

ФОРМИРОВАНИЕ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ НА РОССИЙСКОМ РЫНКЕ АКЦИЙ ПРИ ПОМОЩИ НЕПАРАМЕТРИЧЕСКОГО МЕТОДА – ДЕРЕВА РЕШЕНИЙ

Буянова Е.А.¹, Саркисов А.Р.²

Формирование инвестиционного портфеля является одним из ключевых вопросов современной теории финансов. Главной задачей в построении оптимального портфеля является создание модели, способной успешно прогнозировать поведение доходности актива. Однако построить такую модель далеко не просто – во множестве эмпирических работ авторы сталкивались с ситуацией, когда модель успешно подходит для объяснения исторических данных, но не способна прогнозировать будущие показатели доходности. Одной из причин подобной проблемы является то, что порой в таких работах авторы имеют либо слишком короткие ряды данных, либо данные слишком «шумные». Поэтому, используя классические параметрические методы, авторы и сталкиваются с подобными трудностями.

В данной работе был проведен анализ акций, входящих в базу расчета индекса ММВБ, при помощи непараметрического метода – Classification and Regression Tree (CART), и на основе полученных данных был построен оптимальный портфель. Использование данного метода позволило анализировать не только детерминанты доходности акций, относящиеся к классу макроэкономических факторов или факторов технического анализа, но и факторов фундаментального анализа, по которым всегда присутствует ограниченное число наблюдений. При этом в данной работе используется не вполне стандартное определение оптимального портфеля. Во многих работах под оптимальностью подразумевают превышение доходности построенного портфеля над доходностью рыночного портфеля. В общем случае данное определение некорректно, т.к. ввиду большей диверсификации рыночный портфель обладает меньшим риском и, как следствие, меньшей доходностью. Поэтому в данной работе оптимальным портфелем называется портфель с заданным риском, доходность которого не ниже доходности рыночного портфеля с таким же параметром риска. Формулировка «не ниже» говорит о том, что в случае наличия слабой формы эффективности рынка оптимальный портфель должен совпасть с рыночным, если он будет базироваться лишь на исторических данных.

Также в рамках данного исследования был проведен сравнительный анализ степени влияния того или иного фактора на доходность акций. Было выявлено, что основными детерминантами доходности акций российских компаний является momentum, bid-ask spread и цена на нефть.

Ключевые слова: оптимальный портфель, фундаментальный анализ, технический анализ, отбор бумаг в портфель, непараметрические методы

JEL: G10, G11, G17, C14

Обзор литературы

Формирование оптимального инвестиционного портфеля является одним из ключевых вопросов современной теории финансов. Основу данной области положили работы таких ученых, как Г. Марковиц, У. Шарп, а также М. Скоулс и Ф. Блэк (Edwin and Gruber, 1997).

Данные работы представляли собой теоретическое обоснование существования оптимального портфеля и возможности его построения. Однако ни в одной модели не указан универсальный набор факторов, при помощи которого можно составлять оптимальный инвестиционный портфель.

Таким образом, вопрос выбора факторов для анализа ценных бумаг остается открытым и затрагивается во многих эмпирических и теоретических исследованиях. Первые эмпирические исследования, посвященные отбору ценных бумаг в портфель, основывались на использовании параметрических методов (МНК, GLM и т.д.). Основным недостатком данных методов являлись

1. Канд. физ.-матем. наук, доцент, преподаватель департамента финансов НИУ ВШЭ.

2. Аспирант департамента финансов НИУ ВШЭ.

жесткие требования к данным, причем как количественные, так и качественные. Ведь само название «параметрические методы» говорит о том, что при их использовании у параметров модели должны быть известны основные параметры распределения: среднее и дисперсия. Для корректного определения параметров распределения необходимо иметь достаточно длинный ряд данных, очищенный от шумов. Финансовые данные зачастую не соответствуют подобным критериям, поэтому во многих случаях использование параметрических методов в финансовых исследованиях является некорректным.

На данный момент наиболее актуальными являются непараметрические модели оценки, т.к. данные методы подходят для работы с «шумными» данными и ограниченными выборками. В работе Бримен (Breiman, 1984) на примере метода деревьев регрессий были описаны основные преимущества непараметрических методов:

- Непараметрические модели не требуют предварительного отбора параметров. Нет проблемы ошибочной спецификации.
- Непараметрические модели нечувствительны к монотонным преобразованиям независимых переменных.
- Результаты робастны к выбросам в выборке.

В области построения оптимального портфеля при помощи непараметрических моделей следует выделить следующие два направления исследований. На ранних этапах развития данного направления исследователи пытались показать эффективность использования непараметрических методов, используя лишь стандартные показатели фундаментального анализа. В более поздних исследованиях осуществлялись попытки получить оптимальную комбинацию факторов как технического, так и фундаментального анализа, которая бы максимально подходила для конкретного рынка финансовых активов.

Ниже представлена выборка из основных работ, посвященных использованию непараметрических методов для решения проблемы составления портфеля или прогнозирования цены актива.

Чаварнакул и Инке (Chavarnakul and Enke, 2009) в рамках эконометрической модели, основанной на методе деревьев регрессий, формировали портфель, состоящий из акций, котирующихся на DAX. В рамках своей работы авторы отобрали ряд факторов технического анализа, которые наряду с фундаментальными факторами эмитентов позволили построить портфель с доходностью выше рынка.

Но непараметрические методы – это не только метод деревьев регрессий. Так, например, Алварез-Диаз и Хаммоудех (Alvarez-Diaz and Hammoudeh, 2013) тестировали surrogate data method в качестве инструмента для прогноза цен акций на DJIM и DJIA. В итоге авторы получили вывод о том, что, используя данный непараметрический метод, можно эффективно прогнозировать стоимость акций на DJIA в рамках 15-недельного горизонта, а на DJIM – в рамках 9-недельного горизонта.

Kernel method был использован в работе Ширайши и Танигучи (Shiraishi and Taniguchi, 2012) для оценки дисперсии доходностей акций на Токийской фондовой бирже. Исследование авторов показало, что при помощи данного непараметрического метода можно эффективно прогнозировать поведение акций и строить оптимальный портфель. При этом авторы отмечают, что эффективность того или иного непараметрического метода может быть различна в зависимости от структуры данных.

Так называемый I(d) processes (Bierens, 1997) был использован в работе Зонг и Чанг (Zhong and Chang, 2014), в которой авторы строили модель оптимальной диверсификации при построении портфеля ценных бумаг. В ходе своего исследования авторы выяснили, что фондовые рынки США (Dow Jones 30 и S&P 500) являются попарно коинтегрированными с фондовыми рынками стран БРИКС.

Таким образом, различные типы непараметрических моделей широко используются в современной литературе по портфельному анализу и служат хорошим инструментом как для опи-

сания исторических данных, так и для прогноза стоимости активов. Из этого следует, что при помощи непараметрических методов может быть разрешена проблема «Fitting vs Forecasting», которая заключается в том, что модели, которые обычно хорошо описывают исторические данные, имеют низкую предсказательную способность, и наоборот.

Описание метода дерева решений

Метод дерева решений является непараметрическим методом, который распределяет наблюдения в соответствующие группы при помощи классификационного алгоритма. Для использования метода необходимо составить базу данных по характеристикам с целью классификации будущих наблюдений. При этом следует отметить, что финальные группы, в одну из которых будет классифицировано каждое последующее наблюдение, должны быть строго определены и их количество должно быть конечным. Методология данного исследования основана на работе Браймана (Breiman, 1987).

При классификации каждого нового наблюдения рассматриваются его основные параметры, и на их основе наблюдению присваивается определенный класс. Таким образом, строится дерево регрессий с узлами и ветвями.

Допустим, для каждого наблюдения сформирован вектор параметров X , который используется для определения класса каждого наблюдения. Определим набор различных классов $C = \{1, 2, \dots, J\}$. Таким образом, существует J различных классов. Далее необходимо определить правило, согласно которому уже имеющиеся наблюдения классифицируются в один из J классов. Для финансовых активов удобно распределять наблюдения по классам в зависимости от соотношения показателя доходности по наблюдению (R_t) и установленного порогового уровня доходности \bar{R} . Например, правило, которое было использовано при классификации в данной работе:

$$\begin{cases} R_t > \bar{R}, y_t = & \{\text{покупка актива}\} \\ R_t < -\bar{R}, y_t = & \{\text{продажа актива}\} \\ -\bar{R} \leq R_t \leq \bar{R}, y_t = & \{\text{сохранение актива}\} \end{cases}$$

При построении прогноза на основании новых данных, необходимо выделить из имеющихся наблюдений «обучающую выборку» $L = \{(x_1, j_1), \dots, (x_N, j_N)\}$, на основании которых алгоритм будет обрабатываться.

Классификация наблюдений представляет собой бинарную процедуру и заключается в разбиении исходного множества на подмножества $X_1^i, X_2^i, \dots, X_n^i$, где i – порядок итерации расщепления на групп. При этом целью данного разбиения является получение при каждой последующей итерации более однородных групп, чем при предыдущей.

Однако сложно сформировать оптимальное количество операций по расщеплению исходного множества наблюдений на соответствующие подмножества. Если в процессе расчета будет осуществлено слишком малое количество итераций, то в итоге получатся неоднородные группы. Если же итераций будет слишком много, то в итоге в финальных группах будет находиться по одному наблюдению.

Ввиду того что при каждой итерации необходимо увеличение однородности группы, введем для каждого узла t в дереве размера T меру неоднородности $g(t)$. На каждом этапе разбиения наблюдения делятся в пропорциях α и β , в зависимости от того, в какой узел – t_α или t_β – попали. Ниже представлена базовая схема разбиения узла :

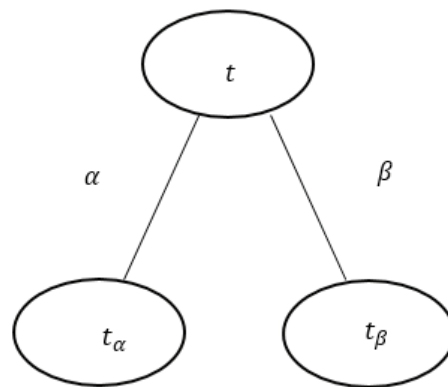


Рисунок 1. Базовая схема разбиения узла

Таким образом, при каждой итерации согласно следующему выражению:

$$\Delta g(s, t) = g(t) - \alpha g(t_\alpha) - \beta g(t_\beta), \quad (1)$$

где $\Delta g(s, t)$ – снижение неоднородности после итерации s ,

$g(t), g(t_\alpha), g(t_\beta)$ — неоднородность узла t, t_α, t_β , и соответственно,

α, β – пропорции разделения узла.

Рассматривая в целом дерево регрессий, понимаем, что общая неоднородность задается следующим выражением:

$$G(T) = \sum_{t=1}^{\tilde{T}} G(t) = \sum_{t=1}^{\tilde{T}} g(t) p(t), \quad (2)$$

где: $G(T)$ — неоднородность всего дерева,

\tilde{T} — заданное количество узлов,

$p(t)$ — пропорции разбиения узла.

Из выражений (1) и (2) очевидно следует, что при проведении последующей операции разбиения узла общая неопределенность нового дерева размерности будет рассчитываться как:

$$G(T') = \sum_{T' - \{t\}} G(t) + G(t_\alpha) + G(t_\beta), \quad (3)$$

где T' новая размерность дерева.

Логично, что в итоге необходимо решить задачу минимизации общей неопределенности дерева. Однако данную задачу можно трансформировать в задачу максимизации уменьшения неоднородности при каждой последующей итерации s :

$$\max_{s \in S} \Delta G(s, t) = \max_{s \in S} G(T) - G(T') = \max_{s \in S} G(T) - G(t_\alpha) + G(t_\beta), \quad (4)$$

где $\Delta G(s, t)$ – изменение общей неоднородности дерева после итерации s .

Следовательно, можно сформировать правило для определения необходимости в последующей итерации по разбиению узла, введя пороговое значение для уменьшения неоднородности:

$$\max_{s \in S} \Delta G(s, t) > \varepsilon, \quad (5)$$

где $\varepsilon > 0$ – пороговое значение для уменьшения неоднородности.

Введя правило, согласно которому останавливается процесс разбиения узлов, необходимо определить, как для каждого терминального¹ узла $t \in \tilde{T}$ определяется класс $j(t)$. Если количество наблюдений, относящихся к классу j в узле t наибольшее, то данному узлу присваивается класс j . Таким образом, терминальному узлу присваивается класс j , если в нем наибольшее

1. Терминальным узлом называется узел, который впоследствии не разбивается.

значение имеет условная вероятность $p(j|t)$ того, что наблюдение, попавшее в узел t , относится к классу j .

Введем для каждого наблюдения величину $r(t)$, определяющую вероятность того, что объект классифицирован некорректно:

$$r(t) = 1 - \max_j p(j|t), \quad (6)$$

где: $r(t)$ — вероятность некорректной классификации объекта в узле t ,

$p(j|t)$ — условная вероятность того, что наблюдение, попавшее в узел t .

Далее необходимо ввести «штраф» $C(i|j)$ за каждое неверно классифицированное наблюдение, принадлежащее к классу i и попавшее в класс j .

$$\begin{cases} C(i|j) > 0, i \neq j \\ C(i|j) = 0, i = j \end{cases} \quad (7)$$

где $C(i|j)$ — «штраф» за каждое неверно классифицированное наблюдение.

Следовательно, для каждого узла можно выписать задачу минимизации ошибки при классификации:

$$r(t) = \min_i \sum_j C(i|j) p(j|t). \quad (8)$$

И данное выражение для всего дерева:

$$R(T) = \sum_t R(t) = \sum_t r(t) p(t), \quad (9)$$

где вероятность некорректной классификации в рамках всего дерева.

Выше была введена функция $g(t)$ как мера неоднородности каждого узла. В данной работе использовалась функция $g(t)$ на основе Индекса Джини, задаваемая следующим выражением:

$$g(t) = \sum_{j \neq i} p(j|t) p(i|t), \quad (10)$$

где $p(j|t)$ — условная вероятность того, что наблюдение, попавшее в узел t , относится к классу j

$p(i|t)$ — условная вероятность того, что наблюдение, попавшее в узел t , относится к классу i

Описанные способы определения размера дерева и точности классификации являются необходимыми, но не достаточными условиями оптимальности построенного дерева. Ниже описана процедура «обрезки» дерева с целью нахождения оптимального размера.

На основе вышеизложенного алгоритма строится дерево максимальной размерности T_{max} . Далее введем выражение, отражающее «издержки размера дерева»:

$$R_\alpha(T) = R(T) + \alpha |\bar{T}|, \quad (11)$$

где: $R(T)$ — вероятность некорректной классификации в рамках всего дерева,

$\alpha |\bar{T}|$ — издержки количества терминальных узлов.

Если будет построено очень простое дерево, то издержки размера дерева будут высокими за счет первого слагаемого $R(T)$ (ошибок классификации) которое рассчитывается согласно выражению (9). Если же будет построено слишком большое дерево, то издержки будут высокими за счет второго слагаемого ($\alpha > 0$), а $|\bar{T}|$ — количество терминальных узлов в дереве.

Далее из всего множества деревьев различной размерности необходимо исключать варианты, в которых терминальные узлы не приводят к качественному улучшению модели.

Введем функцию $s_r(t)$, отражающую силу связи узла в дереве:

$$s_r(t) = \frac{R(t) - R(T_{rt})}{|T_{rt}| - 1}, \quad (12)$$

где: $s_r(t)$ — сила узла в дереве,

T_{rt} — количество узлов, связанных с узлом t ,

$R(T_{rt})$ — вероятность некорректной классификации в узлах, связанных с узлом t .

Следующим шагом является нахождение и удаление узлов, у которых значение $s_r(t)$ является минимальным:

$$s_r(t_r^*) = \min_{t \in T_r} s_r(t). \quad (13)$$

Данная процедура продолжается рекурсивно, и получается упорядоченный ряд деревьев, начиная с максимально большого дерева и заканчивая деревом с одним узлом. Далее на каждом этапе «обрезки дерева» обновляется величина α : $\alpha_{(r+1)} = s_r(t_r^*)$. Таким образом, получаются два ряда: упорядоченный ряд деревьев и ряд издержек α_{r+1} .

Затем при помощи процедуры 1 SE rule осуществляется процедура перекрестной проверки на обучающей выборке $L = \{(x_1, j_1), \dots, (x_N, j_N)\}$ в результате которой определяется оптимальное дерево классификаций с минимальной величиной $R_\alpha(T)$.

Построение оптимального портфеля

Прежде чем описывать методику построения оптимального портфеля, следует четко определить, что является критерием оптимальности. Во многих работах под оптимальностью подразумевают превышение доходности построенного портфеля над доходностью рыночного портфеля. В общем случае данное определение некорректно, т.к. ввиду большей диверсификации рыночный портфель обладает меньшим риском и, как следствие, меньшей доходностью.

Поэтому в данной работе оптимальным портфелем называется портфель с заданным риском, доходность которого не ниже доходности рыночного портфеля с таким же параметром риска. Формулировка «не ниже» говорит о том, что в случае наличия слабой формы эффективности рынка оптимальный портфель должен совпасть с рыночным, если он будет базироваться лишь на исторических данных.

В рамках данной работы рассматривался процесс формирования портфеля, состоящего из обыкновенных акций компаний, входящих в базу расчета индекса ММВБ, т.к. они являются наиболее ликвидными инструментами на российском рынке акций. В рамках данной работы анализировалась недельная доходность акции, которая в период определялась следующим образом:

$$R_t = \frac{E_t(P_{t+1}) - P_t}{P_t}, \quad (14)$$

где: R_t — доходность акции в период t ,

P_t — цена акции в период t ,

P_{t+1} — цена акции в период $t+1$,

$E_t(P_{t+1})$ — ожидаемая в периоде t цена акции в периоде $t+1$.

Следует отметить, что выбор именно будущей доходности неслучаен, т.к. является наиболее оптимальным при прогнозировании доходности, потому что на практике алгоритму придется строить именно ожидаемую цену акции при расчете доходности.

Пороговое значение доходности \bar{R} выбиралось для каждой акции индивидуально и лежало в пределах [0%; 5%] – на основе средних показателей доходности за рассматриваемый период.

Ниже представлена таблица со списком переменных, которые использовались в построении деревьев классификации для каждой акции:

Перечень переменных, использованных при анализе

Переменная ¹	Тип	Описание
Цена нефти	Макроэкономическая	Котировка нефти марки Brent
Инфляция	Макроэкономическая	Месячная инфляция рассчитывается на основе индексов потребительских цен
ROE	Фундаментальная	Return on Equity
EPS/P	Фундаментальная	Earnings per Share к цене акции
ΔEPS/P	Фундаментальная	3-х месячное изменение Earnings per Share к цене акции
Sales/P	Фундаментальная	Выручка к цене акции
Momentum	Техническая	$M_t = P_t - P_{t-T}, T$
MA/P	Техническая	$MA(T) = \frac{\sum_{t-T}^t P_t}{T}, T$
MA St. Error	Техническая	Стандартное отклонение MA
Bid-Ask spread,%	Техническая	$\frac{Ask Price - Bid Price}{Ask Price}$
Количество сделок	Техническая	Среднее дневное количество сделок за неделю

Отдельно хотелось бы остановиться на переменной Momentum. В рамках данного анализа под Momentum подразумевается относительное место доходности определенной акции в период t .

Первичное формирование портфеля осуществлялось по следующей методике:

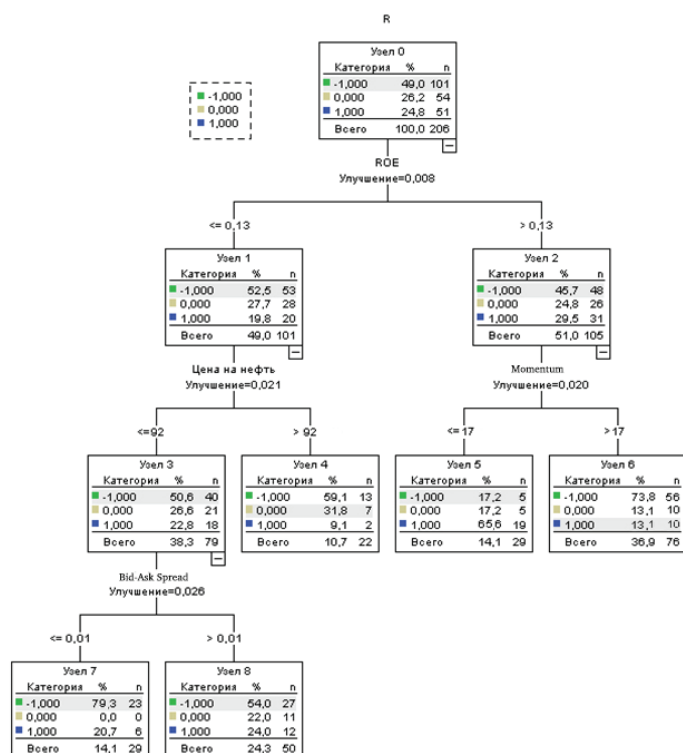
1. Для каждой компании, входящей в базу расчета индекса ММВБ, формировалась матрица входных параметров X_i , включающая себя данные по переменным, представленным в таблице 1, за период с января 2008 г. по декабрь 2014 г. Следует отметить, что по фундаментальным переменным были доступны лишь квартальные данные.
2. На основе данных недельных доходностей за период с января 2008 г. по декабрь 2014 г. формировался вектор классов для каждого наблюдения $C_i = \{(\text{покупка актива}), (\text{продажа актива}), (\text{сохранение актива})\}$ согласно следующему правилу :

$$\left[\begin{array}{l} R_t > \bar{R}, y_t = \{ \text{покупка актива} \} \\ R_t < -\bar{R}, y_t = \{ \text{продажа актива} \} \\ -\bar{R} \leq R_t \leq \bar{R}, y_t = \{ \text{сохранение актива} \} \end{array} \right.$$

Пороговое значение доходности \bar{R} выбиралось для каждой акции индивидуально, на основании средних показателей доходности за рассматриваемый период.

3. На основании векторов C_i и X_i для каждой компании строилось дерево регрессий. Ниже представлен пример дерева регрессий для компании ОАО «Акрон».

1. Данные взяты за период с января 2008 г. по декабрь 2014 г.

Рисунок 2. Дерево регрессий для ОАО «Акрон»¹

- На основе построенных деревьев отбирались акции компаний, которым на момент января 2015 г. был присвоен класс. Разберем данный пункт на примере представленного выше дерева регрессий для компании «Акрон». Если на момент формирования отчета показатель ROE компании превышает 13%, а показатель Momentum был больше 17, то мы попадаем в терминальный узел, в котором большинство наблюдений класса. Следовательно, при формировании портфеля акция «Акрона» попадет в пул компаний, акции которых будут приобретены.
- Определялся вес акций, по которым был присвоен класс {покупка актива}, в портфеле. Вес для отобранных акций определялся путем решения следующей оптимизационной задачи по максимизации дохода при фиксированном риске:

$$\left\{ \begin{array}{l} \max \sum_{i=1}^n w_i r_i \\ \sqrt{\sum_{i=1}^n w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_i w_j k_{ij} \sigma_i \sigma_j} < \bar{\sigma} \\ \sum_{i=1}^n w_i = 1 \\ w_i \geq 0, \end{array} \right. \quad (15)$$

где: r_i — доходность i -й акции,

σ_i — стандартное отклонение доходности i -й акции,

w_i — доля i -й акции в портфеле.

В данной работе параметр предельной склонности к риску задавался экзогенно и равен 10% в годовом исчислении, что соответствует стандартному отклонению индекса ММВБ за анализируемый период.

1. В данном примере -1 означает {покупка актива}; 1 означает {продажа актива}; 0 означает {сохранение актива}.

После первичного формирования инвестиционного портфеля каждые три месяца происходила его ребалансировка, в рамках которой вновь строились деревья регрессии с учетом новых данных. Происходила продажа бумаг компаний, которые были в портфеле и которым в процессе ребалансировки был присвоен класс. Средства, полученные от реализации бумаг, распределялись среди активов с классом {покупка актива} путем решения задачи (15). Период ребалансировки, равный трем месяцам, выбран ввиду того, что каждые три месяца появляются новые данные квартальной отчетности компаний.

Чтобы определить, можно ли при помощи описанной выше методологии построить оптимальный портфель, было проведено сравнение доходности портфеля с доходностью рыночного портфеля аналогичного риска. Ниже представлены результаты тестирования модели на реальных данных за период с января 2015 г. по январь 2016 г.



Рисунок 3. Разницы доходности оптимального и рыночного портфелей

В результате тестирования модели на временном горизонте в 13 месяцев доходность построенного портфеля заданного риска не ниже доходности рыночного портфеля заданного риска, что полностью соответствует данному определению оптимальности.

Но помимо способа построения портфеля дерево регрессий несет в себе и другую важную информацию: относительную важность того или иного признака. Порядок вхождения того или иного признака в структуру дерева не является случайным: в узлах более высокого порядка находятся параметры, позволяющие осуществить разбиение исходного множества на однородные группы. Таким образом, в зависимости от порядка вхождения того или иного фактора в структуру дерева можно провести сравнительный анализ важности того или иного фактора при отборе акций в портфель.

В таблице представлено ранжирование факторов в зависимости от порядка их вхождения в структуру деревьев.

Таблица 2

Ранжирование факторов в зависимости от их значимости

Переменная	Тип	Место фактора в зависимости от порядка вхождения в структуру деревьев
Momentum	Техническая	1
Bid-Ask spread, %	Техническая	2
Цена нефти	Макроэкономическая	3
ROE	Фундаментальная	4
Количество сделок	Техническая	5
ΔEPS/P	Фундаментальная	6
Инфляция	Макроэкономическая	7
Sales/P	Фундаментальная	8

EPS/P	Фундаментальная	9
MA/P	Техническая	10
MA St. Error	Техническая	11

Из таблицы, представленной выше, следует, что наиболее значимой детерминантной доходности российских акций является Momentum акции, уровень ликвидности акции, характеризующийся показателем Bid-Ask spread, а также показатель цены нефти марки Brent.

В целом при анализе влияния факторов можно выдвинуть гипотезу, что российский рынок акций имеет признаки спекулятивного рынка. Для инвесторов одним из наиболее важных факторов является фактор цены на нефть, выступающий главным макроэкономическим индикатором, определяющим состояние российской экономики, а также технические параметры Momentum и параметры ликвидности. Важность параметра Momentum можно трактовать как желание инвестора в краткосрочной/среднесрочной перспективе инвестировать лишь в акции, имеющие высокие относительные показатели доходности в прошлом. Значимость же параметра ликвидности (Bid-Ask spread, %) можно трактовать как желание инвесторов иметь в портфеле наиболее ликвидные акции, чтобы в случае ухудшения состояния экономики осуществить их незамедлительную продажу.

Также следует отметить относительную незначимость фундаментальных показателей, характеризующих деятельность компаний, а также незначимость показателей технического анализа, кроме Momentum.

Заключение

В рамках данной работы был рассмотрен вопрос формирования оптимального инвестиционного портфеля на российском рынке акций при помощи непараметрического метода дерева регрессий. В данной работе оптимальным портфелем назывался портфель с заданным риском, доходность которого не ниже доходности рыночного портфеля с таким же риском.

В результате на основе системы факторов, включающих факторы макроэкономические, технические и фундаментальные, при помощи метода дерева регрессий был построен оптимальный портфель акций российских компаний. Среднее превышение доходности построенного портфеля над рыночным составляло 6%. Таким образом, была показана эффективность непараметрического метода дерева решений как инструмента построения портфеля акций на российском рынке.

Также был проведен сравнительный анализ характеристик, определяющих уровень доходности российских акций. Согласно полученным данным, наиболее значимыми являются следующие факторы: Momentum, Bid-Ask Spread, а также цена на нефть марки Brent. Также была выявлена незначимость факторов фундаментального анализа.

В дальнейшем в качестве расширения представленной модели необходимо учитывать выплаты дивидендов компаниями, а также наличие издержек по ребалансировке портфеля. Также можно рассмотреть вопрос оптимальности портфеля, построенного при помощи метода CART при различных значениях несклонности к риску.

Список литературы

1. Álvarez-Díaz, R., Shawkat, H., and Rangan, G. (2013), "Detecting predictable non-linear dynamics in Dow Jones industrial average and Dow Jones Islamic market indices using nonparametric regressions", *The North American Journal of Economics and Finance*, vol. № 29, pp. 22–35.
2. Breiman, L., Friedman, J., Olshen, A., and Stone, J. (1987), "Classification and regression trees", The Wadsworth Statistics/Probability Series.
3. Chavarnakul, T., and Enke, D. (2009), "A hybrid stock trading system for intelligent technical analysis-based equivolume charting", *Neurocomputing*, vol. № 72, pp. 3517–3528.

4. Diebold, F., and Mariano, R. (1995), “Comparing predictive accuracy”, *Journal of Business and Economic Statistics*, vol. № 13, pp. 253–263.
5. Edwin, E., and Gruber, M. (1997), “Modern portfolio theory, 1950 to date” , *Journal of Banking and Finance*, vol. № 21, pp. 1743–1759.
6. Zhong, M., and Chang, T. (2014), “International Equity Diversification Between the United States and Brics Countries”, *Romanian Journal of Economic Forecasting*, vol. №1, pp. 123–138.
7. Richard, B. (2009), “*Statistical Learning from a Regression Perspective*, Springer Series in Statistics”. New York: Springer-Verlag.
8. Shirashi, E., et al. (2012), “Statistical estimation of portfolios for dependent financial returns”, *Advances in Decision Sciences*, vol. № 34, pp. 226–229.
9. Qin, Z., Li, X., and Ji, X. (2009), “Portfolio selection based on fuzzy cross-entropy”, *Journal of Computational and Applied Mathematics*, vol. № 26, pp. 79–87.

CONSTRUCTING OF OPTIMAL PORTFOLIO ON RUSSIAN STOCK MARKET USING NONPARAMETRIC METHOD – CLASSIFICATION AND REGRESSION TREE

Buyanova Elena,

*docent, Candidate of Physico-Mathematical Sciences,
assistant professor of the Department of Finance NRU HSE*

Sarkisov Artur,

Ph.D. student of the Department of Finance at NRU HSE

Abstract

Constructing of optimal portfolio is one of the most important subject in the modern finance theory. The main goal in constructing of optimal portfolio is to create model, which will be able to forecast dynamic of the asset's return. However, the problem of creating such model is still actual. In many works there are models, which perfectly explained past dynamic of the asset's return, but don't match to forecast it. One of the reason of that is sometime there are no appropriate data for analysis (not enough observations or «noisy» data). In this case traditional parametric methods will give biased estimation.

In this paper we used nonparametric method (classification and regression tree) for analyzing 50 stocks, which are included in calculation base of MICEX stock index. Using of this method allowed us to use not only macroeconomic and technical factors, but also factors with limited data set (factors of fundamental analysis). As a result we constructed optimal portfolio with average return on 7% higher than market portfolio with the same risk. It should be mentioned that in this research we don't use traditional definition of optimal portfolio which implies comparison of constructed portfolio with market portfolio. In general case this comparison is not correct because market portfolio is more diversified and, hence, have less risk. Therefore we use another definition of optimal portfolio: optimal portfolio is the portfolio with certain risk which return is not less than the return of market portfolio with the same risk.

Also in this research we conducted a comparative analysis of the influence of the factors on stock return. As the result we showed that Russian stock market has features of speculative market, because the most important factors for stock return of Russian stocks are momentum, bid – ask spread and oil price.

Keywords: optimal portfolio, fundamental analysis, technical analysis, portfolio selection, nonparametric methods

JEL: G10, G11, G17, C14

References

1. Álvarez-Díaz, R., Shawkat, H., and Rangan, G. (2013), “Detecting predictable non-linear dynamics in Dow Jones industrial average and Dow Jones Islamic market indices using nonparametric regressions”, *The North American Journal of Economics and Finance*, vol. № 29, pp. 22–35.
2. Breiman, L., Friedman, J., Olshen, A., and Stone, J. (1987), “*Classification and regression trees*”, The Wadsworth Statistics/Probability Series.
3. Chavarnakul, T., and Enke, D. (2009), “A hybrid stock trading system for intelligent technical analysis-based equivolume charting”, *Neurocomputing*, vol. № 72, pp. 3517–3528.
4. Diebold, F., and Mariano, R. (1995), “Comparing predictive accuracy”, *Journal of Business and Economic Statistics*, vol. № 13, pp. 253–263.
5. Edwin, E., and Gruber, M. (1997), “Modern portfolio theory, 1950 to date”, *Journal of Banking and Finance*, vol. № 21, pp. 1743–1759.

6. Zhong, M., and Chang, T. (2014), “International Equity Diversification Between the United States and Brics Countries”, *Romanian Journal of Economic Forecasting*, vol. №1, pp. 123–138.
7. Richard, B. (2009), “*Statistical Learning from a Regression Perspective*, Springer Series in Statistics”. New York: Springer-Verlag.
8. Shirashi, E., et al. (2012), “Statistical estimation of portfolios for dependent financial returns”, *Advances in Decision Sciences*, vol. № 34, pp. 226–229.
9. Qin, Z., Li, X., and Ji, X. (2009), “Portfolio selection based on fuzzy cross-entropy”, *Journal of Computational and Applied Mathematics*, vol. № 26, pp. 79–87.