В.А. Калягин
НИУ ВШЭ, Нижний Новгород

А.П. Колданов
НИУ ВШЭ, Нижний Новгород

П.А. Колданов
НИУ ВШЭ, Нижний Новгород

П.М. Пардалос
Университет Флориды, США

Статистические процедуры идентификации сетевых структур фондовых рынков[1]

Аннотация. Под сетевой моделью фондового рынка понимается полный взвешенный граф, вершины которого соответствуют рыночным активам, а веса ребер задаются значением некоторой меры зависимости характеристик этих активов. Наиболее распространенная характеристика рыночных активов – их доходность. При анализе сетевых моделей доходностей основной интерес представляют сетевые структуры, содержащие ключевую информацию о рассматриваемой сети. Популярными сетевыми структурами являются максимальное остовное дерево, максимально отфильтрованный планарный граф, граф рынка, клики и независимые множества графа рынка. Задача идентификации сетевой структуры заключается в построении процедуры ее выделения по результатам наблюдений. Важной характеристикой статистических процедур идентификации служит неопределенность идентификации, связанная с конечным объемом выборки. Существенную роль играют совместное распределение доходностей рыночных активов и выбор меры зависимости между ними. Наиболее распространена мера зависимости в виде коэффициента корреляции Пирсона. Широкий класс совместных распределений доходностей активов представляют эллиптические модели. Процедуры, построенные на основе корреляций Пирсона, оказываются неустойчивыми при отклонениях совместного распределения доходностей от нормального в классе эллиптических распределений. Цель настоящей работы – изложение общего подхода к построению устойчивых статистических процедур идентификации сетевых структур фондовых рынков. В качестве меры зависимости используется вероятность совпадения знаков доходностей ценных бумаг. Показано, что в сети вероятностей совпадения знаков одношаговые и многошаговые стандартные процедуры идентификации сетевых структур обладают свойством устойчивости в классе эллиптических распределений. Это позволяет рекомендовать указанные процедуры для практического применения.

Ключевые слова: *фондовый рынок, сетевая (графическая) модель, сетевая структура, коэффициент корреляции Пирсона, вероятность совпадения знаков.*

Классификация JEL: C02.

1. Введение

Сетевые (графические) модели являются популярным средством анализа фондовых рынков (Mantegna, 1999; Boginsky, 2003; Boginski et al., 2005; Boginski et al., 2006; Bonanno et al., 2001; Tumminello et al., 2010). Под сетевой (графической) моделью фондо-

[1] Работа выполнена в лаборатории «Алгоритмы и технологии анализа сетевых структур» НИУ ВШЭ. В.А. Калягин и П.А. Колданов выполняли работу при финансовой поддержке Российского гуманитарного научного фонда (проект 15-32-01052).

Журнал НЭА, №3 (35), 2017, с. 33–52

вого рынка понимается полный взвешенный граф, вершины которого соответствуют рыночным активам, а веса ребер задаются значением некоторой меры зависимости характеристик этих активов. В качестве характеристик могут использоваться доходности, цены, объемы продаж, ликвидность и т.п. В настоящей работе рассматривается сетевая (графическая) модель фондового рынка, построенная по доходностям рыночных активов. Доходности считаются случайными величинами. Представление доходностей как случайных величин лежит в основе современной теории портфеля ценных бумаг, основы которой заложены в работе (Markowitz, 1952). Известны различные модели распределения доходностей. Наибольшее распространение получили многомерные эллиптические распределения (Gupta, 2013) и распределения, заданные различными копула-функциями (Фантаццини, 2011).

Формально, сетевая (графическая) модель фондового рынка порождается сетью случайных величин (random variable network). Сетью случайных величин называется пара (X,γ), где $X=(X_1,\ldots,X_N)$ – случайный вектор, $\gamma=\gamma(X,Y)$ – мера зависимости случайных величин X,Y (Kalyagin et al., 2017). В сетевой (графической) модели, порожденной сетью случайных величин, вес ребра (i,j) задается значением $\gamma(X_i,X_j)$. Для построения сети случайных величин фондового рынка и порожденной ею сетевой (графической) модели вектор $X=(X_1,\ldots,X_N)$ определяется совместным распределением доходностей рыночных активов. При анализе сетевых (графических) моделей основной интерес представляют сетевые структуры (подграфы полного графа), содержащие ключевую информацию о рассматриваемой сети. Наиболее распространенными сетевыми структурами являются максимальное остовное дерево, максимально отфильтрованный планарный граф, граф рынка, клики и независимые множества графа рынка.

Максимальное остовное дерево введено в работе (Mantegna, 1999). Такая структура на фондовом рынке представляет набор связанных между собой активов с максимальной суммарной связью между ними. Вершины с большим числом связей в максимальном остовном дереве можно интерпретировать как наиболее влиятельные активы на рынке. Структура остовного дерева определяет иерархическую организацию рынка по степени влияния активов друг на друга. Анализ таких иерархических структур для различных финансовых рынков проведен в работах (Coronnello et al., 2005; Garas, Argyrakis, 2007; Huang et al., 2009; Jung et al., 2006; Tabak et al., 2010; Tee, 2005; Wang et al., 2012). В частности, в работе (Wang et al., 2012) изучено изменение во времени максимального остовного дерева на рынке валют. Показано, что доллар США и евро являются доминирующими валютами в максимальном остовном дереве рынка валют, но во время кризиса доллар США теряет свою центральную позицию, в то время как евро сохраняет ее. Максимально отфильтрованный планарный граф – это обобщение

Журнал НЭА, №3 (35), 2017, с. 33–52

максимального остовного дерева, и он может быть использован для выявления кластерных структур на рынках (Tumminello et al., 2005).

Граф рынка (Boginsky et al., 2003) представляет подграф сетевой модели с ребрами, веса которых больше заданного порога. В работе (Boginsky et al., 2005) показано, что распределение степеней вершин графа рынка подчиняется степенному закону. Особый интерес вызывают клики и независимые множества графа рынка. Клику в графе рынка образуют наборы активов, все попарные связи между которыми превышают заданный порог (набор тесно связанных между собой активов). Независимое множество в графе рынка составляют наборы активов, все попарные связи между которыми меньше заданного порога (набор активов со слабыми или отрицательными попарными связями). В работе (Boginsky et al., 2006) была исследована динамика клик и независимых множеств графа рынка США и показана ее связь с процессом глобализации. В работе (Визгунов и др., 2012) анализировался граф рынка для фондового рынка России. Оказалось, что максимальную по размеру клику графа рынка образуют небольшое число активов, обеспечивающих 97% объемов торгов. Этот феномен не наблюдается на фондовых рынках развитых стран. Независимое множество графа рынка может быть эффективно использовано при построении оптимальных портфелей инвестиций (Koldanov и др., 2014). Авторами (Koldanov и др., 2014) предложен алгоритм выбора небольшого числа активов для построения эффективного оптимального портфеля. На первом этапе происходит отбор активов по отношению Шарпа. Затем строится граф рынка на выбранном множестве активов и вычисляется независимое множество максимального размера в этом графе. Эффективный фронт портфелей, построенных на активах из независимого множества, близок к эффективному фронту портфелей, построенных по исходному множеству активов. Это дает основание для практической рекомендовации ограниченного числа активов для формирования эффективных портфелей инвестиций (portfolio selection problem).

Во всех упомянутых работах в качестве меры зависимости доходностей используется выборочный коэффициент корреляции Пирсона. Основное внимание этих работ сосредоточено на интерпретации найденных сетевых структур без анализа статистической неопределенности процедур их идентификации. Первой работой, в которой задача идентификации сетевых структур рассматривается как задача теории статистических решений, является работа (Koldanov et al., 2013). В работах (Bautin et al., 2013a, 2014) было выявлено, что неопределенность процедур идентификации сетевых структур, основанных на корреляциях Пирсона, существенно возрастает при отклонении совместного распределения доходностей от нормального. В работе (Bautin et al., 2013b) предложена новая мера зависимости, связанная с вероятностью совпадения знаков случайных величин.

Журнал НЭА, №3 (35), 2017, с. 33–52

В работах (Bautin et al., 2013a, 2014) экспериментально установлено, что процедуры идентификации сетевых структур, основанные на вероятности совпадения знаков, обладают свойством устойчивости. В работе (Kalyagin et al., 2017) предложен математический аппарат анализа устойчивости статистических процедур идентификации сетевых структур в сетевых моделях, порожденных сетями случайных величин.

Целью настоящей работы является изложение общего подхода к построению устойчивых статистических процедур идентификации сетевых структур фондовых рынков. В качестве меры зависимости предлагается использовать вероятность совпадения знаков доходностей ценных бумаг. Показано, что в сети вероятностей совпадения знаков одношаговые и многошаговые стандартные процедуры идентификации сетевых структур обладают свойством устойчивости в широком классе эллиптических распределений. Это позволяет рекомендовать указанные процедуры для практического применения.

Статья организована следующим образом: в разд. 2 введены основные определения и обозначения; в разд. 3 излагается общий подход к построению процедур идентификации сетевых структур с позиций теории решений; в разд. 4 формулируется понятие устойчивости процедур идентификации сетевых структур в классе эллиптических распределений; в разд. 5 описаны практические процедуры идентификации сетевых структур; в разд. 6 изложена схема доказательства устойчивости практических процедур идентификации сетевых структур фондовых рынков; в разд. 7 обсуждаются полученные результаты и дальнейшие направления исследований.

2. Основные определения и обозначения

Рассмотрим сеть случайных величин (X,γ).

Сетевой (графической) моделью называется полный взвешенный граф с N вершинами, в котором вес ребра (i,j) задается мерой зависимости γ между случайными величинами X_i и X_j: $\gamma_{i,j}=\gamma(X_i,X_j)$, $i,j=1,\dots,N$.

Коэффициент корреляции Пирсона является наиболее известной и популярной мерой зависимости в сетевом анализе рынков:

$$\gamma^P_{i,j}=\frac{E((X_i-E(X_i))(X_j-E(X_j)))}{\sqrt{E(X_i)^2-E^2(X_i)}\sqrt{E(X_j^2)-E^2(X_j)}}. \tag{1}$$

Коэффициент корреляции Пирсона как мера зависимости порождает сеть корреляций Пирсона. Аналогично могут быть определены сети корреляций Кендалла, Спирмена, Фехнера и др. В настоящей работе в качестве меры зависимости случайных величин предлагается использовать вероятность совпадения знаков

$$\gamma^S_{i,j}=p^{i,j}=P((X_i-E(X_i))(X_j-E(X_j))>0).$$

Соответствующую сеть мы будем называть *сетью вероятностей совпадения знаков.* Эта сеть связана с сетью корреляций Фехнера соотноше-

Журнал НЭА, № 3 (35), 2017, с. 33–52

нием $\gamma^F_{i,j} = 2\gamma^S_{i,j} - 1$. Для построения сетевой (графической) модели фондового рынка в качестве случайных величин X_i рассматриваются доходности рыночных активов. В качестве меры зависимости между доходностями можно выбирать различные коэффициенты корреляции. Основное внимание уделяется двум сетевым структурам: графу рынка и максимальному остовному дереву рынка.

Граф рынка (MG) строится следующим образом: ребро между вершинами i и j входит в граф рынка, если $\gamma_{i,j} > \gamma_0$, где γ_0 — некоторый заданный порог. Граф рынка отражает структуру наиболее значимых (на уровне γ_0) связей рыночных активов. Для данного значения порога γ_0 граф рынка определяется матрицей смежности $S = (s_{i,j})$, где $s_{i,j} = 0$ при $\gamma_{i,j} \le \gamma_0$ и $s_{i,j} = 1$ при $\gamma_{i,j} > \gamma_0$, $s_{i,i} = 0$, $i, j = 1, \dots, N$. В дальнейшем будем использовать множество G всех $N \times N$ симметричных матриц $G = (g_{i,j})$ с $g_{i,j} \in \{0,\ 1\}$, $g_{i,i} = 0$, $i, j = 1, \dots, N$. Матрицы $G \in G$ являются матрицами смежности всех простых неориентированных графов на N вершинах. Общее число матриц в G равно $L = L_{MG} = 2^M$, где $M = N(N-1)/2$.

Максимальным остовным деревом (MST) сетевой (графической) модели фондового рынка (X, γ) называется остовное дерево (граф без циклов на N вершинах) с максимальным суммарным весом ребер. Максимальное остовное дерево отражает некоторую иерархическую структуру активов фондового рынка с максимальными суммарными связями. Согласно формуле Кэли общее число остовных деревьев в полном графе на N вершинах равно $L = L_{MST} = N^{N-2}$.

3. Задача идентификации сетевых структур фондовых рынков

При практическом построении сетевых структур распределение вектора X и значение $\gamma_{i,j}$ неизвестны. Доступными данными являются наблюдения за доходностями рыночных активов. Под проблемой идентификации сетевой структуры в настоящей работе мы будем понимать задачу построения сетевой структуры по наблюдениям. В качестве модели наблюдений мы используем повторную выборку конечного объема из распределения случайного вектора X: $x(t) = (x_1(t), \dots, x_N(t)),\quad t = 1, \dots, n$.

Проблема идентификации графа рынка может быть сформулирована как статистическая проблема выбора одной из $L = L_{MG}$ гипотез:

$$H_G : \gamma_{i,j} \le \gamma_0, \text{ если } g_{i,j} = 0;\ \gamma_{i,j} > \gamma_0, \text{ если } g_{i,j} = 1;\ i \ne j. \qquad (2)$$

Любая статистическая процедура δ идентификации графа рынка является разбиением выборочного пространства $R^{N \times n}$ на области пространства решений $D = \{d_G, g \in G\}$, где решение d_G соответствует принятию гипотезы H_G, $G \in G$.

Пусть $S = (s_{i,j})$, $Q = (q_{i,j})$, $S, Q \in G$. Обозначим через $w(S,Q)$ потери от принятия решения d_Q, когда гипотеза H_S истинна

Журнал НЭА, №3 (35), 2017, с. 33–52

$w(H_S; d_Q) = w(S,Q),\ S,Q \in G$. Будем считать, что $w(S,S) = 0,\ S \in G$. В соответствии с теорией решений Вальда (Wald, 1950) качество статистической процедуры δ характеризуется функцией риска. Пусть $f(x;\theta)$ – функция плотности случайного вектора $X, \theta \in \Omega$ – параметрическое пространство, Ω_S – область изменения параметров, при котором истинна гипотеза H_S. Функция риска определяется формулой

$$R(S,\theta;\delta) = \sum_{Q \in G} w(S,Q) P_\theta(\delta(x) = d_Q), \theta \in \Omega_S.$$

В теории процедур выбора одного из многих решений и теории процедур множественной проверки гипотез (Lehmann, Romano, 2005; Dudoit et al., 2003) существуют различные критерии качества статистических процедур. Одним из распространенных критериев является вероятность хотя бы одной ошибки первого рода (FWER), а именно вероятность хотя бы одного ложного включения ребра в сетевую структуру. Этот критерий, как и многие другие, можно считать риском функции потерь, определенной подходящим образом. Действительно, рассмотрим функцию потерь $w(S,Q)$, которая принимает два значения: $w(S,Q) = 1$, если хотя бы одно ребро неправильно включено в сетевую структуру, и $w(S,Q) = 0$ – в противном случае. Тогда функция риска равна вероятности хотя бы одной ошибки первого рода.

В задаче идентификации сетевых структур фондовых рынков естественным предположением является предположение аддитивности функции потерь. Аддитивность функции потерь означает, что потери от неправильной идентификации сетевой структуры равны сумме потерь от неправильной идентификации отдельных ребер. В частности, для графа рынка аддитивность функции потерь означает

$$w(S,Q) = \sum_{\{i,j: s_{i,j}=0; q_{i,j}=1\}} a_{i,j} + \sum_{\{i,j: s_{i,j}=1; q_{i,j}=0\}} b_{i,j}.$$

Функция риска в задаче идентификации графа рынка для аддитивной функции потерь может быть записана в виде

$$R(S,\theta,\delta) = \sum_{i=1}^{N}\sum_{j=1}^{N} R(s_{i,j},\theta;\varphi_{i,j}), \tag{3}$$

где $R(s_{i,j},\theta;\varphi_{i,j})$ – риск тестов $\varphi_{i,j}$ для проверки индивидуальных гипотез

$$h_{i,j}: \gamma_{i,j} \le \gamma_0 \ \text{ против } \ k_{i,j}: \gamma_{i,j} > \gamma_0 \ (i,j = 1,\ldots,N; \quad i \ne j). \tag{4}$$

Проблема идентификации MST может быть сформулирована как статистическая проблема выбора одного из $L = L_{MST}$ остовных деревьев. Одной из процедур идентификации MST является процедура Краскала: на первом этапе по наблюдениям вычисляются оценки $\hat{\gamma}_{i,j}$ весов ребер $\gamma_{i,j}$; на втором – значения $\hat{\gamma}_{i,j}$ упорядочиваются по убыванию; на третьем – применяется алгоритм Краскала (Кормен и др., 2005) для построения максимального остовного дерева. На каждом шаге алгоритма выбирается ребро максимального веса из оставшихся ребер и добавляется в структуру, если это не приводит к образованию

Журнал НЭА, №3 (35), 2017, с. 33–52

цикла. Если приводит – ребро отбрасывается и рассматриваются оставшиеся.

Процедура Краскала может быть модифицирована для построения процедуры идентификации планарного максимально отфильтрованного графа (Tumminello et al., 2005).

Задача идентификации сетевых структур заключается в построении статистических процедур идентификации с практически полезными свойствами. К таким свойствам относятся: оптимальность, несмещенность, устойчивость и контроль ошибок процедур в различных классах распределений. Оптимальные и несмещенные процедуры идентификации графа рынка в сети корреляции Пирсона исследованы в (Koldanov et al., 2013). Значение функции риска этих процедур существенно возрастает при отклонениях совместного распределения доходностей от нормального. Это же свойство справедливо для процедур Краскала идентификации максимального остовного дерева и максимального отфильтрованного планарного графа. Актуальной, таким образом, становится задача построения устойчивых к распределениям процедур идентификации сетевых структур фондовых рынков.

4. Сетевые модели в классе эллиптических распределений

Пусть K – класс распределений вектора X, такой что при фиксированной мере зависимости γ сетевые (графические) модели, порожденные сетями случайных величин (X,γ), совпадают, т.е. $\gamma(X_i^{(1)},X_j^{(1)})=\gamma(X_i^{(2)},X_j^{(2)}) \quad \forall X^{(1)},X^{(2)}\in K, \quad \forall i,j=1,\ldots,N$. В этом случае для всех распределений из класса K сетевые структуры в порожденных графических (сетевых) моделях также совпадают. Однако свойства статистических процедур идентификации сетевых структур могут сохранять зависимость от распределения вектора X из класса K. Это связано с возможной зависимостью распределений, определяющих процедуру статистик от распределения $X\in K$. Под устойчивостью статистической процедуры δ в классе K будем понимать независимость функции риска процедуры δ от распределения вектора X из класса K.

При анализе фондового рынка широко используется класс эллиптических распределений, задаваемых плотностью

$$f(x;\theta)=f(x;\mu,\Lambda)=|\Lambda|^{-1/2}\ g\{(x-\mu)'\Lambda^{-1}(x-\mu)\}, \tag{5}$$

где Λ – симметричная положительно определенная матрица, $g(x)\geq 0$, $\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}g(y'y)dy_1\ldots dy_N=1$. Класс эллиптических распределений содержит, в частности, многомерное нормальное распределение и многомерное распределение Стьюдента.

В классе эллиптических распределений можно выделить подкласс $K(\Lambda)$ распределений с фиксированной матрицей Λ. Известно (Anderson, 2003) выражение коэффициента корреляции Пирсона между компонентами вектора, имеющего эллиптическое распределение:

Журнал НЭА, №3 (35), 2017, с. 33–52

$\gamma^P_{i,j} = \lambda_{i,j} / \sqrt{\lambda_{i,i}\lambda_{j,j}}$, $\Lambda = (\lambda_{i,j})$. Таким образом, сетевые (графические) модели, порожденные сетями случайных величин (X, γ^P), $X \in K(\Lambda)$, совпадают. Ниже будет показано, что данный вывод справедлив для сетей случайных величин (X, γ^S), где γ^S – вероятность совпадения знаков, $X \in K(\Lambda)$.

Теорема 1. *Пусть вектор* $X = (X_1, \ldots, X_N)$ *имеет эллиптическое распределение* (5). *Тогда справедливо соотношение*

$$\gamma^S_{i,j} = \frac{1}{2} + \frac{1}{\pi}\arcsin\frac{\lambda_{i,j}}{\sqrt{\lambda_{i,i}\lambda_{j,j}}} = \frac{1}{2} + \frac{1}{\pi}\arcsin\gamma^P_{i,j}. \tag{6}$$

Доказательство. Известно, что $E(X) = \mu$ (Anderson, 2003). Без потери общности будем считать, что $\mu = 0$.
Введем матрицу $A = (a_{i,j}) = \Lambda^{-1}$. Плотность распределения двумерного случайного вектора (X_i, X_j) имеет вид

$$f(x_i, x_j) = |A^{-1}|^{-1/2}\ g(a_{i,i}x_i^2 + 2a_{i,j}x_i(x_j + a_{j,j}x_j^2).$$

Лемма 1. *Вероятности совпадения знаков* $\gamma^S_{i,j} = P(X_iX_j > 0)$ *полностью определяются матрицей* Λ *и не зависят от функции* g.

Доказательство. Рассмотрим матрицу

$$A_{i,j} = \begin{pmatrix} a_{i,i} & a_{i,j} \\ a_{i,j} & a_{j,j} \end{pmatrix}.$$

Так как матрица $A_{i,j}$ положительно определенная, то существует матрица $C = \begin{pmatrix} c_{i,i} & c_{i,j} \\ c_{j,i} & c_{j,j} \end{pmatrix}$, такая что $C'AC = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Положим $U = c_{i,i}X_i + c_{i,j}X_j$, $V = c_{j,i}X_i + c_{j,j}X_j$, тогда случайный вектор (U, V) будет иметь распределение с плотностью $f(u, v) = g(u^2 + v^2)$. Следовательно,

$$P(X_i > 0, X_j > 0) = P\left(\frac{c_{ii}}{c_{jj}}V < U < \frac{c_{ij}}{c_{jj}}V\right) + P\left(\frac{c_{ij}}{c_{jj}}V < U < \frac{c_{ii}}{c_{jj}}V\right) =$$

$$= \begin{cases} \left[\operatorname{arctg}(c_{ij}/c_{jj}) - \operatorname{arctg}(c_{ii}/c_{jj})\right]/2\pi & \text{при } c_{ij}/c_{jj} > c_{ii}/c_{jj}; \\ \left[\operatorname{arctg}(c_{ii}/c_{jj}) - \operatorname{arctg}(c_{ij}/c_{jj})\right]/2\pi & \text{при } c_{ij}/c_{jj} < c_{ii}/c_{jj}. \end{cases}$$

Таким образом, вероятности $P(X_i > 0, X_j > 0)$ полностью определяются матрицей Λ и не зависят от g. Аналогично доказывается, что вероятности $P(X_i < 0, X_j < 0)$ также не зависят от g. Следовательно, вероятности $P(X_iX_j > 0) = P(X_i > 0, X_j > 0) + P(X_i < 0, X_j < 0)$ полностью определяются матрицей Λ и не зависят от g. Лемма доказана.

Продолжим доказательство теоремы. Для двумерного нормального распределения (X_i, X_j) при $E(X_i) = E(X_j) = 0$ известно равенство (Крамер, 1972):

$$P(X_i > 0, X_j > 0) = \frac{1}{4} + \frac{1}{2\pi}\arcsin(\gamma^P_{i,j}).$$

Журнал НЭА, №3 (35), 2017, с. 33–52

В силу симметрии распределения отсюда следует, что

$$\gamma^S_{i,j} = P(X_i X_j > 0) = \frac{1}{2} + \frac{1}{\pi}\arcsin(\gamma^P_{i,j}).$$

По лемме 1 значение $\gamma^S_{i,j}$ полностью определяется матрицей Λ и, следовательно, равенство остается справедливым для всех распределений из класса $K(\Lambda)$. Теорема доказана.

5. Практические процедуры идентификации графа рынка

Для описания процедур идентификации графа рынка рассмотрим множество индивидуальных гипотез (4):

$$h_{i,j} : \gamma_{i,j} \le \gamma_0 \quad \text{против} \quad k_{i,j} : \gamma_{i,j} > \gamma_0 \;\; (i, j = 1,\ldots,N; \quad i \ne j).$$

Будем предполагать, что тесты гипотез известны и имеют вид

$$\varphi_{i,j}(x) = \begin{cases} 0, & T_{i,j}(x) \le c_{i,j}; \\ 1, & T_{i,j}(x) > c_{i,j}, \end{cases} \tag{7}$$

где $T_{i,j}(x)$ – статистики тестов, $i, j = 1,\ldots,N$, $i \ne j$. Предположим что статистики $T_{i,j}$ имеют функцию распределения F_{γ_0} при условии $\gamma_{i,j} = \gamma_0$, $i, j = 1,\ldots,N$. Для заданного уровня значимости $\alpha_{i,j}$ тестов критическое значение $c_{i,j}$ определяется из уравнения $F_{\gamma_0}(c_{i,j}) = 1 - \alpha_{i,j}$.

5.1. Одношаговая процедура

Пусть $\Phi(x)$ матрица вида

$$\Phi(x) = \begin{pmatrix} 1 & \varphi_{12}(x) & \ldots & \varphi_{1N}(x) \\ \varphi_{21}(x) & 1 & \ldots & \varphi_{2N}(x) \\ \ldots & \ldots & \ldots & \ldots \\ \varphi_{N1}(x) & \varphi_{N2}(x) & \ldots & 1 \end{pmatrix}, \tag{8}$$

тогда одношаговая статистическая процедура идентификации графа рынка проводится по формуле

$$\delta(x) = d_Q, \text{ если } \Phi(x) = Q. \tag{9}$$

Другими словами, ребро (i, j) включается в граф рынка, если значение статистики $T_{i,j}(x)$ превышает критическое. Для контроля FWER (вероятность хотя бы одного ложного включения ребра в граф рынка) на уровне α используется поправка Бонферрони (Hochberg, 1987) для уровней значимости индивидуальных тестов: $\alpha_{i,j} = \alpha / M$, $M = C_N^2$. Недостатком процедуры является большое число ложно невключенных ребер.

5.2. Многошаговая процедура Холма

Многошаговая процедура Холма уменьшает число ложно невключенных ребер с сохранением контроля FWER. Процедура Холма (Hochberg, Tamhane, 1987) состоит не более чем из $M = C_N^2$ шагов. На каждом шаге алгоритма одна из оставшихся гипотез отвергается или все оставшиеся гипотезы принимаются.

Журнал НЭА, №3 (35), 2017, с. 33–52

Процедура Холма состоит из следующих шагов:

1) если $\max_{i,j=1,\ldots,N} T_{i,j} \le c_1^H$, то принимаем все гипотезы $h_{i,j}$, $i,j=1,\ldots,N$; иначе если $\max_{i,j=1,\ldots,N} T_{i,j} = T_{i_1,j_1}$ — отвергаем гипотезу h_{i_1,j_1} и переходим на шаг 2;

...

K) пусть $I=\{(i_1,j_1),(i_2,j_2),\ldots,(i_{K-1},j_{K-1})\}$ — множество индексов ранее отвергнутых гипотез. Если $\max_{(i,j)\notin I} T_{i,j} \le c_K^H$, то принимаем все гипотезы $h_{i,j}$, $(i,j)\notin I$, иначе если $\max_{(i,j)\notin I} T_{i,j} = T_{i_K,j_K}$ — отвергаем гипотезу h_{i_K,j_K} и переходим на шаг (K+1);

...

M) пусть $I=\{(i_1,j_1),\ldots,(i_{M-1},j_{M-1})\}$ — множество индексов ранее отвергнутых гипотез и $(i_M,j_M)\notin I$. Если $T_{i_M,j_M} \le c_M^{Hg}$, то принимаем гипотезу h_{i_M,j_M}, иначе отвергаем гипотезу h_{i_M,j_M}.

Для контроля FWER на уровне α критические значения c_K^H процедуры Холма определяются из уравнений $F_{\gamma_0}(c_K^H)=1-\alpha/\left(M-K+1\right)$, $K=1,\ldots,M$.

5.3. Многошаговая процедура Хочберга

Другой практически применяемой многошаговой процедурой множественной проверки гипотез является процедура Хочберга (Hochberg, 1988). Она также состоит не более чем из $M=C_N^2$ шагов. На каждом шаге одна гипотеза принимается или все оставшиеся гипотезы отвергаются:

1) если $T_{i_1,j_1} = \min_{i,j=1,\ldots,N} T_{i,j} > c_1^{Hg}$, то отвергаем все гипотезы $h_{i,j}$, иначе принимаем гипотезу h_{i_1,j_1} и переходим на шаг 2;

...

K) пусть $I=\{(i_1,j_1),\ldots,(i_{K-1},j_{K-1})\}$ — множество ранее принятых гипотез. Если $T_{i_K,j_K} = \min_{i,j=1,\ldots,N;(i,j)\notin I} T_i(x) > c_K^{Hg}$, то отвергаем все гипотезы $h_{i,j}$, $(i,j)\notin I$, иначе принимаем гипотезу h_{i_K,j_K} и переходим на шаг K+1;

...

M) пусть $I=\{(i_1,j_1),\ldots,(i_{M-1},j_{M-1})\}$ — множество ранее принятых гипотез и $(i_M,j_M)\notin I$. Если $T_{i_M,j_M} > c_M^{Hg}$, то отвергаем гипотезу h_{i_M,j_M}, иначе принимаем гипотезу h_{i_M,j_M}.

Критические значения c_K^{Hg} процедуры Хочберга определяются из уравнений

$$F_{\gamma_0}(c_K^{Hg})=1-\alpha/K, \quad K=1,\ldots,M.$$

Если статистики $T_{i,j}(x)$ независимы или положительно зависимы, такой выбор критических значений c_K^{Hg} обеспечивает контроль FWER на уровне α (Sarkar, Chung-Kuei, 1997). В общем случае контроль

Журнал НЭА, №3 (35), 2017, с. 33–52

ошибки не гарантирован. Известно, что для процедуры Хочберга число ложно невключенных ребер, как правило, меньше числа ложно невключенных ребер процедурой Холма (Hochberg, 1988).

5.4. Тесты проверки индивидуальных гипотез

Для сети корреляции Пирсона индивидуальные гипотезы (4) имеют вид: $h_{i,j}:\gamma^P_{i,j}\le\gamma^P_0$. Наиболее распространенные тесты проверки индивидуальных гипотез основаны на статистиках вида (Anderson, 2003):

$$T^P_{i,j}=\sqrt{n}\left(0,5\ln\left(\frac{1+r_{i,j}}{1-r_{i,j}}\right)-0,5\ln\left(\frac{1+\gamma^P_0}{1-\gamma^P_0}\right)\right),$$

где

$$r_{i,j}=\sum_t x_i(t)x_j(t)\Big/\sqrt{\sum_t(x_i(t))^2\sum_t(x_j(t))^2}. \qquad (10)$$

В соответствии с (Anderson, 2003) асимптотическое распределение статистик $T^P_{i,j}$ при условии $\gamma^P_{i,j}=\gamma^P_0$ является стандартным нормальным, поэтому константы для одношаговой статистической процедуры, а также процедур Холма и Хочберга вычисляются из уравнений

$$F(c_{i,j})=1-\alpha_{i,j},\ F(c^H_K)=1-\frac{\alpha}{M-K+1},\ F(c^{Hg}_K)=1-\alpha/K,$$

где $F(x)$ – функция стандартного нормального распределения $N(0,\ 1)$.

Для сети вероятностей совпадения знаков индивидуальные гипотезы имеют вид: $h_{i,j}:\ p^{i,j}\le p_0$. Построим тесты проверки индивидуальных гипотез. Определим индикатор

$$I_{i,j}(t)=\begin{cases}1, & x_i(t)x_j(t)\ge 0;\\ 0, & x_i(t)x_j(t)<0\end{cases}$$

и статистику

$$T^{sg}_{i,j}=\frac{1}{n}\sum_{t=1}^{n}I_{i,j}(t). \qquad (11)$$

Статистика $T^{Sg}_{i,j}$ при $p^{i,j}=p_0$ имеет биномиальное распределение с параметрами n,p_0. Следовательно, константы для одношаговой статистической процедуры, а также процедур Холма и Хочберга определяются как наименьшая из констант, удовлетворяющая соответствующему неравенству:

$$F(c_{i,j};n,p_0)\ge 1-\alpha_{i,j},$$

$$F(c^H_K;n,p_0)\ge 1-\alpha/\left(M-K+1\right),\quad F(c^{Hg}_K;n,p_0)\ge 1-\alpha/K,$$

где $F(c_{i,j};n,p_0)$ – функция биномиального распределения с параметрами n,p_0.

6. Устойчивость процедур идентификации сетевых структур

6.1. Сетевая модель вероятностей совпадения знаков

Устойчивость статистических процедур идентификации сетевых структур будем исследовать в классе $K(\Lambda)$. Для этого возьмем

Журнал НЭА, №3 (35), 2017, с. 33–52

сеть вероятностей совпадения знаков. В этом разделе показана устойчивость процедур идентификации, основанных на статистиках $T_{i,j}^{Sg}$, в сетевых моделях вероятностей совпадения знаков. Рассмотрим задачу идентфикации графа рынка.

Теорема 2. *Для любой функции потерь $w(S,Q)$ функция риска $R(S,\theta;\delta)$ процедур Бонферрони, Холма и Хочберга идентификации графа рынка в сетевой модели вероятностей совпадения знаков не зависит от распределения вектора X из класса $K(\Lambda)$.*

Д о к а з а т е л ь с т в о. Доказательство разобъем на ряд вспомогательных утверждений, имеющих самостоятельное значение.

Лемма 2. *Пусть случайный вектор $(X_1,\dots,X_N)$ имеет эллиптическое распределение с плотностью $f(x;0,\Lambda)=|\Lambda|^{-1/2}\ g(x'\Lambda x)$. Тогда вероятности*

$$p(i_1,\dots,i_N):=P(i_1X_1>0,\ \dots,\ i_NX_N>0),\quad i_k\in\{-1,\ 1\} \tag{12}$$

полностью определяются матрицей Λ и не зависят от функции g для любых $i_k\in\{-1,\ 1\},\quad k=1,\dots,N$.

Д о к а з а т е л ь с т в о. Имеем

$$P(i_1X_1>0,\dots,i_NX_N>0)=\int_{i_kx_k>0}|\Lambda|^{-1/2}\ g(x'\Lambda x)dx_1\dots dx_N. \tag{13}$$

Матрица Λ положительно определенная, поэтому существует матрица C : $C'\Lambda C=I$. Положим $y=C^{-1}x$. Тогда

$$\int_{i_kx_k>0,k=1,\dots,N}|\Lambda|^{-1/2}\ g(x'\Lambda x)dx_1\dots dx_N=\int_D g(y'y)dy_1\dots dy_N, \tag{14}$$

где область D имеет вид

$$0<i_k(c_{k,1}y_1+\ldots+c_{k,N}Y_N)<\infty,\ k=1,\dots,N. \tag{15}$$

Вектор y может быть записан в полярных координатах:

$$\begin{aligned} &y_1=r\sin(\theta_1),\\ &y_2=r\cos(\theta_1)\sin(\theta_2),\\ &\dots\\ &y_{N-1}=r\cos(\theta_1)\cos(\theta_2)\mathrm{K}\cos(\theta_{N-2})\sin(\theta_{N-1}),\\ &y_N=r\cos(\theta_1)\cos(\theta_2)\mathrm{K}\cos(\theta_{N-2})\cos(\theta_{N-1}), \end{aligned} \tag{16}$$

где $-\pi/2\le\theta_i\le\pi/2,\ i=1,\dots,N-2;\ -\pi\le\theta_{N-1}\le\pi,\ 0\le r\le\infty$. Якобиан преобразования (16) равен $r^{N-1}\cos^{N-2}(\theta_1)\dots\cos(\theta_{N-2})$. В полярных координатах область (15) преобразуется в область $R_+^1\times D'$, где D' имеет вид $(k=1,\dots,N)$:

$$0<i_k(c_{k,1}\sin(\theta_1)+\ldots+c_{k,N}\cos(\theta_1)\cos(\theta_2)\dots\cos(\theta_{N-2})\cos(\theta_{N-1}))<\infty. \tag{17}$$

Тогда вероятность $p(i_1,i_2,\dots,i_N)$ может быть записана в виде:

$$\int_{D'}\int_0^\infty r^{N-1}\cos^{N-2}(\theta_1)\dots\cos(\theta_{N-2})g(r^2)drd\theta_1\dots d\theta_{N-1}=$$
$$=\int_{D'}\cos^{N-2}(\theta_1)\dots\cos(\theta_{N-2})d\theta_1\dots d\theta_{N-1}\int_0^\infty r^{N-1}g(r^2)dr.$$

Известно (Anderson, 2003), что $\int_0^\infty r^{N-1}g(r^2)dr=1/C(N)$, где

$$C(N)=\int_{-\pi/2}^{\pi/2}\dots\int_{-\pi/2}^{\pi/2}\int_{-\pi}^{\pi}\cos^{N-2}(\theta_1)\dots\cos(\theta_{N-2})d\theta_1\dots d\theta_{N-1}.$$

Журнал НЭА, № 3 (35), 2017, с. 33–52

Область D' определяется матрицей Λ и не зависит от функции g. Следовательно, вероятности $p(i_1,\ldots,i_N)$ зависят от матрицы Λ и не зависят от функции g. Лемма 2 доказана.

Лемма 3. *Пусть случайный вектор* $(X_1,\ldots,X_N)$ *имеет эллиптическое распределение. Тогда совместное распределение статистик* $T_{i,j}^{Sg}$ *(*$i,j=1,\ldots,N;\ i\neq j$*) полностью определеяется матрицей* Λ *и не зависит от функции* g.

Д о к а з а т е л ь с т в о. Статистику $T_{i,j}^{Sg}$ можно записать как

$$T_{i,j}^{Sg}=\frac{1}{2}+\frac{1}{2n}\sum_{t=1}^{n}\operatorname{sign}(X_i(t))\operatorname{sign}(X_j(t)).$$

Как следует из леммы 2, совместное распределение случайного вектора $\operatorname{sign}(X)=(\operatorname{sign}(X_1),\ldots,\operatorname{sign}(X_N))$ определяется матрицей Λ и не зависит от функции g. Случайные вектора $\operatorname{sign}(X(t))$, $t=1,\ldots,n$ независимы и одинаково распределены. Поэтому совместное распределение случайных величин $\operatorname{sign}(X_i(t))$, $i=1,\ldots,N$; $t=1,\ldots,n$ полностью определяется матрицей Λ и не зависит от функции g. Следовательно, совместное распределение статистик $T_{i,j}^{Sg}$, $i,j=1,\ldots,N;\ i<j$ полностью определяется матрицей Λ и не зависит от функции g. Лемма доказана.

Для завершения доказательства теоремы рассмотрим функцию риска

$$R(S,\theta;\delta)=\sum_{Q\in G}w(S,Q)P_\theta(\delta(x)=d_Q),\quad \theta\in\Omega_S.$$

Процедуры Бонферрони, Холма и Хочберга построены на основе статистик $T_{i,j}^{Sg}$. В соответствии с леммой 3 для каждой из этих процедур вероятности $P_\theta(\delta(x)=d_Q)$ не зависят от распределения вектора X из класса $K(\Lambda)$. Теорема 2 доказана.

Из теоремы 2 следует, что FWER (вероятность хотя бы одного ложного включения ребра) для процедур Бонферрони, Холма и Хочберга не зависит от распределения вектора X из класса $K(\Lambda)$. Этот же результат справедлив и для естественной при идентификации сетевых структур аддитивной функции потерь.

Рассмотрим задачу идентификации максимального остовного дерева в сетевой модели вероятностей совпадения знаков.

Теорема 3. *Для любой функции потерь функция риска процедуры Краскала идентификации MST в сетевой модели вероятностей совпадения знаков не зависит от распределения вектора* X *в классе* $K(\Lambda)$.

Д о к а з а т е л ь с т в о. Для применения процедуры Краскала идентификации MST в сетевой модели вероятностей совпадения знаков на первом шаге в качестве оценки $\gamma_{i,j}^{S}$ будем использовать статистики $T_{i,j}^{Sg}$. На втором – упорядочиваем значения статистик $T_{i,j}^{Sg}$ по убыванию; на третьем – применяем алгоритм Краскала для построения максимального остовного дерева.

Журнал НЭА, №3 (35), 2017, с. 33–52

Пусть M – класс всех остовных деревьев полного графа на N вершинах. Для $S,Q \in M$ обозначим через $w(S,Q)$ потери от выбора остовного дерева Q при истинном остовном дереве S. Положим $w(S,S)=0$. Функция риска имеет вид $R(S,\theta;\delta)=\sum_{Q\in M} w(S,Q)P_\theta(\delta(x)=d_Q)$, $\theta\in\Omega_S$, где d_Q – решение о выборе остовного дерева Q, Ω_S – область изменения параметра θ, в которой истинным является остовное дерево S. По лемме 3 $P_\theta(\delta(x)=d_Q)$ не зависит от распределения вектора X из класса $K(\Lambda)$. Теорема доказана.

6.2. Сетевая модель корреляций Пирсона

Устойчивость статистических процедур идентификации сетевых структур будем исследовать в классе $K(\Lambda)$. Вычислительные эксперименты доказали (Bautin et al., 2014), что одношаговые и многошаговые статистические процедуры идентификации графа рынка и процедура Краскала идентификации MST, построенные на статистиках корреляции Пирсона $T^P_{i,j}$, оказываются неустойчивыми в классе $K(\Lambda)$. В этом разделе мы покажем, что в классе $K(\Lambda)$ максимальные остовные деревья в сетевой модели вероятностей совпадения знаков и в сетевой модели корреляций Пирсона совпадают.

Теорема 4. *Пусть случайный вектор X имеет эллиптическое распределение из класса $K(\Lambda)$. Тогда максимальное остовное дерево в сетевой модели корреляций Пирсона совпадает с максимальным остовным деревом в сетевой модели вероятностей совпадения знаков.*

Доказательство. По теореме 1 для $X\in K(\Lambda)$ справедливо соотношение

$$\gamma^S_{i,j}=\frac{1}{2}+\frac{1}{\pi}\arcsin\gamma^P_{i,j}.$$

Так как функция $\arcsin(x)$ является монотонной, то неравенство $\gamma^P_{i,j}\le\gamma^P_{k,l}$ эквивалентно неравенству $\gamma^S_{i,j}\le\gamma^S_{k,l}$. Поэтому упорядочение ребер в сетевой модели корреляций Пирсона эквивалентно упорядочению ребер в сетевой модели вероятностей совпадения знаков. Следовательно, максимальные остовные деревья, построенные по алгоритму Краскала, в этих сетевых моделях совпадают. Теорема доказана.

Таким образом, процедура Краскала идентификации MST, основанная на статистиках $T^{Sq}_{i,j}$, устойчиво идентифицирует MST как в сетевой модели вероятностей совпадения знаков, так и в сетевой модели корреляции Пирсона.

Исследуем возможность устойчивой идентификации графа рынка в сетевой модели корреляций Пирсона.

Теорема 5. *Пусть случайный вектор X имеет эллиптическое распределение из класса $K(\Lambda)$. Тогда граф рынка с порогом γ^P_0 в сетевой модели корреляций Пирсона совпадает с графом рынка с порогом $\gamma^S_0=0,5+\arcsin(\gamma^P_0)/\pi$ в сетевой модели вероятностей совпадения знаков.*

Журнал НЭА, №3 (35), 2017, с. 33–52

Д о к а з а т е л ь с т в о. Построение графа рынка основано на проведении ребра между вершинами i и j, если $\gamma_{i,j} > \gamma_0$. Для сетевой модели корреляций Пирсона ребро между вершинами i и j добавляется в граф рынка, если $\gamma^P_{i,j} > \gamma^P_0$, а для сетевой модели вероятностей совпадения знаков ребро между вершинами i и j добавляется в граф рынка, если $\gamma^S_{i,j} > \gamma^S_0$. Неравенство $\gamma^P_{i,j} > \gamma^P_0$ эквивалентно неравенству $\gamma^S_{i,j} > \gamma^S_0$ при $\gamma^S_0 = 0,5 + \arcsin(\gamma^P_0)/\pi$. Следовательно, граф рынка в сетевой модели корреляций Пирсона и граф рынка в сетевой модели вероятностей совпадения знаков одинаковы при указанном выборе порогов. Теорема доказана.

Для построения устойчивой процедуры идентификации графа рынка с порогом γ^P_0 в сетевой модели корреляций Пирсона положим $\gamma^S_0 = 0,5 + \arcsin(\gamma^P_0)/\pi$ и применим устойчивую процедуру идентификации графа рынка с порогом γ^S_0 в сетевой модели вероятностей совпадения знаков. Итоговая процедура будет устойчиво идентифицировать граф рынка с порогом γ^P_0 в сетевой модели корреляций Пирсона.

7. Заключение

Основным результатом настоящей работы является новая методология идентификации сетевых структур в сетевых моделях фондовых рынков. Предлагаемая методология приводит к устойчивым процедурам идентификации в широком классе распределений доходностей. В основу методологии положена сетевая модель вероятностей совпадения знаков доходностей. Эта модель представляет самостоятельный интерес и имеет преимущества:

1) вероятность совпадения знаков допускает простую интерпретацию;
2) вероятность совпадения знаков имеет естественное обобщение как мера зависимости на любое число случайных величин;
3) процедуры идентификации сетевых структур, построенные в сетевой модели вероятностей совпадения знаков, обладают свойством устойчивости в практически используемом классе распределений;
4) процедуры идентификации сетевых структур, построенные в сетевой модели вероятностей совпадения знаков, могут быть использованы для построения устойчивых процедур идентификации сетевых структур в других сетевых моделях фондовых рынков.

Таким образом, предложенные в настоящей статье процедуры идентификации сетевых структур, основанные на вероятностях совпадения знаков, могут быть рекомендованы для практического применения.

Журнал НЭА, №3 (35), 2017, с. 33–52

ЛИТЕРАТУРА

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.

Крамер Г. (1972). Математические методы статистики. М.: Наука.

Фантаццини Д. (2011). Моделирование многомерных распределений с использованием копула-функций. I // *Прикладная эконометрика*. Т. 22 (2). P. 98–134.

Anderson T.W. (2003). An Introduction to Multivariate Statistical Analysis. New York: Wiley-Interscience.

Bautin G.A., Koldanov A.P., Pardalos P.M. (2014). Robustness of Sign Correlation in Market Network Analysis // *Springer Optimization and Its Applications*. Vol. 100. P. 25–33.

Bautin G., Kalyagin V.A., Koldanov A.P. (2013a). Comparative Analysis of Two Similarity Measures for the Market Graph Construction // *Springer Proceedings in Mathematics and Statistics*. Vol. 59. P. 29–41.

Bautin G.A., Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M. (2013b). Simple Measure of Similarity for the Market Graph Construction // *Computational Management Science*. Vol. 10. P. 105–124.

Boginski V., Butenko S., Pardalos P.M. (2005). Statistical Analysis of Financial Networks // *Journal Computational Statistics and Data Analysis*. Vol. 48 (2). P. 431–443.

Boginski V., Butenko S., Pardalos P.M. (2006). Mining Market Data: A Network Approach // *Journal Computers and Operations Research*. Vol. 33 (11). P. 3171–3184.

Boginsky V., Butenko S., Pardalos P.M. (2003). On Structural Properties of the Market Graph. In: Nagurney A. (ed.) "*Innovations in Financial and Economic Networks*". Northampton: Edward Elgar Publishing Inc. P. 29–45.

Bonanno G., Lillo F., Mantegna R.N. (2001). High-Frequency Cross-Correlation in a Set of Stocks // *Quantitative Finance*. Vol. 1. P. 96–104.

Coronnello C., Tumminello M., Lillo F., Micciche S., Mantegna R.N. (2005). Sector Identification in a Set of Stock Return Time Series Traded at the London Stock Exchange // *Acta Physica Polonica B*. Vol. 36. P. 2653–2679.

Dudoit S., Shaffer J.P., Boldrick J.C. (2003). Multiple Hypothesis Testing in Microarray Experiments // *Statistical Science*. Vol. 18. No. 1. P. 71–103.

Garas A., Argyrakis P. (2007). Correlation Study of the Athens Stock Exchange // *Physica A*. Vol. 380. P. 399–410.

Hochberg Y., Tamhane A.C. (1987). Multiple Comparison Procedures. Hoboken: John Wiley and Sons, Inc.

Hochberg Y.A. (1988). Sharper Bonferroni Procedure for Multiple Tests of Significance // *Biometrika*. Vol. 75 (4). P. 800–802.

Huang W.-Q., Zhuang X.-T., Yao S. (2009). A Network Analysis of the Chinese Stock Market // *Physica A*. Vol. 388. P. 2956–2964.

Журнал НЭА, №3 (35), 2017, с. 33–52

Jung W.-S., Chae S., Yang J.-S., Moon H.-T. (2006). Characteristics of the Korean Stock Market Correlations // *Physica A*. Vol. 361. P. 263–271.

Kalyagin V.A., Koldanov A.P., Koldanov P.A. (2017). Robust Identification in Random Variables Networks // *Journal of Statistical Planning and Inference*. Vol. 181. No. Feb. P. 30–40.

Koldanov A.P., Kalyagin V.A., Koldanov P., Pardalos P.M. (2013). Statistical Procedures for the Market Graph Construction // *Computational Statistics and Data Analysis*. Vol. 68. P. 17–29.

Koldanov P., Kalyagin V.A., Koldanov A.P., Zamaraev V.A. (2014). Market Graph and Markowitz Model. In: "*Optimization in Science and Engineering (In Honor of the 60th Birthday of Panos M. Pardalos)*". New York: Springer Science, Business Media. P. 301–313.

Lehmann E.L., Romano J.P. (2005). Testing Statistical Hypotheses. New York: Springer.

Mantegna R.N. (1999). Hierarchical Structure in Financial Market // *European Physical Journal. Series B*. Vol. 11. P. 193–197.

Markowitz H.M. (1952). Portfolio Selection // *Journal of Finance*. Vol. 7 (1). P. 77–91.

Sarkar S.K., Chung-Kuei C. (1997). The Simes Method for Multiple Hypothesis Testing with Positively Dependent Test Statistics // *Journal of the American Statistical Association*. Vol. 92. No. 440. P. 1601–1608.

Tabak B.M., Thiago R.S., Cajueiro D.O. (2010). Topological Properties of Stock Market Networks: The Case of Brazil // *Physica A*. Vol. 389. P. 3240–3249.

Tee O.C. (2005). The Impact of Globalisation on the Formulation and Implementation of Monetary Policy in Singapore. In: "*Globalisation and Monetary Policy in Emerging Markets*". BIS Papers Chapters. Vol. 23. Bank for International Settlements. P. 263–268.

Tumminello M., Aste T., Matteo T.D., Mantegna R.N. (2005). A Tool for Filtering Information in Complex Systems // *Proceedings of the National Academy of Sciences*. Vol. 102 (30). P. 10421–10426.

Tumminello M., Lillo F., Mantegna R.N. (2010). Correlation, Hierarchies and Networks in Financial Markets // *Journal of Economic Behavior Organization*. Vol. 75. P. 40–58.

Wald A. (1950). Statistical Decision Function. New York: John Wiley and Sons.

Wang G.J., Chi X., HanF., Sun B. (2012). Similarity Measure and Topology Evolution of Foreign Exchange Markets Using Dynamic Time Warping Method: Evidence from Minimal Spanning Tree // *Physica A*: *Statistical Mechanics and its Applications*. Vol. 391 (16). P. 4136–4146.

Поступила в редакцию 28 января 2017 года

REFERENCES (with English translation or transliteration)

Anderson T.W. (2003). An Introduction to Multivariate Statistical Analysis. New York: Wiley-Interscience.

Bautin G., Kalyagin V.A., Koldanov A.P. (2013a). Comparative Analysis of Two Similarity Measures for the Market Graph Construction. *Springer Proceedings in Mathematics and Statistics*, 59, 29–41.

Bautin G.A., Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M. (2013b). Simple Measure of Similarity for the Market Graph Construction. *Computational Management Science*, 10, 105–124.

Bautin G.A., Koldanov A.P., Pardalos P.M. (2014). Robustness of Sign Correlation in Market Network Analysis. *Springer Optimization and Its Applications*, 100, 25–33.

Boginski V., Butenko S., Pardalos P.M. (2005). Statistical Analysis of Financial Networks. *Journal Computational Statistics and Data Analysis*, 48 (2), 431–443.

Boginski V., Butenko S., Pardalos P.M. (2006). Mining Market Data: A Network Approach. *Journal Computers and Operations Research*, 33 (11), 3171–3184.

Boginsky V., Butenko S., Pardalos P.M. (2003). On Structural Properties of the Market Graph. In: Nagurney A. (ed.) "*Innovations in Financial and Economic Networks*". Northampton: Edward Elgar Publishing Inc, 29–45.

Bonanno G., Lillo F., Mantegna R.N. (2001). High-Frequency Cross-Correlation in a Set of Stocks. *Quantitative Finance*, 1, 96–104.

Coronnello C., Tumminello M., Lillo F., Miccichи S., Mantegna R.N. (2005). Sector Identification in a Set of Stock Return Time Series Traded at the London Stock Exchange. *Acta Physica Polonica B*, 36, 2653–2679.

Dudoit S., Shaffer J.P., Boldrick J.C. (2003). Multiple Hypothesis Testing in Microarray Experiments. *Statistical Science*, 18, 1, 71–103.

Fantazzini D. (2011). Analysis of Multidimensional Probability Distributions with Copula Functions. I. *Applied Econometrics*, 22 (2), 98–134.

Garas A., Argyrakis P. (2007). Correlation Study of the Athens Stock Exchange. *Physica A*, 380, 399–410.

Hochberg Y., Tamhane A.C. (1987). Multiple Comparison Procedures. Hoboken: John Wiley and Sons, Inc.

Hochberg Y.A. (1988). Sharper Bonferroni Procedure for Multiple Tests of Significance. *Biometrika*, 75 (4), 800–802.

Huang W.-Q., Zhuang X.-T., Yao S. (2009). A Network Analysis of the Chinese Stock Market. *Physica A*, 388, 2956–2964.

Jung W.-S., Chae S., Yang J.-S., Moon H.-T. (2006). Characteristics of the Korean Stock Market Correlations. *Physica A*, 361, 263–271.

Kalyagin V.A., Koldanov A.P., Koldanov P.A. (2017). Robust Identification in Random Variables Networks. *Journal of Statistical Planning and Inference*, 181, Feb., 30–40.

Koldanov A.P., Kalyagin V.A., Koldanov P., Pardalos P.M. (2013). Statistical Procedures for the Market Graph Construction. *Computational Statistics and Data Analysis*, 68, 17–29.

Koldanov P., Kalyagin V.A., Koldanov A.P., Zamaraev V.A. (2014). Market Graph and Markowitz Model. In: "*Optimization in Science and Engineering (In Honor of the 60th Birthday of Panos M. Pardalos)*". N.Y.: Springer Science, Business Media, 301–313.

Kramer G. (1972). Mathematical Methods of Statistics. Moscow: Nauka.

Lehmann E.L., Romano J.P. (2005). Testing Statistical Hypotheses. New York: Springer.

Mantegna R.N. (1999). Hierarchical Structure in Financial Market. *European Physical Journal. Series B*, 11, 193–197.

Markowitz H.M. (1952). Portfolio Selection. *Journal of Finance*, 7 (1), 77–91.

Sarkar S.K., Chung-Kuei C. (1997). The Simes Method for Multiple Hypothesis Testing With Positively Dependent Test Statistics. *Journal of the American Statistical Association*, 92, 440, 1601–1608.

Tabak B.M., Thiago R.S., Cajueiro D.O. (2010). Topological Properties of Stock Market Networks: The Case of Brazil. *Physica A*, 389, 3240–3249.

Tee O.C. (2005). The Impact of Globalisation on the Formulation and Implementation of Monetary Policy in Singapore. In: "*Globalisation and Monetary Policy in Emerging Markets*". BIS Papers Chapters. Bank for International Settlements, 23, 263–268.

Tumminello M., Aste T., Matteo T.D., Mantegna R.N. (2005). A Tool for Filtering Information in Complex Systems. *Proceedings of the National Academy of Sciences*, 102 (30), 10421–10426.

Tumminello M., Lillo F., Mantegna R.N. (2010). Correlation, Hierarchies and Networks in Financial Markets. *Journal of Economic Behavior Organization*, 75, 40–58.

Vizgunov A.N., Goldengorin B.I., Zamaraev V.A., Kalyagin 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.

Wald A. (1950). Statistical Decision Function. New York: John Wiley and Sons.

Wang G.J., Chi X., HanF., Sun B. (2012). Similarity Measure and Topology Evolution of Foreign Exchange Markets Using Dynamic Time Warping Method: Evidence from Minimal Spanning Tree. *Physica A*: *Statistical Mechanics and its Applications*, 391 (16), 4136–4146.

Received 28.01.2017

Журнал НЭА, №3 (35), 2017, с. 33–52

V.A. Kalyagin,
National Research University Higher School of Economics, Nijni Novgorod, Russia

A.P. Koldanov
National Research University Higher School of Economics, Nijni Novgorod, Russia

P.A. Koldanov
National Research University Higher School of Economics, Nijni Novgorod, Russia

P.M. Pardalos
University of Florida, USA

Statistical Procedures for Stock Markets Network Structures Identification

Abstract. Network (graphical) model of stock market is a complete weighted graph. Nodes of the graph corresponds to the stocks and weights of edges are given by some measure of dependence between characteristics of the stocks. The most common characteristic of stocks is their return. In the analysis of network (graphical) models of returns of primary interest are network structures (subgraphs of a complete graph), containing key information about the considered network. Popular network structures are the minimum spanning tree, planar maximally filtered graph, market graph, a cliques and an independent set of the market graph. The problem of identification of network structure is to define the structure from observations. An important characteristic of identification statistical procedure is its uncertainty related with finite sample size. Significant role in this play the joint distribution of returns and the choice of measures of dependence between them. The most common measure of dependence is Pearson's correlation. A wide class of joint distributions of stock returns is represented by elliptical models. However, the procedures based on Pearson correlations are non robust when the joint distribution of returns is deviated from the normal in the class of elliptical distributions. The aim of this work is to present a general approach to construction of robust (distribution free) statistical procedures of identification of network structures. It is proposed to use the probability of sign coincidence of stock returns as a measure of dependence. It is shown that the single-step and stepwise standard procedure of identification of network structures based on the probability of sign coincidence are robust in the class of elliptical distributions. It allows to recommend these procedures for practical applications.

Keywords: *financial market, network (graphical) model, network structure, Pearson correlation, probability of sign coincidence.*

JEL classification: C02.