Министерство науки и высшего образования
Российской Федерации
«Московский физико-технический институт
(государственный университет)»
Физтех-школа прикладной математики и информатики
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

Рыбка Елизавета Михайловна

Регрессионные модели с ограничениями и регуляризацией при больших наборах регрессоров. Задача восстановления состава инвестиционного портфеля

03.03.01 — Прикладные математика и физика

Выпускная квалификационная работа
(бакалаврская диссертация)

Научный руководитель:
д.т.н. профессор
Моттль Вадим Вячеславович

Москва
2018

Аннотация

Данная работа посвящена конкретной прикладной задаче восстановления скрытого состава инвестиционного портфеля, представленного временным рядом его периодических доходностей (относительные приращения стоимости портфеля). Эта задача регрессионного оценивания при совокупности дополнительных допущений. Во-первых, предполагается, что коэффициенты регрессии дважды ограничены: отдельным неравенствами неотрицательности вместе с равенством единице их суммы. Во-вторых, предполагается, что число регрессоров намного превышает размер выборки. В-третьих, предполагается, что коэффициенты регрессии отличаются от нуля только в пределах реально существующего малого подмножества большого универсума регрессоров, и поиск этого подмножества (Factor Search) является основной целью обработки данных.

Factor Serach невозможен без использования некоторой априорной информации об ожидаемой структуре портфеля. Исследуется регрессионная модель, построенная в соответствии с теорией оптимального состава портфеля Гарри Марковица. Предлагается способ дополнить технический анализ прошлых доходностей активов экспертным мнением о предпочтительном скрытом составе портфеля. Полученная регрессионная модель сравнивается с необогащенной на модельных портфелях.

Оглавление

1. Введение

Данная работа продиктована конкретной прикладной задачей инвестиционного менеджмента. Математические ограничения основаны на особенностях области и требованиях специалистов. Поэтому для целостности повествования начнем с введения определений из области финансов и переложим их на математический язык.

Финансовый актив (далее просто актив) — некое экономическое средство, которое обеспечивает денежные поступления в форме прямых или скрытых выплат. К последним относятся, например, изменение стоимости недвижимости, акций или самой компании [9]. Как видно из определения актив в первую очередь характеризуется своей стоимостью (естественно меняющейся во времени). Поэтому i-ый актив разумно описывать как временной ряд $(z_{1,i}, \cdots, z_{T,i})$, где $z_{t,i}$ — цена актива в момент времени t.

Доходность — применяемый в экономике относительный показатель эффективности вложений в те или иные активы. Т.е. доходность i-ого актива:

$$x_{t,i} = \frac{z_{t,i} - z_{t-1,i}}{z_{t-1,i}}, \qquad t = 1, \cdots, T$$

В финансах ***портфель*** — это набор активов различной доходности, удерживаемых инвестиционной компанией, хедж-фондом, финансовой организацией или индивидуальным предпринимателем[9]. Таким образом, когда говорят о портфеле, под сути, говорят о том, как капитал вложен в некий набор активов $i = 1, \cdots, n$. Долю капитала, которая вложена конкретно в i-ый актив, будем обозначать β_i. Каждый актив характеризуется доходностью $\mathbf{x}_i$. Она может быть отрицательна в некоторые моменты времени, если актив убыточен. Весь портфель также характеризуется доходностью $\mathbf{y} = (y_1, \cdots, y_T)$, которая складывается из доходностей активов, которые в него входят $\mathbf{y} = \sum_{i=1}^{n} \beta_i x_i$

Переход к доходности обусловлен тем, что различные инвестиционные организации обязаны отчитываться перед регулирующими органами именно по этому показателю. Цена портфеля, как и его состав, является конфиденциальной информацией, доходность находится в общественном доступе.

Обратим внимание, что мы неявно предположили, что долевое распределение капитала не изменяется за рассматриваемый период времени (β_i не зависит от t). Это сужает прикладное применение выстроенной в работе теории. Однако возможность её применения для анализа случаев, когда держатель портфеля не является активным игроком на рынке всё ещё представляет большой практический интерес. Это применимо к так называемые консервативным портфелям.

Отметим, что, вообще говоря, доля капитала, вложенная в i-ый актив, может быть отрицательна $\beta_i < 0$. Эта математическая постановка соответствует случаю, когда деньги или активы могут быть заняты из внешних источников. Подобная стратегия возможна для так называемых хедж-фондов, слаборегулирумых инвестиционных компаний[4]. В данной работе мы не будем затрагивать подобные организации, а ограничимся исследованием только класса взаимных (паевых) инвестиционных фондов, которым разрешено формировать портфели, используя только свой внутренний капитал[4]. Это ограничение хорошо сочетается с предыдущим предположением о стационарности β_i в силу природы инвестиционного стиля типичного для указанных организаций.

Еще одно ограничение, исходит из природы экономического процесса. Если рассматривать капитал просто как некоторое количество наличных денег, то необходимость вложить его полностью в активы не кажется очевидной. Однако в силу того, что всегда существует такое явление как инфляция, капитал никогда не хранится в валюте. Та часть капитала, которая не используется для инвестиций, хранится в безрисковом активе, имеющем среднюю доходность по экономике, т.е. покрывающем инфляцию, но не приносящего никакого реального дохода. Поэтому можно считать, что β_i — доли общего капитала, вложенного *полностью* в ценные бумаги $\sum_{i=1}^{n} \beta_i = 1$

Таким образом, основное уравнение, описывающее долевое распределение портфеля выглядит, следующим образом:

$$\begin{cases} \mathbf{y} = \sum_{i=1}^{n} \beta_i x_{t,i} = \mathbf{X}^T \boldsymbol{\beta} \\ \sum_{i=1}^{n} \beta_i = 1 \\ \beta_i \geqslant 0, \qquad i = 1, \cdots, n \end{cases} \tag{1}$$

, где $\mathbf{y} = (y_1, \cdots y_T) \in \mathbb{R}^T$ — доходности портфеля

$\mathbf{X} = (\mathbf{x}_1, \cdots \mathbf{x}_T) \in \mathbb{R}^{n \times T}$ — доходности активов

$\boldsymbol{\beta} = (\beta_1, \cdots \beta_n) \in \mathbb{R}^n$ — долевое распределения капитала

Менеджеры портфелей, как и любой высокооплачиваемый специалист своей области, скрытны в отношении операций покупки-продажи активов, которые они выполняют, руководствуясь своим экспертным мнением. Дробная структура портфеля $\boldsymbol{\beta}$ не является публичной информацией. Однако, согласно нормативно-правовому регламенту области в общественном доступе находятся временной ряд доходности портфеля $\mathbf{y}$ и временные ряды доходностей всевозможных активов $\mathbf{X}$. Если по имеющейся информации восстановить долевое распределение портфеля, то полученная информация будет представлять большой интерес для сторонних наблюдателей. Наличие реального спроса на подобные интелектуальные решения обеспечивает, например, успех компании Markov Processes

International, Inc. [7]. Их система инвестиционного исследования, анализа и оповещения широко используется институциональными инвесторами, консультантами, управляющими активами и помощниками по пенсионному плану, для совершения инвестиционных исследований, портфельного строительства и оптимизации, анализ производительности, наблюдения за рисками.

2. Постановка задачи

2.1. Анализ инвестиционных портфелей на основе доходности. Модель Уильяма Шарпа

Проблема восстановления скрытого распределения капитала в портфеле на основе общедоступных данных была сформулирована Уильямом Шарпом, лауреатом Нобелевской премии 1990 года по экономике [8].Его метод предполагает анализ временных рядов периодической доходности портфеля, о которых компания обязана отчитываться, совместно с синхронными временными рядами доходности на фондовом рынке активов, предполагаемых к формированию портфеля.

Поскольку фактический портфель может содержать сотни и даже тысячи инструментов, Шарп предложил аппроксимировать полученную доходность портфеля небольшим количеством рыночных индексов[4] представляющих определенные классы активов и инвестиционные стили. Этот подход известен под названием Анализ Стиля на Основе Доходности, Returns Based Style Analysis, RBSA.

Если рассматриваются классы активов, то их ожидаемая доходность обычно оценивается специальными аналитическими компаниями как индекс доходности.

Для последовательных временных моментов составляющих последовательность периодов владения, например, рабочие дни, месяцы и кварталы фондовой биржи, периодические доходности как портфеля, так и активов образуют ряд временных рядов, соответственно y_t и $x_{t,i}$, $i = 1, \ldots, n$. Эти значения можно считать известными, так как любая инвестиционная компания должна регулярно публиковать доходность своего портфеля, а доходности классов активов регулярно публикуются как онлайн, так и в печатных финансовых СМИ. В простейшем случае ежедневной периодичности доходность каждого отдельного актива может быть немедленно рассчитана на основе известных изменений его цены.

В модели Шарпа периодическая доходность портфеля примерно равна линейной комбинации периодических доходностей активов или классов активов с коэффициентами, имеющими значение долей портфеля, вложенных в каждый из них в начале периода, при условии, что весь бюджет полностью потрачен на инвестиции. Т. е. доходность портфеля приближается линейной моделью от доходностей активов его составляющих $\mathbf{y} \cong f(\mathbf{x})$. Отличным средством для нахождения соответствующих оценок коэффициентов регрессии является метод минимизации остаточной суммы квадратов, известный как метод наименьших квадратов.

Итак, математическая постановка задачи восстановления инвестиционного портфеля на основе доходностей, может быть записана следующим образом:

$$\begin{cases} \hat{\boldsymbol{\beta}} = \arg\min \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{n} \beta_i x_{t,i}\right)^2 & = \left(\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}\right)^T \left(\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}\right) \\ \sum_{i=1}^{n} \beta_i & = \mathbf{1}^T \boldsymbol{\beta} = 1 \\ \beta_i \geqslant 0, & i \in \mathbb{I} = \{1, \cdots, n\} \end{cases} \tag{2}$$

Задача регрессии с ограничениями (2) относится к классу квадратичных задач с линейными ограничениями.

2.2. Регуляризация

Вернемся к системе (2). Ее решение $\mathbf{X}\mathbf{X}^T\boldsymbol{\beta} = \mathbf{X}\mathbf{y}$ дает единственную оценку коэффициентов регрессии только если регрессоры линейно назависимы, т.е. $\det \mathbf{X}\mathbf{X}^T \neq 0$; что в свою очередь возможно только для случая $n \leqslant T$. Однако, как уже упоминалось, в рассматриваемом практическом приложении типична ситуация, когда число регрессоров (число возможных активов n) много больше размера выборки (длительности наблюдения T) $n \gg T$. В этом случае можно собрать портфель идентичный по доходности исследуемому, но не имеющему ничего общего с ним по составу. Поэтому необходимо учитывать некие априорные представления об ожидаемой регрессионной модели. Типичным способом введения их в МНК является добавление функции регуляризации $V(\boldsymbol{\beta})$. Эта штрафная функция отражает нежелательность некоторых значений $\boldsymbol{\beta}$. Чем менее правдоподобным предполагается значение $\boldsymbol{\beta}$ из априорных знаний, тем больше должно быть значение $V(\boldsymbol{\beta})$. Конкретный вид штрафной функции мы обсудим в последующих разделах.

$$\begin{cases} \hat{\boldsymbol{\beta}} & = \arg\min \left[V(\boldsymbol{\beta}) + \left(\mathbf{y} - \mathbf{X}^T\boldsymbol{\beta}\right)^T \left(\mathbf{y} - \mathbf{X}^T\boldsymbol{\beta}\right) \right] \\ \mathbf{1}^T\boldsymbol{\beta} & = 1 \\ \beta_i \geqslant 0, & i \in \mathbb{I} = \{1, \cdots, n\} \end{cases} \tag{3}$$

2.3. Factor-Search

Итак, множество возможных активов велико $\mathbb{I} = \{1, \ldots, n\}$. Одну из основных целей обработки данных по доходности можно сформулировать через Factor-Search — нахождение реально входящих в состав портфеля активов среди очень большого количества потенциальных факторов-активов. Т.е. через введение понятия активного подмножества $\hat{\mathbb{I}} \subset \mathbb{I}$ такого, что оценки коэффициентов регрессии в модели портфеля на основе доходности (2), положительны внутри активного подмножества, и равны нулю вне него. Factor-Search — это поиск $\hat{\mathbb{I}}$. Можно далее сказать, что ищется модель портфолио вида:

$$\begin{cases} y_t = \sum_{i \in \hat{\mathbb{I}}} \beta_i x_{t,i}, \quad t = 1, \cdots, T \\ \sum_{i \in \hat{\mathbb{I}}} \beta_i = 1 \\ \beta_i > 0, i \in \hat{\mathbb{I}} \end{cases} \tag{4}$$

Эту задачу можно называть выбором регрессионной модели с двойным ограничением. В любом случае, выбор регрессии подразумевает своего рода регуляризацию, уменьшение свободы изменения коэффициентов регрессии при минимизации остатка по критерию наименьших квадратов (2). Несмотря на то, что ограничения равенства и неравенства сами по себе несут в себе довольно большую регуляризацию [3], только они недостаточны для наделения критерия свойством Factor-Search при сравнении временных рядов доходности портфеля с большими наборами индексов фондового рынка. Поэтому в следующем Разделе 2.4 учитываются, помимо естественных ограничений, дополнительные априорные предложения по рациональности состава искомого неизвестного портфеля с точки зрения его полезности как финансового инструмента. В Разделе 3 мы рассмотрим конкретные виды регуляризации приводящие к Factor-Search.

2.4. Априорные предположения о составе портфеля и его оптимальность

Разумно предполагать разумность портфеля. Будем исходить из предположения, что портфель, временной ряд доходности $(y_t, t = 1, \ldots, T)$ которого анализируется, был составлен командой его менеджеров с целью получения *большей доходности при приемлемом уровне риска.*

Доходности активов $(\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,n}), t = 1, \ldots, T)$ будем считать совокупностью реализаций случайного вектора $\mathbf{x}$ с математическим ожиданием $\overline{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\} \in \mathbb{R}^n$ и ковариационной матрицей $\mathbf{\Sigma} = \mathbb{E}\{(\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T\}$. В этом случае доходность портфеля $y_t = \boldsymbol{\beta}^T \mathbf{x}_t$ есть также реализация стационарного случайного процесса, причем имеющего математическое ожидание $\overline{y} = \boldsymbol{\beta}^T \overline{\mathbf{x}}$ и дисперсию $\sigma^2 = \boldsymbol{\beta}^T \mathbf{\Sigma} \boldsymbol{\beta}$. Эти величины имеют финансовую интерпретацию:

- $\overline{y}$ — ожидаемая доходность
- σ^2 — создает риск потерь составленного портфеля

При составлении портфеля балансируются два факта. С одной стороны, чем выше ожидаемая доходность $\overline{y}$, тем лучше. С другой, увеличение дисперсии, приводит к увеличению вероятности на практике получить низкую или отрицательную доходность.

Гарри Марковиц в с своей теории[5], принесшей ему Нобелевскую премию, утверждает, что портфель$(\boldsymbol{\beta} \geqslant \mathbf{0}; \mathbf{1}^T \boldsymbol{\beta} = 1)$ можно считать рациональным только, когда:

- $\boldsymbol{\beta}^T \mathbf{\Sigma} \boldsymbol{\beta} \to \min; \quad \overline{\mathbf{x}}^T \boldsymbol{\beta} = const$
- $\overline{\mathbf{x}}^T \boldsymbol{\beta} \to \max; \quad \boldsymbol{\beta}^T \mathbf{\Sigma} \boldsymbol{\beta} = const$

По сути это *равновесие по Парето* — такое состояние системы, при котором значение каждого частного показателя, характеризующего систему, не может быть улучшено без ухудшения других.

Оба этих варианта покрываются следующей формулой:

$$U(\boldsymbol{\beta}|\mu) = \mu\overline{\mathbf{x}}^T\boldsymbol{\beta} - \boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta} \to \max \tag{5}$$

Или

$$V(\boldsymbol{\beta}|\mu) = \boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta} - \mu\overline{\mathbf{x}}^T\boldsymbol{\beta} \to \min \tag{6}$$

Примем следующею терминологию:

$U(\boldsymbol{\beta}|\mu)$ — Portfolio Utility, портфельная функция полезности;

$\mu \in [0; +\infty)$ — Risk Tolerance, толерантность к риску;

$\frac{1}{\mu}$ — Risk Aversion, неприятие риска;

Портфели, которые не отвечают условию (6) не являются обоснованными по теории Марковица. Но используя только эту теорию нельзя ничего сказать о толерантности к риску μ. Это внутренняя психологическая характеристика инвестора. μ является гиперпараметром данной функции регуляризации (6).

3. Регуляризация для Factor-Search

3.1. Central Regularization

Как упоминалось выше (Раздел 2.3) регуляризация $\boldsymbol{\beta} \geqslant \mathbf{0}; \mathbf{1}^T\boldsymbol{\beta} = 1$ не приводит к эффекту factor-search. Основная причина — сильная коррелированность регрессоров. Это приводит к едва различимости по доходности существенно разных по составу комбинаций активов ($x_i,\ i \in \hat{\mathbb{I}} \subset \mathbb{I}$).

Эту трудность можно существенно ослабить, если априорно предположить, что регрессоры искомой комбинации слабо коррелируют между собой. Это предположение обосновано, т.к. при составлении портфеля менеджеры, как правило, стремятся к диверсификации. Чем меньше корреляция между случайными величинами, тем меньше дисперсия их линейной комбинации с фиксированной нормой вектора коэффициентов. Таким образом, для осуществления этого априорного предположения, мы должны минимизировать:

$$\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} \to \min; \quad \boldsymbol{\beta}^T \boldsymbol{\beta} = const$$

Однако, это только один вид возможных априорных предположений об искомом подмножестве регрессоров. В общем случае любая положительно определенная матрица может выражать интуицию пользователя. Эта причина приводит к *квадратичной регуляризации*:

$$\begin{cases} \boldsymbol{\beta}^T \mathbf{G} \boldsymbol{\beta} + c \left(\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}\right)^T \left(\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}\right) \to \min(\boldsymbol{\beta}) \\ \mathbf{1}^T \boldsymbol{\beta} = 1 \\ \boldsymbol{\beta} \geqslant \mathbf{0} \end{cases} \tag{7}$$

Матрица квадратичной функции регуляризации $\mathbf{G}$ указывает на некоторое априорно предпочтительное направление в n-размерном пространстве регрессоров. Точка минимума квадратичной регуляризирующей функции — нулевой вектор $\boldsymbol{\beta} = \mathbf{0}$, центр пространства. Поэтому такую регуляризацию назовем *центральной регуляризацией*.

3.2. Biased Regularization

Однако направление может быть не единственным видом априорной информации, может быть известно и конкретное значение вектора коэффициентов регрессии. Например, в случае если мы знаем, что количество коэффициентов регрессии не равных нулю, полученное из (7), больше числа на самом деле активных регрессоров. Это может быть сделано следующим образом:

$$(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{G}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \to \min, \quad \boldsymbol{\beta}_0 < 0 \tag{8}$$

, когда смещение в отрицательный квадрант (запрещенную неравенством $\boldsymbol{\beta} \geqslant \mathbf{0}$), поощряет нулевые значения коэффициентов регрессии. Этот вариант регуляризации будем называть *смещенной регуляризацией.* Чем глубже $\boldsymbol{\beta}_0$ находится в запретной зоне, тем сильнее поощряется функцией регуляризации (8) уменьшение числа активных регрессоров. Если раскрыть эту функцию и убрать константную часть, то получим: $\boldsymbol{\beta}^T \mathbf{G}\boldsymbol{\beta} - \boldsymbol{\beta}_0^T \mathbf{G}\boldsymbol{\beta} \to \min$. В итоге приходим к следующей задаче регрессии с ограничениям:

$$\begin{cases} \boldsymbol{\beta}^T \mathbf{G}\boldsymbol{\beta} + \boldsymbol{\beta}_0^T \mathbf{G}\boldsymbol{\beta} + c\left(\mathbf{y} - \mathbf{X}^T\boldsymbol{\beta}\right)^T \left(\mathbf{y} - \mathbf{X}^T\boldsymbol{\beta}\right) \to \min(\boldsymbol{\beta}) \\ \mathbf{1}^T\boldsymbol{\beta} = 1 \\ \boldsymbol{\beta} \geqslant \mathbf{0} \end{cases} \tag{9}$$

3.3. Risk-Aversion Regularization

Теперь легко показать, что функция регуляризации (6), построенная на основе теории Марковица рациональности портфелей, есть частный случай смещенной регуляризации (biased regularization). Перепишем в явном виде:

$$V(\boldsymbol{\beta}|\mu) = \boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta} - \mu\overline{\mathbf{x}}^T\boldsymbol{\beta} = (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T\mathbf{G}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + const \qquad (10)$$

$\boldsymbol{\beta}_0 = -\frac{\mu}{2}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}$,
$\mathbf{G} = \boldsymbol{\Sigma}$

Поведение матрицы $\boldsymbol{\Sigma}$ будет рассмотрено в экспериментальной части (Раздел 5).

4. Регуляризация на основе мнений экспертов

4.1. Аргументация

Как пояснялось во Введении, мы рассматриваем консервативные портфели. Этот вид рассчитан на длительный период, обычно от пяти до десяти лет, и состоящий, в основном, из крупных, авторитетных компаний с устойчивыми прогнозами роста и сравнительно низким риском. Т.е при составлении портфеля помимо технической информации, которую можно получить из анализа последовательности доходностей активов, учитывается и экспертная информация. Таким образом, рационально и необходимо включение в модель некоторого регуляризатора отвечающего этой информации, что мы и сделаем в следующем параграфе.

4.2. Expert Based Regularization

Общий вид разумных функций регуляризации это (9). Мы рассматривали случай когда G зависят от данных. Теперь попробуем взять их из сторонних источников. Будем основывать построение Expert-Based Regularization на аналогии с Risk-Aversion (10). **Предлагается** — дополнить оценки ковариационной матрицы и среднего вектора доходностей активов $(\mathbf{\Sigma}, \overline{\mathbf{x}})$ матрицей и вектором той же структуры $(\mathbf{B}; \mathbf{z})$, формализующих мнение эксперта о разумном составе портфеля.

Пусть эксперт выразил свое мнение относительно:

Волатильности доходностей активов	Несовместности пар активов	Предпочтительном составе портфеля
$\mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} \in \mathbb{R}^n$	$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \dots & r_{1n} \\ & 1 & \dots & r_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}$	$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$

Агрегированная матрица: $\mathbf{B} = [\mathrm{Diag}(\mathbf{d})]^T \mathbf{R} [\mathrm{Diag}(\mathbf{d})]$

Фактически эти матрицы определяют суждение эксперта о скрытом портфеле:

$$\begin{cases} \boldsymbol{\beta}^T \mathbf{B} \boldsymbol{\beta} - \mu \mathbf{z}^T \boldsymbol{\beta} \to \min(\boldsymbol{\beta}) \\ \mathbf{1}^T \boldsymbol{\beta} = 1; \qquad \boldsymbol{\beta} \geqslant \mathbf{0} \end{cases} \tag{11}$$

$0 \leqslant \mu < \infty$ степень недоверия к доходностям активов

Обратим внимание, что матрица $\mathbf{B}$ имеет ту же структуру, что и матрица $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_{nn} \end{pmatrix} = \begin{pmatrix} \sigma_1\sigma_1\rho_{11} & \dots & \sigma_1\sigma_n\rho_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_n\sigma_1\rho_{n1} & \dots & \sigma_n\sigma_n\rho_{nn} \end{pmatrix} =$$
$$= \begin{pmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_{21} & \dots & \sigma_1\sigma_n\rho_{1n} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_n\sigma_1\rho_{n1} & \dots & \dots & \sigma_n^2 \end{pmatrix} \tag{12}$$

σ_{ij} — ковариация случайных величин, заданных векторами реализаций

$\mathbf{x}_i$ и $\mathbf{x}_j$; $\;i, j = \overline{1, n}$

σ_i — среднеквадратичное отклонение $\mathbf{x}_i$

ρ_{ij} — коэффициент корреляции $\mathbf{x}_i$ и $\mathbf{x}_j$. Всегда верно $\rho_{ij} \in [-1; 1]$. Последний переход в (12) сделан в силу известного равенства $\rho_{i,i} = 1$

Наше предложение заключается в том, чтобы вместо вычисления σ_{ij}, ρ_{ij} через значения векторов доходностей $\mathbf{x}_i$ использовать d_i, r_{ij} заданные экспертами. Для этого поймем, какой смысл эти величины принимают в регуляризаторе:

$$\boldsymbol{\beta}^T \mathbf{G} \boldsymbol{\beta} \to min(\boldsymbol{\beta}) \tag{13}$$

Предпочтительнее активы β_i, соответствующие малым значениям d_i. Т.е. d_i — есть *неприятие* i-го актива; насколько, по мнения эксперта, не вероятно включение i-го регрессора в активное множество $\hat{\mathbb{I}}$. Также легко видеть, что r_{ij} имеет смысл неприятия одновременного вхождения i-го и j-го; их *несовместимости*. (Для удобства ограничим значение r_{ij}, задаваемых экспертом, промежутком $[0; 1]$)

Таким образом, мы ввели следующую матрицу, определяющую функцию ошибок, основанную на мнение экспертов:

$$\mathbf{B} = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix} = \begin{pmatrix} d_1^2 & d_2 d_1 r_{12} & \dots & d_1 d_n r_{1n} \\ d_1 d_2 r_{12} & d_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ d_1 d_n r_{1n} & \dots & \dots & d_n^2 \end{pmatrix} \tag{14}$$

А сама функция ошибки имеет вид:

$$W(\boldsymbol{\beta}) = \boldsymbol{\beta}^T \mathbf{B} \boldsymbol{\beta} = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{i,j} \beta_i \beta_j \tag{15}$$

Более того, мы предложили удобный и интерпретируемый для эксперта способ формирования матрицы $\mathbf{B}$ через σ_i, r_{ij}.

Проанализируем введенные величины.

Утверждение 1. $b_{i,j} = b_{j,i} \geqslant 0$. *Т.е.* $\mathbf{B}$ *симметрична.*

Утверждение 2. $\mathbf{B}$ *положительно определена в* $\mathbb{R}^{n+}$

Это утверждение очевидно, если обратить внимание на то, что слагаемые соответствующей квадратичной формы (15) неотрицательны при $\beta \in \mathbb{R}^{n+}$

Утверждение 3. $W(\boldsymbol{\beta})$ *выпукла в* $\mathbb{R}^{n+}$

Т.к. является квадратичной формы положительно определенной матрицы в $\mathbb{R}^{n+}$.

Вектор $\mathbf{z}$ имеет смысл аналогичный вектору $\overline{\mathbf{x}}$. Он выражает желание включить в портфель отдельный актив. Чем выше средняя доходность i-го актива $\overline{x}_i$, тем желательнее он с точки зрения технического анализа. Аналогично большее значение z_i означает предпочтительность i-го актива с точки зрения эксперта. Т.е. Expert Based Regularization есть баланс между волатильностью активов, их несовместимостью и их предпочтительностью.

Итоговая задача оптимизации приобретает вид:

$$
\begin{cases}
\boldsymbol{\beta}^T \underbrace{\Big((1-\alpha)\boldsymbol{\Sigma} + \alpha\mathbf{B}\Big)}\boldsymbol{\beta} - \mu \underbrace{\Big((1-\alpha)\overline{\mathbf{x}} + \alpha\mathbf{z}\Big)^T}\boldsymbol{\beta} + \\
\qquad\qquad + c\left(\mathbf{y} - \mathbf{X}^T\boldsymbol{\beta}\right)^T\left(\mathbf{y} - \mathbf{X}^T\boldsymbol{\beta}\right) \to \min(\boldsymbol{\beta}) \\
\mathbf{1}^T\boldsymbol{\beta} = 1 \\
\boldsymbol{\beta} \geqslant \mathbf{0}
\end{cases}
\tag{16}
$$

Коэффициент α отражает мнение о том? чем в большей степени руководствовался менеджер данного конкретного портфеля при его составлении — объективными данными или репутационными. По этой причине назовем его *Pliability, внушаемостью.*

Отметим также тот факт, что в вышеописанном алгоритме построения $\mathbf{B}$, $\mathbf{z}$ элементы выбирались из промежутка $[0; 1]$, однако когда мы переходим к смешенному виде регуляризации (16) необходимо вводить дополнительную нормировку для приведение $\mathbf{B}$ и $\boldsymbol{\Sigma}$, $\overline{\mathbf{x}}$ и $\mathbf{z}$ к одному масштабу.

5. Экспериментальная часть

5.1. Описание данных

В качестве набора регрессоров используются 650 временных рядов месячных индексов фондового рынка

$$\mathbf{x}_t = (x_{t,i};\ i = 1, \ldots, n) \in \mathbb{R}^n, \quad n = 650$$

каждый длиной 20 лет, состоящий из 240 значений доходности $t = 1, \dots, T$, T=240:

$$\mathbf{X} = (\mathbf{x}_t,\ t = 1, \dots, T) \in \mathbb{R}^{n \times T}, \quad n \times T = 650 \times 240$$

Помещенные в порядке убывания, собственные значения их ковариационной матрицы $\mathbf{\Sigma}$ быстро убывают $\lambda_1 = 1639.8,\ \lambda_2 = 1076.2,\ \dots$, $\lambda_{20} = 134.0 < 0.01\lambda_1, \lambda_{50} = 32.2 < 0.002\lambda_1$ что свидетельствует о сильной корреляции между индексами.

5.2. Исследование Expert Based Regulariszation

Первый эксперимент направлен на то, чтобы показать отбор регрессоров при экспертном задании $\mathbf{B},\ \mathbf{z}$.

Предположим, что эксперт разбил это множество на $m = 15$ групп, объединяющих активы, сходные по некоторым социально-деловым представлениям: $\mathbb{I} = \mathbb{I}_1 \cup \dots \cup \mathbb{I}_m,\ \mathbb{I}_k \cap \mathbb{I}_l = \varnothing, k \neq l$. Эксперт предполагает, что в портфель вряд ли входят более одного представителя каждой группы. В терминах матрицы $\mathbf{R}$ несовместимости пар активов это означает, что штраф $r_{i,j}$ для i, j из одной группы должен быть очень большим, например 0.9999, поскольку $r_{i,i} = 1$.

Строим матрицу $\mathbf{B}$ согласно описанному Разделе 4.2 методу, вектор $\mathbf{z}$ задаем так, что чем меньше порядковый номер индекса, тем он предпочтительнее.

Находим оптимальный состав портфеля (набор бета-коэффициентов $\boldsymbol{\beta}^*$) с точки зрения экспертного мнения, решая следующую оптимизационную задачу:

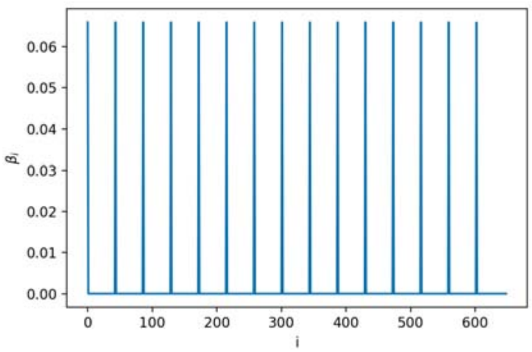

Рис. 1. Долевое распределение портфеля при неприятии активов из одной группы при отсутствии предпочтений между группами; $\mu = 0$

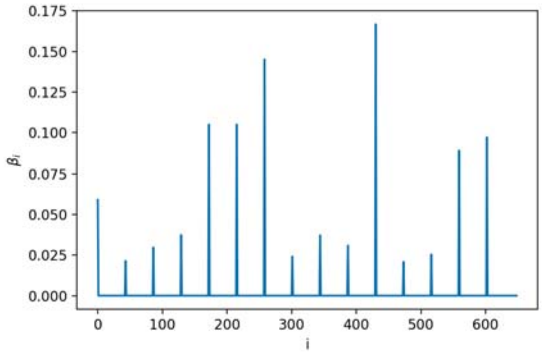

Рис. 2. Долевое распределение портфеля при неприятии активов из одной группы при наличии предпочтений между группами; $\mu = 0$

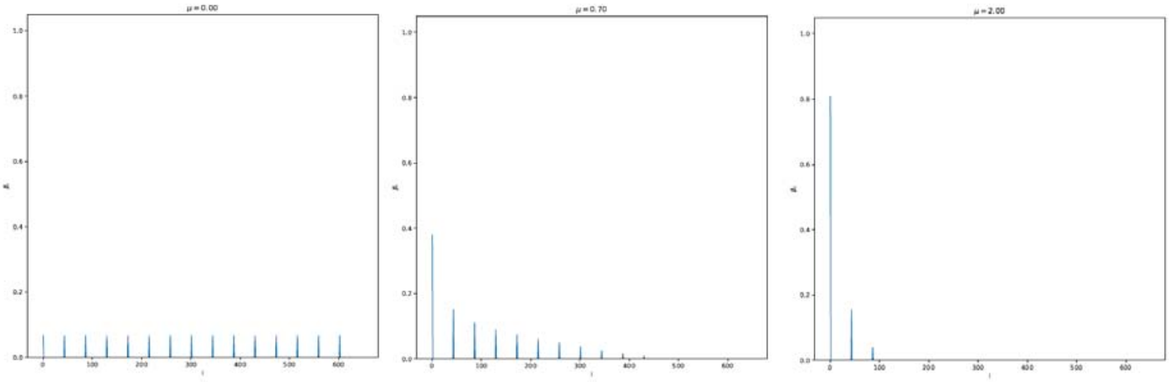

Рис. 3. Долевое распределение экспертного портфеля для различных $\mu = (0;\ 0.7;\ 2)$

$$\begin{cases} \boldsymbol{\beta}^* = \boldsymbol{\beta}^T \mathbf{B} \boldsymbol{\beta} - \mu \mathbf{z}^T \boldsymbol{\beta} \\ \mathbf{1}^T \boldsymbol{\beta} = 1; \qquad \boldsymbol{\beta} \geqslant \mathbf{0} \end{cases} \tag{17}$$

Состав полученных портфелей представлен на Рис.1 и Рис.2. Влияние μ на состав портфеля на Рис.3.

5.3. Сравнение моделей

Из вышеописанного на экспериментальное рассмотрение мы выносим следующие регрессионные модели:

- Naiv (2)

- Risk Aversion (10)
- Expert Based (16)

Суть основного эксперимента заключается в исследования способности различных регуляризаторов восстанавливать некоторый *истинный* набор бета-коэффициентов $\boldsymbol{\beta}^* = (\beta_1^*, \dots, \beta_n^*)$ задающих разумный инвестиционный портфель, составленный с учетом экспертно мнения.

Такой портфель для фиксированных $\mu \in [0, \infty), \alpha \in [0, 1]$, задается:

$$\begin{cases} \boldsymbol{\beta}^*_{\alpha,\mu} = \arg\min\left(\alpha\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta} + (1-\alpha)\boldsymbol{\beta}^T\mathbf{B}\boldsymbol{\beta} - \mu\overline{\boldsymbol{x}}^T\boldsymbol{\beta}\right) \\ \mathbf{1}^T\boldsymbol{\beta} = 1 \\ \boldsymbol{\beta} \geqslant \mathbf{0} \end{cases} \tag{18}$$

Введем дополнительное обозначение: $\mathbb{I}^*_{\alpha,\mu} = \{i : \quad \beta^*_{\alpha,\mu;\ i} > 0\ \}$. Именно это множество ищет Factor-Search. Обратим внимание (Рис.4), что для любых возможных α, μ количество активных регрессоров $\hat{n} = |\mathbb{I}^*_{\alpha,\mu}|$ значительно меньше общего числа регрессоров n.

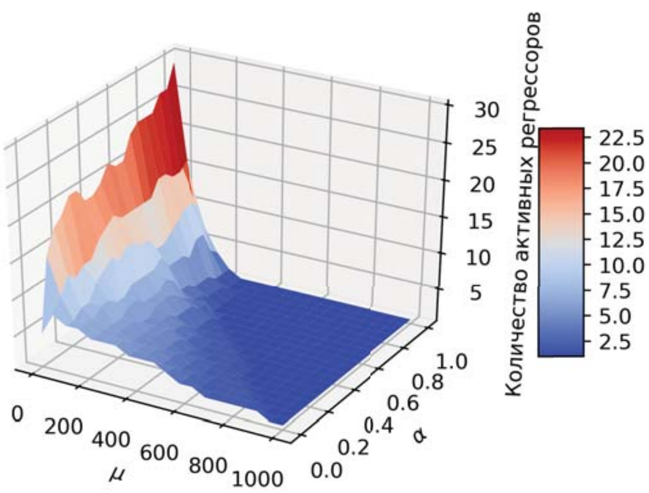

Рис. 4. Зависимость $|\mathbb{I}^*_{\alpha,\mu}|$ от μ, α

Рис. 5. Долевое распределение модельного портфеля для различных μ, α

Для истинного состава портфеля $\boldsymbol{\beta}^*_{\alpha,\mu}$ вычисляем временной ряд доходностей гипотетического портфеля

$$\mathbf{y}_{\alpha,\mu} = (y_{\alpha,\mu;t},\ t = 1, \ldots T) \in \mathbb{R}^T$$

$$y_{\alpha,\mu;t} = \sum_{i \in \mathbb{I}^*_{\alpha,\mu}} \beta^*_{\alpha,\mu;i} x_{t,i} + \xi_t \qquad (19)$$

как последовательность случайных величин с 10%-ой шумовой дисперсией $\sigma^2(\xi_t) = 0.1 \sum_{i \in \mathbb{I}^*_{\alpha,\mu}} \beta^*_{\alpha,\mu;i} x_{t,i}$ в зависимости от соответствующей последовательности коэффициентов регрессии.

Значения коэффициентов регрессии, полученные при оценки различными моделями регрессии будем называть *расчетными* и обозначать $\hat{\boldsymbol{\beta}}_{\alpha,\mu}$.

Мера качества: Точность совпадения активного множества расчетных коэффициентов с активным множеством истинных коэффициентов И Percentage Least Squares Error

$$PLSE = \frac{\sum_{t=1}^{T}\left(y_t - \sum_{i \in I_{\alpha,\mu}} \hat{\beta}_{\alpha,\mu;i} x_{t,i}\right)^2}{\sum_{t=1}^{T}\left(\sum_{i \in I_{\alpha,\mu}} \hat{\beta}_{\alpha,\mu;i} x_{t,i}\right)^2} \times 100 \qquad (20)$$

Результаты работы трех исследуемых регрессионных моделей представлены на Рис.6 в виде мозаичных диаграмм. Параметр c определялся по LOO.

Как можно видеть из диаграмм для любой рассмотренной комбинации μ, α есть существенное несоответствие между фактическим и оценочным составами портфеля для случаев, когда не учитывается мнение экспертов. Это свидетельствует об абсолютной необходимости дополнительной регуляризации на основе твердой уверенности в том, что анализируемый портфель был построен с учетом и технических и экспертных суждений.

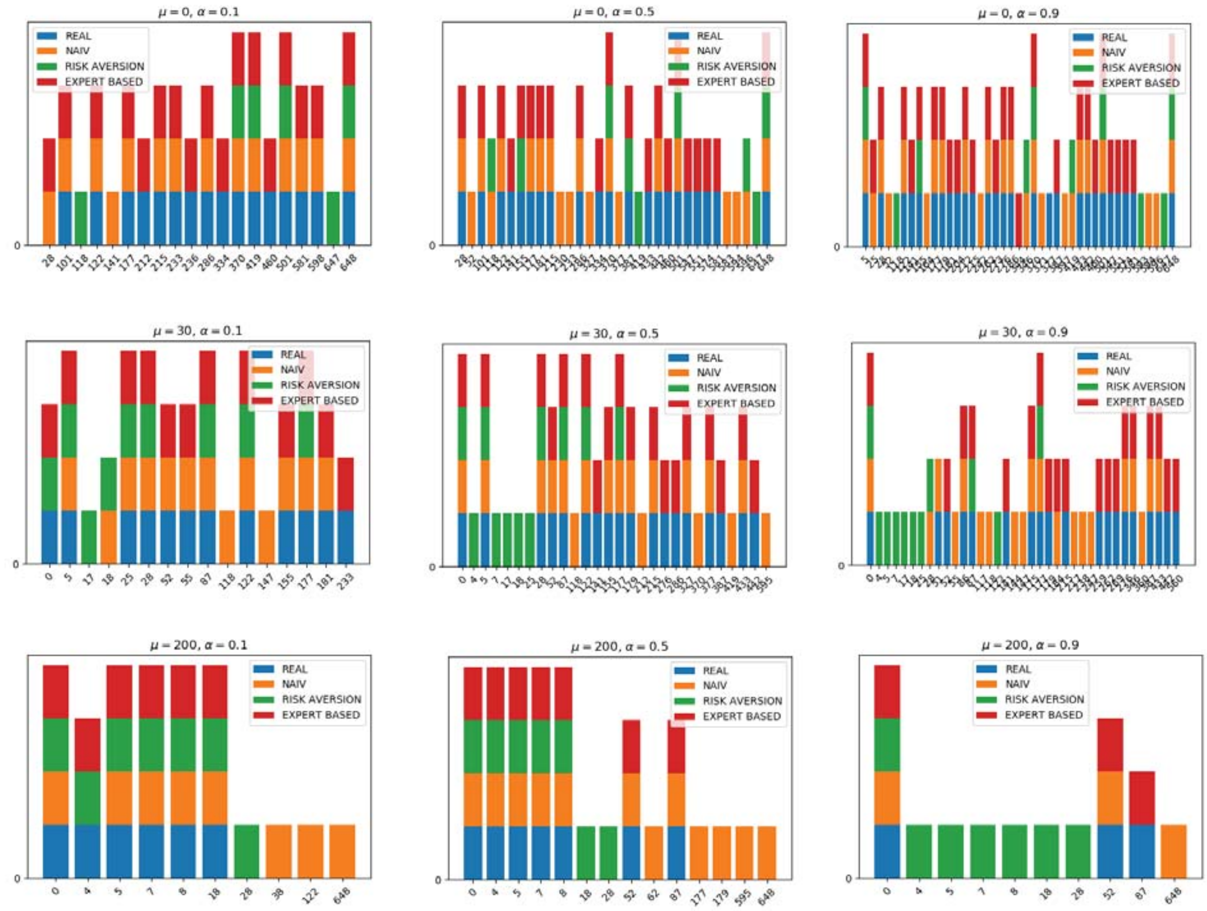

Рис. 6. По оси x отложены номера активных коэффициентов. Строки диаграмм снизу вверх: Множество активных истинных коэффициентов; Naiv Factor-Search; Risk Aversion Factor-Search; Expert Based Factor-Search;

6. Заключение

В работе были исследованы различные модели регрессии с ограничениями и регуляризацией при больших наборах регрессоров применительно к задаче восстановления состава инвестиционного портфеля. Было показано, что, если при построении портфеля помимо информации, полученной из анализа рядов доходностей, использовались личные представления администратора о виде оптимального портфеля, его невозможно восстановить ни стандартным методом МНК, ни общепринятым для данной задачи подходом Risk-Parity. Был разработан способ учета экспертной информации, было проиллюстрировано его применение. Лучше всего на модельных портфелях себя показал разработанный метод с комбинированной регуляризацией Expert Based Regularization.

В данный момент кажется целесообразным выделить два основных направления дальнейшего развития решений рассмотренных в данной работе.

Во-первых, нужно научиться анализировать портфели хедж-фондов. В рассмотренных алгоритмах мы ограничивались паевыми фондами, т. е. случаем $\beta i \geqslant 0$. И хотя это ограничение является абсолютно естественным для одного типа фондов, представляется все же необходимым построить методы анализа и для второго типа: класса фондов, для которых возможны любые значения β_i

Во-вторых, в данной работе предполагалась стационарность долевого распределения капитала. Опять же это сужает возможности применения до исследования консервативных портфелей. Это большой класс, на котором сфокусирован интерес многих аналитиков. Но все же всего лишь подмножество множества всевозможных инвестиционных портфелей.

Большим итогом предстоящей работы видится разработка вычислительно эффективного комбинированного метода для анализа любого портфеля любого фонда, без необходимости введения каких-либо сужений и ограничений.

Список литературы

1. Bishop C. M. Pattern Recognition and Machine Learning. Springer, Series: Information Science and Statistics, 2006. P. 740.
2. Ruszczyinski A. Nonlinear Optimization. Princeton University Press, 2006. P. 448.
3. Gill P. E., Murray W., Wright M. Practical Optimization. Academic, London, 1981.
4. Tyson E. Investing For Dummies. John Wiley and Sons, 2011. P. 410.
5. Markowitz H. Portfolio Selection // The Journal of Finance. 1952. — Mar. Vol. 7, no. 1. P. 77–91.
6. Sharpe W. F. Asset allocation: Management style and performance measurement // The Journal of Portfolio Management. 1992. — Winter. P. 7–19.
7. Markov M. Seeing Through Walls. Bringing Greater Transparency To Mutual Fund and Hedge Fund Analysis. 2005.
8. Sharpe W. F. Capital asset prices with and without negative holdings / Nobel Lecture. 1990. — December 7.
9. Melicher R., Norton E. Introduction to Finance: Markets, Investments, and Financial Management. John Wiley and Sons, 2011. P. 590.
10. Krasotkina O., Markov M., Mottl V. Constrained Regularized Regression Model Search in Large Sets of Regressors.
11. Seredin O., Mottl V., Krasotkina O. Compactness Hypothesis, Potential Functions, and Rectifying Linear Space in Machine Learning // Key Ideas in Learning Theory from Inception to Current State: Emmanuel Braverman's Legacy, Springer. 2018.