

О ПРИМЕНЕНИИ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ В ЗАДАЧАХ ФИНАНСОВОГО МОНИТОРИНГА

© 2015 А.С. Денисенко, Г.О. Крылов, И.А. Корнев

Федеральная служба по финансовому мониторингу, г.Москва

Поступила в редакцию 30.07.2015

Эффективное развитие национальной экономики требует своевременного реагирования и мониторинга обстановки. В задаче оценки эффективности предпринимаемых экономических и законодательных мер могут помочь объективные характеристики исследуемых объектов. Ранжирование объектов (компаний) по степени их бизнес-активности позволит выделять и проводить мониторинг динамики её изменения. В работе исследован процесс и результаты анализа методом главных компонент данных по финансовым потокам компаний и аффилированных с ними лиц отдельной отрасли экономики, синтезированы рейтинговые оценки активности в отраслях экономики в разрезе регионов. Результаты интегрированы в процесс визуально-сетевого анализа отрасли экономики. Ключевые слова: Метод главных компонент, противодействие отмыванию преступных доходов, управление в социально-экономических системах, национальная экономическая безопасность, факторный анализ, финансовый мониторинг.

1. ПРОБЛЕМАТИКА

В качестве исходного массива были использованы статистические показатели по финансовым потокам компаний рыбопромышленного комплекса. Кроме того, в массив признаков включены сведения о транзакциях руководителей компаний, их учредителей. Требуется получить количественные оценки об уровне финансовой активности компаний в заданной отрасли экономики региона.

Каждый из объектов задан упорядоченным набором показателей. Практика часто сталкивается с подобными задачами, поскольку результаты наблюдения, как правило, представлены векторами:

$$X = (X_1, X_2, \dots, X_p), \text{ где } X_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})^T \quad (1)$$

– вектор значений анализируемых признаков (свойств) $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$, присущих i -му объекту.

Здесь мы сталкиваемся с фундаментальной проблемой: в математике операция сравнения векторов по принципу «больше-меньше» не определена. Следовательно, требуется скаляризация векторных показателей.

Разрешением ситуации может являться переход от $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ к $z^{(1)}, z^{(2)}, \dots, z^{(p)}$, $p \ll p$ [1]. Существует ряд методов по снижению размерности исходных данных. Часть информации теряется, однако грубые оценки оказываются вполне достаточными и позволяют быть корректно интерпретированными. Этому способствуют следующие предпосылки [2]:

- дублирование информации коррелированных признаков;
- низкая информативность признаков, имеющих малую вариабельность;
- возможность агрегирования, т.е. простого или «взвешенного» суммирования по некоторым признакам.

2. КРАТКИЙ ОБЗОР МЕТОДОВ СНИЖЕНИЯ РАЗМЕРНОСТИ

Методы, основанные на экспертных оценках (экспертно-статистические методы)

Основной идеей методов данной группы является скалярное снижение размерности многокритериальной системы.

Главным недостатком метода является то обстоятельство, что получение интегрального показателя связано не со статистической информацией об объектах, но на использовании экспертных оценок анализируемого свойства. Экспертные оценки являются субъективными в силу ряда причин, среди которых: политизированность экспертов, их квалификация и личные симпатии, что неизбежно сказывается на конечном результате.

Многомерное шкалирование

Многомерное шкалирование представлено набором методов, позволяющих по заданной информации о мерах различия (близости) между объектами рассматриваемой совокупности приписывать каждому из этих объектов вектор характеризующих его количественных показателей; при этом размерность искомого координатного пространства задается заранее, а «погружение»

Денисенко Андрей Сергеевич, аспирант НИЯУ МИФИ.
E-mail: real_lkr@mail.ru

Крылов Григорий Олегович, доктор физико-математических наук, профессор Финансового университета при Правительстве Российской Федерации.

Корнев Иван Александрович, аспирант НИЯУ МИФИ.

в него анализируемых объектов производится таким образом, чтобы структура взаимных различий (близостей) между ними, измеренных с помощью приписываемых им вспомогательных координат, в среднем наименее отличалась бы от заданной в смысле того или иного функционала качества. Процедуры многомерного шкалирования применяются, когда данные заданы в виде матрицы попарных расстояний между объектами или удаленностей или их порядковых отношений.

В этом случае исследователь располагает в качестве массива исходных статистических данных матрицей размера $n \times n$, если рассматриваются характеристики попарной близости объектов или $p \times p$ (если рассматриваются характеристики попарной близости признаков вида

$$\rho = (\rho_{ij}), \quad i, j = 1, 2, \dots, m, \quad m = n \text{ или } m = p. \quad (2)$$

где величины ρ_{ij} интерпретируются либо как расстояния между объектами (признаками) i и j , либо как ранги, задающие упорядочение этих расстояний. Задача многомерного шкалирования состоит в том, чтобы «погрузить» наши объекты (признаки) в такое p' -мерное пространство ($p' < \min(p, n)$), то есть так выбрать координатные оси $Oz^{(1)}, \dots, Oz^{(p')}$ чтобы исходная геометрическая конфигурация совокупности анализируемых точек-объектов (или точек-признаков), заданных с помощью (1) или (2), оказалась бы наименее искаженной в смысле некоторого критерия средней «степени искажения» $D(Z)$ взаимных попарных расстояний.

Метод экстремальной группировки признаков и метод корреляционных плеяд

Метод экстремальной группировки предполагает объединение совокупности исходных показателей в заданное число таких p' групп $S_1, \dots, S_{p'}$, что признаки, принадлежащие одной группе, были бы взаимокоррелированы сравнительно сильно, в то время как признаки, принадлежащие к разным группам, были бы коррелированы слабо. Одновременно решается задача замены каждой (i -й) группы сильно взаимокоррелированных исходных показателей одним вспомогательным «равнодействующим» показателем $z^{(i)}$, который, естественно, должен быть в тесной корреляционной связи с признаками своей группы. Метод корреляционных плеяд, так же как и метод экстремальной группировки, предназначен для нахождения таких групп признаков — «плеяд», когда корреляционная связь, т. е. сумма модулей коэффициентов корреляции между параметрами одной группы (внутриплеядная связь) достаточно велика, а связь между параметрами из разных групп (межплеядная) — мала. По определенному правилу по корреляционной матрице признаков образуют чертеж — граф, который затем с помощью различных приемов разбивают на подграфы. Элементы, соответствующие каждому из подграфов, и образуют плеяду.

Факторный анализ

Методы факторного анализа основаны на общей базовой идее, в соответствии с которой структура связей между анализируемыми признаками $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ может быть объяснена тем, что все эти переменные зависят (линейно или как-то иначе) от меньшего числа других, непосредственно не измеряемых («скрытых», «латентных») факторов $f^{(1)}, f^{(2)}, \dots, f^{(p')}$ ($p' < p$), которые принято называть общими (common factors) и которые в большинстве моделей конструируются так, что они оказались взаимно некоррелированными. При этом в общем случае, естественно, не постулируется возможность однозначного (детерминированного) восстановления значений каждого из наблюдаемых признаков $x^{(i)}$ по соответствующим значениям общих факторов $f^{(1)}, f^{(2)}, \dots, f^{(p')}$ (в предположении, что мы их умеем вычислять): допускается, что каждый из исходных признаков зависит также и от некоторой «специфической» (для него) остаточной случайной компоненты $u^{(i)}$, которая и обуславливает статистический характер связи между $x^{(i)}$ с одной стороны, и $f^{(1)}, f^{(2)}, \dots, f^{(p')}$ — с другой.

Модель факторного анализа объясняет структуру связей между исходными показателями $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ тем, что поведение каждого из них статистически зависит от одного и того же набора общих факторов $f^{(1)}, f^{(2)}, \dots, f^{(p')}$

$$x^{(j)} - \mu^{(j)} = \sum_{v=1}^{p'} q_{i.v} f^{(v)} + u^{(j)} \quad (j = 1, 2, \dots, p),$$

где $q_{i.v}$ — «нагрузка» общего фактора $f^{(v)}$ на исходный показатель $x^{(j)}$, а $u^{(j)}$ — остаточная «специфическая» случайная компонента, причем $M[f^{(v)}] = 0$, $M[u^{(j)}] = 0$, $D[f^{(v)}] = 1$, и $f^{(1)}, f^{(2)}, \dots, f^{(p')}$, $u^{(1)}, u^{(2)}, \dots, u^{(p)}$ попарно не коррелированы.

Метод главных компонент

Во многих задачах обработки многомерных наблюдений и, в частности, в задачах классификации исследователя интересуют в первую очередь лишь те признаки, которые обнаруживают наибольшую изменчивость (наибольший разброс) при переходе от одного объекта к другому. С другой стороны, для описания состояния объекта, не обязательно использовать какие-либо из исходных, непосредственно замеренных на нём признаков. Так, например, для определения специфики фигуры человека при покупке одежды достаточно назвать значения двух признаков (размер, рост), являющихся производными от измерений ряда параметров фигуры. Если при конкретизации постановки задачи снижения размерности в качестве класса допустимых преобразований определить всевозможные линейные

ортогональные нормированные комбинации исходных показателей, т.е. [3]

$$z^{(j)}(X) = c_{j1}(x^{(1)} - \mu^{(1)}) + \dots + c_{jp}(x^{(p)} - \mu^{(p)});$$

$$\sum_{v=1}^p c_{jv}^2 = 1, \quad j = 1, 2, \dots, p;$$

$$\sum_{v=1}^p c_{jv} c_{kv} = 0, \quad j, k = 1, 2, \dots, p \quad j \neq k,$$

здесь $m^{(n)} = E[x^{(n)}]$ – математическое ожидание $x^{(n)}$, а в качестве меры информативности $I_p(Z)$ вы-

брать выражение
$$I_p(Z|X) = \frac{\sum_{i=1}^b D_{Z_i}}{\sum_{j=1}^p D_{X_j}}$$
 (здесь D

означает знак вычисления дисперсии соответствующей случайной величины), то мы перейдём к p -главным компонентам и определим метод главных компонент.

Таким образом, МГК является совокупностью линейных ортогональных преобразований. Целью метода является снижение размерности многомерных данных, сохраняя максимально возможную информативность. При анализе данных методом главных компонент анализу подвержен только массив исходных данных (разновидность методов «без учителя», без каких-либо указаний целевых параметров).

Первой главной компонентой $z^{(1)}(X)$ исследуемой системы показателей $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})'$ называется такая нормированно-центрированная линейная комбинация этих показателей, которая среди всех прочих нормированно-центрированных линейных комбинаций переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ обладает наибольшей дисперсией. И далее: k -й главной компонентой ($k = 2, \dots, p$) исследуемой системы показателей X называется такая нормированно-центрированная линейная комбинация этих показателей, которая не коррелирована с $k - 1$ предыдущими главными компонентами и среди всех прочих нормированно-центрированных и не коррелированных с предыдущими $k - 1$ главными компонентами линейных комбинаций переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ обладает наибольшей дисперсией. Главные компоненты могут использоваться при решении следующих основных типов задач анализа данных [4]:

1) упрощение, сокращение размерностей анализируемых моделей статистического исследования зависимостей или классификации с целью облегчения счета и интерпретации получаемых статистических выводов;

2) наглядное представление (визуализация) исходных многомерных данных, получаемое с помощью их проецирования в пространство, натянутое на первую, первые две или первые три главные компоненты;

3) предварительная ортогонализация объясняющих переменных в задачах построения регрессионных зависимостей как средство устранения мультиколлинеарности;

4) сжатие объемов хранимой статистической информации;

Вышеприведённые достоинства положительно выделяют этот метод среди прочих методов снижения размерности.

За пределами рассмотрения остались методы снижения размерности путём выбора наиболее информативных показателей при использовании различных статистических методов, таких как дискриминантный анализ и регрессионный анализ, так как они не являются универсальными.

Интерпретация метод главных компонент заключается в переходе к новой системе координат, где осями служат главные компоненты распределения. Простейший двумерный случай представлен на графике.

Указанные свойства положительно выделяют метод главных компонент перед другими методами снижения размерности.

Геометрическая интерпретация МГК состоит в перемещении начала координат в центр рассеяния. Главными компонентами являются полуоси гиперэллипсоида. График (рис. 1) демонстрирует простейший двумерный случай.

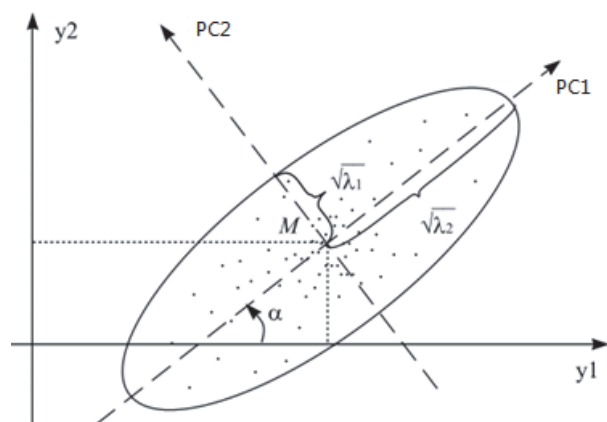


Рис. 1. Геометрическая интерпретация метода главных компонент

При реализации метода главных компонент важнейшее значение имеет интерпретация полученных результатов, интерпретация найденных главных компонент.

Ниже представлена корреляционная матрица исходных признаков (табл. 1).

Наблюдается сильная корреляция между переменными, характеризующими финансовую составляющую деятельности руководителя компании. На основе этой матрицы получены дисперсии 5 главных компонент, а также коэффициенты корреляции (табл. 2) исходных показателей с главными компонентами (внутренними факторами).

Таблица 1. Корреляционная матрица

		Company				Manager					Founders				
		Population	a	b	c	foreign currency operations	a	b	c	d	foreign currency operations	a	b	c	foreign currency operations
Company	Population	1.000000	0.017993	0.043409	-0.052575	-0.012553	-0.009673	-0.006475	-0.018519	-0.001852	0.010396	-0.042311	-0.014918	0.010174	-0.022738
	a	0.017993	1.000000	0.845319	0.396953	0.309282	0.105980	0.043497	0.033846	0.178458	0.142993	0.087031	0.007086	0.044338	0.016662
	b	0.043409	0.845319	1.000000	0.089849	0.144344	0.019274	0.006598	-0.002307	0.171165	0.087083	0.067717	-0.001809	0.040594	0.010907
	c	-0.052575	0.396953	0.089849	1.000000	0.237506	0.262899	0.237499	0.237172	0.150517	0.087902	0.131567	0.008697	0.073889	0.070716
	foreign currency operations	-0.012553	0.309282	0.144344	0.237506	1.000000	0.037634	0.000318	-0.000812	0.113701	0.273046	0.043717	0.009901	0.014931	-0.003401
Manager	a	-0.009673	0.105980	0.019274	0.262899	0.037634	1.000000	0.852542	0.755321	0.324605	0.049782	0.039017	0.037767	0.070561	-0.005291
	b	-0.006475	0.043497	0.006598	0.237499	0.000318	0.852542	1.000000	0.867069	0.241134	0.000636	0.016513	0.046427	0.071425	-0.001841
	c	-0.018519	0.033846	-0.002307	0.237172	-0.000812	0.755321	0.867069	1.000000	0.282004	-0.004281	0.016222	0.047733	0.064703	-0.005537
	d	-0.001852	0.178458	0.171165	0.150517	0.113701	0.324605	0.241134	0.282004	1.000000	0.294059	0.051774	0.011936	0.072116	0.084071
	foreign currency operations	0.010396	0.142993	0.087083	0.087902	0.273046	0.049782	0.000636	-0.004281	0.294059	1.000000	0.171806	0.000601	0.221683	0.133497
Founders	a	-0.042311	0.087031	0.067717	0.131567	0.043717	0.039017	0.016513	0.016222	0.051774	0.171806	1.000000	0.558683	0.559171	0.166116
	b	-0.014918	0.007086	-0.001809	0.008697	0.009901	0.037767	0.046427	0.047733	0.011936	0.000601	0.558683	1.000000	0.480678	0.016364
	c	0.010174	0.044338	0.040594	0.073889	0.014931	0.070561	0.071425	0.064703	0.072116	0.221683	0.559171	0.480678	1.000000	0.259270
	foreign currency operations	-0.022738	0.016662	0.010907	0.070716	-0.003401	-0.005291	-0.001841	-0.005537	0.084071	0.133497	0.166116	0.016364	0.259270	1.000000

Таблица 2. Оценки коэффициентов корреляции между исходными признаками и главными компонентами

		PC1	PC2	PC3	PC4	PC5
Company	a	0,383232	0,555072	0,630701	0,295075	-0,018248
	b	0,265293	0,514337	0,582604	0,389123	0,174214
	c	0,484944	0,175408	0,210193	-0,042923	-0,406711
	foreign currency operations	0,212330	0,343185	0,321652	-0,339027	-0,249639
Manager	a	0,838183	-0,379216	-0,048119	0,032028	0,028212
	b	0,833784	-0,446771	-0,092744	0,088446	0,023651
	c	0,809568	-0,435356	-0,091851	0,069343	0,017623
	d	0,499476	0,083103	0,135004	-0,377423	0,246909
	foreign currency operations	0,233785	0,390075	0,036852	-0,660422	0,169638
Founder	a	0,257548	0,573406	-0,556946	0,137181	-0,065386
	b	0,195081	0,413292	-0,600666	0,335825	-0,040844
	c	0,284392	0,516553	-0,582977	0,015005	0,092575
	foreign currency operations	0,108180	0,258319	-0,208605	-0,367998	0,083155
	population	-0,024475	0,005908	0,062949	0,107544	0,822029

В результате применения метода главных компонент выделены 12 главных компонент. Диаграмма иллюстрирует вклад каждой компоненты в общую дисперсию данных (рис. 2).

Согласно критерию Кайзера, первые 5 главных компонент являются наиболее информативными. Они описывают более 73% общей дисперсии в данных.

Согласно матрице факторных нагрузок можно сделать следующие предварительные выводы.

Анализ исходных данных позволил провести интерпретацию факторов.

1-я главная компонента биполярна. Наблюдается её положительная корреляция со всеми переменными, характеризующими её финансовую активность. Наибольшая положительная

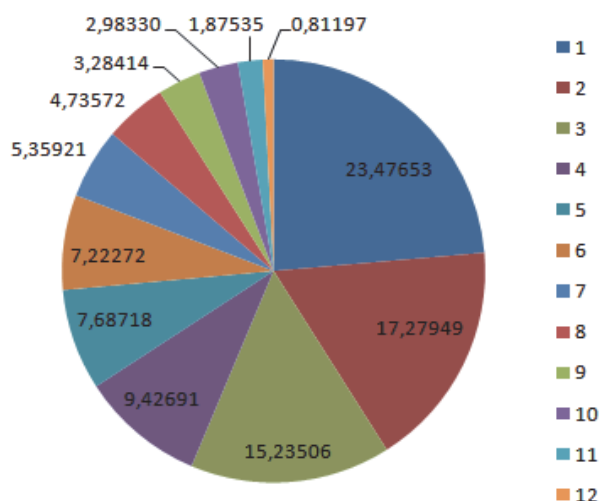


Рис. 2. Вклад главных компонент в общую дисперсию

корреляция наблюдается с признаками руководителя (0,5 – 0,84), и малая отрицательная – с численностью населения в регионе регистрации компании. Подобная корреляция может быть интерпретирована следующим образом: при равных показателях финансового потока, чем меньше население, тем выше на душу населения региона активность компании и аффилированных с ней лиц.

2-я главная компонента биполярна: положительно коррелирует с переменными учредителя и безналичными переводами на добывающую компанию, и, вместе с тем, отрицательно с руководителем. Имеет место предположение о том, переменная отражает характер распределения поступающих в отрасль средств: чем больше средств поступает на учредителя и добывающую компанию, тем меньше их поступает на руководителя. Меньшее значение главной компоненты характеризует большее поступление средств на руководителя.

3-я главная компонента также биполярна: имеет отрицательную корреляцию с переменными учредителя и руководителя, и положительную – с переменными добывающей компании. Следовательно, чем больше значение главной компоненты, тем больше объёмы поступлений на счета добывающей компании и меньше поступлений в адрес её руководителя и учредителей.

4-я ГК биполярна: характеризует валютные операции всех участников, причём валютные операции коррелируют с операциями с наличностью, что закономерно, поскольку определённая часть операций руководителя с наличными является валютно-обменными операциями.

5-я ГК преимущественно коррелирует с переменной населения региона.

Таким образом, 1-я главная компонента, характеризующая 23% дисперсии, является внутренним фактором, который описывает

именно ту часть дисперсии, которая является информативной с точки зрения отражения уровня активности компании, её руководителя и учредителей, с учётом численности населения региона регистрации компании. Её вклад в общую дисперсию – 23%.

Решение обратной факторной задачи

Синтезируем рейтинговые оценки уровня бизнес-активности (финансово-хозяйственной активности) компаний с использованием 1-й главной компоненты.

Каждая главная компонента является неким новым общим свойством всех объектов исследуемой выборки. Каждая компонента является функцией особенностей каждого из изучаемых объектов. Обычно исследователь имеет дело с ситуацией, когда одна главная компонента связана с одним или несколькими признаками. Таким образом, особенно важно получение значений главной компоненты для каждого наблюдения. Это позволит ранжировать и классифицировать объекты по полученным рейтинговым оценкам.

Обратимся к модели метода главных компонент и развернём равенство:

$$y'_j = \sum_{r=1}^n a_{jr} f_r \text{ для } j\text{-ого признака:}$$

$$y_j = a_{j1}f_1 + a_{j2}f_2 + \dots + a_{jn}f_n;$$

Выразим значения главных компонент через значения признаков. Для r -й компоненты [24]:

$$f_r = \frac{1}{\lambda} (a_{1r}y_1 + a_{2r}y_2 + \dots + a_{nr}y_n).$$

Ниже представлены синтезированные рейтинговые оценки уровня бизнес-активности каждой компании в отрасли, полученные с использованием 1-й главной компоненты по формуле, приведённой выше.

Таблица 3. Матрица счетов (1-я главная компонента)

№	Company	Score
1	ООО Компания 1	69.35441
2	ООО Компания 2	22.06919
3	ОАО Компания 3	15.75709
4	ООО Компания 4	11.84502
5	ОАО Компания 5	10.41410
6	ООО Компания 6	9.53931
	ООО Компания 7	9.29094

Идентификация компаний, имеющих наивысшую активность

1-я главная компонента характеризует уровень бизнес-активности. Чем выше её значение для отдельной компании, тем выше уровень её активности.

Интерес представляют компании, имеющие наивысшую активность. Научный смысл задачи заключается в определении доверительного интервала для мат.ожидания. Поскольку по определению метода главных компонент $E[f^{(v)}] = 0$, оценка доверительного интервала для среднего μ выборки может быть оценена по формуле:

$$\mu \pm t_{\text{набл.}} \frac{s}{\sqrt{n}} = 0,00 \pm 0,069.$$

Диапазон между границами доверительного интервала содержит около 90% компаний. Эти компании имеют наименьшую активность. Следовательно, оставшиеся 10% компаний имеют наибольшую активность.

Анализ природы распределения идентифицированной совокупности

Положим за гипотезу о зависимости количества компаний от их рейтинговых оценок (H_0) предположение, что выборка может быть асимптотически аппроксимирована логнормальной функцией:

$$y = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

Гипотеза подтверждается критерием хи-квадрат и критерием согласия Колмогорова. График (рис. 3) демонстрирует, что кривая

логнормальной теоретической функции асимптотически аппроксимирует дискретные эмпирические данные.

Другим известным методом иллюстрации корректности гипотезы является график К-К (квантиль-квантиль, рис. 4). График ниже иллюстрирует квантили эмпирических данных и теоретической логнормальной функции.

Физический смысл логнормального распределения тесно связано с повторяющимся делением целого на части [5].

В качестве классической задачи, иллюстрирующей природу распределения, можно привести следующую. Предположим, землевладелец делит имущество среди наследников. Далее эти наследники делят то же имущество среди своих наследников, и так из поколения в поколение. Какое распределение будет иметь размер состояния через несколько поколений? В среднем состояние становится, конечно, меньше, нам интересен характер распределения. Если T_i – случайная величина размера состояния в i -м поколении, то u_{i+1} – доля состояния отца, которая достаётся случайно выбранному наследнику следующего поколения. Если пропорции деления наследства в некоторой мере случайны и существует требуемая независимость между поколениями, то на основании изложенного можно ожидать, что распределение будет асимптотически логнормальным. Аналогичную аргументацию можно применить при рассмотрении повторяющегося деления любой количественно выраженной величины. Например, общество при изучении делится на все более мелкие группы, численность которых в пределе имеет приближённо логнормальное

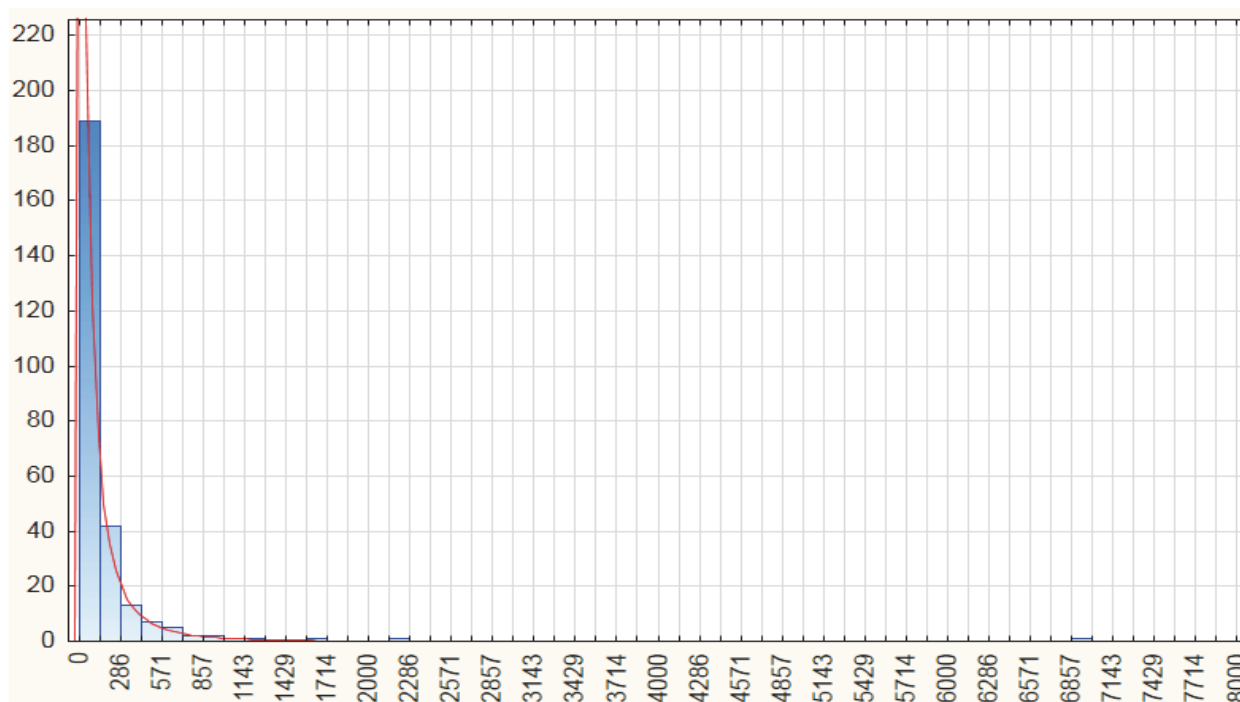


Рис. 3. Асимптотическая аппроксимация эмпирических данных логнормальной функцией

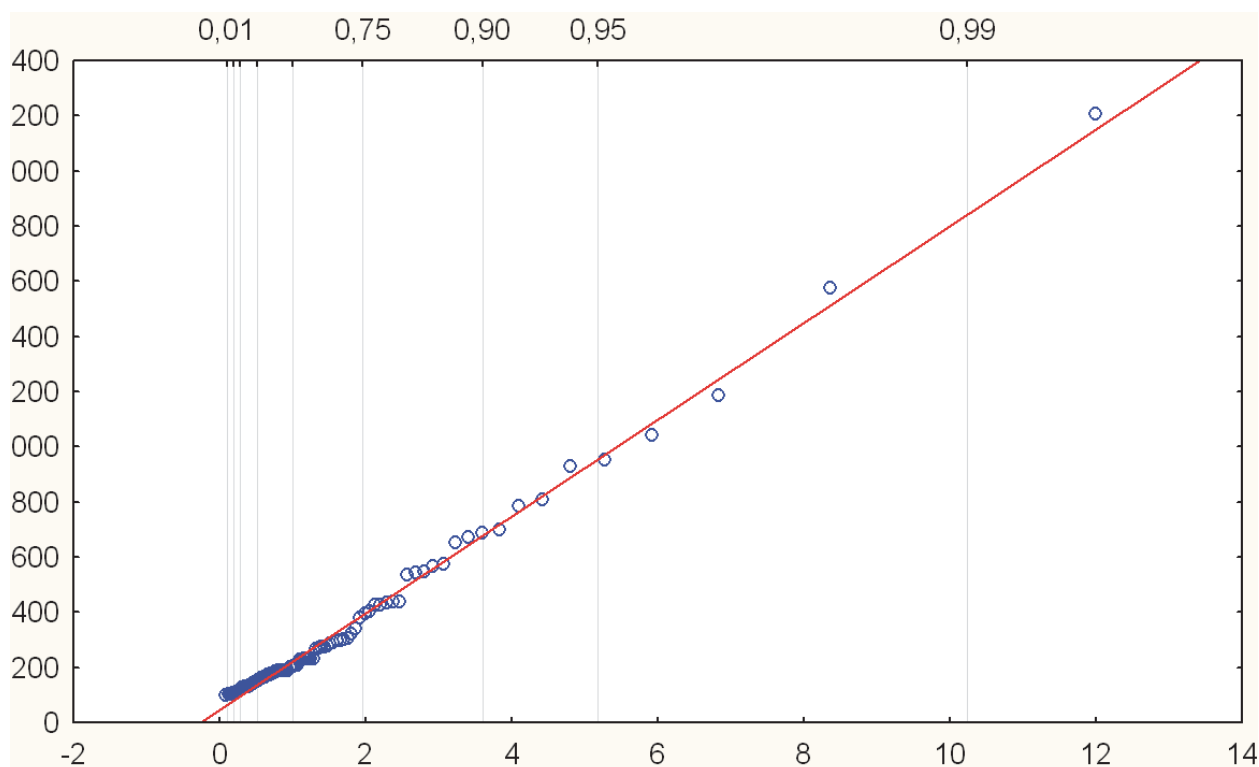


Рис. 4. Квантили эмпирических данных и теоретической логнормальной функции

распределение. Как отмечено Айтинсоном и Брауном (1957), эти идеи приводят к разработке теории группирования, в соответствии с которой группы получают путём последовательного деления. Например, при изучении контингента рабочих они вначале подразделяются по уровню квалификации, по полу, затем по уровню физического труда и т.д. Распределение численности рассматриваемых групп может быть асимптотически логнормальным.

Таким образом, чем выше уровень бизнес-активности представляет группа компаний, тем меньше компаний в этой группе. Убывание количества компаний в группах с ростом рейтинговых оценок может быть асимптотически аппроксимировано логнормальной функцией.

4. ИСПОЛЬЗОВАНИЕ РЕЗУЛЬТАТОВ

Использование результатов анализа данных методом главных компонент при визуально-сетевом анализе

Экономические отношения между компаниями анализируемой отрасли визуализированы в виде неориентированного графа. Компании объединены в кластеры в зависимости от их специализации в отрасли. Наиболее активные компании отмечены красным цветом на схеме ниже (рис. 5).

Результаты МГК на детальной схеме при проведении анализа отдельных объектов представлены на рисунке ниже (рис. 6).

Положительным эффектом от полученных ре-

зультатов также является повышение оперативности анализа объектов – аналитику не требуется выявлять наиболее активных объектов «вручную» – они уже отмечены на схеме.

Таким образом, процесс анализа отрасли может быть существенно упрощён за счёт препроецессорной обработки данных методом главных компонент.

Применение метода главных компонент позволило раскрыть роли и положение компаний идентифицированной совокупности в схемах цепочек собственников (рис. 7).

Успешное применение метода главных компонент для синтеза рейтинговых оценок уровня бизнес-активности позволило получить оценки регионального уровня (макроаналитические оценки).

Финансовые потоки данных по компаниям были агрегированы по регионам их регистрации. В результате анализа корреляционной матрицы выделены три главных компоненты. Для решения задачи синтеза интегральных оценок по регионам была выбрана 1-я главная компонента (как проекции показателей по каждому региону на 1-ю главную компоненту).

Визуальное представление информации является одним из наиболее эффективных путей её восприятия. По этой причине проведено цветовое кодирование оценок и их представление на карте региона (рис. 8). Карта содержит «снимок» активности компаний региона за определённый период. Регулярный мониторинг данных и их последующий анализ методом главных компо-

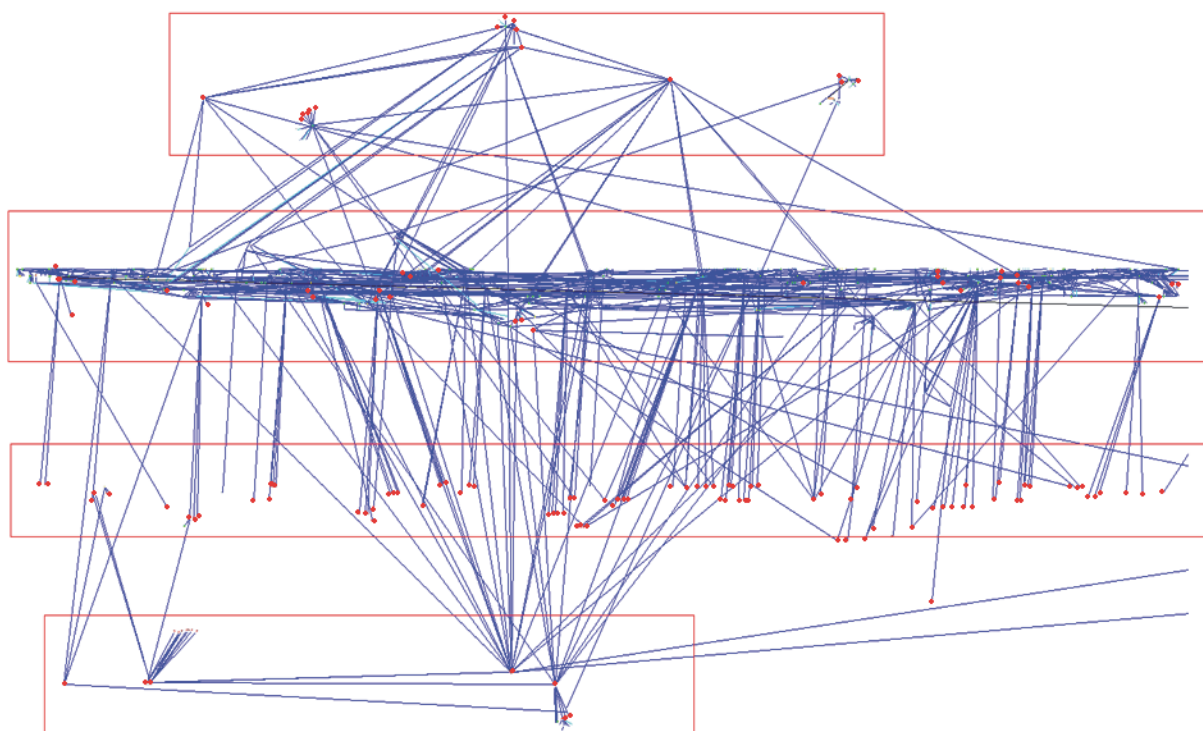


Рис. 5. Использование результатов анализа данных МГК при визуальном-сетевом анализе отрасли

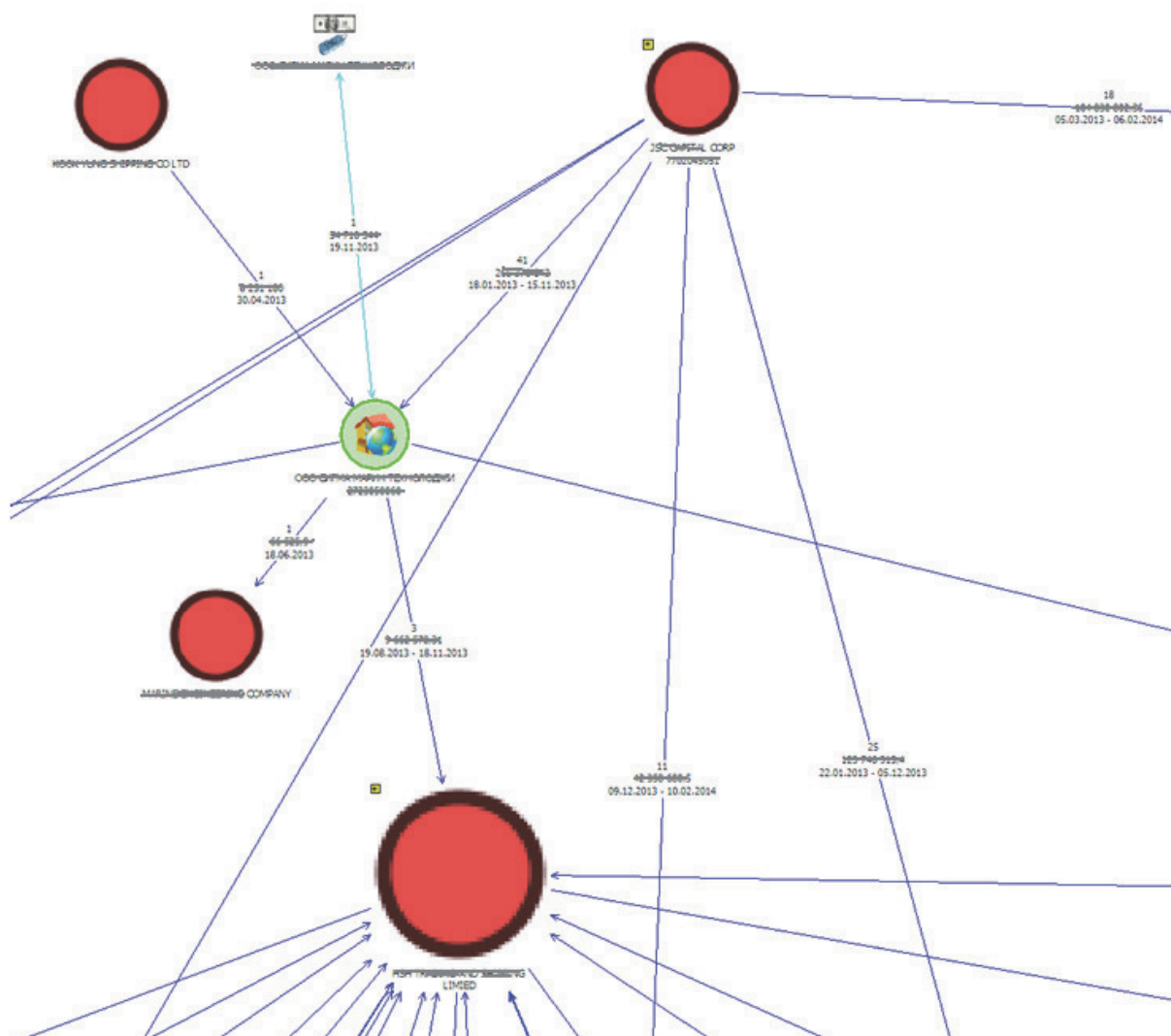


Рис. 6. Наиболее активные компании (результаты МГК) отмечены красным на схеме

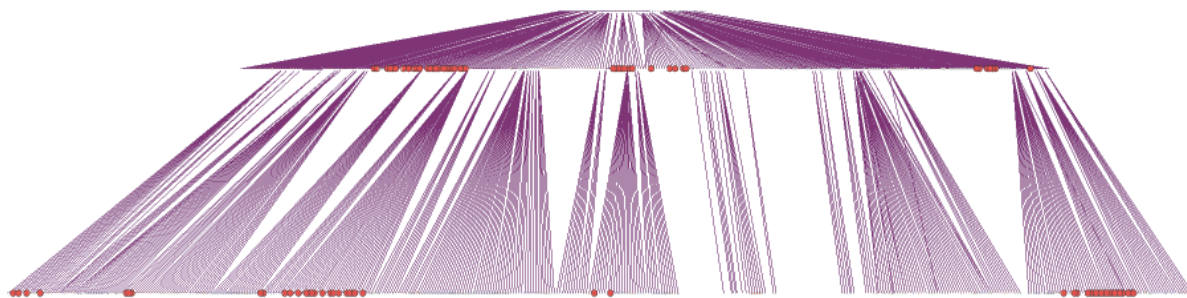


Рис. 7. Роли компаний на схеме цепочек собственников



Рис. 8. Цветовое кодирование оценок регионов

нент позволит проводить анализ динамики интегральных показателей активности, что, в свою очередь, способствует решению задачи анализа эффективности предпринимаемых экономических и законодательных мер.

5. ВЫВОДЫ

В данной статье авторами проанализировано применение метода главных компонент для анализа данных о финансовых потоках компаний определённой отрасли. В ходе анализа выделены, интерпретированы главные компоненты. Обоснован выбор компоненты для решения обратной факторной задачи, в результате чего синтезиро-

ваны рейтинговые оценки бизнес-активности компаний. Кроме того, идентифицированы компании, имеющие наибольшую активность в отрасли. Авторами проанализирована природа их распределения в результате чего установлено, что зависимость между количеством компаний в группе и уровнем их активности (значением их рейтинговых оценок) может быть асимптотически аппроксимировано логнормальной функцией.

Получены интегральные оценки активности на уровне регионов в результате анализа методом главных компонент данных по региональным финансовым потокам.

В результате проведённого исследования авторы пришли к выводу об эффективности и

целесообразности применения метода главных компонент в качестве препроцессорного обработчика в решении различных прикладных экономических задач.

СПИСОК ЛИТЕРАТУРЫ

1. Бартоломью Д. Стохастические модели социальных процессов. М.: Финансы и статистика, 1985. 296 с.
2. Dubrov A.M., Mhitarian V.S., Troshin L.I.

Multidimensional statistic methods, М.: Finance and statistics, 1998.

3. Dubrov A.M. Data processing with the principal components analysis. М.: Statistics, 1978.-130p.
4. Andrukowich P.F. a. o. Abstract painting as a specific – Generale – Language. A Stat. Appr. To the problem// Metron XXIX. 1971. N 1–2.
5. Колмогоров А. Н. О логарифмически нормальном законе распределения размеров частиц при дроблении. Докл. АН СССР. 1941. Т. 31, № 2. С. 99-101.

ADAPTATION OF PRINCIPAL COMPONENT ANALYSIS TO FINANCIAL MONITORING TASKS

© 2015 A.S. Denisenko, G.O. Krylov, I.A. Kornev

Federal Financial Monitoring Service, Moscow

National economics needs well-timed reacting and situation monitoring for effective growing. Objective characteristics of prototype system may help in an evaluation task of effectivity of acceptable economical and legislative measures. The ranking of objects (companies) according to their business activity will allow to choose and carry out monitoring of dynamics of their changing. The process and results of the principal component analysis of data on financial streams of companies and affiliated with them organisations in separate sector of economy are investigated and the ranking evaluations of activity in economic sectors are synthesized in this work. The results have been integrated into the process of visual-net analysis of economical sector.

Keywords: Principal component analysis, anti-money laundering, management in social economic systems, national economic security, factor analysis, financial monitoring.

Andrey Denisenko, Postgraduate, National Research Nuclear University MPhI. E-mail: real_lkr@mail.ru

Grigory Krylov, Doctor of Physical and Mathematical Sciences, Professor of Financial University under the Government of the Russian Federation.

Ivan Kornev, Postgraduate, National Research Nuclear University MPhI.