

Бексултанов А.А.,
Сулайманова Г.У.,
ИНДО при БГУ им.К.Карасаева

КЛАСТЕРНЫЙ АНАЛИЗ В ПОРТФЕЛЬНОМ ИНВЕСТИРОВАНИИ

Понятие кластерного анализа

При анализе и прогнозировании социально-экономических явлений исследователь довольно часто сталкивается с многомерностью их описания. Это происходит при решении задачи сегментирования рынка, построении типологии стран по достаточно большому числу показателей, прогнозирования конъюнктуры рынка отдельных товаров, изучении и прогнозировании экономической депрессии и многих других проблем.

Кластерный анализ позволяет рассматривать достаточно большой объем информации и резко сокращать, сжимать большие массивы социально-экономической информации, делать их компактными и наглядными.

Важное значение кластерный анализ имеет применительно к совокупностям временных рядов, характеризующих экономическое развитие (например, общехозяйственной и товарной конъюнктуры). Здесь можно выделять периоды, когда значения соответствующих показателей были достаточно близкими, а также определять группы временных рядов, динамика которых наиболее схожа.

В задачах социально-экономического прогнозирования весьма перспективно сочетание кластерного анализа с другими количественными методами (например, с регрессионным анализом).

Как и любой другой метод, кластерный анализ имеет определенные недостатки и ограничения: В частности, состав и количество кластеров зависит от выбираемых критериев разбиения. При сведении исходного массива данных к более компактному виду могут возникать определенные искажения, а также могут теряться индивидуальные черты отдельных объектов за счет замены их характеристиками обобщенных значений параметров кластера. При проведении классификации объектов игнорируется очень часто возможность отсутствия в рассматриваемой совокупности каких-либо значений кластеров.

В кластерном анализе считается, что:

- а) выбранные характеристики допускают в принципе желательное разбиение на кластеры;
 - б) единицы измерения (масштаб) выбраны правильно.
- Выбор масштаба играет большую роль. Как правило, данные нормализуют вычитанием среднего и делением на стандартное отклонение, так что дисперсия оказывается равной единице.

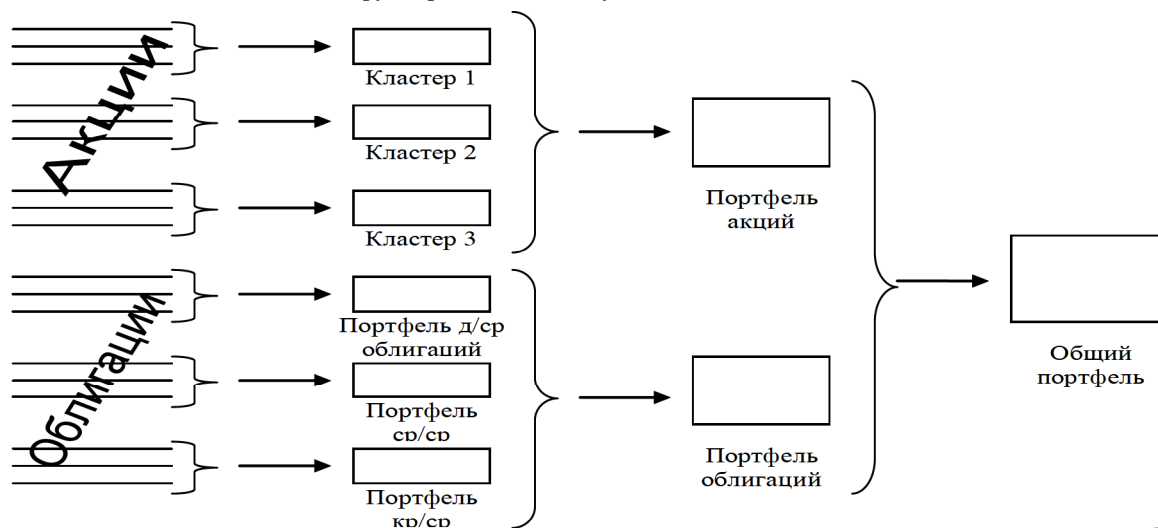
Кластерный анализ в портфельном инвестировании.

Общеизвестно, что изменение курсовой стоимости и дивидендов различных ценных бумаг не только в Кыргызстане, но и во всем мире зависит от ряда внутренних и международных факторов экономического и неэкономического характера. Эти факторы могут быть взаимосвязаны в различной степени, а тенденции изменения их динамики способны отличаться друг от друга в достаточно сильной степени. Следовательно, изменение стоимости инвестиционного портфеля в результате сложения различных тенденций с большой вероятностью оказывается достаточно сложной и практически непредсказуемой, если использовать обычный регрессионный анализ. Основные факторы воздействия влияют на различные ценные бумаги не только с разной эффективностью, но зачастую и в прямо противоположных направлениях.

В свете вышесказанного, перед инвесторами возникают следующие проблемы:

- 1) Определение с максимальной степенью точности существенных факторов и их влияние на курс ценных бумаг;
- 2) Составление научно-обоснованного прогноза динамики поведения этих ценных бумаг, основываясь на изучении данных факторов;
- 3) Составление на основе полученных сведений о фондовом рынке оптимального инвестиционного портфеля, позволяющего максимизировать прибыль от вложений при заданной степени риска.

Рис.1 Группировка ценных бумаг со сходными тенденциями



Как теоретики, так и практики, занимающиеся оптимизацией портфеля ценных бумаг, регулярно сталкиваются с трудностями, когда перед ними возникает практически неизбежная задача разбиения множества существующих ценных бумаг на различные группы с относительно однородной структурой. Краеугольным камнем проблемы является вопрос подбора и согласования выбранных факторов так, чтобы их представление в многомерной системе координат достаточно точно производило разбиение на кластеры, характеризующиеся максимально схожими тенденциями. При этом нужно учитывать, что даже если бы и удалось подобрать точные коэффициенты для существующих количественных факторов, всегда найдутся не менее важные качественные показатели, выразить которые в количественной форме практически невозможно. В связи с этим принято группирование ценных бумаг на основе существующих индустриальных и прочих классификаций, а также отталкиваясь от априорной доходности (ex ante).

Разбиение множества ценных бумаг на отдельные кластеры в зависимости от динамики доходности осуществляется следующим образом: данные по доходности ценных бумаг на протяжении базы прогноза компонуются в общую матрицу вида:

$$\begin{pmatrix} R_{10} & R_{11} & \dots & R_{1T} \\ R_{20} & R_{21} & \dots & R_{2T} \\ \dots & \dots & \dots & \dots \\ R_{m0} & R_{m1} & \dots & R_{mT} \end{pmatrix} [1]$$

где R_{km} – доходность по k -й ценной бумаге за m -й период, $k \in [1; N]$; $m \in [0; T]$.

Далее, разбиение на кластеры происходит через вычисление евклидова расстояния между ценными бумагами p и q по формуле

$$r_{p,q} = \sqrt{\sum_{m=1}^T \left(\frac{R_{pm} - R_{qm}}{\sigma_{Rm}} \right)^2},$$

где m – номер периода,

σ_{Rm} – среднее квадратическое отклонение доходности за период m .

Критическая величина разбиения предполагается равной квадратному корню из количества периодов T , то есть средней величине евклидова расстояния:

$$r_{\text{крит}} = \bar{r}_{p,q} = \sqrt{T}.$$

Преимущество данной методики заключается, во-первых, в том, что она позволяет с крайне высокой степенью точности группировать ценные бумаги со сходными тенденциями в изменении доходности на протяжении всего периода, определяющего базу прогноза, что дает основания рассчитывать на сохранение подобной тенденции и в дальнейшем.

Вторым ее преимуществом является возможность полной автоматизации, что значительно облегчает работу, позволяя использовать современные вычислительные средства, а также обрабатывать однородную информацию, получаемую из электронных баз

данных. Поэтому она может быть без особых затруднений внедрена не только в компьютерных системах отдельных фирм, занимающихся инвестированием, но также и на соответствующих ресурсах сети интернет.

Пусть имеются N факторов $X_1 \dots X_N$, предположительно влияющих на доходность инвестиционного портфеля. При вводе в уравнение регрессии фактора X_i показатель R -квадрат принимает некоторое определенное значение. Выберем фактор, при котором оно будет наибольшим:

$$P_1^2 = R_1^2 = r_{yx_1}^2,$$

где P_1^2 – коэффициент последовательной детерминации для данного фактора,

r_{yx_1} – парный коэффициент корреляции между доходностью и этим фактором.

Теперь вводится в полученное уравнение регрессии второй фактор таким образом, чтобы значение R -квадрат снова оказалось максимально возможным, и затем рассчитываем второй коэффициент последовательной детерминации:

$$P_2^2 = R_2^2 - R_1^2 = r_{yx_2x_1}^2 - r_{yx_1}^2.$$

Аналогичным образом рассчитываем следующие коэффициенты:

$$P_3^2 = R_3^2 - R_2^2 = r_{yx_3x_1x_2}^2 - r_{yx_2x_1}^2,$$

$$P_4^2 = R_4^2 - R_3^2 = r_{yx_4x_1x_2x_3}^2 - r_{yx_3x_1x_2}^2 \text{ и т.д.}$$

Базовый отбор факторов продолжается до тех пор, пока величина получаемых коэффициентов последовательной детерминации не станет меньше некоторого критического значения. Учитывая, что в механизм расчета скорректированной величины R -квадрат входит поправка на возрастание энтропии при вводе новых факторов, ее прирост на каждой итерации алгоритма должен быть положительным и, следовательно, критическое значение r должно быть больше нуля.

Данный метод позволяет отобрать из всех имеющихся факторов именно те, которые оказывают наибольшее влияние на доходность рассматриваемых ценных бумаг. Это позволяет существенно понизить размерность модели, создаваемой на основе методики, ускорить вычисления и при этом отбросить данные, не имеющие большого влияния на интересующие нас показатели. Как правило, от выявленных главных компонент зависит не менее 85% общей дисперсии, что лишний раз показывает эффективность выбранного метода анализа.

Теперь, когда определены методы отбора факторов и технология разбиения множества ценных бумаг на отдельные кластеры, можно приступить непосредственно к построению методики оптимизации инвестиционного портфеля. Учитывая, что в настоящее время внедрение любой экономической методики невозможно без автоматизации, существует алгоритм, по которому надлежит производить операции для получения искомого результата: оптимизированного набора ценных бумаг, позволяющих получить максимальную прибыль при заданном уровне риска.

На первом этапе определяются исходные массивы данных, которые подлежат математической обработке.

В начале имеются следующими исходными данными: $S_1, S_2, \dots, S_N$ – рассматриваемое множество ценных бумаг;

$$\begin{bmatrix} R_{10} & R_{11} & \dots & R_{1T} \\ R_{20} & R_{21} & \dots & R_{2T} \\ \dots & \dots & \dots & \dots \\ R_{N0} & R_{N1} & \dots & R_{NT} \end{bmatrix} -$$

матрица доходности ценных бумаг S_1-S_N за периоды $[0; T]$, где R_{ij} – доходность по ценной бумаге i за j -й период;

$$\begin{bmatrix} X_{10} & X_{11} & \dots & X_{1T} \\ X_{20} & X_{21} & \dots & X_{2T} \\ \dots & \dots & \dots & \dots \\ X_{N0} & X_{N1} & \dots & X_{NT} \end{bmatrix} -$$

матрица факторов $XI-XK$ за периоды $[0; T]$, где X_{ij} – значение фактора Xi за j -й период; sn – оценка риска предполагаемого портфеля ценных бумаг.

Теперь необходимо определить доли $m_1, \dots, m_N$ имеющихся в инвестиционном портфеле ценных бумаг с целью максимизации доходности в следующем периоде при заданном уровне риска:

$$\begin{cases} \sum_{i=1}^N m_i \times ER_i \rightarrow \max \\ E\sigma_{\pi} = \sigma_{\pi}^* \end{cases},$$

где уровень доходности Ri вычисляется как отношение ожидаемой в отчетный период стоимости ценной бумаги Si к курсовой стоимости в момент формирования портфеля за вычетом единицы.

Так, доходность за месяц в момент времени $t=1$ вычисляется следующим образом:

$$R_{i1} = \frac{Pr_{i1}}{Pr_{i0}} - 1.$$

В случае, когда инвестор не имеет возможностей продавать ценные бумаги без покрытия, вводится дополнительное условие: $my > 0$, где y – номер соответствующей ценной бумаги.

Вывод: принято группирование ценных бумаг на основе существующих индустриальных и прочих классификаций.

Алгоритм оптимизации портфеля с применением кластерного анализа

Предлагаемый алгоритм можно условно разбить на четыре основные стадии:

- 1) Разбиение множества ценных бумаг на отдельные кластеры;
- 2) Определение факторов, влияющих на доходность составляющих каждого кластера. Расчет факторных весов. Построение уравнения регрессии;
- 3) Прогнозирование динамики выбранных факторов;
- 4) Вычисление ожидаемой доходности и степени риска для каждой ценной бумаги;

5) Определение оптимального набора ценных бумаг и их долевого веса в инвестиционном портфеле для обеспечения максимизации доходности.

Теперь можно рассмотреть эти стадии подробнее:

1. Разбиение множества ценных бумаг на отдельные кластеры.

Эта стадия начинается с формирования таблицы евклидовых расстояний между имеющимися ценными бумагами:

Таблица 1 – Таблица евклидовых расстояний

Ценные бумаги:	S_1	S_2	$\dots S_i \dots$	S_N
S_1	-	$r_{1,2}$	$r_{1,i}$	$r_{1,N}$
S_2		-	$r_{2,i}$	$r_{2,N}$
$\dots S_i \dots$			-	$r_{i,N}$
S_N				-

Расстояния вычисляются по формуле

$$r_{i,j} = \sqrt{\sum_{t=1}^T \left(\frac{R_{it} - R_{jt}}{\sigma_{Rt}} \right)^2} \quad [2]$$

Две ценные бумаги с наименьшим расстоянием объединяются в кластер, доходность которого вычисляется как средняя арифметическая доходностей этих ценных бумаг, после чего процедура расчета повторяется. Процесс объединения в кластеры прекращается, когда минимальное расстояние между группами превысит критическое значение:

$$r_{кр.} = \sqrt{T}.$$

В результате описанной процедуры, вместо случайного множества ценных бумаг, мы получаем набор упорядоченных кластеров, объединенных на основе общих тенденций в динамике доходности. При этом достигаются сразу две важные цели: во-первых, значительно сокращается количество переменных, что в заметной степени упрощает вычисления, а во-вторых, уменьшается доля воздействия случайных факторов, которые могут в отдельные моменты коррелировать с доходностью отдельных ценных бумаг. В рамках кластера за счет произведенной диверсификации вероятность случайных совпадений уменьшается во много раз, что дает возможность гораздо более ясно определить факторы, реально воздействующие на доходность.

2. Определение факторов, влияющих на доходность составляющих каждого кластера. Расчет факторных весов. Построение уравнения регрессии.

Для того, чтобы вычислить величину влияния каждого фактора на соответствующий кластер ценных бумаг, представим доходность по кластерам в следующем виде:

$$R'_t = F_1 \times X_{1t} + F_2 \times X_{2t} + \dots + F_k \times X_{kt} + \varepsilon_t,$$

где F_i – коэффициент фактора Xi в уравнении множественной регрессии,

ε_t – ошибка в период времени t . При этом величина T должна значительно (не менее чем в пять раз) превышать количество факторов k .

Значимые факторы отбираются при помощи описанного выше метода с применением коэффициентов последовательной детерминации. Факторы отбираются последовательно, а выбор определяется путем максимизации коэффициента.

$$R_i^2 = R_i^2 - R_{i-1}^2 = r_{yx_i, x_1, \dots, x_{i-1}}^2 - r_{yx_i, x_1, \dots, x_{i-1}}^2$$

Процесс добавления факторов продолжается до тех пор, пока максимальный скорректированный коэффициент последовательной детерминации не окажется отрицательной величиной. Для любого выбранного количества факторов коэффициенты $F1, F2, \dots, Fk$ рассчитываются таким образом, чтобы минимизировать сумму квадратов ошибок регрессии за период базы прогноза:

$$\sum_{t=0}^T \epsilon_t^2 \rightarrow \min.$$

Этой цели можно достигнуть путем математических преобразований матрицы факторных весов. В настоящее время существует ряд программных пакетов, позволяющих производить данные расчеты с высокой скоростью и за короткое время.

3. Прогнозирование динамики выбранных факторов

Результатом вышеуказанных вычислений является получение формул множественной регрессии для каждого кластера, с помощью которых, опираясь на статистические данные о динамике факторов, можно получить прогноз развития доходности кластеров на последующий период и оценить величину существующего риска. Преимущество прогнозирования факторов по сравнению с прогнозированием курсов отдельных ценных бумаг состоит в наличии значительно большего количества авторитетных исследований по движению макроэкономических факторов, а также статистических сводок органов государственного регулирования.

4. Вычисление ожидаемой доходности и степени риска для каждой ценной бумаги.

В большинстве моделей, опирающихся на САРМ, для ценных бумаг рассчитывается бета-коэффициент, отражающий взаимосвязь между динамикой доходности изучаемой ценной бумаги и существующими рыночными тенденциями. Простая линейная регрессия по отношению к рыночной динамике может оказаться слишком неточной, так как не позволяет учитывать специфические факторы, оказывающие на данную ценную бумагу влияние весомее, чем на фондовый рынок в целом. Поэтому для более подробного изучения прибегают к более эффективным средствам, в частности: к факторному анализу. Без сопоставления с существующими тенденциями велик риск усиления влияния случайных факторов. Таким образом, для получения достоверного результата методика анализа рынка ценных бумаг должна совмещать оба вышеописанных подхода.

Достаточно высокая эффективность прогнозирования, основанная на использовании бета-коэффициента показывает, что между отдельными ценными бумагами и состоянием фондового рынка в целом наблюдается существенная зависимость, которую можно использовать для проведения оценки будущей доходности. При этом корреляция доходности ценных бумаг со средней доходностью по кластеру значительно

выше, чем с рынком в целом. Поэтому в данной методике бета-коэффициент каждой отдельной ценной бумаги рассчитывается, опираясь на не рыночный индекс, а относительно кластера:

$$\beta_i = r_{ic} \times \frac{\sigma_i}{\sigma_c},$$

где r_{ic} – коэффициент корреляции между доходностью ценной бумаги и средней доходностью кластера, к которому она принадлежит,

σ_i и σ_c – соответственно их среднеквадратические отклонения.

После расчета бета-коэффициента доходность каждой из исследуемых ценных бумаг можно будет выразить при помощи следующего уравнения регрессии:

$$R_{it} = \alpha_i + \beta_i \times \bar{R}_{ct} + \epsilon_{it},$$

а ожидаемая в следующем периоде доходность будет равна.

При этом коэффициент неопределенности для каждой ценной бумаги равняется

$$E_T R_{i,T+1} = \alpha_i + \beta_i \times E_T \bar{R}_{c,T+1}.$$

$$\text{а величина риска} - \sigma_i = \sqrt{\frac{\sum_{t=1}^T (E_T R_{it} - R_{it})^2}{T}}.$$

5. Определение оптимального набора ценных бумаг и их долевого веса в инвестиционном портфеле для максимизации доходности.

После всех проведенных преобразований получена для каждой ценной бумаги величину ожидаемой доходности и оценку имеющегося риска. Теперь задача сводится к тому, чтобы определить долевой вес этих ценных бумаг в инвестиционном портфеле с целью максимизации прибыли при заданном уровне риска sn .

Как известно, множество эффективных портфелей расположено на так называемой эффективной границе, не ниже точки минимизации риска. Следовательно, в случае наличия определенности относительно желаемого уровня риска оптимальная точка для заданного набора ценных бумаг может быть определена однозначно:

Основываясь на данных, полученных на трех предыдущих этапах, исходные формулы выглядят следующим образом:

$$\sum_{i=1}^N (E_T R_{i,T+1} \times m_i) \rightarrow \max,$$

причем

$$\sum_{i=1}^N m_i = 1;$$

$$\sqrt{\sum_{i=1}^N \sum_{j=1}^N m_i m_j V_{ij}} = \sigma_{sn} \quad [2]$$

Как уже отмечалось, в случае необходимости добавляется условие не отрицательности долей m_i .

Вывод: полученная задача легко решается как при помощи стандартно используемых вычислительных методов, так и большинством математических и экономических программных пакетов (MathCAD, SAS, Solver for MS Excel и т.д.).

Литература

1. Витин А. Мобилизация финансовых ресурсов для инвестиций // Вопросы Экономики, № 7. 1994.
2. Бажилина Э. Инвестициям - новые цели // Экономика и жизнь. - 1998. - № 13.
3. Филатов В. Проблемы инвестиционной политики в индустриальной экономике переходного типа // Вопросы Экономики, № 7. - 1994.
4. Вардуль Н. Предварительный диагноз — асфиксия // Коммерсантъ. 24 октября 1995.
5. Красавина Л. Н. Международные валютно-кредитные и финансовые отношения. - М., 1994. Гл. 9.
6. Мартынов А. Активизация инвестиционной политики. - М., // Экономист, № 9, 1997.
7. Миловидов В. Иностранные инвестиционные фонды в России: первые итоги деятельности // Рынок ценных бумаг. 1995. № 13.
8. Милюков А. Сначала туда, где есть деньги, идут товары. За ними - иностранный инвестор // Деловое Поволжье. -1997. - № 44.
9. Сейфульмулюков И. Иностранные инвестиции в добывающих отраслях // Российский экономический журнал. 1992. № 11.
10. Смородинская Н., Капустин А. Свободные экономические зоны: мировой опыт и российские перспективы // Вопросы экономики. 1994. № 12.
11. Фадеев А., Рукин А. Инвестиционные портфели // Рынок ценных бумаг. 1995. № 14.
12. Цветков Н. На мировом рынке инвестиционных капиталов // Инвестиции в России. -1997. - № 11-12.