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

**И.А. Пугач**
МФТИ

Выпускная квалификационная работа магистра

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 неизвестна. Известны лишь **периодические доходности** $y_t$ (относительные приращения стоимости портфеля), которые компания обязана публиковать

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

Известны цены и доходности активов $x_{t,i}$

Модель доходности портфеля $y_t \cong \sum_{i=1}^{n} \beta_i x_{t,i}$

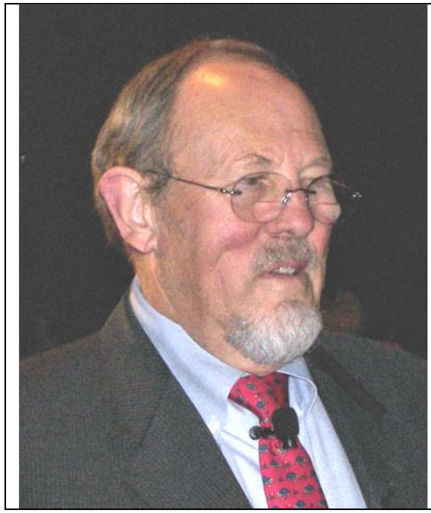
Для паевых фондов (пенсионных)

$\sum_{i=1}^{n} \beta_i = 1$ – бюджетное ограничение

$\beta_i \geq 0$ – нет заёмного капитала

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

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



**William F. Sharpe**, Professor of Finance
Лауреат Нобелевской премии по экономике 1990 года

Предложил аппроксимировать доходность портфеля небольшим числом биржевых индексов, представляющих разные стили инвестирования, решая задачу оценивания линейной регрессии при линейных ограничениях.

**Задача квадратичного программирования:**

$$\begin{cases} (\hat{\beta}_1,\dots,\hat{\beta}_n): \sum_{t=1}^{N}\left(y_t - \sum_{i=1}^{n}\beta_i x_{t,i}\right)^2 \to \min, \\ \sum_{i=1}^{n}\beta_i = 1,\ \beta_i \ge 0. \end{cases}$$

# Задача оценивания скрытого состава инвестиционного портфеля – Factor Search

| | |
|---|---|
| Модель Шарпа | $\sum_{t=1}^{T}\left(y_t-\sum_{i=1}^{n}\beta_i x_{t,i}\right)^2\rightarrow\min,\quad \underbrace{\sum_{i=1}^{n}\beta_i=1}_{\text{долевое распределение капитала}},\ \underbrace{\beta_i\geq 0}_{\text{запрет заемного капитала}},\ i=1,\ldots,n.$ |
| $y_t,\ t=1,\ldots,T$ | известные доходности портфеля (дни, месяцы, кварталы, годы) |
| $x_{t,i},\ i\in\mathbb{I}=\{1,\ldots,n\}$ | 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 … |
| | $\sum_{t=1}^{T}\left(y_t-\sum_{i\in\hat{\mathbb{I}}}\beta_i x_{t,i}\right)^2\rightarrow\min,\quad \sum_{i\in\hat{\mathbb{I}}}\beta_i=1,\ \beta_i\geq 0,\ i\in\hat{\mathbb{I}}$ |
| | … 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 $\left(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j \;\; i,j=1,\ldots,n\right)$ – 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<br>Средние доходности активов $\bar{x}_i$<br>и итоговая доходность портфеля $\bar{y}$ |

Вопрос:

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

# Теория Марковица о выборе инвестиционного портфеля

| | | | |
|---|---|---|---|
| Минимум риска при фиксированной доходности | $\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^T x_j)\beta_i\beta_j \to \min \\ \sum_{i=1}^{n}\bar{x}_i\beta_i = const \end{cases}$ | Минимум доходности при фиксированном риске | $\begin{cases} \sum_{i=1}^{n}\bar{\boldsymbol{x}}_i\beta_i \to \max \\ \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j = const \end{cases}$ |

Это 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

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

# .Селективность критерия восстановления структуры портфеля

Задача квадратичного программирования

Полиномиальная вычислительная сложность по числу регрессоров

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

| | |
|---|---|
| Параметр регуляризации $\mu \ge 0$ | Финансовый смысл – скрытая психология инвестора:<br>$\mu = 0$ – риск неприемлем (absolute Risk Aversion)<br>$\mu \to \infty$– риск поощряется (absolute Risk Tolerance) |

Критерий очень селективен – оставляет мало биржевых активов

Экспериментальный пример:

| | | | | | |
|---|---|---|---|---|---|
| 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 биржевых индексов | $n = 650$ | | | | |

# Специальный алгоритм решения задачи квадратичного программирования.

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

Преобразуем критерий к агрегированной форме:

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

Разделение 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 расщепление переменных (удвоение их числа):

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

Поочередная оптимизация.    Линейная вычисл. сложность по числу регрессоров

# Эксперименты

| | |
|---|---|
| Массив данных: | $n = 650$ реальных биржевых индексов месячной доходности<br>$T = 240$ месяцев (20 лет) |

Модельные портфели, оптимальные по принципу баланса доходность-риск с разными коэффициентами Risk Aversion

| Номер модельного портфеля | | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
|---|---|---|---|---|---|---|
| Толерантность к риску | | $\mu_1 = 0$ | $\mu_2 = 10$ | $\mu_3 = 30$ | $\mu_4 = 200$ | $\mu_5 = 1000$ |
| Количество активных индексов | $n_\mu = \lvert \mathbb{I}_\mu \rvert$<br>$\mathbb{I}_\mu = \left\{ i : \beta_i^\mu > 0 \right\}$ | $m_1 = 13$ | $m_2 = 11$ | $m_3 = 10$ | $m_4 = 7$ | $m_5 = 1$ |
| Всего биржевых индексов | | $n = 650$ | | | | |

Каждый портфель представлен временным рядом месячной доходности

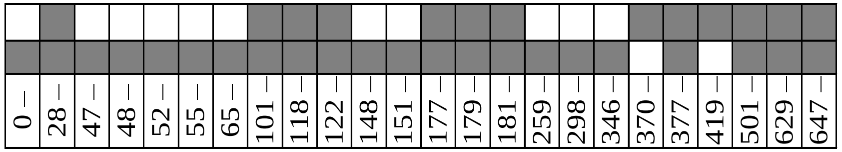
$$y_t^k = \sum_{i=1}^{n} \beta_i^k x_{t,i} + \text{шум}\ 10\%, \qquad t = 1, \ldots, T$$

Задача: восстановить набор активных индексов $\mathbb{I}_\mu$

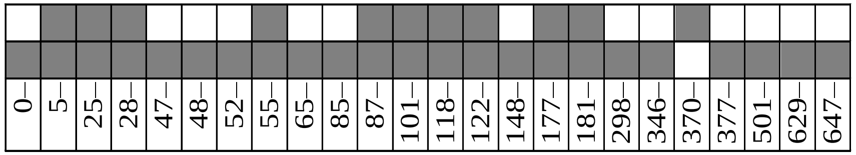
(и фактическое долевое распределение капитала между ними)

# Слепой фактор-поиск

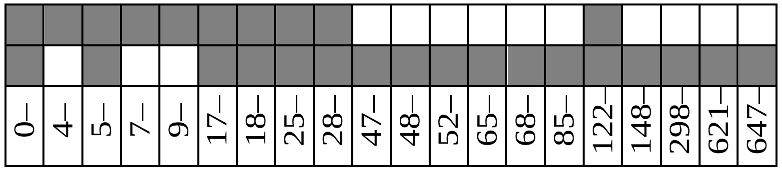
$$\begin{cases} \sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{n}\beta_i x_{t,i}\right)^2 \to \min \\ \sum_{i=1}^{n}\beta_i = 1,\ \ \beta_i \ge 0,\ i = 1,\ldots,n \end{cases}$$



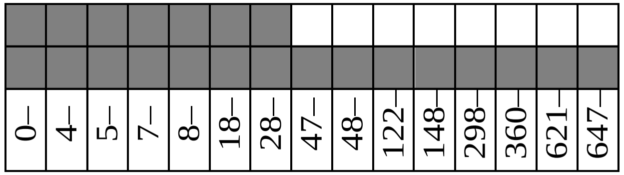

$\mu = 0$ total 24, wrong 13, 54% errors



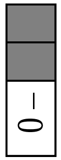

$\mu = 10$ total 24, wrong 14, 58% errors




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



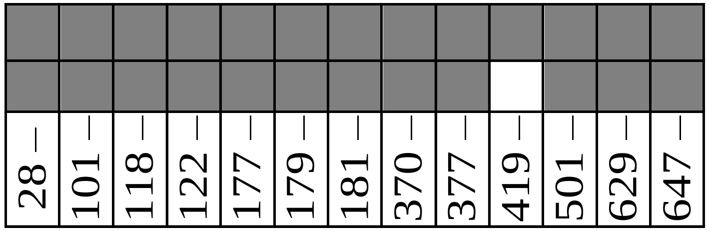

$\mu = 200$ total 15, wrong 8, 53% errors



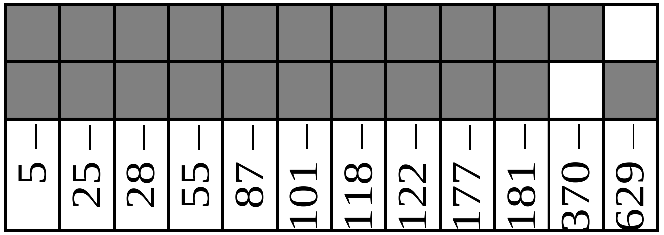

$\mu = 1000$ wrong 0, 0% errors

# Фактор-поиск с априорными предположениями:

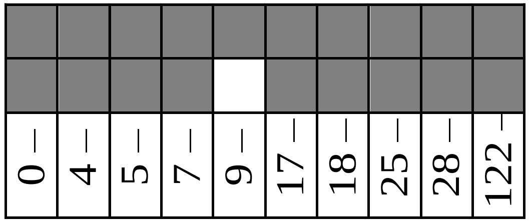
$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)\beta_i\beta_j - \mu\sum_{i=1}^{n}\bar{\boldsymbol{x}}_i\beta_i + c\sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{n}\beta_i x_{t,i}\right)^2 \to \min \\ \sum_{i=1}^{n}\beta_i = 1,\ \beta_i \geq 0 \quad, i = 1,\ldots,n \end{cases}$$



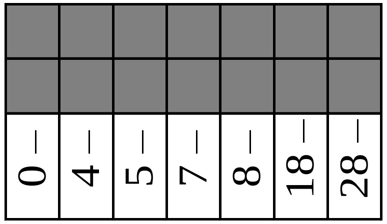

$\mu = 0$ total 13, wrong 1, 8% errors



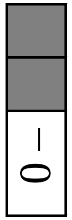

$\mu = 10$ total 12, wrong 2, 16% errors




$\mu = 30$ total 10 wrong 1, 10% errors




$\mu = 200$ total 7 wrong 0, 0% errors




$\mu = 1000$ wrong 0, 0% errors

# 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).