

На основе аппарата байесовских сетей получена графическая информация взаимодействия курсов мировых валют. Описаны алгоритмы определения скелета байесовской сети, а также определения частичной ориентации ребер графа.

© Н.А. Гупал, С.С. Ржепецкий,
2010

УДК 519.681

Н.А. ГУПАЛ, С.С. РЖЕПЕЦКИЙ

ПРИМЕНЕНИЕ АППАРАТА БАЙЕСОВСКИХ СЕТЕЙ ДЛЯ АНАЛИЗА РЫНКА ВАЛЮТ

Введение. В работе рассматривается применение аппарата байесовских сетей для анализа рынка валют и проверяется гипотеза об идентичности динамических связей на международных фондовом и валютном рынках. Наиболее общим аппаратом для извлечения структур описания объектов из данных является аппарат байесовских сетей. За последние 10 лет выпущено более двух десятков монографий по байесовским сетям.

Байесовская сеть состоит из направленного ориентированного ациклического графа и заданного распределения условных вероятностей в каждой вершине. Байесовские сети применяются как для задач извлечения внутренней структуры моделируемой системы, так и для последующего моделирования самой системы и решения задач предсказания и распознавания [1].

В работе [2] исследована схема интеграции 9 основных мировых фондовых рынков, выведена общая схема взаимодействия фондовых рынков, которая дает информацию о динамике передачи информации об изменении цен на акции между рынками, т. е. о взаимном влиянии рынков. Рассматривались фондовые рынки Австралии, Японии, Гонконга, Германии, Великобритании, Франции, Швейцарии, США и Канады. Цель исследования – изучение связей между рынками и общей схемы обмена финансовой информацией на рынке. Основным инструментом в [2] рассматривался аппарат байесовских сетей, а также математическое и экономическое обоснование их применимости для описанной задачи. Основным результатом [2] показан на диаграмме, которая содержит итоговый граф полученной байесовской сети (рис. 1).

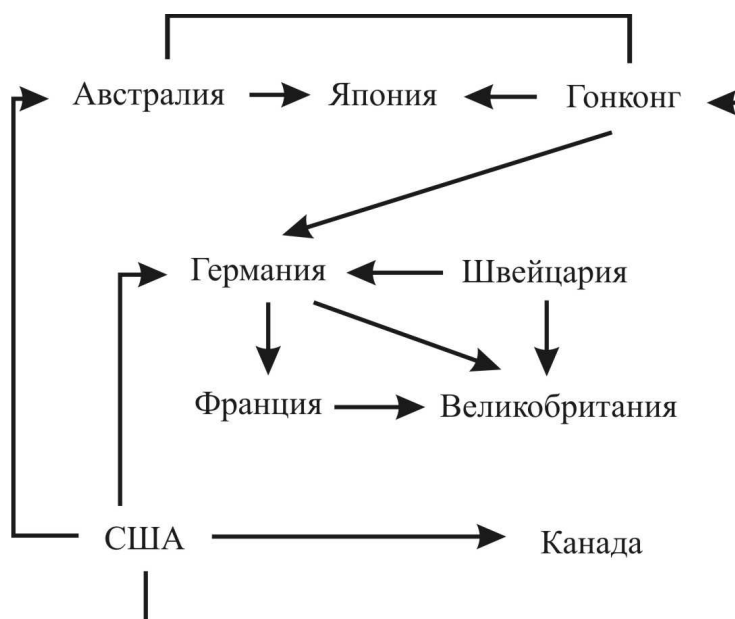


РИС. 1. Граф модели байесовской сети взаимодействия фондовых рынков

Естественной и логичной является гипотеза, о том, что граф байесовской сети динамического взаимодействия цен на мировые валюты не должен существенно противоречить графу на рис. 1. Эта гипотеза основана на том, что и цена на валюту, и состояние фондового рынка того или иного государства определяются более глобальными, общими для обоих рынков макроэкономическими показателями этого государства. Непосредственная проверка данной гипотезы позволит получить информацию о взаимодействии цен на рынке валют и продемонстрировать возможность и эффективность применения байесовских сетей для анализа этого рынка.

Исследовались курсы валют относительно некоторой базовой валюты. Исследование проводилось для набора валют аналогичного набора фондовых бирж в [2]. Всего рассмотрено взаимодействие 7 валют (AUD – австралийский доллар, NZD – новозеландский доллар, USD – доллар США, CHF – швейцарский франк, GBP – британский фунт, JPY – японская йена, CAD – канадский доллар) по отношению к двум базовым валютам. Выборка данных состояла из дневных курсов валют (по цене закрытия) на международной валютной бирже Forex за период с 1 января 2001 г. по 1 января 2008 года [3]. Изначально данные представлялись строками значений отношения курсов двух валютных пар: EURUSD 1999.10.01, 1.06790, 1.06840, 1.06790, 1.06810, 8950.

Каждая строка содержала следующую информацию: идентификатор валютной пары, дата, цена открытия торгов, максимальная цена торгов за день, минимальная цена торгов за день, цена закрытия торгов, количество сделок. Анализ проводился на основании некой «базовой» валюты. Например, для «базовой» валюты EUR были взяты значения валютных пар EURUSD, EURJPY, EURGBP и т. д. Таким образом, использование «базовой» валюты позволило сравнить курсы исследуемых валют.

Для вывода байесовской модели из данных применялся вариант алгоритма РС [4], модифицированный согласно идеям, изложенным в [5, 6]. В качестве статистического критерия тестирования на условную независимость при работе алгоритма использовался известный критерий Фишера. Графы выведенных моделей байесовских сетей показаны на рис. 2.

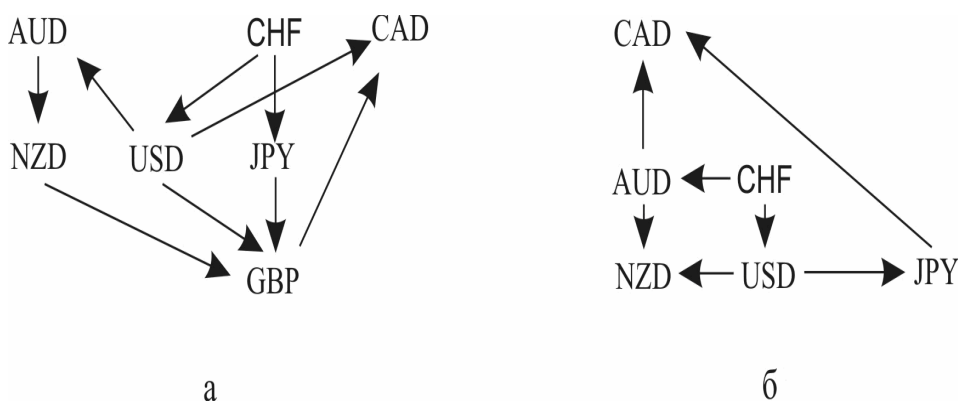


РИС. 2. Граф модели байесовской сети взаимодействия фондовых рынков:
 а – базовая валюта – EUR; б – базовая валюта – GBP

Основные результаты о взаимодействии фондовых рынков, полученные в [2]:

- фондовый рынок Канады полностью зависим от рынка США;
- Великобритания не порождает никакой новой информации на мировом рынке;
- Швейцария является основным источником влияния на рынки Европы.

Результаты для фондового рынка были подтверждены и для рынка валют (см. рис. 2), что свидетельствует в пользу гипотезы о схожести интеграционных процессов на рынках. При этом рынок валют имеет и ряд отличий. Основное отличие состоит в том, что если для фондового рынка на графе модели байесовской сети (см. рис. 1) можно четко выделить три геополитических региона, обменивающихся между собой информацией через «буферные» биржи

(Гонконг, Австралия, Германия), то в случае рынка валют мы имеем два центра (США и Швейцария), которые влияют сразу на весь мировой рынок.

Таким образом, с помощью байесовских сетей была получена информация о динамике взаимодействия курсов мировых валют. Показана применимость моделей байесовских сетей для анализа рынка валют.

Кратко опишем основную идею известного РС-алгоритма для нахождения скелета байесовских сетей. Графовые модели для вероятностных моделей являются в последнее время популярным инструментом, позволяющим анализировать и учитывать условные независимости между случайными переменными. Основные элементы графовых моделей – вершины, представляющие собой переменные, и ребра, соединяющие вершины и представляющие собой факты об условной зависимости между переменными.

Особенно интересным для решения прикладных задач является случай байесовских сетей, представленных ориентированными ациклическими графами (АОГ). Вычислительная сложность построения графа байесовской сети из данных довольно высока. Показано, что в общем случае количество возможных АОГ – суперэкспонента от количества вершин графа. Предлагались различные подходы к решению этой задачи, например, сужение класса рассматриваемых графов до класса деревьев [1], или применения жадных алгоритмов поиска. В настоящее время один из наиболее эффективных подходов – это сепарационный подход, позволяющий строить более точные графы, чем жадные алгоритмы, и не ограничивающийся каким-то одним классом АОГ.

РС-алгоритм основывается на сепарационном подходе и позволяет находить АОГ байесовской сети с точностью до класса эквивалентности. Алгоритм начинает работу с полного неориентированного графа, в котором все вершины соединены между собой, и рекурсивно удаляет ребра с помощью статистических тестов на условную независимость. В результате получается неориентированный граф, который затем частично ориентируется с помощью данных о d -сепарации (условной независимости), установленных во время выполнения алгоритма. В наихудшем случае время выполнения РС-алгоритма является экспоненциальным (от количества вершин), хотя для графов с разреженными матрицами смежности (с малым количеством ребер) время работы алгоритма полиномиальное.

Суть сепарационного подхода состоит в том, что если для некоторых двух переменных установлен факт об их условной независимости, относительно любого набора других переменных, то на графе ребро между ними отсутствует. Таким образом, на основании всех возможных фактов об условной независимости можно точно восстановить скелет графа, т. е. задача определения скелета сети сводится к наиболее быстрому нахождению всех фактов условной независимости между парами переменных графа.

РС-алгоритм для нахождения скелета сети. Используя псевдокод, опишем первую часть алгоритма РС – нахождение скелета.

Алгоритм 1

1: **ВВОД:** Набор переменных V и выборка данных с набором значений каждой из этих переменных

2: **ВЫВОД:** Скелет графа байесовской сети C , и набор S фактов о d-сепарации (необходим для ориентации ребер скелета)

3: Сформировать полный неориентированный граф \tilde{C} на V

4: $l = -1; C = \tilde{C}$

5: **repeat**

6: $l = l + 1$

7: **repeat**

8: Выбрать новую упорядоченную пару вершин i, j смежных в C , таких, что $|adj(C(i) \setminus \{j\})| \geq l$

9: **repeat**

10: Выбрать новый набор вершин $k \subseteq adj(C, i) \setminus \{j\}$, такое, что $|k| = l$

11: **if** i и j являются условно независимыми при условии k **then**

12: Удалить ребро между i и j

13: Обозначить полученный граф как C

14: Записать k в $S(i, j)$ и $S(j, i)$

15: **end if**

16: **until** ребро между i и j было удалено либо исчерпаны все $k \subseteq adj(C, i) \setminus \{j\}$, такие, что $|k| = l$

17: **until** все упорядоченные пары смежных вершин i и j , таких, что $|adj(C, i) \setminus \{j\}| \geq l$ не будут проверены на условную независимость

18: **until** для каждой пары смежных вершин i и j : $|adj(C, i) \setminus \{j\}| < l$

В приведенном алгоритме обозначение $adj(C, i)$ означает набор вершин, смежных с вершиной i , для графа заданного матрицей смежности C . Заметим, что способ, с помощью которого осуществляется проверка на условную независимость в одиннадцатой строчке алгоритма, не является строго фиксированным, хотя наиболее часто применяются методы χ -квадрат и точный тест Фишера.

При установлении скелета алгоритм РС перебирает все пары смежных вершин проводя проверку на условную независимость относительно l вершин, смежных к паре выбранных. Известно, что если максимальное число ребер входящих в некую вершину графа, равно q , то $l_{\max} \in \{q-1, q\}$.

Таким образом, эффективность алгоритма РС достигается за счет того, что тесты на условную независимость проводятся не для всех, а только для смежных пар вершин, а также благодаря тому, что в первую очередь проводятся тесты меньших рангов (l), являющиеся более простыми и быстрыми. Показано, что такая процедура гарантирует точное нахождение скелета при условии, что количество строк данных стремится к бесконечности.

Для графа с количеством вершин n и количеством ребер k максимальное количество статистических тестов на условную независимость в РС-алгоритме задается неравенством $W(n) \leq \frac{n^2(n-2)^k}{(k-1)!}$. Следует отметить, что сложность

каждого теста на условную независимость также экспоненциальная в зависимости от количества значений переменных (для дискретного случая) и ранга теста (l).

РС-алгоритм для частичной ориентации ребер графа. В алгоритме 1 заполнялась матрица S , хранящая сепараторы вершин, между которыми алгоритм удалял ребра. Эта информация была излишней для нахождения скелета графа, но необходима для определения ориентации ребер.

Основная идея алгоритма заключается в том, что для цепочки переменных $a-b-c$, таких что a и c не являются смежными вершинами, а b не входит в сепаратор $S(a, c)$, возможна лишь одна ориентация, называемая коллайдером, $a \rightarrow b \leftarrow c$. Таким образом, сначала ориентируются все коллайдеры, которые можно найти с помощью S , а затем ориентируются все остальные ребра в тех случаях, где ориентация однозначна в виду того, что противоположное направление ребра привело бы к появлению нового коллайдера. Приведем псевдокод алгоритма нахождения частичной ориентации скелета графа (приведение графа к классу эквивалентности).

Алгоритм 2

ВВОД: Скелет графа G_{skel} , набор сепараторов S

ВЫВОД: Частично ориентированный граф G , находящийся в классе эквивалентности искомого графа

for всех пар несмежных вершин i и j с общим соседом k **do**

if $k \notin S(i, j)$ **then**

 Заменить $i-k-j$ в G_{skel} на $i \rightarrow k \leftarrow j$ (единственная конфигурация, при которой k не сепарирует i и j)

end if

end for

В итоговом графе необходимо ориентировать столько вершин, сколько возможно, последовательно применяя следующие правила:

П1 Ориентировать $j-k$ в $j \rightarrow k$, если $i \rightarrow j$ и вершины i и k не являются смежными

П2 Ориентировать $i-j$ в $i \rightarrow j$, если существует цепочка $i \rightarrow k \rightarrow j$

П3 Ориентировать $i-j$ в $i \rightarrow j$, если существуют две цепочки $i-k \rightarrow j$ и $i-l \rightarrow j$, а вершины l и k не являются смежными

П4 Ориентировать $i-j$ в $i \rightarrow j$, если существуют две цепочки $i-k \rightarrow l$ и $k-l \rightarrow j$, а вершины l и k не являются смежными

Заключение. На основе аппарата байесовских сетей была получена графическая информация о динамике взаимодействия курсов мировых валют. Показана применимость моделей байесовских сетей для анализа рынка валют. Кратко описаны алгоритмы определения скелета байесовской сети и определения частичной ориентации ребер графа.

Н.А. Гупал, С.С. Ржепецкий

ЗАСТОСУВАННЯ АПАРАТА БАЙЕСІВСЬКИХ МЕРЕЖ ДЛЯ АНАЛІЗУ РИНКУ ВАЛЮТ

За допомогою апарата байесівських мереж отримано графічну інформацію щодо динаміки взаємодії курсів світових валют. Описано алгоритми побудови скелету байесівської мережі, а також часткової орієнтації ребер графа.

N.A. Gupal, S.S. Rzhpetskiy

APPLICATION OF BAYESIAN NETWORK TECHNIQUE TO THE CURRENCY MARKET ANALYSIS

On the basis of bayesian network technique, a graphic information of co-operation of world currency exchange rates is obtained. The algorithms of determination of skeleton of bayesian network as well as determination of partial orientation of edges of a graph are described.

1. *Pearl J.* Probabilistic reasoning in intelligent systems: networks of plausible inference. – San Mateo: Morgan Kaufmann, 1988. – 552 p.
2. *Bessler D., Yang J.* The international price transmission in stock index futures markets // *Economic Inquiry*. – 2004. – **42**, N 6. – P. 370–386.
3. *Диллинговий центр Forex «Alpari»* <http://www.alpari.ru>
4. *Kalisch M., Buhlmann P.* Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm // *J. of Machine Learning Research*. – 2008. – N 8. – P. 613–636.
5. *Балабанов А.С.* Минимальные сепараторы в структурах зависимостей. Свойства и идентификация // *Кибернетика и системный анализ*. – 2008. – № 6. – С. 17 – 32.
6. *Балабанов А.С.* Формирование минимальных *d*-сепараторов в системе зависимостей // Там же. – 2009. – № 5. – С. 38 – 50.

Получено 22.12.2009

Об авторах :

Гупал Никита Анатольевич,

инженер Института кибернетики имени В.М. Глушкова НАН Украины,

Ржепецкий Сергей Сергеевич,

младший научный сотрудник Института кибернетики имени В.М. Глушкова НАН Украины.