamma_0, P^*$ ) меньшего размера.

Результаты статистического моделирования по построению ДМР( $\gamma_0, P^*$ ) в знаковой сети и сети Кендалла, в отличие от сети Пирсона, практически одинаковы и не зависят от числа истинных гипотез. А именно вероятность  $P^* = 0,9$ , что все ребра  $(i, j): \gamma_{i,j} \geq \gamma_0$  в знаковой сети и в сети Кендалла попадут в построенное множество, контролируется как при изменении параметра смеси  $\varepsilon$ , так и при изменении параметра  $\gamma_0$ . То есть построенное множество является ДМР( $\gamma_0, P^*$ ).

Неожиданным результатом является устойчивость ДМР( $\gamma_0, P^*$ ) в сети Кендалла, несмотря на неустойчивость индивидуального теста (9), показанную в (Koldanov, 2019a). Это можно объяснить грубостью неравенства Бонферонни. В то же время в сети Пирсона грубость неравенства Бонферонни не приводит к устойчивым доверительным множествам.

Полученные результаты позволяют рекомендовать процедуры построения доверительных множеств, основанные на коэффициентах корреляции Пирсона, при больших порогах  $\gamma_0$ . Это имеет практическое значение при выделении групп сильносвязанных активов фондового рынка, т.е. при построении клик графа рынка (Boginski, Butenko, Pardalos, 2006). В то же время при малых порогах, т.е. при выделении групп несвязанных или слабосвязанных активов фондового рынка, целесообразно использовать коэффициенты корреляции Фехнера или Кендалла. Как известно, группы слабосвязанных активов или независимые множества графа рынка представляют интерес при построении диверсифицированных портфелей инвестиций (Koldanov et al., 2014).

Несмотря на эквивалентность истинных графов рынка в рассматриваемых сетях случайных величин с эллиптическим распределением, выборочные графы рынка могут различаться. Доверительные множества ДМР( $\gamma_0, P^*$ ) для ребер графа рынка в знаковой сети и сети Кендалла различаются лишь неправильно включенными ребрами, так как содержат все ребра с весами не менее  $\gamma_0$  с вероятностью не менее  $P^*$ .

Предложена процедура проверки согласия с эллиптической моделью распределения доходностей, основанная на числе элементов в симметрической разнице графов рынка в различных сетях случайных величин. Применение этой процедуры на уровне  $\alpha = 0,1$  к анализу

реальных данных фондового рынка США показывает, что гипотеза об эллиптической модели распределения доходностей для несильно разреженных графов не отвергается.

#### ЛИТЕРАТУРА / REFERENCES

- Визгунов А.Н., Гольденгорин Б.И., Замараев В.А., Калягин В.А., Колданов А.П., Колданов П.А., Пардалос П.М.** (2012). Применение рыночных графов к анализу фондового рынка России // *Журнал Новой экономической ассоциации*. № 3 (15). С. 66–81. [Vizgunov A.N., Gol'dengorin B.I., Zamaraev V.A., Kaljagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M. (2012). Applying market graphs for Russian stock market analysis. *Journal of the New Economic Association*, 3 (15), 66–81 (in Russian).]
- Калягин В.А., Колданов А.П., Колданов П.А., Пардалос П.М.** (2017). Статистические процедуры идентификации сетевых структур фондовых рынков // *Журнал Новой экономической ассоциации*. № 3 (35). С. 33–52. [Vizgunov A.N., Koldanov A.P., Koldanov P.A., Pardalos P.M. (2017). Statistical procedures for stock markets network structures identification. *Journal of the New Economic Association*, 3 (35), 33–52 (in Russian).]
- Феллер В.В.** (1984). Введение в теорию вероятностей и ее приложения. Пер. с англ. М.: Мир. [Feller W.W. (1984). An Introduction to probability theory and its applications. Translated from the English. Moscow: Mir (in Russian).]
- Anderson T.W.** (2003). *An introduction to multivariate statistical analysis*. New York: Wiley-Interscience.
- Bodnar T.** (2013). *Elliptically contoured models in statistics and portfolio theory*. T. Bodnar, F.K. Gupta, T. Varga (eds.). New York: Springer.
- Boginski V., Butenko S., Pardalos P.M.** (2005). Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, 48 (2), 431–443.
- Boginski V., Butenko S., Pardalos P.M.** (2006). Mining market data: A network approach. *Computers & Operations Research*, 33 (11), 3171–3184.
- Chicportiche R., Bouchaud J.-P.** (2012). The joint distribution of stock returns is not elliptical. *International Journal of Theoretical and Applied Finance*, 15, 3, 1250019.
- Daniels H.E., Kendall M.G.** (1947). The significance of rank correlations where parental correlation exists. *Biometrika*, 34, 3/4, 197–208.
- Finner H., Giani G.** (1994). Closed subset selection procedures for selecting good populations. *Journal of Statistical Planning and Inference*, 38, 179–200.
- Garas F., Argyrakis P.** (2007). Correlation study of the Athens stock exchange. *Physica A*, 380, 399–410.
- Gunawardena A.D.A., Meyer R.R., Dougan W.L., Monaghan P.E., Choton-Basu P.E.M.** (2012). Optimal selection of an independent set of cliques in a market graph. *International Proceedings of Economics Development and Research*, 29, 281–285.
- Gupta S.S., Panchapakesan S.** (2002). *Multiple decision procedures. Theory and methodology of selecting and ranking populations*. SIAM.
- Hero A., Rajaratnam B.** (2012). Hub discovery in partial correlation graphs. In: *IEEE Transactions on Information Theory*, 58, 9, 6064–6078, Sept. DOI: 10.1109/TIT.2012.2200825

- Hochberg Y., Tamhane A.** (1987). Multiple comparison procedures. N.Y.: John Wiley & Sons, Inc.
- Hoffding W.** (1947). On the distribution of the rank correlation coefficient when the variates are not independent. *Biometrika*, 34, 3/4, 183–196.
- Holm S.** (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6 (2), 65–70.
- Huang W.Q., Zhuang X.T., Yao S.A.** (2009). A network analysis of the Chinese stock market. *Physica A*, 388, 2956–2964.
- Jordan M.I.** (2004). Graphical models. *Statist. Sci.*, 19, 1, 140–155.
- Kalyagin V.A., Koldanov A.P., Koldanov P., Pardalos P.M.** (2018). Optimal decision for the market graph identification problem in a sign similarity network. *Annals of Operations Research*, 266, 1–2, 313–327.
- Kalyagin V.A., Koldanov A.P., Koldanov P.A.** (2017). Robust identification in random variable networks. *Journal of Statistical Planning and Inference*, 181, 30–40.
- Koldanov A.P., Kalyagin V.A., Koldanov P.A., Pardalos P.M.** (2013). Statistical procedures for the market graph construction. *Computational Statistics & Data Analysis*, 68, 17–29.
- Koldanov P., Kalyagin V.A., Koldanov A.P., Zamaraev V.A.** (2014). Market graph and markowitz model. In: *Optimization of science and engineering (In Honor of the 60th birthday of Panos M. Pardalos)*. N.Y.: Springer, Ch. 15, 301–313.
- Koldanov P., Lozgageva N.** (2016). Multiple testing of sign symmetry for stock return distributions. *International Journal of Theoretical and Applied Finance*, 19, 8, 1650049.
- Koldanov P.A.** (2019a). Statistics of individual tests for market graph identification in market network. *Advances in Computer Science Research*, 50–55.
- Koldanov P.A.** (2019b). Testing new property of elliptical model for stock returns distribution. arXiv:1907.10306 [stat.AP]
- Kruskal W.H.** (1958) Ordinal measures of association. *Journal of the American Statistical Association*, 53, 284, 814–861.
- Lehmann E.L.** (1957). A theory of some multiple decision problems, I. *The Annals of Mathematical Statistics*, 28, 1–25.
- Namaki A., Shirazi A.H., Raei R., Jafari G.R.** (2011). Network analysis of a financial market based on genuine correlation and threshold method. *Physica A*, 390, 3835–3841.
- Semenov D., Koldanov P.** (2018). Rejection graph for multiple testing of elliptical model for market network. In: *Computational aspects and applications in large-scale networks. Springer Proceedings in Mathematics & Statistics*, 247, 221–234.
- Shirokikh J., Pastukhov G., Boginski V., Butenko S.** (2013). Computational study of the US stock market evolution: A rank correlation-based network model. *Computational Management Science*, 10 (2–3), 81–103.
- Tse C.K., Liu J., Lau F.C.M.** (2010). A network perspective of the stock market. *Journal of Empirical Finance*, 17, 659–667.

Поступила в редакцию 02.02.2020

Received 02.02.2020

**A.P. Koldanov**National Research University Higher School of Economics,  
Nizhnii Novgorod, Russia**P.A. Koldanov**National Research University Higher School of Economics,  
Nizhnii Novgorod, Russia**D.P. Semenov**National Research University Higher School of Economics,  
Nizhnii Novgorod, Russia

## Confidence set for connected stocks of stock market<sup>2</sup>

**Abstract.** The problem of analysis of pairwise connections between stocks of financial market by observations on stock returns is considered. Such problem arise in stock market network analysis. It is assumed that joint distribution of stock returns belongs to the wide class of elliptical distributions. Classical Pearson correlation, Fechner correlation and Kendall correlation are used as measure of dependence. The construction problems of sets of stocks with strong connections between its returns are investigated. The construction problems of sets of stocks with strong connections between its returns are investigated. To construct such sets the multiple hypotheses testing procedures on values of correlations are used. The properties of these statistical procedures are investigated by simulations. The simulation results show that procedures based on individual Fechner and Kendall tests lead to such sets of stocks with given confidence probability unlike procedure based on Pearson individual tests which do not control the confidence probability. At the same time it is emphasized that for Student distribution the constructed set is nearly the same to the confidence set. The procedure of consistency testing with elliptical model is proposed and exemplified. The peculiarities of the model are discussed.

**Keywords:** *Network model of stock market, threshold graph, Pearson correlation, Kendall correlation, Fechner correlation, sufficient set, multiple hypotheses testing procedures.*

JEL Classification: C02.

DOI: 10.31737/2221-2264-2021-50-2-1

---

<sup>2</sup> The article was prepared as a result of research within the Program of fundamental research in the National Research University "Higher School of Economics". The work of the first two authors was partly granted by Russian Foundation for Basic Research (project 18-07-00524). The work of the third author was partly granted by Russian Foundation for Basic Research (project 19-31-90088).