

ИССЛЕДОВАНИЕ ИНСТИТУЦИОНАЛЬНОЙ СТРУКТУРЫ БАНКОВСКОЙ СИСТЕМЫ РОССИИ МЕТОДОМ КЛАСТЕРНОГО АНАЛИЗА

Лабусов М.В.

Финансовый университет при Правительстве РФ

Москва, Россия

Научный руководитель Орлова И.В.

Аннотация: Статья посвящена изучению институционального состава банковского сектора экономики РФ. Основным методом, использованным для изучения представленной совокупности данных, является кластерный анализ. В работе показан процесс структурирования банковских кредитных учреждений на основе анализа ключевых количественных показателей их деятельности, представлены полученные однородные группы, проведена интерпретация полученных результатов и проанализирована адекватность модели.

Ключевые слова: банковские учреждения, кластер, иерархический кластерный анализ, кластерный анализ методом k – средних, двухэтапный кластерный анализ.

THE STUDY OF THE INSTITUTIONAL STRUCTURE OF THE RUSSIAN BANKING SYSTEM BY THE METHOD OF CLUSTER ANALYSIS

Labusov M. V.

Financial University under the Government of the Russian Federation

Moscow, Russia

Scientific leader: Orlova I.V.

Abstract: The article is devoted to the study of institutional composition of the banking sector of the Russian economy. The main method, applied for the analysis of the represented data aggregate, is cluster analysis. The work contains the process of structuring banking credit institutions on the basis of analyzing the key quantitative indicators of their activity, the representation of the homogeneous groups formed, the interpretation of the received results and the analysis of the model adequacy.

Key words: banking institutions, cluster, hierarchical cluster analysis, k-means cluster analysis, two-step cluster analysis.

Введение

Объектом данного исследования является рынок институтов, предоставляющих банковские услуги (система банков и кредитных организаций) на территории Российской Федерации. Цель исследования – разбиение объекта исследования на однородные группы с соот-

ветствующими характерными признаками с использованием нескольких программных продуктов и последующая интерпретация полученных результатов. Данные для исследования были получены из Web – базы данных СПАРК (www.spark-interfax.ru) - крупнейшей базы данных по компаниям России. Выбор данной базы данных в качестве источника информации объясняется, во-первых, актуальностью и полнотой приведённых в ней данных и, во-вторых, широким спектром предоставляемых аналитических возможностей.

Размер загруженной с сайта выборки (октябрь 2014 г.) после удаления противоречивых данных составил 14994 записи о 833 банках и кредитных организациях. Московские кредитно-банковские учреждения составляют большую часть выборки. О каждой единице выборки известны следующие данные: регистрационный номер, наименование, регион, адрес, ИНН, ОКПО, валюта баланса, активы, прибыль до налогообложения, депозиты частных лиц, вложения в государственные ценные бумаги, вложения в негосударственные ценные бумаги, средства в банках (брутто), кредиты небанковскому сектору, просроченные кредиты небанковскому сектору, кредиты физическим лицам, обязательства перед банками и иностранные обязательства – всего 18 пунктов.

В качестве основного инструментария в процессе исследования были использованы 2 программы прикладного статистического анализа: IBM SPSS Statistics 22 и STATISTICA 10. Последовательное рассмотрение процесса кластеризации начнём с анализа в программном продукте SPSS.

В части кластеризации программа SPSS даёт возможность проведения иерархической кластеризации, кластеризации k – средними и двухэтапного кластерного анализа [3]. В качестве начального (наиболее общего) метода применим иерархический кластерный анализ. При первом использовании данного метода программа разбивает всю выборочную совокупность на 2 кластера. Такое разбиение носит малую точность и информативность. Для определения оптимального количества кластеров решающее значение имеет показатель под заголовком «Коэффициент» - расстояние между двумя кластерами, определённое на основании выбранной дистанционной меры (в данном случае в качестве меры выбран квадрат Евклидова расстояния, т.к. выборка осуществлена из совокупности, имеющей многомерное нормальное распределение, а компоненты вектора наблюдений однородны по физическому смыслу и одинаково важны для классификации[2]). Оптимальным считается число кластеров, равное разности общего количества наблюдений и номера шага, после которого коэффициент увеличивается скачкообразно. В нашем случае оптимальное число кластеров будет равно $833 - 828 = 5$.

827	9	15	5007165521354 735600,000	824	820	830
828	4	5	5288902006003 873800,000	826	0	829
829	3	4	2084893818968 5514000,000	0	828	830
830	3	9	3318880817128 4304000,000	829	827	831
831	2	3	1879570310589 13700000,000	0	830	832
832	1	2	3165258237863 793300000,000	0	831	0

Таблица 1. Порядок агломерации (кластеров). Составлено автором.

Теперь проведём новый иерархический анализ с уже заданным числом кластеров. В результате его проведения в первом кластере оказался 1 банк («Сбербанк»), во втором – 1 банк (банк ВТБ), в третьем – 1 банк («Газпромбанк»), в четвёртом – 5 банков, в пятом – все остальные банки. Таким образом, в данном случае иерархический анализ как метод кластеризации непригоден для указанной задачи.

Для получения более точного результата применим другой метод кластеризации – двухэтапный кластерный анализ. Он более привлекателен по сравнению с традиционными методами кластерного анализа в связи с возможностью работы с различными типами данных, использованием различных критериев кластеризации и масштабируемостью. При использовании автоматического определения числа кластеров, как и в методе иерархической кластеризации, формируется лишь 2 кластера: в первый входят крупные банки с госучастием («Сбербанк», банк ВТБ, «Газпромбанк», «Россельхозбанк», «Внешэкономбанк» и др.), во второй кластер попадают все остальные банки. При изначально заданном количестве кластеров (в нашем случае воспользуемся полученным оптимальным числом кластеров из прошлого метода - 5) получаем следующее разбиение:

Кластеры	N	% объединенных	% общего итога
1	9	1,1%	1,1%
2	236	28,3%	28,3%
3	199	23,9%	23,9%
4	133	16,0%	16,0%
5	256	30,7%	30,7%
Объединённый	833	100,0%	100,0%
Всего	833		100,0%

Таблица 2. Распределение для кластеров на основе проведения двухэтапного кластерного анализа. Составлено автором.

1. Первый кластер – крупные банки с госучастием («Сбербанк», банк ВТБ, «Газпромбанк», «Россельхозбанк», «Внешэкономбанк») и наиболее крупные частные банки («Альфа-Банк» и Банк ФК «Открытие»). Эти банки имеют наибольшие по сравнению с другими банками валюту баланса, активы, депозиты частных лиц, средства на счетах в других банках. Они наиболее активно кредитуют небанковский сектор. Их обязательства перед различными экономическими субъектами также велики. Они ведут активную деятельность с зарубежными клиентами и партнёрами.

2. Второй кластер – крупные коммерческие банки, наибольшая доля которых находится в Москве («Райффайзенбанк», «Промсвязьбанк», банк «Русский Стандарт», «Ситибанк», «Дойче Банк» «Связьбанк» и др.) и Санкт-Петербурге (банк «Россия», банк «Санкт-Петербург»), а также Центрально-Чернозёмном экономическом районе и Поволжье. Их отличает большой объём активов и депозитов частных лиц, а также кредитов физическим лицам. Эти банки в своём большинстве не являются убыточными.

3. Третий кластер – средние коммерческие банки, расположенные в основном в Москве («Московский индустриальный банк», «Совкомбанк», банк «Ренессанс Кредит» и др.), Центральном экономическом районе, а также в Поволжье, на Урале и в Сибири. Для таких банков характерны не очень большой объём активов и депозитов частных лиц, меньший объём кредитов, чем у банков предыдущего кластера. Некоторые из таких банков не имеют достаточно средств для приобретения государственных и негосударственных ценных бумаг.

4. Четвёртый кластер – небольшие московские («ИШБанк», банк «Унифин», «Чайна Констракшн Банк» и др.) и региональные банки (банк «Спурт» и «Энергобанк» в г. Казань, «Уралтрансбанк» в г. Екатеринбург, «Нико-Банк» в г. Оренбург и др.). Для них характерна работа преимущественно с физическими лицами и некрупными юридическими лицами. Объём их внешнеэкономической деятельности незначителен.

5. Пятый кластер – мелкие московские (банк «Адмиралтейский», «Милбанк», банк «Океан», «Пробанк» и др.) и региональные банки («Тамбовкредитпромбанк» в г. Тамбов, банк «Долинск» в г. Южно-Сахалинск, «ГТ Банк» в г. Майкоп» и др.). Для этих банков характерен наименьший по сравнению с банками из других кластеров объём активов и обязательств, а также валюта баланса. Многие из этих банков не приобретают ценные бумаги. Довольно существенная часть этих банков была убыточна по состоянию на июль 2014 г.

Таким образом, программа SPSS позволила осуществить процесс кластерного анализа с образованием 5 кластеров, пригодных для дальнейшего исследования. В то же время программа не смогла сформировать кластеры достаточно высокого качества.

В качестве второго программного продукта для осуществления кластерного анализа используем статистическую программу Statistica 10. Эта программа позволяет осуществить кластерный анализ на основе иерархического метода и метода k – средних. В качестве начального метода, как и в программе SPSS, используем метод иерархической кластеризации.

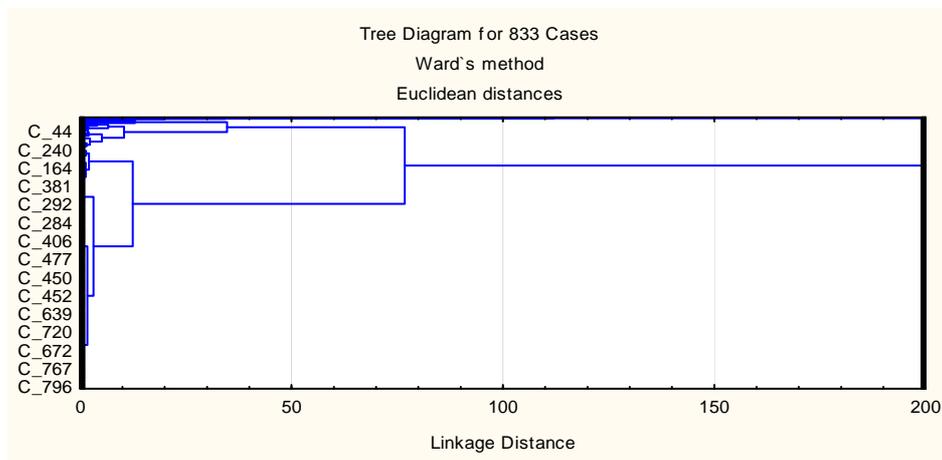


Рисунок 1. Дендрограмма иерархической кластеризации в программе Statistica. Составлено автором.

Равно как и в программе SPSS, в ходе кластерного анализа в программе Statistica вся выборочная совокупность разбивается на 2 кластера. Безусловно, такой результат мало пригоден для дальнейшей интерпретации.

Следовательно, необходимо использовать ещё один традиционный метод кластерного анализа – метод кластеризации k – средними. Этот метод состоит в том, что программа вычисляет расстояние от каждого объекта до центров кластеров и приписывает объекты к ближайшему кластеру. Процесс повторяется до тех пор, пока центры тяжести не перестанут «мигрировать в пространстве» [4]. В конце проведения кластеризации методом k – средних

происходит разбиение выборочной совокупности на 5 кластеров.

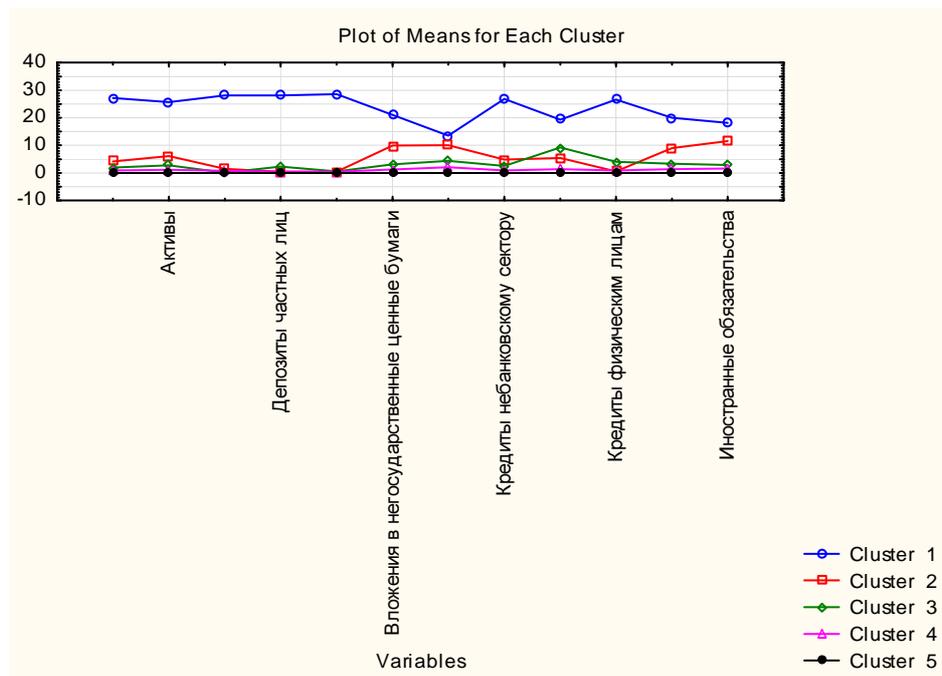


Рисунок 2. График средних для каждого кластера в программе Statistica. Составлено автором.

Однако новая кластеризация также не отличается высокой точностью: только первый кластер (в который попал лишь «Сбербанк») существенно отличается от остальных кластеров по указанным на рисунке характеристикам. Второй кластер (в который попали банк ВТБ, «Газпромбанк» и «Внешэкономбанк») существенно уступает в количественных характеристиках первому кластеру. Третий (банк ВТБ 24, Банк Москвы и «Россельхозбанк»), четвёртый («Альфа-Банк», АКБ НКЦ, банк «Открытие», «Промсвязьбанк», «Юникредитбанк» и «Райффайзенбанк») и пятый (все остальные банки) кластеры слабо различаются между собой почти по всем показателям.

Заключение

Таким образом, кластерный анализ, проведённый в программе Statistica, дал ещё менее точные результаты, чем аналогичный анализ, проведённый в программе SPSS. Такой результат вызван тем, что метод двухэтапного кластерного анализа, использованный в программе SPSS, является наиболее подходящим для исследования больших выборочных совокупностей, к которым и относится исходная. Ошибки также могут быть связаны с наличием взаимосвязанных характеристик, по которым проводилась кластеризация, а также существенными количественными различиями анализируемых данных. Несмотря на возможные ошибки и неточности, поставленная цель исследования была выполнена: была проведена классификация выборочной совокупности и было выделено 5 групп однородных элементов с соответствующими характеристиками, различающихся между собой и пригодных для интер-

претации. Полученная модель отвечает условию адекватности и может использоваться для дальнейших исследований.

Литература

1. Мхитарян В.С., Архипова М.Ю., Сиротин В. П. Эконометрика: Учебно-методический комплекс// М.: Изд. центр ЕАОИ, 2008. – 144 с.
2. Орлова И. Многомерный статистический анализ в экономических задачах: компьютерное моделирование в SPSS// М.: Вузовский учебник, 2014. – 310 с.
3. Халафян А. Statistica 6: Математическая статистика с элементами теории вероятностей// М.: Бином, 2010. – 496 с.
4. Доугерти К. Введение в эконометрику: Учебник: Пер. с английского// М.: ИНФРА-М, 2010. – 418 с.
5. Сайт <http://www.spark-interfax.ru>