

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»

Пугач Илья Александрович

**Идентификации модели инвестиционного портфеля
по временному ряду его доходностей и большому
массиву доходностей биржевых активов**

511656 — Математические и информационные технологии

Выпускная квалификационная работа
(магистерская диссертация)

Научный руководитель:
д.т.н., профессор В. В. Моттль

Москва, 2018

Аннотация. Рассматривается практическая задача восстановления скрытого состава портфеля инвестиционной компании (фонда), представленного временным рядом его периодических доходностей. Эта задача относится к классу задач регрессионного оценивания. Коэффициенты регрессии имеют смысл долевого распределения капитала на множестве биржевых активов, представленных временными рядами их доходностей. Специфика совместного регрессионного анализа массива биржевых данных определяется самой задачей идентификации модели портфеля, заключающейся в поиске небольшого подмножества биржевых активов, в которые действительно вложен капитал, в огромном множестве «всех» биржевых активов. С точки зрения регрессионного анализа это равносильно предположению, что коэффициенты регрессии отличаются от нуля только в пределах реально существующего малого подмножества в большом универсуме регрессоров, и основной целью обработки данных является именно поиск этого подмножества, а не аппроксимация целевого сигнала. Такую задачу мы называем задачей Factor Search. Однако такой поиск проблематичен, если нет априорной информации об ожидаемой структуре портфеля. Мы предлагаем регуляризованную регрессионную модель доходности инвестиционного портфеля, основанную на предположении, что анализируемый портфель рационально составлен его администрацией в соответствии с теорией оптимального состава портфеля Гарри Марковица, а именно как компромисс между двумя взаимно противоречащими требованиями его прибыльности и риска банкротства.

Ключевые слова. Анализ инвестиционных портфелей на основе доходности, состав инвестиционного портфеля, разреженная регрессия с ограничениями, априорная регуляризация, квадратичная оптимизация с ограничениями, портфель Марковица, неприятие риска, оценивание гиперпараметров, кросс-валидация по отдельным объектам (leave-one-out).

Оглавление

Оглавление	3
1 Введение.	4
2 Анализ информации о доходности инвестиционных портфелей	5
2.1 Модель Уильяма Шарпа для оценивания состава портфеля на основе доходностей	5
2.2 Проблема поиска подмножества биржевых активов, фактически составляющих исследуемый портфель	6
2.3 Априорные предположения о составе анализируемого портфеля	7
2.4 Существующие методы решения задачи восстановления состава портфеля	9
3 Регуляризация фактор-поиска в моделях ограниченной регрессии	12
3.1 Центральная регрессия	12
3.2 Смещенная регуляризация	13
3.2.1 Контроль селективности в фактор-поиске	13
3.2.2 Управление толерантностью к риску в портфельных моделях	14
4 Алгоритм итерационного пути для ограниченного регуляризованного фактор-поиска.	14
4.1 Сепарабельная оптимизация	14
4.1.1 Сепарабельная целевая функция	14
4.1.2 Сепарабельная оптимизация с ограничениями	15
4.2 Последовательность сепарабельных задач квадратичной оптимизации вместо одной несепарабельной	18
4.2.1 Разделение переменных	18
4.2.2 Две симметричные сепарабельные задачи квадратичного программирования	19
4.3 Итерационный алгоритм	19
5 Экспериментальное исследование метода оценки задачи регрессии с ограничениями и регуляризацией	20
5.1 Организация эксперимента	20
5.2 Слепой фактор-поиск	21
5.3 Фактор-поиск на основе априорной информации	22
6 Публикации по материалам работы	23
7 Положения, выносимые на защиту	24
Список литературы	24

1 Введение.

Практическая задача восстановления скрытого состава портфеля инвестиционной компании (инвестиционного фонда), представленного временным рядом его периодических доходностей, относится к классу задач регрессионного оценивания, но требует учета совокупности целого ряда дополнительных допущений.

Во-первых, коэффициенты регрессии имеют здесь смысл долевого распределения капитала на множестве биржевых активов, представленных временными рядами их доходностей, поэтому сумма коэффициентов априори равна единице.

Во-вторых, предполагается, что число активов (регрессоров) намного превышает размер выборки, т.е. число рассматриваемых последовательных периодов владения.

В-третьих, мы ограничиваемся здесь только паевыми фондами (Mutual Funds), которым запрещено заимствовать капитал, что выражается ограничением неотрицательности коэффициентов регрессии.

Наконец, четвертое и важнейшее предположение определяется самой задачей идентификации модели портфеля как задачей поиска небольшого подмножества биржевых активов, в которые действительно вложен капитал, в огромном множестве «всех» биржевых активов.

С точки зрения регрессионного анализа это равносильно предположению, что коэффициенты регрессии отличаются от нуля только в пределах реально существующего малого подмножества в большом универсуме регрессоров, и основной целью обработки данных является именно поиск этого подмножества, а не аппроксимация целевого сигнала. Такую задачу мы называем задачей фактор-поиска (Factor Search). Однако фактор-поиск проблематичен, если нет априорной информации об ожидаемой структуре портфеля.

Мы предлагаем регуляризованную регрессионную модель доходности инвестиционного портфеля, основанную на предположении, что анализируемый портфель рационально составлен его администрацией в соответствии с теорией оптимального состава портфеля Гарри Марковица, а именно как компромисс между двумя взаимно противоречащими требованиями его доходности и риска банкротства.

Статья структурирована следующим образом. В Разделе 2 мы кратко рассмотрим практическую задачу анализа инвестиционного портфеля, положившую начало данному исследованию. В Разделе 3 мы исследуем специфику задачи выбора регрессоров при наличии ограничений в больших наборах коррелированных регрессоров. Раздел 4 полностью посвящен эффективному решению задачи оценки регрессии с ограничениями в больших наборах регрессоров. В Разделе 5 мы экспериментально изучаем разработанную математическую и алгоритмическую систему.

2 Анализ информации о доходности инвестиционных портфелей

2.1 Модель Уильяма Шарпа для оценивания состава портфеля на основе доходностей

Пусть капитал анализируемой инвестиционной компании полностью инвестирован в биржевые активы n видов в пропорциях $\beta = (\beta_1, \dots, \beta_n)$, $\sum_{i=1}^n \beta_i = 1$, $\beta_i \geq 0$. Эти доли капитала называются портфелем компании.

Ограничения неотрицательности $\beta_i \geq 0$ выражают предположение, что активы могут быть приобретены только с использованием внутреннего капитала без заимствования денег или активов из внешних источников, или, как принято говорить при управлении портфелем, без удержания коротких (отрицательных) позиций – большинству паевых инвестиционных фондов это не разрешено. Инвестиционные компании, занимающие отрицательные (короткие) позиции $\beta_i < 0$, соответствующие заемным средствам, называются хедж-фондами [15]. В этой статье мы не рассматриваем хедж-фонды.

Менеджеры портфелей, как правило, очень скрытны в отношении того, что они покупают и продают, а дробная структура портфеля $\beta = (\beta_1, \dots, \beta_n)$ обычно скрыта от общественности. Такая информация, будучи восстановленной, представляла бы большой интерес для тех, кто следит за портфелем, в частности, она давала бы инвесторам этого портфеля предварительное оповещение.

Проблема восстановления скрытого распределения капитала в портфеле на основе общедоступных данных была сформулирована Уильямом Шарпом, лауреатом Нобелевской премии 1990 года по экономике [1]. 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 [15] представляющих определенные классы активов и инвестиционные стили. Этот подход известен под названием Анализ Стиля на Основе Доходности, Returns Based Style Analysis, RBSA.

Периодическая доходность – это процент прибыли от инвестиций в финансовый инструмент за определенный период времени. Пусть $z_{beg,i}$ и $z_{end,i}$ цены активов в начале и в конце некоторого временного интервала, называемого периодом владения, соответственно, $z_{beg,p}$ и $z_{end,p}$ будут стоимостью портфеля в целом. Отношение $y = (z_{end,p} - z_{beg,p}) / z_{beg,p}$ называется доходностью портфеля за период владения, а $x_i = (z_{end,i} - z_{beg,i}) / z_{beg,i}$ представляют собой доходности активов. Если рассматриваются классы активов, то их ожидаемая доходность обычно оценивается специальными аналитическими компаниями как индекс доходности.

Для последовательных временных моментов $t = 1, 2, 3, \dots$, 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 and $x_{t,i}$, $i = 1, \dots, n$. 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 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:

$$\sum_{t=1}^T \left(y_t - \sum_{i=1}^n \beta_i x_{t,i} \right)^2 \rightarrow \min, \quad \sum_{i=1}^n \beta_i = 1, \quad \beta_i \geq 0, \quad i = 1, \dots, n. \quad (1)$$

Принцип аппроксимации обычно используется в современном инвестиционном анализе как Анализ Стиля на Основе Доходности (RBSA).

Задача регрессии с ограничениями (1) относится к классу квадратичной оптимизации при линейных ограничениях (задача квадратичного программирования) [18]. Его вычислительная сложность полиномиальна по отношению к числу активов или классов активов, предположительно образующих портфель ($O(n^3)$), но она линейна по отношению к числу периодов времени, на которых были вычислены периодические доходности ($O(T)$).

2.2 Проблема поиска подмножества биржевых активов, фактически составляющих исследуемый портфель

Типичной целью обработки данных по доходности является фактор-поиск – нахождение реально существующего скрытого состава анализируемого портфеля среди очень большого количества потенциальных факторов-активов или классов активов (индексов фондового рынка). Пусть $\mathbb{I} = \{1, \dots, n\}$ обозначает полный набор факторов, тогда требуется найти среди них небольшое активное подмножество $\hat{\mathbb{I}} \subset \mathbb{I}$, $\hat{n} = |\hat{\mathbb{I}}| < n$. Это означает, что оценки выявления факторов, т.е., коэффициентов регрессии в модели портфеля на основе доходности (1), должны быть положительными внутри активного подмножества $\hat{\beta}_i > 0$, $i \in \hat{\mathbb{I}}$, и равными нулю вне его $\hat{\beta}_i = 0$, $i \notin \hat{\mathbb{I}}$. Таким образом, требуется найти модель портфолио вида

$$y_t \cong \sum_{i \in \hat{\mathbb{I}}} \beta_i x_{t,i}, \quad \sum_{i \in \hat{\mathbb{I}}} \beta_i = 1, \quad \beta_i > 0, \quad i \in \hat{\mathbb{I}}.$$

Эту задачу мы называем задачей фактор-поиска (Factor Search). В любом случае, выбор модели регрессии подразумевает ту или иную регуляризацию, а именно, уменьшение свободы варьирования коэффициентов регрессии при

минимизации остатка по критерию наименьших квадратов (1). Несмотря на то, что ограничения типа равенств и неравенств вносят весьма глубокую регуляризацию [2], сами по себе они недостаточны для наделения критерия свойством фактор-поиска при сравнении временных рядов доходности портфеля с большими наборами индексов фондового рынка.

Поэтому в следующем Разделе 2.3 мы предлагаем учесть, помимо естественных ограничений, дополнительные априорные предложения по рациональности состава искомого неизвестного портфеля с точки зрения его полезности как финансового инструмента.

2.3 Априорные предположения о составе анализируемого портфеля

Мы исходим из предположения, что портфель, временной ряд доходности $(y_t, t = 1, \dots, T)$ которого анализируется, был составлен командой его менеджеров с целью получения большей доходности при приемлемом уровне риска.

Пусть доходности активов в последующие периоды времени $(\mathbf{x}_t = (x_{t,1} \dots x_{t,n}), t = 1, \dots, T)$ рассматривается как совокупность реализаций случайного вектора, имеющего математическое ожидание $\bar{\mathbf{x}} = E\{\mathbf{x}\} \in \mathbb{R}^n$ и ковариационную матрицу $\Sigma = E\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T\}$ ($n \times n$). Здесь мы опускаем вопрос о том, являются ли доходности в последовательность $(\mathbf{x}_t, t = 1, \dots, T)$ зависимыми или нет, предполагается, что соответствующий временной ряд стационарен. Тогда доходность портфеля $y_t = \boldsymbol{\beta}^T \mathbf{x}_t$ также следует рассматривать как стационарный случайный процесс, и его доходность на заданном отрезке времени будет случайной величиной с математическим ожиданием $\bar{y} = \boldsymbol{\beta}^T \bar{\mathbf{x}}$, а именно, ожидаемой прибылью (доходностью), и дисперсией $\sigma^2 = \boldsymbol{\beta}^T \Sigma \boldsymbol{\beta}$, которая создает некоторый риск потерь при составлении портфеля.

Чем больше ожидаемая доходность, тем лучше портфель. С другой стороны, чем больше дисперсия доходности, тем больше вероятность низкой и даже отрицательной доходности и тем хуже становится портфель.

В соответствии с теорией оптимального состава портфеля Гарри Марковица [3], рациональным можно считать только тот состав портфеля $(\boldsymbol{\beta} \geq \mathbf{0}, \mathbf{1}^T \boldsymbol{\beta} = 1)$, который обеспечивает наименьшее отклонение для фиксированной ожидаемой доходности $(\boldsymbol{\beta}^T \Sigma \boldsymbol{\beta} \rightarrow \min, \bar{\mathbf{x}}^T \boldsymbol{\beta} = \text{const})$, или наоборот, наибольшую ожидаемую доходность $(\bar{\mathbf{x}}^T \boldsymbol{\beta} \rightarrow \max, \boldsymbol{\beta}^T \Sigma \boldsymbol{\beta} = \text{const})$. Легко видеть, что оба случая охватываются условием

$$U(\boldsymbol{\beta} | \mu) = \mu \bar{\mathbf{x}}^T \boldsymbol{\beta} - \boldsymbol{\beta}^T \Sigma \boldsymbol{\beta} \rightarrow \max, \quad \boldsymbol{\beta} \geq \mathbf{0}, \mathbf{1}^T \boldsymbol{\beta} = 1. \quad (2)$$

Квадратичную функцию $U(\boldsymbol{\beta} | \mu)$ обычно называют портфельной (утилитой) функцией полезности [4], где параметр $0 \leq \mu < \infty$ интерпретируется как толерантность к риску и, соответственно, $1/\mu$ как неприятие риска.

Таким образом, когда математическое ожидание и ковариационная матрица доходности активов $\Sigma = E\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T\}$ ($n \times n$) фиксированы, существует только однопараметрическое семейство состава портфеля, которое может быть разумным с точки зрения невозможности улучшить каждую из них как в ожидаемой доходности, так и в дисперсии доходности.

Кстати, именно этот факт обеспечивает само существование равновесия фондового рынка в сообществе инвесторов, не склонных к риску, и привел, наконец, к известной модели ценообразования на капитальные активы (САРМ) [5], которая была главным обоснованием Нобелевской премии по экономике, присужденной в 1990 году Гарри Марковицу и Уильяму Шарпу.

Портфели, не отвечающие условию (2) не могут быть признаны обоснованными с точки зрения теории Марковица, но нельзя судить о том, какое значение толерантности к риску μ лучше. Это внутренняя психологическая характеристика каждого конкретного инвестора.

В настоящей работе предлагается использовать класс функций полезности (2) в качестве параметрического семейства чрезвычайно сильных априорно обоснованных регуляризационных функций для задач фактор-поиска:

$$V(\boldsymbol{\beta} | \mu) = \boldsymbol{\beta}^T \Sigma \boldsymbol{\beta} - \mu \bar{\mathbf{x}}^T \boldsymbol{\beta} \rightarrow \min. \quad (3)$$

Здесь под $\bar{\mathbf{x}} \in \mathbb{R}^n$ и Σ ($n \times n$) подразумеваются математическое ожидание и ковариационная матрица случайного вектора доходности активов, ожидаемого в начале инвестиционного периода, при условии, что они существенно не изменятся в течение некоторого периода времени в будущем. Коэффициент толерантности к риску $0 \leq \mu < \infty$ является гиперпараметром этой функции регуляризации.

Эти статистические характеристики совокупности активов фондового рынка могут быть получены на достаточно длительном интервале времени $\{t = -(T' - 1), \dots, -1, 0\}$, предшествующем интервалу наблюдения за портфелем $\{t = 1, 2, \dots, T\}$:

$$\bar{\mathbf{x}} = \frac{1}{T'} \sum_{t=-(T'-1)}^0 \mathbf{x}_t, \quad \Sigma = \frac{1}{T'} \sum_{t=-(T'-1)}^0 (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})^T. \quad (4)$$

Однако вычисление $\bar{\mathbf{x}}$ и Σ непосредственно из интервала наблюдения портфеля

$$\bar{\mathbf{x}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t, \quad \Sigma = \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})^T \quad (5)$$

не будет роковой ошибкой, если совокупность активов фондового рынка можно считать относительно стабильной в течение довольно длительного периода времени.

2.4 Существующие методы решения задачи восстановления состава портфеля

При регрессионном анализе широко применяется метод наименьших квадратов. Пусть $t \in \{1, \dots, T\}$ индексы конечного неупорядоченного множества объектов, каждый из которых характеризуется вектором признаков $\mathbf{x}_j = (x_{t,i}, i=1, \dots, n) \in \mathbb{R}^n$ и целевой переменной $y_t \in \mathbb{R}$. Если предположить, что существует линейная регрессия $y \cong f(\mathbf{x})$, тогда минимизация остаточной суммы квадратов является очевидным и в большинстве случаев отличным средством для нахождения соответствующей оценки коэффициентов регрессии:

$$\begin{cases} \hat{\boldsymbol{\beta}} = \arg \min \sum_{t=1}^T (y_t - \boldsymbol{\beta}^T \mathbf{x}_t)^2 = \arg \min (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}), \\ \mathbf{y} = (y_1, \dots, y_T) \in \mathbb{R}^T, \mathbf{X} = (\mathbf{x}_1 \cdots \mathbf{x}_T) (n \times T), \boldsymbol{\beta} = (\beta_i, i=1, \dots, n) \in \mathbb{R}^n. \end{cases}, \quad (6)$$

Этот принцип широко известен как метод наименьших квадратов (МНК).

Понятно, что простое решение соответствующей системы линейных уравнений $\mathbf{X}\mathbf{X}^T \boldsymbol{\beta} = \mathbf{X}\mathbf{y}$ дает единственную оценку коэффициентов регрессии только если регрессоры линейно независимы в наборе данных $|\mathbf{X}\mathbf{X}^T| \neq 0$, что возможно только при $n \leq T$.

Однако для многих приложений характерно, что число регрессоров значительно превышает размер выборки $n \gg T$. В этом случае неизбежно некоторое ограничение свободы выбора вектора коэффициентов регрессии $\boldsymbol{\beta} \in \mathbb{R}^n$. Существует два принципиально разных способа реализации такого сужения.

Первый способ заключается в добавлении функции регуляризации $V(\mathbf{a})$ критерию МНК (6) для использования некоторых априорных знаний об ожидаемой регрессионной модели:

$$\hat{\boldsymbol{\beta}} = \arg \min [V(\boldsymbol{\beta}) + (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})]. \quad (7)$$

Функция регуляризации является штрафом, предназначенным для выражения нежелательности некоторых значений $\boldsymbol{\beta}$ – чем более маловероятным представляется наблюдателю этот вектор коэффициента регрессии, тем больше должен быть штраф $V(\boldsymbol{\beta})$.

Широко известные регуляризации: ridge $V(\boldsymbol{\beta}) = \sum_{i=1}^n \beta_i^2$ [6,7], bridge $\sum_{i=1}^n |\beta_i|^p$ ($p > 0$) [8,9], в частности, LASSO $\sum_{i=1}^n |\beta_i|$ [10], их комбинация Elastic Net $\sum_{i=1}^n (\beta_i^2 + \mu |\beta_i|)$ [11], и более продвинутый SCAD [12] штрафуют отклонение коэффициентов регрессии от нуля. Фактически, то же самое делает регуляризация на основе матрицы $V(\boldsymbol{\beta}) = \boldsymbol{\beta}^T \mathbf{G} \boldsymbol{\beta}$ [13]. Этот факт наталкивает на идею их совместного рассмотрения под названием **центральной регуляризации (central regularization)**

Однако в прикладной задаче анализа инвестиционного портфеля, положившей начало этой работе, требуется математически выразить априорное знание неизвестном составе портфеля, для чего мы будем использовать особый вид **смещенной регуляризации (biased regularization)** $V(\beta) = (\beta - \beta_0)^T G (\beta - \beta_0)$ [14], где вектор $\beta_0 \in \mathbb{R}^n$ это гиперпараметр вместе с матрицей G ($n \times n$).

Но, что особенно важно, способ регуляризации, получающийся при использовании штрафной функции (3), наиболее важен в обсуждаемом приложении. Состав портфеля должен оцениваться как доля капитала над набором активов фондового рынка или классов активов, следовательно, коэффициенты регрессии должны в сумме давать единицу

$$\sum_{i=1}^n \beta_i = 1, \text{ or in vector form, } \mathbf{1}^T \beta = 1, \mathbf{1} = (1 \dots 1)^T \in \mathbb{R}^n.$$

Кроме того, мы ограничимся рассмотрением только класса взаимных (паевых) инвестиционных фондов, которым разрешено формировать портфели, используя только свой внутренний капитал; в отличие от хедж-фондов, которые могут занимать деньги или активы из внешних источников [15]; что математически приводит к предположению, что коэффициенты регрессии не могут быть отрицательными $\beta_i \geq 0, i = 1, \dots, n$, или $\beta \geq \mathbf{0}$.

В терминах регуляризирующей функции (7), это ограничение вряд ли поддается обработке:

$$V(\beta) = \begin{cases} 0, & \beta \geq \mathbf{0}, \mathbf{1}^T \beta = 1; \\ \infty, & \text{otherwise} \end{cases}.$$

Будет удобнее неявно добавить ограничения к критерию МНК(6):

$$\begin{cases} \hat{\beta} = \arg \min (\mathbf{y} - \mathbf{X}^T \beta)^T (\mathbf{y} - \mathbf{X}^T \beta), & \beta \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^T, \mathbf{X} (n \times T). \\ \mathbf{1}^T \beta = 1, \beta \geq \mathbf{0}, \end{cases} \quad (8)$$

Эта задача регрессии с ограничениями многократно рассмотрена в литературе при предположении о большом количестве регрессоров $n \gg T$, но только с ограничениями равенствами $\mathbf{1}^T \beta = 1$ или неравенствами $\beta \geq \mathbf{0}$ и никогда не с обоими.

Прежде всего, следует отметить, что даже при отсутствии знаковых ограничений задача квадратичной оптимизации с n переменными $\mathbf{a} \in \mathbb{R}^n$

$$\begin{cases} \hat{\beta} = \arg \min (\mathbf{y} - \mathbf{X}^T \beta)^T (\mathbf{y} - \mathbf{X}^T \beta), \\ \mathbf{1}^T \beta = 1, \end{cases} \quad (9)$$

неизбежно имеет полиномиальную вычислительную сложность $O(n^3)$. Однако, все решения, которые дают один и тот же вектор невязок $\delta = \mathbf{y} - \mathbf{X}^T \beta \in \mathbb{R}^T$ эквивалентны с точки зрения полученной модели регрессии. Следовательно, только соответствующее T -мерное подпространство \mathbb{R}^n имеет значение, что сравнительно низкоразмерно в силу предполагаемого неравенства $n \gg T$. В [16] показано что, при отсутствии каких-либо ограничений, достаточно найти предельное решение задачи гребневой (ridge) регрессии $\alpha \beta^T \beta + (\mathbf{y} - \mathbf{X}^T \beta)^T (\mathbf{y} - \mathbf{X}^T \beta) \rightarrow \min(\beta), \alpha \rightarrow 0$, в полном соответствии с методологией Тихонова решения некорректных задач [17].

В данной статье предложен алгоритм решения задачи при ограничениях равенства $\mathbf{1}^T \boldsymbol{\beta} = 1$ (9). Алгоритм имеет линейную вычислительную сложность $O(n)$ относительно числа переменных и полиномиальную $O(T^3)$ по размеру выборки. В дальнейшем он служит базой, лежащей в основе наших алгоритмов решения полностью ограниченной задачи (8) при больших наборах регрессоров.

Задача регрессии с ограничениями (8) является частным случаем в классе задач квадратичного программирования [18].

$$\begin{cases} \boldsymbol{\beta}^T \mathbf{G} \boldsymbol{\beta} + \mathbf{u}^T \boldsymbol{\beta} \rightarrow \min(\boldsymbol{\beta} \in \mathbb{R}^n), \\ \mathbf{B} \boldsymbol{\beta} = \mathbf{b} \in \mathbb{R}^k, \mathbf{D} \boldsymbol{\beta} \leq \mathbf{d} \in \mathbb{R}^m, k < n, m \leq n, \end{cases} \quad (10)$$

где матрицы $\mathbf{B}(k \times n)$ и $\mathbf{D}(m \times n)$ а также векторы \mathbf{b} и \mathbf{d} указываются согласно предварительным знаниям. Наш частный случай (8) характеризуется особенно простыми ограничениями, одним равенством $\mathbf{1}^T \boldsymbol{\beta} = 1$, $k=1$, и n неравенствами на знаки $\boldsymbol{\beta} \geq \mathbf{0}$, $m=n$.

Нет необходимости использовать общепринятые общие итерационные методы, такие как процедуры внутренние или внешние точки, для решения задачи ограниченной регрессии (8). Возможность конечного решения этой задачи возникает из того факта, что после нахождения решения $\hat{\boldsymbol{\beta}} = (\beta_i \geq 0, i=1, \dots, n)$ и определения подмножества активных регрессоров $\{i : \hat{\beta}_i > 0\}$, оптимальные значения активных ненулевых коэффициентов регрессии совместно являются решениями задачи наименьших квадратов без ограничений в активном подмножестве. Может показаться нереальным найти его среди огромного количества всех подмножеств за $\{1, \dots, n\}$, но общий поиск не требуется из-за особых свойств задачи оптимизации (8).

Метод точного штрафа, рассмотренный в [19,20], дает решение этой задачи в ограниченной форме. Суть метода в замене задачи условной оптимизации на безусловную функцию штрафа

$$\mathcal{E}_\mu(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}) + \mu \left[(\mathbf{1}^T \boldsymbol{\beta} - 1) + \sum_{i=1}^n \max(0, -\beta_i) \right].$$

Понятно, что эта функция является мажорантой исходной целевой функции и совпадает с ней в точке решения $\hat{\boldsymbol{\beta}} = (\mathbf{1}^T \hat{\boldsymbol{\beta}}, \hat{\beta}_i \geq 0, i=1, \dots, n)$:

$$\mathcal{E}_\mu(\boldsymbol{\beta}) \geq (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}), \quad \mathcal{E}_\mu(\hat{\boldsymbol{\beta}}) = (\mathbf{y} - \mathbf{X}^T \hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}^T \hat{\boldsymbol{\beta}}).$$

Если параметр штрафа $\mu > 0$ достаточно велик, любой минимизатор мажоранты $\mathcal{E}_\mu(\boldsymbol{\beta})$ решает задачу (8).

Однако абсолютное большинство публикаций по оценке ограниченной регрессии рассматривает только знаковые неравенства $\boldsymbol{\beta} \geq \mathbf{0}$ и опускает равенство $\mathbf{1}^T \boldsymbol{\beta} = 1$. В ряде публикаций, в частности в [21], показано, что знаковые ограничения наделяют критерий наименьших квадратов четко выраженным свойством нахождения разреженной (sparse) регрессионной модели, даже без дополнительных средств уменьшения числа регрессоров.

Простая процедура, обеспечивающая неотрицательность коэффициентов регрессии и называемая алгоритмом активного множества, предлагается в [22] и модифицируется в [23]. Алгоритм основан на том, что после нахождения решения $\hat{\beta} = (\beta_i \geq 0, i=1, \dots, n)$ и определения подмножества активных регрессоров $\{i : \hat{\beta}_i > 0\}$, оптимальные значения активных ненулевых коэффициентов регрессии совместно являются решением неограниченной задачи наименьших квадратов в активном подмножестве. Итерационный алгоритм чередующихся наименьших квадратов начинается с исходного допустимого набора коэффициентов регрессии, и на каждом шаге переменные выбираются и удаляются из активного множества таким образом, что оно строго уменьшается. Решение будет найдено после конечного числа итераций простой неограниченной линейной регрессии на активном подмножестве. Однако вычислительная сложность каждого шага является кубической относительно размера текущего активного подмножества, и число итераций может быть довольно большим.

Однако все вышеперечисленные алгоритмы имеют полиномиальную вычислительную сложность относительно числа регрессоров n . Поэтому в этой статье мы разрабатываем быстро сходящийся итерационный алгоритм, каждая итерация которого имеет линейную вычислительную сложность $O(n)$. Идея алгоритма – декомпозиция задачи условной минимизации (8) в последовательность сепарабельных задач оптимизации, каждая из которых содержит только одну переменную, так что они могут быть решены по отдельности параллельно.

3 Регуляризация фактор-поиска в моделях ограниченной регрессии

3.1 Центральная регрессия

Как упоминалось выше в Разделе 2.2, естественные ограничения равенства и неравенства могут оказаться недостаточными для надления критерия регрессионной оценки (8) свойством фактор-поиска. Это будет явно видно, если мы запишем (8) в полускалярной форме:

$$\begin{cases} (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}) = \left(\mathbf{y} - \sum_{i=1}^n \beta_i \mathbf{x}_i \right)^T \left(\mathbf{y} - \sum_{i=1}^n \beta_i \mathbf{x}_i \right) \rightarrow \min(\beta_1, \dots, \beta_n), \\ \sum_{i=1}^n \beta_i = 1, \beta_i \geq 0, i = 1, \dots, n. \end{cases} \quad (11)$$

Основным препятствием является корреляция между регрессорами $\mathbf{x}_i = (x_{1,i} \dots x_{T,i}) \in \mathbb{R}^T$, $i = 1, \dots, n$, что делает едва различимыми различные комбинации $[\mathbf{x}_i, i \in \hat{\mathbb{I}} \subset \mathbb{I} = \{1, \dots, n\}]$.

Эта трудность может быть существенно ослаблена, если априори предположить, что регрессоры в искомой комбинации слабо коррелируют между собой. Чем меньше корреляция между случайными переменными $(x_i, i = 1, \dots, n)$, тем меньше дисперсия их линейной комбинации $y = \sum_{i=1}^n \beta_i x_i$ с фиксированной нормой вектора коэффициентов $\sum_{i=1}^n \beta_i^2$. Таким образом, если

мы хотим найти подмножество слабо коррелированных регрессоров, мы должны минимизировать квадратичную форму $\beta^T \Sigma \beta \rightarrow \min$, $\beta^T \mathbf{1} = \text{const}$.

Это только один вид возможных априорных предположений о искомом подмножестве регрессоров, в общем случае любая положительно определенная матрица может выражать интуицию пользователя, и она не должна обязательно зависеть от данных. Эта причина приводит к квадратичной регуляризации

$$\begin{cases} \beta^T \mathbf{G} \beta + c(\mathbf{y} - \mathbf{X}^T \beta)^T (\mathbf{y} - \mathbf{X}^T \beta) \rightarrow \min(\beta_1, \dots, \beta_n), \\ \mathbf{1}^T \beta = 1, \beta \geq \mathbf{0}. \end{cases} \quad (12)$$

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

3.2 Смещенная регуляризация

3.2.1 Контроль селективности в фактор-поиске

Однако может случиться так, что априорные знания предполагают не только направление, но и конкретное значение вектора коэффициентов регрессии. Ярким примером такой ситуации является априори известный факт, что число активных регрессоров, для которых коэффициенты регрессии положительны $\beta_i > 0$, меньше числа, полученного в результате решения исходной задачи с ограничениями(12), и требуется принудительно присвоить некоторым из них нулевые значения.

Это может быть сделано путем выбора квадратичной функции регуляризации

$$(\beta - \beta_0)^T \mathbf{G} (\beta - \beta_0) \rightarrow \min, \quad \beta_0 < \mathbf{0},$$

когда смещение в отрицательном квадранте, которое запрещено ограничениями неравенства $\beta \geq \mathbf{0}$, поощряет нулевые значения коэффициентов регрессии. Мы называем эту квадратичную функцию **смещенной регуляризацией**.

Чем глубже точка β_0 выталкивается в запретную зону, тем сильнее будет склонность регуляризационной функции $(\beta - \beta_0)^T \mathbf{G} (\beta - \beta_0) \rightarrow \min$ к уменьшению числа активных регрессоров. Расширение квадратичной формы приводит к регуляризации $\beta^T \mathbf{G} \beta - \beta_0^T \mathbf{G} \beta \rightarrow \min$. Если мы положим $\beta_0 = -\mu \mathbf{1}$, получим $\beta^T \mathbf{G} \beta + \mu \mathbf{1}^T \mathbf{G} \beta \rightarrow \min$.

Итак, мы пришли к задаче регрессии с ограничениями

$$\begin{cases} \beta^T \mathbf{G} \beta + \mu \mathbf{1}^T \mathbf{G} \beta + c(\mathbf{y} - \mathbf{X}^T \beta)^T (\mathbf{y} - \mathbf{X}^T \beta) \rightarrow \min(\beta_1, \dots, \beta_n), \\ \mathbf{1}^T \beta = 1, \beta \geq \mathbf{0}. \end{cases} \quad (13)$$

Чем больше параметр селективности $\mu \geq 0$, тем сильнее свойство критерия (13) уменьшать подмножество активных регрессоров.

Легко видеть некоторую аналогию с Elastic Net $\sum_{i=1}^n (\beta_i^2 + \mu |\beta_i|)$ (Раздел 2.4). Если положить $\mathbf{G} = \mathbf{I}$, второе штрафное слагаемое в (13) будет $\mu \mathbf{1}^T \boldsymbol{\beta}$, что, в силу ограничений неотрицательности $\beta_i = |\beta_i|$, привело бы к

$$\mu \mathbf{1}^T \boldsymbol{\beta} = \mu \sum_{i=1}^n \beta_i = \mu \sum_{i=1}^n |\beta_i|.$$

Однако эта аналогия не совсем прямолинейна $\mathbf{1}^T \mathbf{G} \boldsymbol{\beta} \neq \mathbf{1}^T \boldsymbol{\beta}$, иначе ограничение равенства $\mathbf{1}^T \boldsymbol{\beta} = 1$ уничтожит эффект избирательности.

Параметр селективности $\mu \geq 0$ должен быть подобран с учетом набора данных (\mathbf{y}, \mathbf{X}) из принципа максимизации обобщающей способности модели, которая должна быть измерена с помощью соответствующей процедуры проверки.

3.2.2 Управление толерантностью к риску в портфельных моделях

Другим примером необходимости смещенной регуляризации является оценка состава портфеля при априорном допущении рационального баланса между ожидаемой доходностью и дисперсией доходности. Действительно, квадратичная функция регуляризации (3) может быть представлена как $V(\boldsymbol{\beta} | \mu) = \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} - \mu \bar{\mathbf{x}}^T \boldsymbol{\beta} = (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \boldsymbol{\Sigma} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + const$, где

$\boldsymbol{\beta}_0 = \arg \min [\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} - \mu \bar{\mathbf{x}}^T \boldsymbol{\beta}]$. Условие $\nabla_{\boldsymbol{\beta}} (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} - \mu \bar{\mathbf{x}}^T \boldsymbol{\beta}) = \mathbf{0} \in \mathbb{R}^n$ приводит к равенству

$$\boldsymbol{\beta}_0 = \frac{\mu}{2} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}}.$$

Мы получили смещенную функцию регуляризации

$$(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \boldsymbol{\Sigma} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \rightarrow \min, \quad \boldsymbol{\beta}_0 = \frac{\mu}{2} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}},$$

что приводит к задаче регрессии с ограничениями

$$\begin{cases} \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} - \mu \bar{\mathbf{x}}^T \boldsymbol{\beta} + c(\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}) \rightarrow \min(\beta_1, \dots, \beta_n), \\ \mathbf{1}^T \boldsymbol{\beta} = 1, \quad \boldsymbol{\beta} \geq \mathbf{0}. \end{cases} \quad (14)$$

Коэффициент толерантности к риску $0 < \mu < \infty$ является скрытой характеристикой инвестора, составлявшего анализируемый портфель, и должен оцениваться вместе с составом портфеля. $\boldsymbol{\beta} \in \mathbb{R}^n$. Это гиперпараметр модели, и, чтобы его оценить, нужно применить процедуру верификации; например, критерий leave-one-out рассмотренный в следующем разделе.

4 Алгоритм итерационного пути для ограниченного регуляризованного фактор-поиска.

4.1 Сепарабельная оптимизация

4.1.1 Сепарабельная целевая функция

С математической точки зрения, как критерии фактор-поиска общей регрессионной задачи с ограничениями (13), так и продиктованная приложением

задача с целью нахождения рациональной модели портфеля (14) имеют общую структуру

$$\begin{cases} \boldsymbol{\beta}^T \mathbf{G} \boldsymbol{\beta} + \mu \mathbf{d}^T \boldsymbol{\beta} + c(\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}) \rightarrow \min(\beta_1, \dots, \beta_n), \\ \mathbf{1}^T \boldsymbol{\beta} = 1, \boldsymbol{\beta} \geq \mathbf{0}, \end{cases} \quad (15)$$

где \mathbf{G} произвольная положительно определенная матрица и, соответственно, $\mathbf{d} = \mathbf{1}^T \mathbf{G}$ или $\mathbf{d} = -\bar{\mathbf{x}}$. Это задача квадратичного программирования.

Достаточно рассмотреть его агрегированную форму

$$\begin{cases} \boldsymbol{\beta}^T \mathbf{V} \boldsymbol{\beta} + (\mu \mathbf{d} + c\mathbf{u})^T \boldsymbol{\beta} \rightarrow \min(\boldsymbol{\beta}), & \text{где } \mathbf{V} = \mathbf{G} + c \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^T (n \times n)(n \times n), \\ \mathbf{1}^T \boldsymbol{\beta} = 1, \boldsymbol{\beta} \geq \mathbf{0}, & \text{и } \mathbf{u} = -2 \mathbf{X} \mathbf{y} \in \mathbb{R}^T. \end{cases} \quad (16)$$

Из скалярной записи целевой функции явно видно, что она попарно сепарабельна, т. е. является суммой простых функций двух видов, каждая из которых зависит не более чем от двух переменных:

$$J(\beta_1, \dots, \beta_n) = \sum_{i=1}^n \sum_{l=1}^n \underbrace{b_{il} \beta_i \beta_l}_{2 \text{ переменные}} + \sum_{i=1}^n \underbrace{(\mu d_i + c u_i) \beta_i}_{1 \text{ переменная}} \rightarrow \min, \quad (17)$$

$$\sum_{i=1}^n \beta_i = 1, \beta_i \geq 0, i = 1, \dots, n.$$

В частности, любая квадратичная функция является попарно сепарабельной. Именно парная сепарабельность приводит к системе n линейных уравнений как способу квадратичной оптимизации. Вычислительная сложность решения этой системы полиномиальна $O(n^3)$.

4.1.2 Сепарабельная оптимизация с ограничениями

Еще более конкретно, когда матрица $\mathbf{V}(n \times n)$ диагональна $\mathbf{V} = \text{Diag}(b_{11} \dots b_{nn})$, что вряд ли реально в наших задачах, то целевая функция становится просто сепарабельной, без условия попарной сепарабельности, т. е. полностью представима как сумма функций одной переменной:

$$J(\beta_1, \beta_2, \dots, \beta_n) = \sum_{i=1}^n (b_{ii} \beta_i^2 + (\mu d_i + c u_i) \beta_i).$$

Если ограничений нет, то поиск его точки минимума сводится к поиску точек минимум одиночных составляющих, которые являются квадратичными:

$$J(\beta_1, \dots, \beta_n) = \sum_{i=1}^n (\beta_i - \tilde{\beta}_i)^2 + \text{const}, \quad \tilde{\beta}_i = \arg \min (b_{ii} \beta_i^2 + (\mu d_i + c u_i) \beta_i) = -\frac{\mu d_i + c u_i}{2 b_{ii}}. \quad (18)$$

Понятно, что вычислительная сложность этой задачи оптимизации линейна относительно числа переменных $O(n)$.

Линейная вычислительная сложность будет сохранена в случае ограничений-неравенств:

$$\begin{cases} J(\beta_1, \dots, \beta_n) = \sum_{i=1}^n (b_{ii} \beta_i^2 + (\mu d_i + c u_i) \beta_i) \rightarrow \min(\beta_1, \dots, \beta_n), \\ \beta_i \geq 0, \quad i = 1, \dots, n. \end{cases} \quad (19)$$

Решение (18) определяется условием:

$$\hat{\beta}_i = \begin{cases} \tilde{\beta}_i, & \tilde{\beta}_i > 0, \\ 0, & \tilde{\beta}_i \leq 0. \end{cases} \quad (20)$$

Рассмотрим теперь сепарабельную задачу как с ограничениями-неравенствами, так и с одним общим равенством:

$$\begin{cases} J(\beta_1, \beta_2, \dots, \beta_n) = \sum_{i=1}^n (b_{ii}\beta_i^2 + (\mu d_i + cu_i)\beta_i) \rightarrow \min(\beta_1, \beta_2, \dots, \beta_n), \\ \sum_{i=1}^n \beta_i = 1, \beta_i \geq 0, \quad i = 1, \dots, n. \end{cases} \quad (21)$$

На первый взгляд, уже невозможно обеспечить линейную вычислительную сложность таким же простым способом, как в (19)-(20), потому что задача потеряла свою сепарабельность. Тем не менее, это возможно путем некоторого усложнения алгоритма.

Введем обозначение $\mathbb{I} = \{1, \dots, n\}$ для всего набора индексов регрессоров. Пусть задача (21) решена и $(\hat{\beta}_1, \dots, \hat{\beta}_n)$ ее решение. В силу ограничений на знаки, часть коэффициентов регрессии β_i будет простаивать, т. е. коэффициенты будут получать нулевые значения в результате минимизации критерия. Для обозначения множества положительных (активных) коэффициентов будем использовать символ

$$\begin{cases} \hat{\mathbb{I}} = \{i : \hat{\beta}_i > 0\} \subseteq \mathbb{I} = \{1, \dots, n\} - \text{подмножество всех активных индексов,} \\ \mathbb{I} \setminus \hat{\mathbb{I}} = \{i : \hat{\beta}_i = 0\} - \text{подмножество остальных (неактивных) индексов,} \end{cases} \quad (22)$$

В соответствии с постановкой задачи решение удовлетворяет условию

$$\sum_{i=1}^n \hat{\beta}_i = \sum_{i \in \hat{\mathbb{I}}} \hat{\beta}_i = 1. \quad (23)$$

Если в точке минимума (21) рассматривать только активное подмножество, оптимальные коэффициенты регрессии удовлетворяют условию:

$$\begin{cases} (\hat{\beta}_i, i \in \hat{\mathbb{I}}) = \operatorname{argmin} J(\beta_1, \beta_2, \dots, \beta_n) = \sum_{i=1}^n (b_{ii}\beta_i^2 + (\mu d_i + cu_i)\beta_i), \\ \sum_{i \in \hat{\mathbb{I}}} \beta_i = 1. \end{cases} \quad (24)$$

Это задача с одним общим ограничением равенства.

Если мы обозначим $\eta \in \mathbb{R}$ единственный множитель Лагранжа при ограничении равенства, функция Лагранжа будет квадратичной:

$$L(\beta_i, i \in \hat{\mathbb{I}}, \eta) = \sum_{i \in \hat{\mathbb{I}}} (b_{ii}\beta_i^2 + (\mu d_i + cu_i)\beta_i) - \eta \left(\sum_{i \in \hat{\mathbb{I}}} \beta_i - 1 \right) \rightarrow \begin{cases} \min(\beta_i, i \in \hat{\mathbb{I}}), \\ \partial/\partial\eta = 0. \end{cases} \quad (25)$$

Для фиксированного множителя Лагранжа, оптимальные значения переменных определяются равенствами:

$$\frac{\partial}{\partial\beta_i} L(\beta_i, i \in \hat{\mathbb{I}}, \eta) = 2b_{ii}\beta_i + \mu d_i + cu_i - \eta = 0, \quad \hat{\beta}_i(\eta) = \frac{\eta - \mu d_i - cu_i}{2b_{ii}}, \quad i = 1, \dots, n. \quad (26)$$

Это означает, что если бы значение множителя Лагранжа было известно для усеченной задачи (24), то все оптимальные коэффициенты регрессии для полной задачи (21) определялись бы независимо друг от друга, что обеспечивало бы линейную вычислительную сложность алгоритма оптимизации:

$$\hat{\beta}_i(\eta) = \frac{1}{2b_{ii}} \begin{cases} \eta - (\mu d_i + cu_i), & \mu d_i + cu_i < \eta, \\ 0, & \mu d_i + cu_i \geq \eta, \end{cases} \quad i = 1, \dots, n. \quad (27)$$

Условие (23) полностью определяет значение множителя Лагранжа:

$$\sum_{i=1}^n \hat{\beta}_i = \sum_{i: \mu d_i + cu_i < \eta} \hat{\beta}_i = \sum_{i: \mu d_i + cu_i < \eta} \frac{\eta - \mu d_i - cu_i}{2b_{ii}} = \sum_{i: \mu d_i + cu_i < \eta} \frac{\eta}{2b_{ii}} - \sum_{i: \mu d_i + cu_i < \eta} \frac{\mu d_i}{2b_{ii}} - \sum_{i: \mu d_i + cu_i < \eta} \frac{cu_i}{2b_{ii}} = 1, \quad (28)$$

откуда следует уравнение $\eta \sum_{i: \mu d_i + cu_i < \eta} \frac{1}{2b_{ii}} - \sum_{i: \mu d_i + cu_i < \eta} \frac{\mu d_i + cu_i}{2b_{ii}} = 1$, т.е.

$$\hat{\eta} = \frac{1 + \sum_{i: \mu d_i + cu_i < \hat{\eta}} ((\mu d_i + cu_i)/2b_{ii})}{\sum_{i: \mu d_i + cu_i < \hat{\eta}} (1/2b_{ii})}. \quad (29)$$

Таким образом, решение начальной задачи оптимизации (21) полностью определено в соответствии с (27):

$$\begin{cases} \hat{\beta}_i = \frac{1}{2b_{ii}} \begin{cases} \hat{\eta} - (\mu d_i + cu_i), & \mu d_i + cu_i < \hat{\eta}, \\ 0, & \mu d_i + cu_i \geq \hat{\eta}, \end{cases} \\ i = 1, \dots, n, \end{cases} \quad \hat{\eta} = \frac{1 + \sum_{i: \mu d_i + cu_i < \hat{\eta}} ((\mu d_i + cu_i)/2b_{ii})}{\sum_{i: \mu d_i + cu_i < \hat{\eta}} (1/2b_{ii})}. \quad (30)$$

Легко видеть, что равенство (23) всегда выполняется, если η удовлетворяет уравнению (29). Осталось только найти способ решения этого уравнения.

Расставим коэффициенты $(\mu d_i + cu_i)$ в критерии (21) в порядке возрастания

$$(\mu d_1 + cu_1) < (\mu d_2 + cu_2) < \dots < (\mu d_n + cu_n), \quad (31)$$

и рассмотрим возрастающую функцию

$$z(\eta) = \sum_{i: \mu d_i + cu_i < \eta} \frac{\eta - (\mu d_i + cu_i)}{2b_{ii}} \geq 0, \quad -\infty < \eta < \infty. \quad (32)$$

Понятно, что $z(\eta) = 0$ при $\eta \leq \mu d_1 + cu_1$, и что $z(\eta) < 1$ пока $\eta < \hat{\eta}$ (29), и, наконец, $z(\eta) > 1$ когда $\eta > \hat{\eta}$. Найдем целое число \hat{n} такое что $z(u_{\hat{n}}) < 1$ и $z(u_{\hat{n}+1}) \geq 1$.

Как только это будет сделано, мы полностью определили:

- множество активных регрессоров в точке минимума задачи с обоими видами ограничений (21) $\hat{\mathbb{I}} = \{1, \dots, \hat{n}\} \subseteq \mathbb{I}$;
- значение множителя Лагранжа $\hat{\eta}$ (30);
- оптимальные значения всех переменных

$$\hat{\beta}_i = \frac{1}{2b_{ii}} \begin{cases} \hat{\eta} - (\mu d_i + cu_i), & (\mu d_i + cu_i) < \hat{\eta}, \\ 0, & (\mu d_i + cu_i) \geq \hat{\eta}. \end{cases}, \quad i = 1, \dots, n. \quad (33)$$

Мы сумели получить искомую линейную сложность.

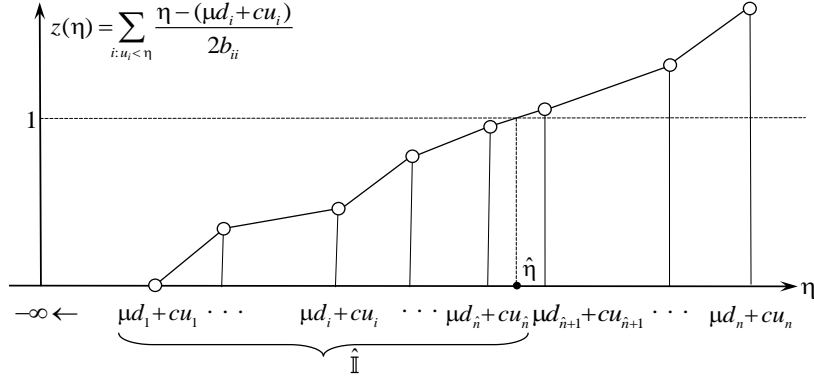


Рисунок 1. Нахождение множителя Лагранжа и подмножества активных регрессоров.

4.2 Последовательность сепарабельных задач квадратичной оптимизации вместо одной несепарабельной

4.2.1 Разделение переменных

Давайте еще раз рассмотрим агрегированную задачу оптимизации (16). Положительно определенная матрица квадратичной формы $\mathbf{B} = (b_{il}, i, l = 1, \dots, n)$ в общем случае не является диагональной. Ее всегда можно представить как сумму двух матриц, одна из которых диагональная:

$$\mathbf{B} = \mathbf{B}_d + \mathbf{B}_{nd}, \quad \mathbf{B}_d = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & b_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{nn} \end{pmatrix}, \quad \mathbf{B}_{nd} = \begin{pmatrix} 0 & b_{12} & \cdots & b_{1n} \\ b_{21} & 0 & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & 0 \end{pmatrix}. \quad (34)$$

Задача (16) переписывается в форме:

$$\begin{cases} \boldsymbol{\beta}^T \mathbf{B}_d \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{B}_{nd} \boldsymbol{\beta} + (\boldsymbol{\mu} \mathbf{d} + \mathbf{c} \mathbf{u})^T \boldsymbol{\beta} \rightarrow \min(\boldsymbol{\beta}), \\ \mathbf{1}^T \boldsymbol{\beta} = 1, \boldsymbol{\beta} \geq \mathbf{0}, \end{cases} \quad (35)$$

Нам будет удобно рассмотреть альтернативную задачу квадратичного программирования с удвоенным числом переменных, которая получается из исходной задачи путем разбиения каждой переменной на двоянную пару:

$$\begin{cases} J(\boldsymbol{\beta}', \boldsymbol{\beta}'') = \\ \frac{1}{2} \boldsymbol{\beta}'^T \mathbf{B}_d \boldsymbol{\beta}' + \frac{1}{2} \boldsymbol{\beta}''^T \mathbf{B}_d \boldsymbol{\beta}'' + \boldsymbol{\beta}'^T \mathbf{B}_{nd} \boldsymbol{\beta}'' + \frac{1}{2} (\boldsymbol{\mu} \mathbf{d} + \mathbf{c} \mathbf{u})^T \boldsymbol{\beta}' + \frac{1}{2} (\boldsymbol{\mu} \mathbf{d} + \mathbf{c} \mathbf{u})^T \boldsymbol{\beta}'' \rightarrow \min(\boldsymbol{\beta}', \boldsymbol{\beta}''), \\ \mathbf{1}^T \boldsymbol{\beta}' = 1, \mathbf{1}^T \boldsymbol{\beta}'' = 1, \boldsymbol{\beta}' \geq \mathbf{0}, \boldsymbol{\beta}'' \geq \mathbf{0}. \end{cases} \quad (36)$$

Видно, что задачи (35) и (36) совпадают, если $\boldsymbol{\beta}' = \boldsymbol{\beta}''$.

Чтобы убедиться, что квадратичная целевая функция $J(\boldsymbol{\beta}', \boldsymbol{\beta}'')$ в расщепленной задаче (36) выпукла относительно удвоенного аргумента, достаточно рассмотреть ее гессиан:

$$\begin{pmatrix} \boldsymbol{\beta}' \\ \boldsymbol{\beta}'' \end{pmatrix} \in \mathbb{R}^{2n}, \quad \mathbf{H} = \begin{pmatrix} \nabla_{\boldsymbol{\beta}'}^2 J(\boldsymbol{\beta}', \boldsymbol{\beta}'') & \nabla_{\boldsymbol{\beta}'} \nabla_{\boldsymbol{\beta}''} J(\boldsymbol{\beta}', \boldsymbol{\beta}'') \\ \nabla_{\boldsymbol{\beta}''} \nabla_{\boldsymbol{\beta}'} J(\boldsymbol{\beta}', \boldsymbol{\beta}'') & \nabla_{\boldsymbol{\beta}''}^2 J(\boldsymbol{\beta}', \boldsymbol{\beta}'') \end{pmatrix} = \begin{pmatrix} \mathbf{B}_d & \mathbf{B}_{nd} \\ \mathbf{B}_{nd} & \mathbf{B}_d \end{pmatrix} (2n \times 2n). \quad (37)$$

Легко показать, что эта положительно определенная матрица имеет $2n$ собственных значений, только n из которых могут быть различны $\lambda_1 > 0, \dots, \lambda_n > 0$, и это собственные значения положительно определенной матрицы \mathbf{V} ($n \times n$).

Таким образом, задача квадратичного программирования (36) является выпуклой. Идея итерационного алгоритма, которую мы рассмотрим ниже, чередует два отдельных шага минимизации по отношению к β' и β'' .

4.2.2 Две симметричные сепарабельные задачи квадратичного программирования

Альтернативная задача квадратичного программирования (36) предполагает два взаимосвязанных шага, в каждом из которых одна из двух версий расщепленной переменной $\beta' \in \mathbb{R}^n$ и $\beta'' \in \mathbb{R}^n$ действует как основная переменная, тогда как другая остается константой.

$$\begin{cases} J'(\beta' | \beta'') = \frac{1}{2} \beta'^T \mathbf{V}_d \beta' + \beta''^T \mathbf{V}_{nd} \beta' + \frac{1}{2} (\mu \mathbf{d} + c \mathbf{u})^T \beta' \rightarrow \min(\beta' | \beta''), \\ \mathbf{1}^T \beta' = 1, \beta' \geq \mathbf{0}, \end{cases} \quad (38)$$

$$\begin{cases} J''(\beta'' | \beta') = \frac{1}{2} \beta''^T \mathbf{V}_d \beta'' + \beta'^T \mathbf{V}_{nd} \beta'' + \frac{1}{2} (\mu \mathbf{d} + c \mathbf{u})^T \beta'' \rightarrow \min(\beta'' | \beta'), \\ \mathbf{1}^T \beta'' = 1, \beta'' \geq \mathbf{0}. \end{cases} \quad (39)$$

Скалярная запись связанных задач (38) и (39) показывает, что каждая из них сепарабельна:

$$J'(\beta'_1, \dots, \beta'_n | \beta'') = \frac{1}{2} \sum_{i=1}^n (b_{ii} \beta_i'^2 + u_i' \beta_i'), \quad J''(\beta''_1, \dots, \beta''_n | \beta') = \frac{1}{2} \sum_{i=1}^n (b_{ii} \beta_i''^2 + u_i'' \beta_i''), \quad (40)$$

где u_i' и u_i'' определяются равенствами:

$$u_i' = 2 \sum_{l=1, l \neq i, l: \beta_l' > 0}^n b_{il} \beta_l'' + (\mu d_i + c u_i), \quad u_i'' = 2 \sum_{l=1, l \neq i, l: \beta_l'' > 0}^n b_{il} \beta_l' + (\mu d_i + c u_i). \quad (41)$$

Сепарабельность обеих задач (38)-(39) предполагает идею итеративного решения задачи в целом (35) путем итерационного чередования двух сепарабельных задач квадратичного программирования относительно β' и β'' .

4.3 Итерационный алгоритм

В качестве начального приближения $\beta'^0 = \beta''^0 = \tilde{\beta}$ мы принимаем решение задачи без ограничений на знаки. Далее в соответствии с алгоритмом (31)-(33) описанном в Разделе 4.1.2

1) Шаг относительно β' :

$$\begin{cases} (\beta_i'^{k+1}) = \arg \min_{\beta_i'} J'(\beta'_1, \dots, \beta'_n | \beta''^k) = \arg \min_{\beta_i'} \frac{1}{2} \sum_{i=1}^n (b_{ii} \beta_i'^2 + u_{\beta''^k, i} \beta_i'), \quad i = 1, \dots, n, \\ \sum_{i=1}^n \beta_i' = 1, \beta_i' \geq 0; \end{cases}$$

вычисление сумм $\sum_{l=1, l \neq i, l: \beta_l'' > 0}^n b_{il} \beta_l''$, $i = 1, \dots, n$ в (41) справа.

2) Шаг относительно β'' :

$$\begin{cases} (\beta_i^{n^{k+1}}) = \arg \min_{\beta_i^n} J^n(\beta_1^n, \dots, \beta_n^n | \beta^{n^{k+1}}) = \arg \min_{\beta_i^n} \frac{1}{2} \sum_{i=1}^n (b_{ii} \beta_i^{n^2} + u_{\beta^{k+1}, i} \beta_i^n), & i = 1, \dots, n, \\ \sum_{i=1}^n \beta_i^n = 1, \beta_i^n \geq 0; \end{cases}$$

вычисление сумм $\sum_{l=1, l \neq i, l: \beta_l^{k+1} > 0} b_{il} \beta_l^{k+1}$, $i = 1, \dots, n$ в (41) слева.

3) Следующее значение искомого вектора: $\beta^{k+1} = \frac{1}{2} (\beta^{k+1} + \beta^{n^{k+1}})$.

Условие останова: $(\beta^{k+1} - \beta^{n^{k+1}})^T (\beta^{k+1} - \beta^{n^{k+1}}) < \varepsilon$.

Так как на каждом шаге гарантируется уменьшение целевой функции

$$J(\beta^{k+1}, \beta^{n^{k+1}}) < J(\beta^k, \beta^{n^k}) \text{ пока } \beta^k \neq \beta^{n^k},$$

и область поиска выпуклая, достаточным условием сходимости является положительная определенность гессиана (37).

5 Экспериментальное исследование метода оценки задачи регрессии с ограничениями и регуляризацией

5.1 Организация эксперимента

Целью экспериментов является изучение возможности процедуры регуляризованного фактор-поиска, описанной в Разделах 3 и 4, оценивание реальной модели sparse-регрессии из большого набора коррелированных переменных. В качестве набора регрессоров мы использовали 650 временных рядов месячных индексов фондового рынка

$$\mathbf{x}_i = (x_{i,t}, i = 1, \dots, n) \in \mathbb{R}^n, n = 650, \quad (42)$$

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

$$\mathbf{X} = (\mathbf{x}_i, t = 1, \dots, T) (n \times T). \quad (43)$$

Помещенные в порядке убывания, собственные значения их ковариационной матрицы Σ (5) быстро убывают $\lambda_1 = 16369.8$, $\lambda_2 = 1435.2$, $\lambda_3 = 1076.2$, $\lambda_{20} = 134.0 < 0.01\lambda_1$, $\lambda_{50} = 32.2 < 0.002\lambda_1$, что свидетельствует о сильной корреляции между индексами.

Идея экспериментов заключается в попытке восстановить некоторый фактический набор бета-коэффициентов $(\beta_1^*, \dots, \beta_m^*)$ составляющих разумный инвестиционный портфель, с точки зрения теории Марковица для данной толерантности к риску $0 \leq \mu < \infty$ (Раздел 2.3).

Согласно правилу (3)

$$\begin{cases} (\beta_{\mu,1}^*, \dots, \beta_{\mu,n}^*) = \arg \min (\beta^T \Sigma \beta - \mu \bar{\mathbf{x}}^T \beta), \\ \mathbf{1}^T \beta = 1, \beta \geq \mathbf{0}, \end{cases}$$

мы вычислили 5 наборов β -коэффициентов, в дальнейшем называемых подлинными разумными составами портфеля для следующих значений коэффициента толерантности к риску:

$$\mu_1 = 0, \mu_2 = 10, \mu_3 = 30, \mu_4 = 200, \mu_5 = 1000. \quad (44)$$

Как и ожидалось, итоговые портфели оказались очень разреженными в множестве 650 фондовых индексов. Это означает, что подмножество активных индексов $\mathbb{I}_\mu^* = \{i: \beta_{\mu,i}^* > 0\}$ содержит гораздо меньшее число индексов, чем все множество индексов $\mathbb{I} : m = |\mathbb{I}_\mu^*| \ll n = |\mathbb{I}| = 650$:

Толерантность к риску	$\mu_1 = 0$	$\mu_2 = 10$	$\mu_3 = 30$	$\mu_4 = 200$	$\mu_5 = 1000$
Количество активных индексов	$m_1 = 13$	$m_2 = 11$	$m_3 = 10$	$m_4 = 7$	$m_5 = 1$

Для каждого подлинного разумного состава портфеля $(\beta_{\mu,i}^*, i \in \mathbb{I}_\mu^*)$, мы вычислили временной ряд доходностей гипотетического портфеля

$$\mathbf{y}_\mu = (y_{\mu,t}, t=1, \dots, T) \in \mathbb{R}^T, \quad y_{\mu,t} = \sum_{i \in \mathbb{I}_\mu^*} \beta_{\mu,i}^* x_{t,i} + \xi_t, \quad (45)$$

как последовательность случайных величин с 10%-ой шумовой дисперсией $\sigma^2(\xi_t) = 0.1 \left((1/T) \sum_{i \in \mathbb{I}_\mu^*} \beta_{\mu,i}^* x_{t,i} \right)$ в зависимости от соответствующей последовательности коэффициентов регрессии. Эти пять временных рядов доходностей и вся совокупность индексов (43)

$$\mathbf{y}_\mu = (y_{\mu,t}, t=1, \dots, T), \quad \mu = \mu_q, \quad q=1, \dots, 5, \quad \mathbf{X} = (\mathbf{x}_t, t=1, \dots, T), \quad (46)$$

служат основой нашего эксперимента.

5.2 Слепой фактор-поиск

Слепой фактор-поиск предполагает, что мы пытаемся восстановить предположительно подлинную регрессионную модель по критерию чистых нерегуляризованных наименьших квадратов с ограничениями-равенствами и неравенствами (11).

$$(\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}) \rightarrow \min(\boldsymbol{\beta}), \quad \mathbf{1}^T \boldsymbol{\beta} = 1, \quad \boldsymbol{\beta} \geq \mathbf{0}, \quad (47)$$

применяется к каждому из пяти смоделированных временных рядов $\mathbf{y}_\mu = (y_{\mu,t}, t=1, \dots, T)$ (46). После нахождения вектора расчетных коэффициентов регрессии $\hat{\boldsymbol{\beta}}_\mu$ within в подмножестве активных регрессоров $\hat{\mathbb{I}}_\mu$, для него был вычислен показатель ошибки наименьших квадратов (PLSE)

$$PLSE = \frac{(1/T) \sum_{t=1}^T \left(y_t - \sum_{i \in \hat{\mathbb{I}}_\mu} \hat{\beta}_{i,\mu} x_{t,i} \right)^2}{(1/T) \sum_{t=1}^T \left(\sum_{i \in \hat{\mathbb{I}}_\mu} \hat{\beta}_{i,\mu} x_{t,i} \right)^2} \times 100. \quad (48)$$

Результаты этого анализа представлены на Рисунке 2. Мозаичные диаграммы показывают подлинные и оценочные подмножества активных регрессоров (биржевых индексов), пронумерованные в данном наборе индексов (42) по мере уменьшения средней доходности $\bar{x}_1 > \dots > \bar{x}_i > \dots > \bar{x}_n$. Каждая диаграмма снабжена PLSE (48).



Из Рис 2 хорошо видно, что, за исключением тривиального случая полной толерантности к риску $\mu \rightarrow \infty$, более 50% индексов, будучи активными в фактическом или оценочном портфеле, ошибочно идентифицируются методом поиска слепого фактора. Существенное несоответствие между фактическим и оценочным составами портфеля свидетельствует о необходимости дополнительной регуляризации на основе твердой уверенности в том, что анализируемый портфель относится к классу разумных портфелей Марковица.

5.3 Фактор-поиск на основе априорной информации

Индикатор $PLSE$ на Рис 2 показывает, что каждый из анализируемых временных рядов доходности приближен почти на уровне ошибок, близком к 10%-ной дисперсии шума в моделируемом портфеле (45). Тем не менее, оценочные наборы активных регрессоров существенно отклоняются от моделируемых. Этот факт наталкивает на идею немного подтолкнуть поиск к разумным подмножествам, в нашем случае, разумным композициям портфолио. Эта идея которая выражается уменьшением функции регуляризации $V(\beta | \mu) = \beta^T \Sigma \beta - \mu \bar{x}^T \beta$ (3) и увеличением транспонированной полезности портфеля (47).

Вместо нерегуляризованного дважды ограниченного критерия наименьших квадратов (47), как в Разделе 5.2, мы применили к каждому из пяти смоделированных временных рядов $y_\mu = (y_{\mu,t}, t = 1, \dots, T)$ (46), $\mu = \mu_1, \dots, \mu_5$ (44), пять критериев, дополнительно обогащенных слагаемым априорной регуляризацией с разными коэффициентами $\mu = \mu_m, m = 1, \dots, 5$:

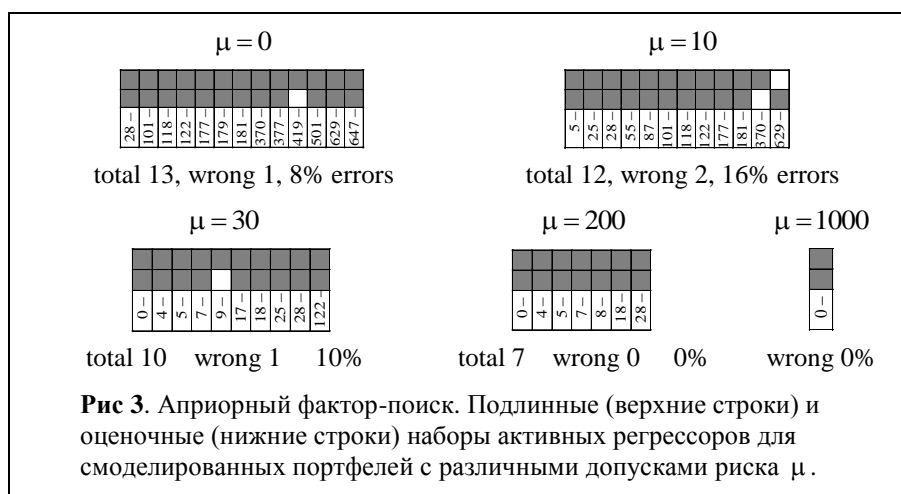
$$\beta^T \Sigma \beta - \mu_m \bar{x}^T \beta + \frac{1}{T} (y_\mu - X^T \beta)^T (y_\mu - X^T \beta) \rightarrow \min(\beta), \mathbf{1}^T \beta = 1, \beta \geq 0. \quad (49)$$

Таким образом, получили пять расчетных векторов коэффициентов $(\hat{\beta}_\mu | \alpha_m)$ в подмножестве активных регрессоров $(\hat{\Pi}_\mu | \mu_m)$. Для каждого из них мы вычислили индикатор leave-one-out ($LOO_\mu | \mu_m$) регрессионной модели в этом соответствующем активном множестве только при ограничении равенства $\sum_{i=1}^n \beta_i = 1$. Наконец, мы выбрали коэффициент толерантности к риску, наиболее соответствующий данному конкретному моделируемому временному ряду доходностей y_μ по критерию LOO

$$m^* = \arg \min(LOO_\mu | \mu_m), \quad m = 1, \dots, 5.$$

Соответствующий вектор регрессии $(\hat{\beta}_\mu | \mu_m)$ и активное подмножество $(\hat{\Pi}_\mu | \mu_m)$ рассматриваются в качестве расчетной модели анализируемого портфеля y_μ .

Эти результаты представлены на Рис. 3. Очевидно, что априорная регуляризация принципиально улучшила качество фактор-поиска.



6 Публикации по материалам работы

O. Krasotkina, M. Markov, V. Mottl, D. Babichev, I. Pugach, A. Morozov. Constrained Regularized Regression Model Search in Large Sets of Regressors. *Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science*, Springer, 2018 (to appear).

Пугач И.А., Моттль В.В., Красоткина О.В. Алгоритмы анализа динамики стиля инвестиций по большому числу биржевых индексов. Тезисы докладов 18-й Всероссийской конференции Математические методы распознавания образов ММРО-2017. Москва, Торус Пресс, 2017, с. 68.

7 Положения, выносимые на защиту

- 1) Формальная постановка задачи оценивания скрытого состава инвестиционного портфеля как математической задачи поиска малого числа фактически активных регрессоров в большом множестве потенциальных регрессоров (Factor Search).
- 2) Способ регуляризации регрессионной модели доходности портфеля на основе предположения о целях инвестора, сформировавшего портфель.
- 3) Однопараметрическое семейство регуляризаторов в регрессионной модели портфеля по принципу Парето между взаимно противоречивыми требованиями максимизации доходности и минимизации риска потери капитала.
- 4) Алгоритм оценивания состава портфеля с линейной вычислительной сложностью относительно большого числа биржевых индексов.

Список литературы

1. Sharpe, W.F. Asset allocation: Management style and performance measurement. *The Journal of Portfolio Management*, Winter 1992, pp. 7-19.
2. Meinshausen, N. Sign-constrained least squares estimation for high-dimensional regression. *Electronic Journal of Statistics*, 2013, Vol. 7, pp. 1607-1631.
3. Markowitz, Harry. Portfolio Selection. *The Journal of Finance*, Mar., 1952, Vol. 7, No. 1, pp. 77-91.
4. Sharpe, W.F. Capital asset prices with and without negative holdings. Nobel Lecture, December 7, 1990. https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1990/sharpe-lecture.pdf
5. Sharpe, W.F. Capital asset prices: a theory of market equilibrium under conditions of risk. *The Journal of Finance*, September 1964, Vol. XIX, No. 3, pp. 425-442.
6. Hoerl, A.E., Kennard, D.J. Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 1962, 58, pp. 54-59.
7. Vinod H.D., Ullah A. Recent advances in regression methods. *Statistics: Textbooks and Monographs*, New York: Marcel Dekker Inc., 41, 1981.
8. Frank, I.E., Friedman, J.H.. A statistical view of some chemometrics regression tools, *Technometrics*, 1993, 35, pp. 109-148.
9. Fu, W.J. Penalized regression: The bridge versus the LASSO. *Journal of Computational and Graphical Statistics*, 1998,7, pp. 397-416.
10. Tibshirani, Robert (1996). Regression Shrinkage and Selection via the lasso. *Journal of the Royal Statistical Society. Series B (methodological)*. Wiley. 58 (1): 267–88.
11. Zou, Hui; Hastie, Trevor (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (statistical Methodology)*. Wiley. 67 (2): 301–20. Zou, Hui; Hastie, Trevor (2005). "Regularization and Variable Selection via the Elastic Net". *Journal of the Royal Statistical Society. Series B (statistical Methodology)*. Wiley. 67 (2): 301–20.
12. Fan, J., Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, December 2001, Vol. 96, Theory and Methods, No. 456, pp. 1348-1360.

13. Sham M. Kakade, Shai Shalev-Shwartz, Ambuj Tewari. Regularization Techniques for Learning with Matrices. Cornell Univesity. Submitted on 4 Oct 2009, last revised 17 Oct 2010. <https://arxiv.org/pdf/0910.0610.pdf>
14. Kienzle, W. and Chellapilla, K. Personalized handwriting recognition via biased regularization. Proceedings of the 23rd International Conference on Machine Learning (ICML 2006), Pittsburgh, Pennsylvania, USA, June 25-29, 2006, pp. 457-464.
15. Tyson, Eric. Investing For Dummies. John Wiley & Sons, 2011, 410 p.
16. Wang, X., Dunson, D., Leng., C. No penalty no tears: Least squares in high-dimensional linear models. Proceedings of the 33rd International Conference on Machine Learning ICML'16. New York, USA, June 19-24, 2016, Vol. 48, pp. 1814-1822.
17. Tikhonov A.N., Arsenin V.Y. Solution of Ill-posed Problems. Washington, Winston & Sons, 1977.
18. Ruszczyński, Andrzej. Nonlinear Optimization. Princeton University Press, 2006, 448 p.
19. Zhou, Hua, Lange, Kenneth. A Path Algorithm for Constrained Estimation. Cornell Univesity. Submitted on 19 Mar 2011. <http://citeseerx.ist.psu.edu/index.jsessionid=91A8D239D45E9E841FF9C74FF30C5214>
20. Silvia M. H. Janesch Lucio Tunes Santos. Exact Penalty Methods with Constrained Subproblems. Invesigacion Operativa, June 1997. <http://www-2.dc.uba.ar/alio/io/pdf/1997/paper-4.pdf>
21. Meinshausen, Nicolai. Sign-constrained least squares estimation for high-dimensional regression. Electronic Journal of Statistics, 2013, Vol. 7, pp. 1607-1631.
22. Gill, P. E., Murray, W., Wright, M.H. Practical Optimization, Academic, London, 1981.
23. Rasmus Bro, Sijmen de Jong. A fast non-negativity-constrained least squares algorithm. Journal of Chemometrics, 1997, Vol. 11, pp. 393-401.