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

И.А. Пугач
МФТИ

Научный руководитель

Д.т.н, профессор В.В. Моттль

Доходности активов и инвестиционного портфеля

Инвестиционный портфель: Совокупность видов активов $i=1,...,n$, в которые вложен капитал инвестиционной компании (обычно акций или облигаций).

Последовательность периодов владения – дней, месяцев, кварталов $t=1,2,...,T$

Доходность портфеля (return) за период владения.

Стоимость портфеля в начале $z_{0,t}^{(p)}$ и в конце $z_{1,t}^{(p)}$ очередного периода владения

Эта стоимость неизвестна (тайна инвестиционной компании).

Известны лишь ***периодические доходности*** (относительные приращения стоимости портфеля) которые компания обязана публиковать

$$y_t = \frac{z_{конец,t}^{(p)} - z_{\text{начало},t}^{(p)}}{z_{\text{начало},t}^{(p)}}$$

Динамическое долевое распределение капитала:

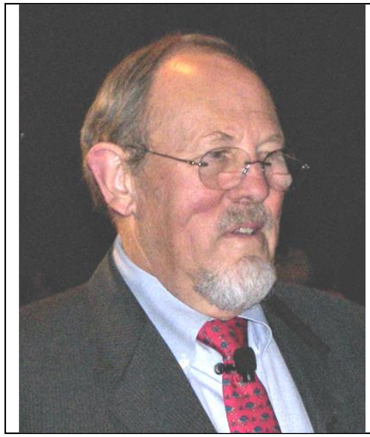
Последовательность векторов $\boldsymbol{\beta}_t = (\beta_{t,1},...,\beta_{t,n}) \in \mathbb{R}^n$, выражающий доли общего объема капитала, вложенные в каждый из видов активов.

Известные цены и доходности активов $x_{t,i} = \dfrac{z_{конец,t}^{(i)} - z_{\text{начало},t}^{(i)}}{z_{\text{начало},t}^{(i)}}$	Модель доходности портфеля $y_t \cong \sum_{i=1}^{n} \beta_i x_{t,i}$

Для паевых фондов (пенсионных) $\sum_{i=1}^{n} \beta_i = 1$ – бюджетное ограничение, $\beta_i \geq 0$ – нет заёмного капитала

Если фонду разрешено заимствование капитала, т.е. допускается $\beta_{t,i} < 0$, то при умелом управлении доходность портфеля может быть существенно увеличена. Однако тогда резко возрастает и риск разорения, поэтому социально значимым фондам это запрещено.

Задача оценивания состава инвестиционного портфеля (стиля инвестирования) по известной информации о доходности



William F. Sharpe, Professor of Finance, Emeritus
Graduate School of Business, Stanford University

Лауреат Нобелевской премии по экономике 1990 года (совместно с **Harry M. Markowitz**):

– Capital Asset Pricing Model (CAPM)
– **Returns Based Style Analysis (RBSA)**
for evaluating the style and performance of investment funds

Доступная информация: $\left[(y_t,\, x_{t,i},\, i=1,\ldots,n),\, t=1,2,3,\ldots \right]$ – временные ряды, образованные доходностями портфеля и активов, в которые предположительно инвестирован капитал, для последовательных периодов владения, например, рабочих дней биржи, месяцев, кварталов.

Уильям Шарп предложил аппроксимировать доходность портфеля небольшим числом биржевых индексов, представляющих разные стили инвестирования, решая задачу оценивания линейной регрессии при линейных ограничениях.	**Задача квадратичного программирования:** $\begin{cases} (\hat{\beta}_1,\ldots,\hat{\beta}_n): \sum_{t=1}^{N} \left(y_t - \sum_{i=1}^{n} \beta_i x_{t,i} \right)^2 \to \min, \\ \sum_{i=1}^{n} \beta_i = 1,\ \beta_i \ge 0. \end{cases}$

Задача оценивания скрытого состава инвестиционного портфеля – Factor Search

Модель Шарпа	$\sum_{t=1}^{T}\left(y_t-\sum_{i=1}^{n}\beta_i x_{t,i}\right)^2\to\min,\quad \underbrace{\sum_{i=1}^{n}\beta_i=1}_{\text{долевое распределение капитала}},\ \underbrace{\beta_i\ge 0}_{\text{запрет заемного капитала}},\ i=1,\dots,n.$
$y_t,\ t=1,\dots,T$	известные доходности портфеля (дни, месяцы, кварталы, годы)
$x_{t,i},\ i\in\mathbb{I}=\{1,\dots,n\}$	доходности очень большого числа биржевых активов, $n\gg T$
$\hat{\mathbb{I}}\subset\mathbb{I},\quad \hat{n}=\lvert\hat{\mathbb{I}}\rvert\ll n=\lvert\mathbb{I}\rvert$	Найти относительно небольшое подмножество активов, из которых фактически состоит портфель …
	$\sum_{t=1}^{T}\left(y_t-\sum_{i\in\hat{\mathbb{I}}}\beta_i x_{t,i}\right)^2\to\min,\quad \sum_{i\in\hat{\mathbb{I}}}\beta_i=1,\ \beta_i\ge 0,\ i\in\hat{\mathbb{I}}$
	… и фактическое долевое распределение капитала между ними $\hat{\beta}_i,\ i\in\hat{\mathbb{I}}$

Это задача поиска иголки в стоге сена

Регуляризация – использование априорной информации о возможном составе

$\sum_{t=1}^{T}\left(y_t-\sum_{i=1}^{n}\beta_i x_{t,i}\right)^2 \to \min$	Минимизация остаточной суммы квадратов: аппроксимация массива данных

Основная идея данной работы

Мыслить как инвестор. Какой портфель он предпочел бы?

1)Диверсификация вложений капитала

$\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j \to \min$	Предпочтение малокоррелированных активов $\left(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j \quad i,j=1,\ldots,n\right)$ – ковариационная матрица их доходностей

2)Стремление увеличить доходность портфеля

$\bar{y}=\sum_{i=1}^{n}\beta_i\bar{x}_i \to \max$	Предпочтение доходных активов Средние доходности активов $\bar{x}_i$ и итоговая доходность портфеля $\bar{y}$

Вопрос:

Как соотносятся эти два требования? Как совместить их друг с другом?

Теория Марковица о выборе инвестиционного портфеля

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j \to \min \qquad \sum_{i=1}^{n}\bar{\boldsymbol{x}}_i\beta_i \to \max$$

$$\sum_{i=1}^{n}\bar{\boldsymbol{x}}_i\beta_i = const \qquad \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j = const$$

Эти два требования взаимно противоречивые требования.
Кто не рискует, тот не пьет щампанского!
Risk Tolerance – индивидуальная характеристика каждого инвестора.

Утверждение данной работы:

$\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j - \mu\sum_{i=1}^{n}\bar{\boldsymbol{x}}_i\beta_i \to \min$	Регуляризующая функция, обеспечивающая баланс Марковица

Итоговый критерий анализа портфеля

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j - \mu\sum_{i=1}^{n}\bar{\boldsymbol{x}}_i\beta_i + c\sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{n}\beta_i x_{t,i}\right)^2 \to \min \\ \sum_{i=1}^{n}\beta_i = 1,\ \beta_i \ge 0 \ \ , i = 1,\dots,n \end{cases}$$

.

Специальный алгоритм решения задачи квадратичного программирования.

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j - \mu\sum_{i=1}^{n}\bar{\boldsymbol{x}}_i\beta_i + c\sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{n}\beta_i x_{t,i}\right)^2 \to \min \\ \sum_{i=1}^{n}\beta_i = 1,\ \beta_i \ge 0 \quad, i = 1,\dots,n \end{cases}$$

Преобразуем критерий к агрегированной форме:

$$\begin{cases} \boldsymbol{\beta}^{\mathrm{T}}\mathbf{B}\boldsymbol{\beta} + (\mu\mathbf{d} + c\mathbf{u})^{\mathrm{T}}\boldsymbol{\beta} \to \min(\boldsymbol{\beta}), \\ \mathbf{1}^{\mathrm{T}}\boldsymbol{\beta} = 1,\ \boldsymbol{\beta} \ge \mathbf{0}, \end{cases}$$

Разделяем матрицу на диагональную и оставшуюся часть: $\mathbf{B} = \mathbf{B}_d + \mathbf{B}_{nd}$

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

И удваиваем число переменных:

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

Селективность критерия

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j - \mu\sum_{i=1}^{n}\overline{\boldsymbol{x}}_i\beta_i + c\sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{n}\beta_i x_{t,i}\right)^2 \to \min \\ \sum_{i=1}^{n}\beta_i = 1,\ \beta_i \ge 0\ \ ,\ i = 1,\dots,n \end{cases}$$

Если вычислять β только на основании регуляризации, то мы получим:

$$\begin{cases} (\beta^*_{\mu,1},\dots,\beta^*_{\mu,n}) = \arg\min\left(\boldsymbol{\beta}^T\boldsymbol{\Sigma\beta} - \mu\overline{\mathbf{x}}^T\boldsymbol{\beta}\right), \\ \mathbf{1}^T\boldsymbol{\beta} = 1, \boldsymbol{\beta} \ge \mathbf{0}, \end{cases}$$

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

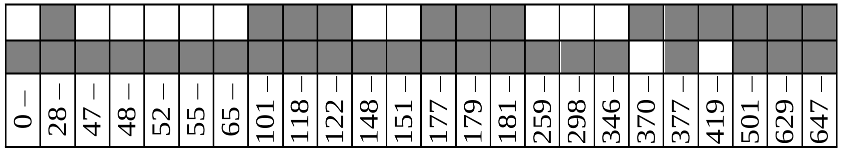
Толерантность к риску	$\mu_1 = 0$	$\mu_2 = 10$	$\mu_3 = 30$	$\mu_4 = 200$	$\mu_5 = 1000$
Количество активных индексов	$m_1 = 13$	$m_2 = 11$	$m_3 = 10$	$m_4 = 7$	$m_5 = 1$
Всего индексов	$n = 650$				

Эксперименты

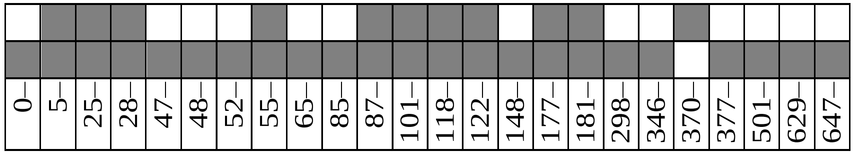
$n = 650$ биржевых индексов, $T = 240$ месяцев (10 лет)

Слепой фактор-поиск

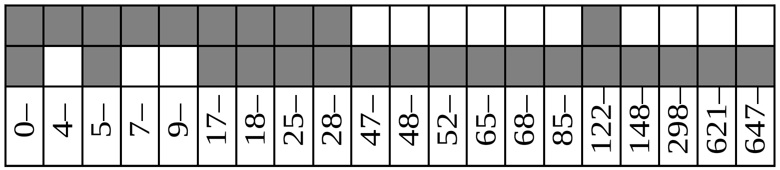
$$\begin{cases} \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{n} \beta_i x_{t,i} \right)^2 \to \min \\ \sum_{i=1}^{n} \beta_i = 1, \ \beta_i \geq 0 \ , i = 1, \dots, n \end{cases}$$



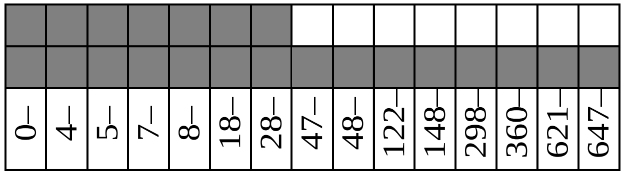
$\mu = 0$ total 24, wrong 13, 54% errors



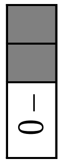
$\mu = 10$ total 24, wrong 14, 58% errors



$\mu = 30$ total 20, wrong 13, 65%



$\mu = 200$ total 15, wrong 8, 53%



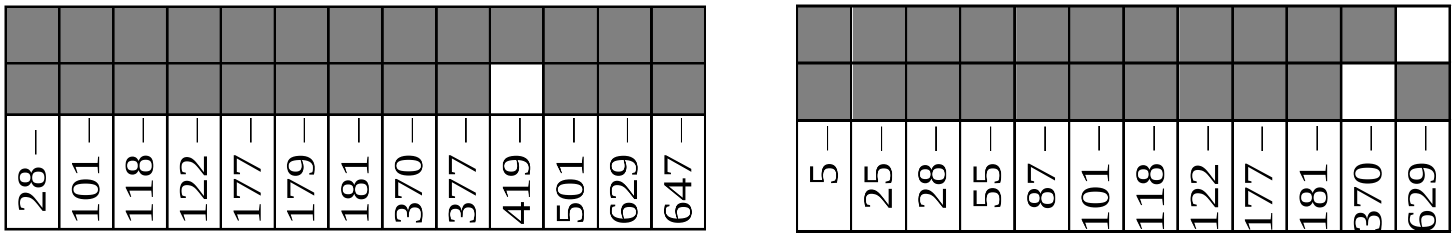
$\mu = 1000$ wrong 0%

Эксперименты

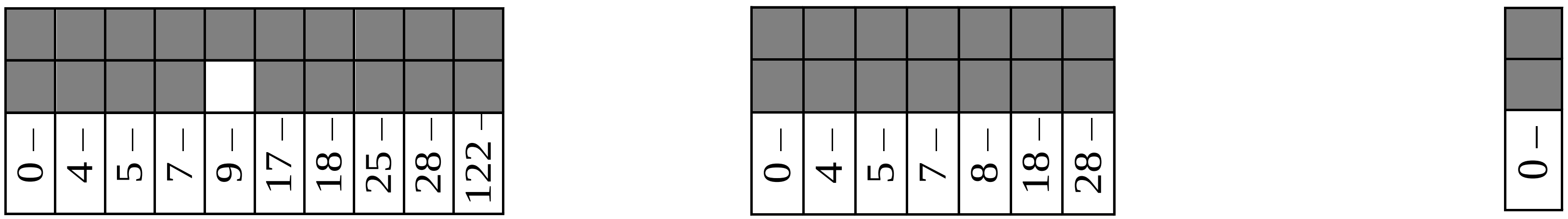
$n = 650$ биржевых индексов, $T = 240$ месяцев (10 лет)

Фактор-поиск с априорными предположениями:

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j - \mu\sum_{i=1}^{n}\bar{\boldsymbol{x}}_i\beta_i + c\sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{n}\beta_i x_{t,i}\right)^2 \to \min \\ \sum_{i=1}^{n}\beta_i = 1,\ \beta_i \ge 0 \quad , i = 1,\dots,n \end{cases}$$



$\mu = 0$ total 13, wrong 1, 8% errors $\mu = 10$ total 12, wrong 2, 16% errors



$\mu = 30$ total 10 wrong 1 10% $\mu = 200$ total 7 wrong 0 0% $\mu = 1000$ wrong 0%

Спасибо за внимание

Публикации по материалам диссертации:

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.

Krasotkina O., Markov M., Mottl V., Babichev D., Pugach I, Morozov A.. 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).