Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра математических методов гадания и хиромантии

Квасов Андрей Федорович

Использование критерия обоснованности модели для подбора структурных параметров в задаче восстановления скрытой стратегии управления инвестиционным портфелем

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Научный руководитель:
к.ф.-м.н., доцент
О.В. Красоткина

Содержание

Аннотация

В данной выпускной квалификационной работе проводилось сравнение работы методов по подбору структурных параметров модели в задаче восстановления скрытой стратегии управления инвестиционным портфелем. Эта задача имеет различные подходы, которые частично описаны в данной работе. Помимо уже предложенных ранее методов в работе рассмотрена возможность применения критерия обоснованности к оценке инвестиционных портфелей инвестиционных фондов.

1 Введение

Анализ структуры инвестиционного портфеля - одна из актуальных задач управления финансовыми активами. Задачи получения скрытых данных о распределении долей активов составляющих инвестиционный портфель сводятся к построению моделей зависящих от структурных параметров. Одна из таких моделей — модель нестационарной регрессии со структурным параметрам волатильности искомой нестационарной модели.

В выпускной квалификационной работе исследовались методы оценки таких моделей на предмет наилучшей интерпретации наблюдаемых данных, то есть исследовались методы подбора структурных параметров моделей, в частности кросс-валидационные методы, критерий Акаике и критерий обоснованности. В ходе работы была разработана вероятностная постановка задачи оценивания нестационарной регрессии, получен критерий обоснованности предложенной вероятностной модели, разработана процедура его максимизации по структурным параметрам. Рассмотренная процедура подбора параметров сравнивалась с существующими подходами с помощью экспериментального исследования на реальных данных инвестиционных компаний.

Работа состоит из следующих разделов. В разделе 2 рассматривается постановка задачи восстановления скрытой стратегии управления инвестиционным портфелем, показано, что она является частным случаем модели нестационарной регрессии с использованием структурного параметра волатильности , а также рассмотрен метод поиска коэффициентов регрессии в задаче нестационарной регрессии и сформулирована задача подбора структурных параметров и показана ее актуальность.

В разделе 3 приводится описание существующих методов оценивания структурных параметров модели нестационарной регрессии, дается вывод критерия обоснованности, предлагается алгоритм его оптимизации.

В пункте 4 описаны реальные данные, на которых проводились эксперименты, предложен внешний критерий оценки качества методов подбора структурного параметра λ, проводится экспериментальное сравнение предложенного подхода и уже существующих методов. Приведен анализ полученных результатов.

2 Анализ предметной области и постановка задачи

2.1 Задача восстановления скрытой стратегии управления инвестиционным портфелем

Инвестиционный портфель инвестиционной компании — это совокупность финансовых инструментов, в которые вложены средства вкладчиков, как например, акции, облигации или другие ценные бумаги. При этом инвестиционная компания, в частности, инвестиционный фонд, выступает в плане посредника: в отношении приобретенных финансовых активов и получаемой от них прибыли инвесторы получают определенные права. В свою очередь инвесторы могут выступать в разных ролях, в том числе как акционеры, то есть они являются владельцами компании и соответственно владеют активами, принадлежащими этой компании. Существует как минимум два преимущества вложения средств инвестором в такие компании: экономия на масштабе и профессиональное управление активами.

Деятельность инвестиционной компании состоит в увеличении стоимости инвестиционного портфеля, при этом она, вообще говоря, не обязана выкладывать информацию о составе портфеля в открытый доступ или выдавать ее по запросу акционерам компании, так как состав портфеля считается профессиональной тайной. Поэтому многие конкуренты, контролирующие государственные организации и другие аналитики желают обладать информацией о составе инвестиционного портфеля инвестиционной компании.

Модель инвестиционного портфеля была предложена Г. Марковицем [1]. В данной модели элементарной единицей данных является период владения между двумя моментами времени t и $t+1$, под которыми можно понимать дни, недели или месяцы. При этом для каждого периода владения фонд обязан публиковать свою доходность $y_t = \frac{W_{t+1}-W_t}{W_t}$, где W_t, W_{t+1} - рыночная стоимость ценных бумаг портфеля в моменты времени t и $t+1$.

Также известны котировки акций на фондовом рынке выраженные в показателях своих доходностей: x_t^i — доходность i-ого актива в момент времени t, определенная аналогичным образом с доходностью портфеля.

Пусть количество акций каждого типа в портфеле m_t^i и цена одной акции z_t^i в момент времени t. Тогда стоимость портфеля выражается, как $W_t = \sum_{i=1}^n m_t^i z_t^i$, а его доходность $y_t = \frac{\sum_{i=1}^n m_{t+1}^i z_{t+1}^i - \sum_{i=1}^n m_t^i z_t^i}{\sum_{i=1}^n m_t^i z_t^i}$. Можно показать, что, если во время периода владения в портфель не поступали средства и средства не извлекались, то доходность портфеля выражается через доходности активов $x_t^i = \frac{z_{t+1}^i - z_t^i}{z_t^i}$ как линейная комбинация доходностей активов с весами, представляющими собой доли актива в портфеле $y_t = \sum_{i=1}^n \frac{m_t^i z_t^i}{\sum_{i=1}^n m_t^i z_t^i}(z_{t+1}^i - z_t^i)/z_t^i + e_t == \sum_{i=1}^n \beta_t^i x_t^i$. Так как требование изолированности портфеля от внешней среды практически никогда не выполняется, то $y_t = \sum_{i=1}^n \beta_t^i x_t^i + e_t$, где e_t нормальный белый шум с нулевым средним и некоторой дисперсией. При этом на возможные значения долей активов налагается ограничения типа равенства $\sum_{i=1}^n \beta_t^i = 1, \ t = 1, \dots, N$. Помимо этого ограничения возможен ввод ограничений типа неравенств, при котором мы ограничиваем область значений активов только положительными величинами $\beta_t^i > 0$. В данной задаче рассматривается частный случай типов инвестиционных фондов – хедж-фонды, при котором данный вид ограничений не накладывается.

Согласно данной постановке ставится задача поиска долей активов β_t^i в инвестиционном портфеле на период времени $t = 1, \dots, N$ с $i = 1, \dots, n$ возможными классами активов, при известной информации о доходности портфеля y_t за этот период и, находящихся в открытом доступе, индексах x_t^i каждого из рассматриваемых активов на данный период времени.

2.2 Иерархическая вероятностная модель для восстановления динамической структуры инвестиционного портфеля

Рассматривается задача оценивания нестационарной регрессии с параметрами $\beta_1, \dots, \beta_N$, наблюдаемыми величинами $(\mathbf{x}_t, y_t)_{t=1}^N$ и гиперпараметрами α, δ.

Рассматривается *модель наблюдения*, в которой целевая переменная y_t получена с помощью линейной функции от наблюдаемых векторов $\mathbf{x}_t$ и соответствующей ей последовательности векторных коэффициентов регрессии $\boldsymbol{\beta}_t$ с прибавлением нормальных шумовых компонент e_t:

$$\begin{aligned} & y_t = \mathbf{x}_t^\intercal \boldsymbol{\beta}_t + e_t, \ e_t \sim \mathcal{N}(0, \delta^{-1}), \mathbb{D}[e_s e_l] = 0, s \neq l \\ & p(y_t | \mathbf{x}_t, \delta) = \mathcal{N}(\mathbf{x}_t^\intercal \boldsymbol{\beta}_t, \delta^{-1}), \ t = 1, \dots, N \end{aligned} \quad (1)$$

Структурный параметр $\delta > 0$ является мерой точности нормальной случайной величины y_t.

Считая наблюдения независимыми случайными величинами их совместная плотность распределения $p(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}, \delta)$ выражается через многомерную нормальную функцию:

$$p(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \delta) = \prod_{t=1}^{N} \mathcal{N}(y_t | \mathbf{x}_t^\intercal \boldsymbol{\beta}_t, \delta^{-1}) \quad (2)$$

, где $\mathbf{X} = \{\mathbf{x}_t\}_{t=1}^N$, $\mathbf{y} = \{y_t\}_{t=1}^N$ совокупности наблюдений, а $\boldsymbol{\beta} = \{\boldsymbol{\beta}_t\}_{t=1}^N$, - совокупность параметров модели.

В такой постановке модели наблюдения с разными векторами коэффициентов в разные моменты времени, задача оказывается некорректно поставленной. Число оцениваемых параметров, то есть размерность вектора $\boldsymbol{\beta}$ в ней в n раз превосходит число наблюдений, поэтому необходима дополнительная регуляризация, в качестве которой выступает априорная модель случайной последовательности векторных коэффициентов регрессии - *модель динамики.*

Модель динамики согласовывает получаемые из модели наблюдения коэффициенты регрессии $\boldsymbol{\beta}$, определяя нормальную зависимость между векторами $\boldsymbol{\beta}_t$ в соседние моменты времени, через матрицу динамики изменения долей активов V_t и матрицу сглаживания $(\alpha \mathrm{U}_t)^{-1}$ матрицу ковариации компонент белого шума ξ_t:

$$\begin{aligned} &\boldsymbol{\beta}_t = \mathrm{V}_t \boldsymbol{\beta}_{t-1} + \xi_t,\ \xi_t \sim \mathcal{N}(0, (\alpha \mathrm{U}_t)^{-1}),\ \mathbb{D}[\xi_s \xi_l] = 0, s \neq l \\ &p(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \alpha) = \mathcal{N}(\mathrm{V}_t \boldsymbol{\beta}_{t-1}, (\alpha \mathrm{U}_t)^{-1}),\ t = 2, \ldots, N \end{aligned} \tag{3}$$

Матрица сглаживания U_t диагональная и имеет на своей диагонали фиксированные значения u_t — меры точности каждой компоненты. В данной работе $u_t = 1,\ \forall t = 1, \ldots, N$, то есть матрица сглаживания равна единичной $\mathrm{U}_t = \mathrm{I}$. Тогда информация о степени сглаживания полностью выражается через отношение $\lambda = \alpha/\delta$. Диагональная матрица V_t - матрица динамики изменения долей активов, также полагается равной единичной матрице.

При условной зависимости параметров $\boldsymbol{\beta}_t$ только от вектора коэффициентов регрессии $\boldsymbol{\beta}_{t-1}$ предыдущего наблюдения, получаем марковский случайный процесс и соответствующую ему совместную априорную плотность распределения коэффициентов:

$$p(\boldsymbol{\beta}|\alpha) = \prod_{t=2}^{N} p(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \alpha) = \prod_{t=2}^{N} \mathcal{N}\big(\boldsymbol{\beta}_t | \mathrm{V}_t \boldsymbol{\beta}_{t-1}, (\alpha \mathrm{U}_t)^{-1}\big) \tag{4}$$

Применительно к задаче оценивания структуры инвестиционного портфеля, используются ограничения типа равенства на сумму коэффициентов регрессии $\sum_{i=1}^{n} \beta_t^i = 1,\ t = 1, \ldots, N.$

Используем формулу Байеса для определения апостериорного распределения на параметры модели $\boldsymbol{\beta}$:

$$p(\boldsymbol{\beta}|\mathbf{X},\mathbf{y},\alpha,\delta) = \frac{p(\mathbf{y}|\mathbf{X},\boldsymbol{\beta},\delta)p(\boldsymbol{\beta}|\alpha)}{p(\mathbf{y}|\mathbf{X},\alpha,\delta)} \propto p(\mathbf{y}|\mathbf{X},\boldsymbol{\beta},\delta)p(\boldsymbol{\beta}|\alpha)$$

В дальнейшем, считая нашу вероятностную модель дискриминативной - то есть на данные наблюдений $\mathbf{X}$ не накладывается вероятностное распределение, для удобства будем опускать совокупность наблюдений $\mathbf{X}$ в условных распределениях, к примеру $p(\mathbf{y}|\mathbf{X},\boldsymbol{\beta},\delta) = p(\mathbf{y}|\boldsymbol{\beta},\delta)$.

С помощью рассмотренной выше функции правдоподобия выборки $p(\mathbf{y}|\mathbf{X},\boldsymbol{\beta},\delta)$ и априорного распределения на параметры $p(\boldsymbol{\beta}|\alpha)$ применим *метод максимального правдоподобия* (*Maximum Likelihood*) к апостериорному распределению на параметры модели $\boldsymbol{\beta}$ при условии наблюдаемых данных $(\mathbf{X},\mathbf{y})$:

$$\begin{aligned}\boldsymbol{\beta}_{ML} &= \underset{\boldsymbol{\beta}:\sum_{i=1}^n \beta_t^i=1}{\operatorname{argmax}}\; p(\boldsymbol{\beta}|\mathbf{y},\alpha,\delta) = \underset{\boldsymbol{\beta}:\sum_{i=1}^n \beta_t^i=1}{\operatorname{argmax}}\; p(\mathbf{y}|\boldsymbol{\beta},\delta)p(\boldsymbol{\beta}|\alpha) = \\ &= \underset{\boldsymbol{\beta}:\sum_{i=1}^n \beta_t^i=1}{\operatorname{argmax}}\; \mathcal{N}(y_1|\mathbf{x}_1^\intercal\boldsymbol{\beta}_1,\delta^{-1})\prod_{t=2}^{N}\mathcal{N}(y_t|\mathbf{x}_t^\intercal\boldsymbol{\beta}_t,\delta^{-1})\mathcal{N}\big(\boldsymbol{\beta}_t|\mathrm{V}_t\boldsymbol{\beta}_{t-1},(\alpha\mathrm{U}_t)^{-1}\big)\end{aligned} \tag{5}$$

Проводя логарифмирование апостериорной вероятности, получаем критерий для нахождения векторов коэффициентов в задаче нестационарной регрессии применительно к анализу состава инвестиционного портфеля:

$$\begin{cases} -\frac{\delta}{2}\sum_{t=1}^N(y_t-\boldsymbol{\beta}_t^\intercal\mathbf{x}_t)^2 - \frac{\alpha}{2}\sum_{t=2}^N(\boldsymbol{\beta}_t-\mathrm{V}_t\boldsymbol{\beta}_{t-1})^\intercal\mathrm{U}_t(\beta_t-\mathrm{V}_t\beta_{t-1}) \to \max\limits_{\boldsymbol{\beta}} \\ \sum_{i=1}^n \beta_t^i = 1,\ t=1,\dots,N \end{cases} \tag{6}$$

Использовав введенный раннее $\lambda = \frac{\alpha}{\delta}$ - структурный параметр гладкости, перепишем задачу поиска параметров модели $\boldsymbol{\beta}$, как задачу квадратичной минимизации с ограничениями типа равенства:

$$\begin{cases} \sum_{t=1}^N(y_t-\boldsymbol{\beta}_t^\intercal\mathbf{x}_t)^2 + \lambda\sum_{t=2}^N(\boldsymbol{\beta}_t-\mathrm{V}_t\boldsymbol{\beta}_{t-1})^\intercal\mathrm{U}_t(\beta_t-\mathrm{V}_t\beta_{t-1}) \to \max\limits_{\boldsymbol{\beta}} \\ \sum_{i=1}^n \beta_t^i = 1,\ t=1,\dots,N \end{cases} \tag{7}$$

Впрочем ограничения типа равенств могут быть безболезненно исключены включении в рассмотрение доходности безрискового актива x_t^0. Распишем модель наблюдения:

$$\mathbf{y} = \boldsymbol{\beta}_t^{\intercal}\mathbf{x}_t + e_t = \sum_{i=0}^{n}\beta_t^i x_t^i + e_t = \sum_{i=1}^{n}\beta_t^i x_t^i + e_t + \beta_t^0 x_t^0 = \{\beta_t^0 = 1 - \sum_{i=1}^{n}\beta_t^i\} =$$
$$= \sum_{i=1}^{n}\beta_t^i x_t^i + e_t + x_t^0 - \sum_{i=1}^{n}\beta_t^i x_t^0 \quad \Rightarrow \quad y_t - x_t^0 = \sum_{i=1}^{n}\beta_t^i(x_t^i - x_t^0) + e_t \tag{8}$$

Поэтому для исключения ограничений достаточно перейти к новым наблюдаемым значениям доходностей $y_t := y_t - x_t^0$ и $x_t^i := x_t^i - x_t^0$. Это приводит к окончательному виду критерия:

$$\sum_{t=1}^{N}(y_t - \boldsymbol{\beta}_t^{\intercal}\mathbf{x}_t)^2 + \lambda\sum_{t=2}^{N}(\boldsymbol{\beta}_t - \mathrm{V}_t\boldsymbol{\beta}_{t-1})^{\intercal}\mathrm{U}_t(\beta_t - \mathrm{V}_t\beta_{t-1}) \to \max_{\boldsymbol{\beta}} \tag{9}$$

Решение задачи поиска параметров модели $\boldsymbol{\beta}$ известна и изложена в секции 2.3.

2.3 Алгоритм динамического программирования для восстановления динамической структуры инвестиционного портфеля

Минимизируемая функция в критерии 9 является парно-сепарабельной и раскладывается на узловые функции $\psi_s(\boldsymbol{\beta}_s) = (y_s - \boldsymbol{\beta}_s^{\intercal}\mathbf{x}_s)^2,\ s = 1, \ldots, N$ и функции связи $\gamma_s(\boldsymbol{\beta}_{s-1}, \boldsymbol{\beta}_s) = \lambda(\boldsymbol{\beta}_s - \mathrm{V}_s\boldsymbol{\beta}_{s-1})^{\intercal}\mathrm{U}_s(\beta_s - \mathrm{V}_s\beta_{s-1}),\ s = 2, \ldots, N$. Таким образом целевая функция общего критерия $J(\boldsymbol{\beta}) = J(\boldsymbol{\beta}_1, \ldots \boldsymbol{\beta}_N)$ принимает вид:

$$J(\boldsymbol{\beta}_1, \ldots \boldsymbol{\beta}_N) = \sum_{t=1}^{N}\psi_t(\boldsymbol{\beta}_t) + \sum_{t=2}^{N}\gamma_t(\boldsymbol{\beta}_{t-1}, \boldsymbol{\beta}_t) \to \min_{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N} \tag{10}$$

Рассмотрим метод динамического программирования применяемый для парно-сепарабельных функций [2] и состоящий в подсчете последовательности левых и правых функций Беллмана [4] $\widetilde{J}_t^-(\boldsymbol{\beta}_t)$ и $\widetilde{J}_t^+(\boldsymbol{\beta}_t)$. Они связаны соответственно с левыми и правыми частичными критериями

Беллмана $J_t^-(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_t)$ и $J_t^+(\boldsymbol{\beta}_t, \ldots, \boldsymbol{\beta}_N)$, как функция минимума от части изначальной суммы функций связи и узловых функций вычисленных таким образом:

$$\widetilde{J}_t^-(\boldsymbol{\beta}_t) = \min_{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{t-1}} J_t^-(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_t);$$
$$J_t^-(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_t) = \sum_{s=1}^{t} \psi_s(\boldsymbol{\beta}_s) + \sum_{s=2}^{t} \gamma_s(\boldsymbol{\beta}_{s-1}, \boldsymbol{\beta}_s) \tag{11}$$
$$\widetilde{J}_t^+(\boldsymbol{\beta}_t) = \min_{\boldsymbol{\beta}_{t+1}, \ldots, \boldsymbol{\beta}_N} J_t^+(\boldsymbol{\beta}_t, \ldots, \boldsymbol{\beta}_N)$$
$$J_t^+(\boldsymbol{\beta}_t, \ldots, \boldsymbol{\beta}_N) = \sum_{s=t}^{N} \psi_s(\boldsymbol{\beta}_s) + \sum_{s=t}^{N-1} \gamma_s(\boldsymbol{\beta}_s, \boldsymbol{\beta}_{s+1}) \tag{12}$$

Важным свойством как левых, так и правых функций Беллмана является возможность их вычисления с помощью рекуррентных соотношений:

$$\widetilde{J}_t^-(\boldsymbol{\beta}_t) = \psi_t(\boldsymbol{\beta}_t) + \min_{\boldsymbol{\beta}_{t-1}} \big[\gamma_t(\boldsymbol{\beta}_{t-1}, \boldsymbol{\beta}_t) + \widetilde{J}_{t-1}^-(\boldsymbol{\beta}_{t-1})\big],\ t = 2, \ldots, N$$
$$\widetilde{J}_1^-(\boldsymbol{\beta}_1) = \psi_1(\boldsymbol{\beta}_1) \tag{13}$$
$$\widetilde{J}_t^+(\boldsymbol{\beta}_t) = \psi_t(\boldsymbol{\beta}_t) + \min_{\boldsymbol{\beta}_{t+1}} \big[\gamma_{t+1}(\boldsymbol{\beta}_t, \boldsymbol{\beta}_{t+1}) + \widetilde{J}_{t+1}^+(\boldsymbol{\beta}_{t+1})\big],\ t = N-1, \ldots, 1$$
$$\widetilde{J}_N^+(\boldsymbol{\beta}_N) = \psi_N(\boldsymbol{\beta}_N) \tag{14}$$

Каждая из левых функций Беллмана $\widetilde{J}_t^-(\boldsymbol{\beta}_t)$ вычисляется по рекуррентной формуле в прямом направлении при $t = 1, \ldots, N$, а каждая из правых функций Беллмана $\widetilde{J}_t^+(\boldsymbol{\beta}_t)$ - в обратном направлении при $t = N, \ldots, 1$, предлагая схему динамического программирования под названием "вперед и навстречу". Она необходима для определения так называемых маргинальных узловых функций $\hat{J}_t(\boldsymbol{\beta}_t)$, которые показывают зависимости целевой функции $J(\boldsymbol{\beta})$ от каждой переменной $\boldsymbol{\beta}_t$ в отдельности, когда остальные переменные принимают условно оптимальное значение в зависимости от $\boldsymbol{\beta}_t$. Эти функции также выражаются через функции Беллмана:

$$\hat{J}_t(\boldsymbol{\beta}_t) = \min_{\boldsymbol{\beta}_s,\ s\neq t, s=1,\ldots,N} J(\boldsymbol{\beta}_1, \ldots \boldsymbol{\beta}_N)$$

$$\hat{J}_t(\boldsymbol{\beta}_t) = \begin{cases} \psi_1(\boldsymbol{\beta}_1) + \min\limits_{\boldsymbol{\beta}_2}\big[\gamma_2(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) + \widetilde{J}_2^{+}(\boldsymbol{\beta}_2)\big],\ t = 1, \\ \min\limits_{\boldsymbol{\beta}_{t-1}}\big[\gamma_t(\boldsymbol{\beta}_{t-1}, \boldsymbol{\beta}_t) + \widetilde{J}_{t-1}^{-}(\boldsymbol{\beta}_{t-1})\big] + \psi_t(\boldsymbol{\beta}_t) + \min\limits_{\boldsymbol{\beta}_{t+1}}\big[\gamma_{t+1}(\boldsymbol{\beta}_t, \boldsymbol{\beta}_{t+1}) + \\ \qquad + \widetilde{J}_{t+1}^{+}(\boldsymbol{\beta}_{t+1})\big],\ t = 2, \ldots, N-1, \\ \min\limits_{\boldsymbol{\beta}_{N-1}}\big[\gamma_N(\boldsymbol{\beta}_{N-1}, \boldsymbol{\beta}_N) + \widetilde{J}_{N-1}^{-}(\boldsymbol{\beta}_{N-1})\big] + \psi_N(\boldsymbol{\beta}_N),\ t = N \end{cases} \tag{15}$$

В точке встречи t левой и правой функции Беллмана маргинальные узловые функции дают оптимальное значение в искомой переменной в этой точке $\hat{\boldsymbol{\beta}}_t = \operatorname{argmin} \hat{J}_t(\boldsymbol{\beta}_t)$.

Чтобы процедура динамического программирования "вперед и навстречу" была численно реализуема для парно-сепарабельной функции, необходимо, чтобы левые и правые функции Беллмана принадлежали параметрическому семейству, замкнутому относительно узловых функций и функций связи, тогда прямой проход и проход навстречу будут вычислять все задействованные функции через рекуррентный вывод параметров функций.

Целевая функция представляет из себя квадратичную парно-сепарабельную функцию, так как каждый член в сумме является квадратичной функцией от вектора параметров $\boldsymbol{\beta}_t$:

$$\gamma_t(\boldsymbol{\beta}_{t-1}, \boldsymbol{\beta}_t) = \lambda(\boldsymbol{\beta}_t - \mathrm{V}_t\boldsymbol{\beta}_{t-1})^{\intercal}\mathrm{U}_t(\beta_t - \mathrm{V}_t\beta_{t-1}) = (\boldsymbol{\beta}_t - \mathrm{V}_t\boldsymbol{\beta}_{t-1})^{\intercal}\widetilde{\mathrm{U}}_t(\beta_t - \mathrm{V}_t\beta_{t-1}),$$

$$\text{где } \widetilde{\mathrm{U}}_t = \lambda\mathrm{U}_t (\text{в дальнейшем обозначим за } \mathrm{U}_t) \tag{16}$$

$$\psi_t(\boldsymbol{\beta}_t) = (y_t - \boldsymbol{\beta}_t^{\intercal}\mathbf{x}_t)^2 = b_t^0 + (\boldsymbol{\beta}_t - \boldsymbol{\beta}_t^0)^{\intercal}\mathrm{Q}_t^0(\boldsymbol{\beta}_t - \boldsymbol{\beta}_t^0),$$

$$\text{где } b_t^0 = 0,\ \mathrm{Q}_t^0 = \mathbf{x}_t\mathbf{x}_t^{\intercal},\ \boldsymbol{\beta}_t^0 = (y_t/\mathbf{x}_t^{\intercal}\mathbf{x}_t)\mathbf{x}_t \tag{17}$$

В таком случае функции Беллмана примут вид квадратичных функций $\widetilde{J}_t^{-}(\boldsymbol{\beta}_t) = \widetilde{b}_t^{-} + (\boldsymbol{\beta}_t - \widetilde{\boldsymbol{\beta}}_t^{-})^{\intercal}\widetilde{\mathrm{Q}}_t^{-}(\boldsymbol{\beta}_t - \widetilde{\boldsymbol{\beta}}_t^{-})$, $\widetilde{J}_t^{+}(\boldsymbol{\beta}_t) = \widetilde{b}_t^{+} + (\boldsymbol{\beta}_t - \widetilde{\boldsymbol{\beta}}_t^{+})^{\intercal}\widetilde{\mathrm{Q}}_t^{+}(\boldsymbol{\beta}_t - \widetilde{\boldsymbol{\beta}}_t^{+})$, что дает возможность на каждом шаге рекуррентно пересчитывать параметры функций Беллмана $(\widetilde{b}_t^-, \widetilde{\boldsymbol{\beta}}_t^-, \widetilde{\mathrm{Q}}_t^-)$ $(\widetilde{b}_t^+, \widetilde{\boldsymbol{\beta}}_t^+, \widetilde{\mathrm{Q}}_t^+)$, имеющие смысл значения функции в точке минимума, точки минимума и соответствующей квадратичной формы:

$$\begin{cases} \widetilde{b}_1^- = 0,\ \widetilde{\boldsymbol{\beta}}_1^- = \boldsymbol{\beta}_1^0,\ \widetilde{\mathrm{Q}}_1^- = \mathrm{Q}_1^0,\ t = 1; \\ \widetilde{\mathrm{Q}}_t^- = \mathrm{Q}_t^0 + \mathrm{U}_t\mathrm{V}_t(\widetilde{\mathrm{Q}}_{t-1}^- + \mathrm{V}_t^\intercal\mathrm{U}_t\mathrm{V}_t)\widetilde{\mathrm{Q}}_{t-1}^-\mathrm{V}_t^{-1}, \\ \widetilde{\boldsymbol{\beta}}_t^- = (\widetilde{\mathrm{Q}}_t^-)^{-1}\big[\mathrm{Q}_t^0\boldsymbol{\beta}_t^0 + \mathrm{U}_t\mathrm{V}_t(\widetilde{\mathrm{Q}}_{t-1}^- + \mathrm{V}_t^\intercal\mathrm{U}_t\mathrm{V}_t)^{-1}\widetilde{\mathrm{Q}}_{t-1}^-\widetilde{\boldsymbol{\beta}}_{t-1}^-\big], \\ \widetilde{b}_t^- = b_t^0 + \widetilde{b}_{t-1}^- + (\boldsymbol{\beta}_t^0 - \widetilde{\boldsymbol{\beta}}_t^-)^\intercal\mathrm{Q}_t^0\boldsymbol{\beta}_t^0 + (\widetilde{\boldsymbol{\beta}}_{t-1}^- - \mathrm{V}_t^{-1}\widetilde{\boldsymbol{\beta}}_t^-)^\intercal\mathrm{U}_t\mathrm{V}_t(\widetilde{\mathrm{Q}}_{t-1}^- + \mathrm{V}_t^\intercal\mathrm{U}_t\mathrm{V}_t)^{-1}\widetilde{\mathrm{Q}}_{t-1}^-\widetilde{\boldsymbol{\beta}}_{t-1}^-, \\ \qquad t = 2, \ldots, N \end{cases} \tag{18}$$

$$\begin{cases} \widetilde{b}_N^+ = 0,\ \widetilde{\boldsymbol{\beta}}_N^+ = \boldsymbol{\beta}_N^0,\ \widetilde{\mathrm{Q}}_N^+ = \mathrm{Q}_N^0,\ t = N; \\ \widetilde{\mathrm{Q}}_t^+ = \mathrm{Q}_t^0 + \mathrm{V}_{t+1}^\intercal\mathrm{U}_{t+1}(\widetilde{\mathrm{Q}}_{t+1}^+ + \mathrm{U}_{t+1})^{-1}\widetilde{\mathrm{Q}}_{t+1}^+\mathrm{V}_{t+1}, \\ \widetilde{\boldsymbol{\beta}}_t^+ = (\widetilde{\mathrm{Q}}_t^+)^{-1}\big[\mathrm{Q}_t^0\boldsymbol{\beta}_t^0 + \mathrm{V}_{t+1}^\intercal\mathrm{U}_{t+1}(\widetilde{\mathrm{Q}}_{t+1}^+ + \mathrm{U}_{t+1})^{-1}\widetilde{\mathrm{Q}}_{t+1}^+\widetilde{\boldsymbol{\beta}}_{t+1}^+\big], \\ \widetilde{b}_t^+ = b_t^0 + \widetilde{b}_{t+1}^+ + (\boldsymbol{\beta}_t^0 - \widetilde{\boldsymbol{\beta}}_t^+)^\intercal\mathrm{Q}_t^0\boldsymbol{\beta}_t^0 + (\widetilde{\boldsymbol{\beta}}_{t+1}^+ - \mathrm{V}_{t+1}\widetilde{\boldsymbol{\beta}}_t^+)^\intercal\mathrm{U}_{t+1}(\widetilde{\mathrm{Q}}_{t+1}^+ + \mathrm{U}_{t+1})^{-1}\widetilde{\mathrm{Q}}_{t+1}^+\widetilde{\boldsymbol{\beta}}_{t+1}^+, \\ \qquad t = N-1, \ldots, 1; \end{cases} \tag{19}$$

Маргинальные узловые функции будут также квадратичными, таким образом при пересчете параметров можно найти точку минимума $\hat{\boldsymbol{\beta}}_t$:

$$\begin{aligned} \hat{\mathrm{Q}}_t &= \mathrm{V}_{t+1}^\intercal\mathrm{U}_{t+1}(\widetilde{\mathrm{Q}}_{t+1}^+ + \mathrm{U}_{t+1})^{-1}\widetilde{\mathrm{Q}}_{t+1}^+\mathrm{V}_{t+1} + \mathrm{Q}_t^0 + \mathrm{U}_t\mathrm{V}_t(\widetilde{\mathrm{Q}}_{t-1}^- + \mathrm{V}_t^\intercal\mathrm{U}_t\mathrm{V}_t)\widetilde{\mathrm{Q}}_{t-1}^-\mathrm{V}_t^{-1}, \\ \hat{\boldsymbol{\beta}}_t &= (\hat{\mathrm{Q}}_t)^{-1}\big[\mathrm{V}_{t+1}^\intercal\mathrm{U}_{t+1}(\widetilde{\mathrm{Q}}_{t+1}^+ + \mathrm{U}_{t+1})^{-1}\widetilde{\mathrm{Q}}_{t+1}^+\widetilde{\boldsymbol{\beta}}_{t+1}^+ + \mathrm{Q}_t^0\boldsymbol{\beta}_t^0 + \mathrm{U}_t\mathrm{V}_t(\widetilde{\mathrm{Q}}_{t-1}^- + \mathrm{V}_t^\intercal\mathrm{U}_t\mathrm{V}_t)\widetilde{\mathrm{Q}}_{t-1}^-\widetilde{\boldsymbol{\beta}}_{t-1}^-\big] \end{aligned} \tag{20}$$

Процедура метода "вперед и навстречу"вычисляет искомые оптимальные параметры $\hat{\boldsymbol{\beta}}_t$ для каждого $t = 1, \ldots, N$ за линейное время от длины вектора неизвестных.

2.4 Необходимость оценивания структурных параметров в модели нестационарной регрессии

В исходном критерии (9) параметр λ имеет роль коэффициента регуляризатора, который соотносит результаты минимизации критерия в модели наблюдений и критерия в модели динамики. Модель наблюдения выдает размытые локальные решения векторов коэффициентов нестационарной регрессии $\boldsymbol{\beta}$, в то время как модель динамики согласовывает их путем сглаживания. В общем случае, рассматривается как диагональная матрица сглаживания U_t, характеризующую различия в сглаживании каждой компоненты, и общий параметр сглаживания λ, но в данной работе был рассмотрен случай, когда компоненты имеют одинаковую степень сглаживания ($\mathrm{U}_t = \mathrm{I}$).

В зависимости от значения λ, решения полученные после процедуры динамического программирования 2.3 начинают больше или меньше зависеть друг от друга в зависимости моментов времени t_1 и t_2. При принимающем большие значения $\lambda \to \infty$ мы получаем более сглаженные во времени коэффициенты, при чем при достаточно больших λ коэффициенты начинают вырождаться $\boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_N$. Такое поведение критерия не желательно, так как приводит к несуществующим моделям инвестиционных фондов. Наоборот, при $\lambda \to 0$ доступна большая вариативность в выборе коэффициентов регрессии и при достаточно малых λ бесконечное множество решений, что связано с большим количеством параметров Nn по сравнению с размеров выборки N. Данная ситуация также нежелательно, так как отвергает модель динамики. Поэтому необходимы методы для оценивания значений структурного параметра λ, чтобы найти оптимальный вариант модели.

В данной работе ставилась проводилась оценка методов подбора структурных параметров в задаче восстановления скрытой стратегии управления инвестиционным портфелем, в частности использование

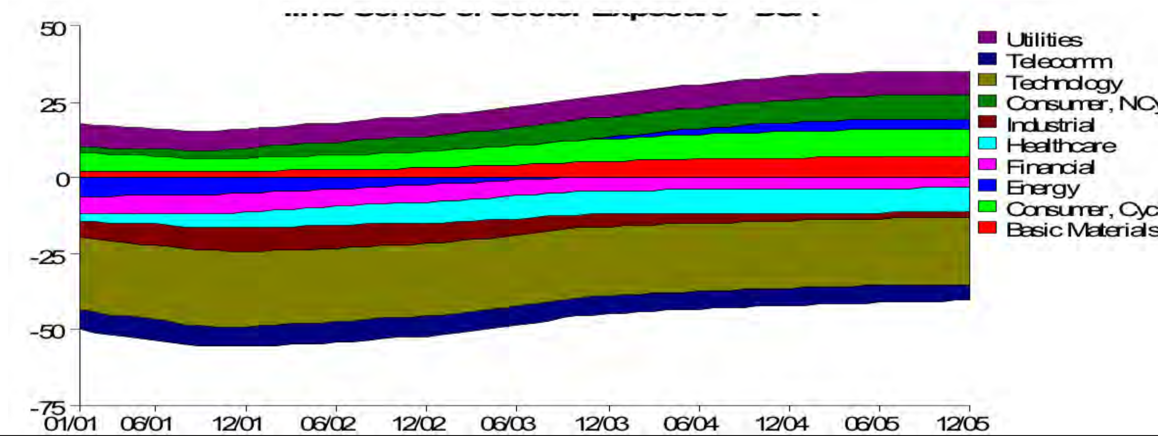
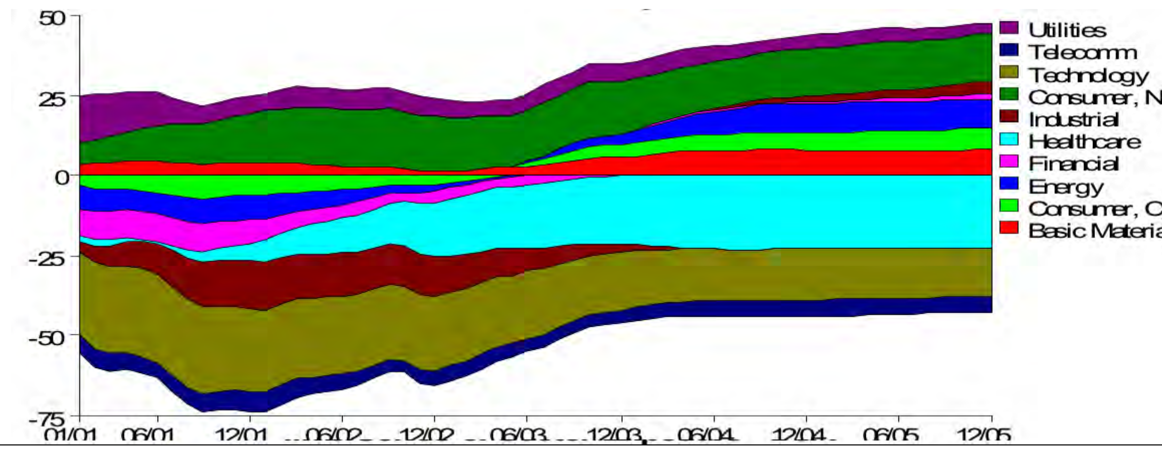
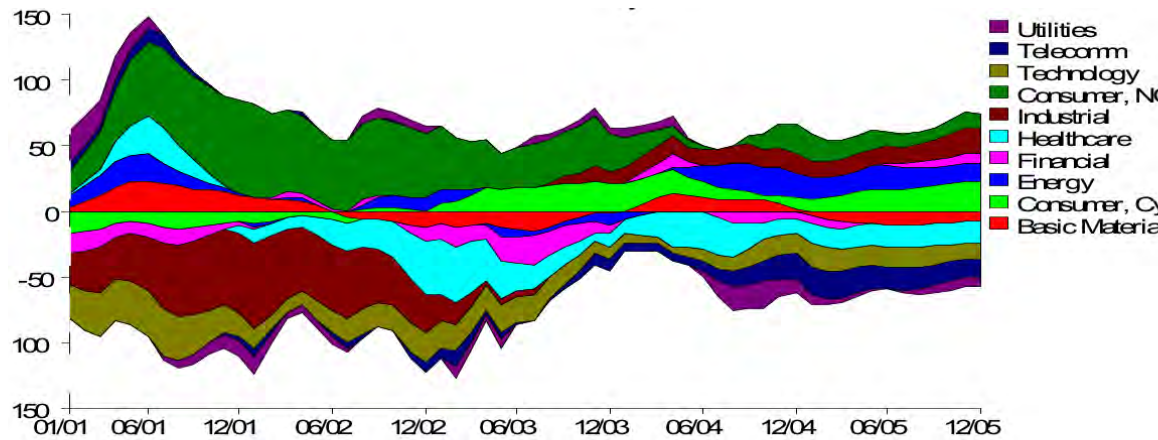
λ'	
$\lambda'' < \lambda'$	
$\lambda''' < \lambda''$	

Таблица 1: Поведение векторов коэффициентов нестационарной регрессии $\boldsymbol{\beta}$ при различных значениях λ

критерия максимального правдоподобия модели. В сравнении с данным критерием были предложены другие методы оценки структурных параметров модели, среди них кросс-валидационные методы и критерий Акаике.

3 Методы оценивания структурных параметров в задаче оценивания нестационарной регрессии

3.1 Кросс-валидационные методы

Для оценки модели нестационарной регрессии, полученной при степени сглаживания λ необходимо сравнить уровень шума наблюдения e_t и изменчивость последовательности векторов коэффициентов регрессии $\boldsymbol{\beta}_t$. Оценкой для уровня шума $e_t = y_t - \mathbf{x}_t^\intercal \boldsymbol{\beta}_t$ служит оценка ее дисперсии $D(\lambda) = \frac{1}{N} \sum_{t=1}^{N} (y_t - \mathbf{x}_t^\intercal \boldsymbol{\beta}_t^{(t)})^2$, причем участвующие здесь вектора коэффициентов регрессии $\boldsymbol{\beta}_t^{(t)}$ для наблюдений $(\mathbf{x}_t, y_t)$ получены путем минимизации критерия сформированного выкидыванием поочередно при $t = 1, \ldots, N$ наблюдений $(\mathbf{x}_t, y_t)$. В таком случае, необходимо с помощью метода динамического программирования провести минимизацию критерия, отличного от исходного критерия, с той лишь разницей, что матрица $\mathrm{Q}_t^0 = 0$, а остальные значения коэффициентов и матриц остаются прежними. Это приводит к минимизации N квадратичных критериев при $t = 1, \ldots, N$:

$$\begin{cases} \sum\limits_{s=1,\ s\neq t}^{N} (y_s - \boldsymbol{\beta}_s^\intercal \mathbf{x}_s)^2 + \lambda \sum\limits_{s=2}^{N} (\boldsymbol{\beta}_s - \mathrm{V}_s \boldsymbol{\beta}_{s-1})^\intercal \mathrm{U}_s (\beta_s - \mathrm{V}_s \beta_{s-1}) \to \max\limits_{\boldsymbol{\beta}} \\ \sum_{i=1}^{n} \beta_s^i = 1,\ s = 1, \ldots, N \end{cases} \tag{21}$$

Последовательность решений этих критериев $(\boldsymbol{\beta}_t^{(t)})_{t=1}^{N}$ соответствуют одному из классов модели, которые различаются значением параметра сглаживания λ, участвующего в критерии. Таким образом, для получения наилучшей модели предлагается рассматривать пробные значения для структурного параметра λ и сравнивать полученные оценки дисперсии шума наблюдений $D(\lambda)$, выбирая ту модель, которой соответствует наименьшая оценка дисперсии $\lambda^* = \operatorname{argmin} D(\lambda)$. Данный метод изложен в работе [2] и называется *методом скользящего контроля* или *Leave One Out* (*LOO*).

Для исследования также использовалась модификация этого метода под названием Leave Half Out (*LHO*). В данном методе решается задача минимизации критерия (9), при исключении одновременно половины наблюдений $(\mathbf{x}_t, y_t)$, соответствующих нечетным индексам, то есть с обнулением $\mathrm{Q}_t = 0$ по тем нечетным индексам t, после чего выбираются вектора $\hat{\boldsymbol{\beta}}_t$ только с нечетными индексами, являющимися оценками на коэффициенты регрессии в изначальной задаче. После чего проводиться аналогичная процедура, но исключаются наблюдения с четными индексами. Полученная последовательность оценок векторов коэффициентов регрессии $\hat{\boldsymbol{\beta}}_t^{(t)}$ участвует в вычислении оценки дисперсии шума наблюдения $\hat{D}(\lambda) = \frac{1}{N}\sum_{t=1}^{N}(y_t - \mathbf{x}_t^\intercal \hat{\boldsymbol{\beta}}_t^{(t)})^2$.

Данные методы отличаются быстротой вычисления и возможностью получения оценок во время решения задачи минимизации исходного критерия (9). Предполагается, что оценка метода *LOO* будет предельно несмещенной при $N \to \infty$, но оптимальная оценка будет заниженной, то есть будет приводить к сильному сглаживанию $\lambda \to \infty$. В методе *LHO* присутствует смещенность оценки дисперсии шума наблюдения, так как при вычислении оценок на коэффициенты регрессии из рассмотрения выбрасывается половина объектов, и при этом она не устойчива к незначительным изменения выборки $(\mathbf{X}_t, \mathbf{y})$.

Методы *LOO* и *LHO* относятся к так называемым *кросс-валидационными* методам, которые не используют информацию о распределении генеральной совокупности, из которой взята выборка наблюдений, а полагаются на эмпирические оценки, вычисленные с помощью этой выборки.

3.2 Оценивание структурных параметров с помощью информационного критерия

Критерий предложенный японским ученым Хиротугу Акаике был обобщен на случай неупорядоченных параметров модели в работах [3] и [6]. В ней предложена формула для подсчета оценки структурного параметра λ - параметра гладкости, который в предложенной в данной работе постановке соответствует диагональным значениям в матрице сглаживания $\mathrm{U}_t = \lambda \mathrm{I}$. Критерий Акаике состоит из двух частей, первая из которых настраивает модель под наилучшее приближение к нашим наблюдаемым данным, а вторая налагает штраф на сложность модели. Модификации критерия Акаике нестационарной регрессии в задаче управления восстановления стратегии управления описана в таком виде:

$$\hat{\lambda}(\mathbf{y}, \mathbf{X}) =$$
$$= \underset{\lambda}{\operatorname{argmin}} \underbrace{\left\{ \frac{1}{2\delta^2} \underbrace{\left(\mathbf{y} - \mathbf{X}^T \hat{\beta}_\lambda(\mathbf{y}, \mathbf{X})\right)^T \left(\mathbf{y} - \mathbf{X}^T \hat{\beta}_\lambda(\mathbf{y}, \mathbf{X})\right)}_{\text{эмпирический риск}} + \underbrace{\Delta(\lambda, \mathbf{X})}_{\text{штраф за сложность}} \right\}}_{\text{структурный риск}}$$

$$\Delta(\lambda, \mathbf{X}) = Tr \left[\mathbf{X}\mathbf{X}^T \left(\mathbf{X}\mathbf{X}^T + \mathbf{B}_\lambda \right)^{-1} \right] \tag{22}$$

Где

$$\mathbf{B}_\lambda = \begin{pmatrix} U_2^\intercal U_2 V_2 & -V_2^\intercal U_2 & 0 & 0 \\ -U_2 V_2 & U_2 + U_3^\intercal U_3 V_3 & -V_3^\intercal U_3 & 0 \\ \ddots & \ddots & \ddots & \vdots \\ 0 & -U_{N-1}V_{N-1} & U_{N-1} + U_N^\intercal U_N V_N & -V_N^\intercal U_N \\ 0 & 0 & -U_N V_N & U_N \end{pmatrix} \in \mathbb{R}^{Nn \times Nn} \tag{23}$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_2 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{x}_3 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{x}_{N-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{x}_N \end{pmatrix} \in \mathbb{R}^{Nn \times N},\ \mathbf{x}_t = (x_t^1, \dots, x_t^n)^\intercal \in \mathbb{R}^n \tag{24}$$

Недостатком данного метода является обязательное требование известного параметра шума модели наблюдения δ. Достоинством является его возможность аналитического выражения и подсчета, не требующего других процедур, в отличие от кросс-валидационных методов.

3.3 Использование критерия максимизации обоснованности модели для оценивания структурных параметров

Для оценки структурного параметра λ и соответствующей ей модели воспользуемся байесовским выводом [5].

Для полного байесовского вывода необходимо задать априорное распределение первого из рассматриваемых во времени векторов $\boldsymbol{\beta}_1$. Положим его распределение нормальным с нулевым вектором матожидания, нулевыми ковариациями компонент и их мерами точности $\rho \cdot \alpha$:

$$p(\boldsymbol{\beta}_1|\rho,\alpha) = \mathcal{N}(\mathbf{0},(\rho\alpha\mathrm{I})^{-1}),\ \rho > 0,\ \alpha > 0 \tag{25}$$

Тогда априорное совместное распределение на параметры нестационарной регрессии $\boldsymbol{\beta}$, выписанное раннее в виде (4), становится равным:

$$p(\boldsymbol{\beta}|\rho,\alpha) = p(\boldsymbol{\beta}_1|\rho,\alpha)\prod_{t=2}^{N} p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1},\alpha) = \\ \mathcal{N}(\boldsymbol{\beta}_1|\mathbf{0},(\rho\alpha\mathrm{I})^{-1})\prod_{t=2}^{N}\mathcal{N}\big(\boldsymbol{\beta}_t|\mathrm{V}_t\boldsymbol{\beta}_{t-1},(\alpha\mathrm{U}_t)^{-1}) \tag{26}$$

Априорное распределение коэффициентов в первый момент времени (25) необходимо лишь для полного определения модели и гиперпараметр ρ достаточно зафиксировать на $\rho = \rho_0 > 0$, потому что, как взаимосвязь между сглаживанием векторов коэффициентов регрессии, так и априорное распределение β_1 при такой постановке зависят от λ. В свою очередь этот параметр участвует в выражении для $\alpha = \lambda\delta$, поэтому в роли структурных параметров модели возьмем δ и λ.

Вновь воспользуемся формулой Байеса, вводя на λ и δ априорное распределение $p(\lambda,\delta) = \mathrm{I}[\lambda > 0, \delta > 0]$, где $\mathrm{I}[x \in \mathcal{X}]$, есть индикатор принадлежности переменной x множеству $\mathcal{X}$. При рассмотрении апостериорного вероятностного распределение на гиперпараметры λ и δ:

$$p(\lambda,\delta,\rho|\mathbf{y}) = \frac{p(\mathbf{y}|\lambda,\delta,\rho)p(\lambda,\delta)}{\int p(\mathbf{y}|\lambda,\delta,\rho)p(\lambda,\delta)} \propto p(\mathbf{y}|\lambda,\delta,\rho) \tag{27}$$

В качестве правдоподобия относительно гиперпараметров выступает так называемая обоснованность модели $p(\mathbf{y}|\lambda,\delta,\rho) = \int p(\mathbf{y}|\boldsymbol{\beta},\delta)p(\boldsymbol{\beta}|\lambda,\delta)\,\mathrm{d}\boldsymbol{\beta}$. Она является неполным правдоподобием, иначе мы бы включали в рассмотрение параметры модели $\boldsymbol{\beta}$, т.е. $p(\mathbf{y}|\boldsymbol{\beta},\lambda,\delta,\rho)$ - функция полного правдоподобия относительно всех скрытых параметров модели. При максимизации полного правдоподобия по всем скрытым переменным, в том числе и по векторам параметрам нестационарной регрессии $\boldsymbol{\beta}$, получаются не ограничивающие значения векторов коэффициентов регрессии структурные параметры λ и δ.

Метод максимальной обоснованности применяется для автоматического подбора гиперпараметров модели, при котором максимизируется неполное правдоподобие относительно структурных параметров модели:

$$(\lambda_{ME}, \delta_{ME}) = \operatorname{argmax} p(\mathbf{y}|\lambda, \delta, \rho) \tag{28}$$

После чего оптимальные параметры модели применяются в задаче настройки параметров нестационарной регрессии $\boldsymbol{\beta}$, в заданной точке

$$\boldsymbol{\beta}_{ML} = \underset{\boldsymbol{\beta}:\sum_{i=1}^{n}\beta_t^i=1}{\operatorname{argmax}} \; p(\boldsymbol{\beta}|\mathbf{y}, \lambda_{ME}, \delta_{ME}, \rho_0) \tag{29}$$

Перейдем непосредственно к вычислению обоснованности.

Перенесем всю информацию о степени гладкости векторов регрессии в структурный параметр λ, приравняв матрицу гладкости U_t к единичной матрице. Также для удобства дальнейшего вывода введем обозначения, при условии, что $\mathbf{x}_t = (x_t^1, \ldots, x_t^n)^\intercal \in \mathbb{R}^n$, $\boldsymbol{\beta}_t = (\beta_t^1, \ldots, \beta_t^n)^\intercal \in \mathbb{R}^n$:

$$\mathrm{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_2 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{x}_3 & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{x}_{N-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{x}_N \end{bmatrix} \in \mathbb{R}^{Nn\times N},\ \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \ldots \\ \boldsymbol{\beta}_N \end{bmatrix} \in \mathbb{R}^{Nn},\ \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N \tag{30}$$

Модель совокупности наблюдений принимает вид нормального многомерного распределения:

$$p(\mathbf{y}|\boldsymbol{\beta}, \delta) = p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \delta) = \prod_{t=1}^{N} p(y_t|\mathbf{x}_t, \boldsymbol{\beta}_t, \delta) =$$
$$= \frac{\delta^{N/2}}{(2\pi)^{N/2}} \exp\left(-\frac{\delta}{2}(\mathbf{y} - \mathbf{X}^\intercal\boldsymbol{\beta})^\intercal(\mathbf{y} - \mathbf{X}^\intercal\boldsymbol{\beta})\right) = \mathcal{N}(\mathbf{y}|\mathbf{X}^\intercal\boldsymbol{\beta}, \delta^{-1}\mathrm{I}) \tag{31}$$

Совместное априорное распределение первого вектора коэффициентов регрессии ($\lambda = \alpha/\delta$):

$$p(\boldsymbol{\beta}|\rho, \lambda, \delta) = p(\boldsymbol{\beta}_1|\rho, \lambda\delta) \prod_{t=2}^{N} p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \lambda\delta) = \mathcal{N}(\boldsymbol{\beta}_1|\mathbf{0}, (\rho\lambda\delta)^{-1}\mathrm{I})\cdot \tag{32}$$

$$\cdot \prod_{t=2}^{N} \mathcal{N}\Big(\boldsymbol{\beta}_t|\mathrm{V}_t\boldsymbol{\beta}_{t-1}, (\lambda\delta)^{-1}\mathrm{I}\Big) =$$

$$= \frac{\rho^{n/2}\lambda^{n/2}\delta^{n/2}}{(2\pi)^{n/2}} \exp\Big(-\frac{\rho\lambda\delta}{2}\boldsymbol{\beta}_1^{\mathsf{T}}\boldsymbol{\beta}_1\Big) \prod_{t=1}^{N} \Big[\frac{\lambda^{n/2}\delta^{n/2}}{(2\pi)^{n/2}} \exp\Big(-\frac{\lambda\delta}{2}(\boldsymbol{\beta}_t - \mathrm{V}_t\boldsymbol{\beta}_{t-1})^{\mathsf{T}}(\boldsymbol{\beta}_t - \mathrm{V}_t\boldsymbol{\beta}_{t-1})\Big)\Big] \tag{33}$$

Рассмотрим совместное априорное распределение при $N = 2$:

$$\begin{aligned}
&p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|\rho, \lambda, \delta) = p(\boldsymbol{\beta}_1|\rho, \lambda, \delta)p(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1, \lambda, \delta) = \\
&= \frac{\rho^{n/2}\lambda^{n/2}\delta^{n/2}}{(2\pi)^{n/2}} \exp\Big(-\frac{\rho\lambda\delta}{2}\boldsymbol{\beta}_1^{\mathsf{T}}\boldsymbol{\beta}_1\Big) \frac{\lambda^{n/2}\delta^{n/2}}{(2\pi)^{n/2}} \exp\Big(-\frac{\lambda\delta}{2}(\boldsymbol{\beta}_2 - \mathrm{V}_2\boldsymbol{\beta}_1)^{\mathsf{T}}(\boldsymbol{\beta}_2 - \mathrm{V}_2\boldsymbol{\beta}_1)\Big) = \\
&= \frac{\rho^{n/2}\lambda^{n}\delta^{n}}{(2\pi)^{n}} \exp\Big(-\frac{\lambda\delta}{2}(\rho\boldsymbol{\beta}_1^{\mathsf{T}}\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2^{\mathsf{T}}\boldsymbol{\beta}_2 - 2\boldsymbol{\beta}_2^{\mathsf{T}}\mathrm{V}_2\boldsymbol{\beta}_1 + \boldsymbol{\beta}_1^{\mathsf{T}}\mathrm{V}_2^{\mathsf{T}}\mathrm{V}_2\boldsymbol{\beta}_1)\Big) = \\
&= \frac{\rho^{n/2}\lambda^{n}\delta^{n}}{(2\pi)^{n}} \exp\Big(-\frac{\lambda\delta}{2}(\boldsymbol{\beta}_1\boldsymbol{\beta}_2)^{\mathsf{T}} \begin{pmatrix} \rho\mathrm{I} + \mathrm{V}_2^{\mathsf{T}}\mathrm{V}_2 & -\mathrm{V}_2^{\mathsf{T}} \\ -\mathrm{V}_2 & \mathrm{I} \end{pmatrix} (\boldsymbol{\beta}_1\boldsymbol{\beta}_2)\Big)
\end{aligned} \tag{34}$$

Используем формулу вычисления определителя блочной матрицы:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A - BD^{-1}C||D| = |A||D - CA^{-1}B| \tag{35}$$

$$|\mathrm{B}_\rho| = \begin{vmatrix} \rho\mathrm{I} + \mathrm{V}_2^{\mathsf{T}}\mathrm{V}_2 & -\mathrm{V}_2^{\mathsf{T}} \\ -\mathrm{V}_2 & \mathrm{I} \end{vmatrix} = |\rho\mathrm{I} + \mathrm{V}_2^{\mathsf{T}}\mathrm{V}_2 - \mathrm{V}_2^{\mathsf{T}}\mathrm{V}_2||\mathrm{I}| = \rho^n \tag{36}$$

Возвращаясь к задаче размерности N:

$$\begin{aligned}
&p(\boldsymbol{\beta}|\rho, \lambda, \delta) = \\
&= \frac{\rho^{n/2}\lambda^{Nn/2}\delta^{Nn/2}}{(2\pi)^{N/2}} \exp\Big[-\frac{\lambda\delta}{2}\Big(\rho\boldsymbol{\beta}_1^{\mathsf{T}}\boldsymbol{\beta}_1 + \sum_{t=2}^{N}(\boldsymbol{\beta}_t^{\mathsf{T}}\boldsymbol{\beta}_t - 2\boldsymbol{\beta}_t^{\mathsf{T}}\mathrm{V}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{\beta}_{t-1}^{\mathsf{T}}\mathrm{V}_t^{\mathsf{T}}\mathrm{V}_t\boldsymbol{\beta}_{t-1})\Big)\Big] = \\
&= \frac{\rho^{n/2}\lambda^{Nn/2}\delta^{Nn/2}}{(2\pi)^{Nn/2}} \exp\Big[-\frac{1}{2}\boldsymbol{\beta}^{\mathsf{T}}\mathrm{B}_{\rho\lambda\delta}\boldsymbol{\beta}\Big]
\end{aligned} \tag{37}$$

, где $\mathrm{B}_{\rho\lambda\delta} \in \mathbb{R}^{Nn\times Nn}$ является блочной трехдиагональной матрицей:

$$\mathrm{B}_{\rho\lambda\delta} = \lambda\delta\mathrm{B}_{\rho} = \lambda\delta \begin{pmatrix} \rho\mathrm{I} + \mathrm{V}_2^\intercal\mathrm{V}_2 & -\mathrm{V}_2^\intercal & \mathbf{0} & \vdots & \mathbf{0} & \mathbf{0} \\ -\mathrm{V}_2 & \mathrm{I} + \mathrm{V}_3^\intercal\mathrm{V}_3 & -\mathrm{V}_3^\intercal & \vdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathrm{V}_3 & \mathrm{I} + \mathrm{V}_4^\intercal\mathrm{V}_4 & \ddots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathrm{V}_{N-1} & \mathrm{I} + \mathrm{V}_N^\intercal\mathrm{V}_N & -\mathrm{V}_N^\intercal \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathrm{V}_N & \mathrm{I} \end{pmatrix}$$

Рекуррентно применяя формулу определителя блочной матрицы с квадратными блоками и выводом для этой матрице размера $2N \times 2N$ (35) получаем, что $|\mathrm{B}_{\rho\lambda\delta}| = \lambda^{Nn}\delta^{Nn}\rho^{n}$.

Таким образом учитывая предыдущие выкладки, получаем что совокупность векторов коэффициентов регрессии имеет многомерное нормальное распределение с нулевым матожиданием и матрицей ковариации равной $\mathrm{B}_{\rho\lambda\delta}^{-1}$:

$$p(\boldsymbol{\beta}|\rho,\lambda,\delta) = \frac{|\mathrm{B}_{\rho\lambda\delta}|^{1/2}}{(2\pi)^{Nn/2}} \exp[-\frac{1}{2}\boldsymbol{\beta}^\intercal\mathrm{B}_{\rho\lambda\delta}\boldsymbol{\beta}] = \mathcal{N}(\boldsymbol{\beta}|\mathbf{0}, \mathrm{B}_{\rho\lambda\delta}^{-1}) \tag{38}$$

Для поиска обоснованности воспользуемся формулой свертки двух нормальных распределений, таких что:

$$\begin{aligned} &p(\mathbf{w}|\mathbf{v}) = \mathcal{N}(\mathbf{w}|\mathrm{A}\mathbf{v}, \Gamma) \\ &p(\mathbf{v}) = \mathcal{N}(\mathbf{v}|\mu, \Sigma) && (39) \\ &p(\mathbf{w}) = \int p(\mathbf{w}|\mathbf{v})p(\mathbf{v})\,\mathrm{d}\mathbf{v} = \int \mathcal{N}(\mathbf{w}|\mathrm{A}\mathbf{v}, \Gamma)\mathcal{N}(\mathbf{v}|\mu, \Sigma)\,\mathrm{d}\mathbf{v} = \\ &\qquad = \mathcal{N}(\mathbf{w}|\mathrm{A}\mu, \Gamma + \mathrm{A}\Sigma\mathrm{A}^\intercal) && (40) \end{aligned}$$

Тогда относительно функции правдоподобия $p(\mathbf{y}|\boldsymbol{\beta},\delta)$ и априорного совместного распределения $p(\boldsymbol{\beta}|\rho,\lambda,\delta)$ получаем, что обоснованность модели $p(\mathbf{y}|\lambda, \delta, \rho)$ равна:

$$p(\mathbf{y}|\lambda, \delta, \rho) = \int p(\mathbf{y}|\boldsymbol{\beta}, \delta) p(\boldsymbol{\beta}|\rho, \lambda, \delta)\, \mathrm{d}\boldsymbol{\beta} = \mathcal{N}(\mathbf{y}|\mathbf{0}, \delta^{-1}\mathrm{I} + \mathrm{X}^{\intercal}\mathrm{B}_{\rho\lambda\delta}^{-1}\mathrm{X}) =$$

$$= \mathcal{N}(\mathbf{y}|\mathbf{0}, \delta^{-1}\mathrm{I} + (\lambda\delta)^{-1}\mathrm{X}^{\intercal}\mathrm{B}_{\rho}^{-1}\mathrm{X}) = \frac{\delta^{N/2}}{(2\pi)^{N/2}|\mathrm{I} + \lambda^{-1}\mathrm{X}^{\intercal}\mathrm{B}_{\rho}^{-1}\mathrm{X}|^{1/2}} \cdot$$

$$\cdot \exp\left(-\frac{\delta}{2}\mathbf{y}^{\intercal}(\mathrm{I} + \lambda^{-1}\mathrm{X}^{\intercal}\mathrm{B}_{\rho}^{-1}\mathrm{X})^{-1}\mathbf{y}\right) \tag{41}$$

Воспользуемся методом покоординатного подъема для максимизации данной функции. Предварительно прологарифмируем ее, тогда точка в которой достигается максимум логарифма обоснованности, равен оптимальной точке и для не прологарифмированной функции:

$$\ln p(\mathbf{y}|\lambda, \delta, \rho) = const + \frac{N}{2}\ln\delta - \frac{1}{2}\ln|\mathrm{I} + \lambda^{-1}\mathrm{X}^{\intercal}\mathrm{B}_{\rho}^{-1}\mathrm{X}| -$$

$$- \frac{\delta}{2}\mathbf{y}^{\intercal}(\mathrm{I} + (\lambda)^{-1}\mathrm{X}^{\intercal}\mathrm{B}_{\rho}^{-1}\mathrm{X})^{-1}\mathbf{y} \tag{42}$$

Найдем частные производные по λ и δ:

$$\frac{\partial}{\partial\delta}\ln p(\mathbf{y}|\lambda, \delta, \rho) = \frac{N}{2}\frac{1}{\delta} - \frac{\mathbf{y}^{\intercal}(\mathrm{I} + \lambda^{-1}\mathrm{X}^{\intercal}\mathrm{B}_{\rho}^{-1}\mathrm{X})^{-1}\mathbf{y}}{2} = 0$$

Воспользуемся тождеством Вубдери:

$$(\mathrm{A} + \mathrm{X}\mathrm{B}\mathrm{X}^{\intercal})^{-1} = \mathrm{A}^{-1} - \mathrm{A}^{-1}\mathrm{X}(\mathrm{B}^{-1} + \mathrm{X}^{\intercal}\mathrm{A}^{-1}\mathrm{X})^{-1}\mathrm{X}^{\intercal}\mathrm{A}^{-1} \Rightarrow$$

$$\Rightarrow (\mathrm{I} + \lambda^{-1}\mathrm{X}^{\intercal}\mathrm{B}_{\rho}^{-1}\mathrm{X})^{-1} = \mathrm{I} - \mathrm{X}^{\intercal}(\lambda\mathrm{B}_{\rho} + \mathrm{X}\mathrm{X}^{T})^{-1}\mathrm{X}$$

$$\delta = \frac{N}{\mathbf{y}^{\intercal}\left(\mathrm{I} - \mathrm{X}^{\intercal}(\lambda\mathrm{B}_{\rho} + \mathrm{X}\mathrm{X}^{T})^{-1}\mathrm{X}\right)\mathbf{y}} \tag{43}$$

$$\frac{\partial}{\partial\lambda}\ln p(\mathbf{y}|\lambda, \delta, \rho) = -\frac{1}{2}\frac{\partial}{\partial\lambda}\ln|\mathrm{I} + \lambda^{-1}\mathrm{X}^{\intercal}\mathrm{B}_{\rho}^{-1}\mathrm{X}| -$$

$$- \frac{\delta}{2}\frac{\partial}{\partial\lambda}\left(\mathbf{y}^{\intercal}(\mathrm{I} - \mathrm{X}^{\intercal}(\lambda\mathrm{B}_{\rho} + \mathrm{X}\mathrm{X}^{T})^{-1}\mathrm{X})\mathbf{y}\right) = 0$$

Воспользуемся формулой дифференцирования матриц по параметру:

$$\frac{\partial(\mathrm{A}\mathrm{X}(z)\mathrm{B})}{\partial z} = \mathrm{A}\frac{\partial\mathrm{X}(z)}{\partial z}\mathrm{B}; \quad \frac{\partial(\mathrm{X}(z))^{-1}}{\partial z} = -\mathrm{X}(z)^{-1}\frac{\partial\mathrm{X}(z)}{\partial z}\mathrm{X}(z)^{-1} \Rightarrow$$

$$\Rightarrow -\frac{\delta}{2}\left[\mathbf{y}^{\intercal}\frac{\partial}{\partial\lambda}\left(\mathrm{I} - \mathrm{X}^{\intercal}(\lambda\mathrm{B}_{\rho} + \mathrm{X}\mathrm{X}^{T})^{-1}\mathrm{X}\right)\mathbf{y}\right] = \frac{\delta}{2}\left[\mathbf{y}^{\intercal}\left(\mathrm{X}^{\intercal}\frac{\partial}{\partial\lambda}(\lambda\mathrm{B}_{\rho} + \mathrm{X}\mathrm{X}^{T})^{-1}\mathrm{X}\right)\mathbf{y}\right]$$

$$= -\frac{\delta}{2}\left[\mathbf{y}^{\intercal}\mathrm{X}^{\intercal}(\lambda\mathrm{B}_{\rho} + \mathrm{X}\mathrm{X}^{T})^{-1}\mathrm{B}_{\rho}(\lambda\mathrm{B}_{\rho} + \mathrm{X}\mathrm{X}^{T})^{-1}\mathrm{X}\mathbf{y}\right] = -\frac{\delta}{2}\bar{\boldsymbol{\beta}}(\lambda)^{\intercal}\mathrm{B}_{\rho}\bar{\boldsymbol{\beta}}(\lambda) \tag{44}$$

Здесь мы обозначили $\bar{\boldsymbol{\beta}}(\lambda) = (\lambda \mathrm{B}_\rho + \mathrm{X}\mathrm{X}^T)^{-1}\mathrm{X}\mathbf{y}$.

Рассмотрим все собственные значения $\mu_i, i = 1, \ldots, N$ положительно определенной симметричной матрицы $\mathrm{X}^\intercal \mathrm{B}_\rho^{-}\mathrm{X} \in \mathbb{R}^{N\times N}$. Тогда собственные значения матрицы $\lambda \mathrm{I} + \mathrm{X}^\intercal \mathrm{B}_\rho^{-1}\mathrm{X}$ равны $\lambda + \mu_i, i = 1 \ldots, N$. Так как определитель квадратной матрицы равен произведению его собственных значений получаем, что:

$$-\frac{1}{2}\frac{\partial}{\partial\lambda}\ln|\lambda\mathrm{I} + \mathrm{X}^\intercal\mathrm{B}_\rho^{-1}\mathrm{X}| = -\frac{1}{2}\frac{\partial}{\partial\lambda}\ln\prod_{i=1}^{N}(\lambda+\mu_i) = -\frac{1}{2}\sum_{i=1}^{N}(\lambda+\mu_i)^{-1} \quad (45)$$

Тогда

$$\frac{\partial}{\partial\lambda}\ln p(\mathbf{y}|\lambda,\delta,\rho) = \frac{N}{2}\frac{1}{\lambda} - \frac{1}{2}\sum_{i=1}^{N}\frac{1}{\lambda+\mu_i} - \frac{\delta}{2}\bar{\boldsymbol{\beta}}^\intercal(\lambda)\mathrm{B}_\rho\bar{\boldsymbol{\beta}}(\lambda) = 0| \cdot 2\lambda$$

$$N - \sum_{i=1}^{N}\frac{\lambda}{\lambda+\mu_i} - \lambda\delta\bar{\boldsymbol{\beta}}^\intercal(\lambda)\mathrm{B}_\rho\bar{\boldsymbol{\beta}}(\lambda) = 0$$

$$\sum_{i=1}^{N}\frac{\mu_i}{\lambda+\mu_i} - \lambda\delta\bar{\boldsymbol{\beta}}^\intercal(\lambda)\mathrm{B}_\rho\bar{\boldsymbol{\beta}}(\lambda) = 0$$

Окончательно формула пересчета для λ :

$$\lambda^{new} = \left(\delta\bar{\boldsymbol{\beta}}^\intercal(\lambda^{old})\mathrm{B}_\rho\bar{\boldsymbol{\beta}}(\lambda^{old})\right)^{-1} \cdot \sum_{i=1}^{N}\frac{\mu_i}{\lambda^{old}+\mu_i} \quad (46)$$

Для подсчета обоснованности определим итерационный алгоритм:

Алгоритм 1: Метод покоординатного подъема в задаче максимизации обоснованности

Исходные данные: $\mathbf{y},\ \mathbf{X},\ \mathrm{B}_\rho,\ \lambda_0,\ \delta_0,\ \rho_0, \epsilon$

Результат: $\lambda_{ME},\ \delta_{ME}$

$\lambda := \lambda_0$, $\delta := \delta_0$;

пока: $|\ln p(\mathbf{y}|\lambda^{new}, \delta^{new}, \rho_0) - \ln p(\mathbf{y}|\lambda^{old}, \delta^{old}, \rho_0)| \le \epsilon$

$\bar{\boldsymbol{\beta}}(\lambda_{old}) = (\lambda_{old}\mathrm{B}_\rho + \mathrm{X}\mathrm{X}^T)^{-1}\mathrm{X}\mathbf{y}$;

$\delta_{new} = N\left(\mathbf{y}^\intercal\mathbf{y} - \mathbf{y}^\intercal\mathrm{X}^\intercal\bar{\boldsymbol{\beta}}(\lambda_{old})\right)^{-1}$;

$\lambda_{new} = \left(\delta\bar{\boldsymbol{\beta}}^\intercal(\lambda_{old})\mathrm{B}_\rho\bar{\boldsymbol{\beta}}(\lambda_{old})\right)^{-1} \cdot \sum_{i=1}^{N}\frac{\mu_i}{\lambda_{old}+\mu_i}$;

$\lambda_{ME} = \lambda_{new},\ \delta_{ME} = \delta_{new}$;

Метод покоординатного подъема гарантирует максимизацию логарифма функции правдоподобия на каждом шаге итерации. Критерием останова является отличие логарифма обоснованности на новой итерации от предыдущей с точностью ϵ. В практическом обосновании метода в книге К. Бишопа "Распознавание образов и машинное обучение" [7] предложенный для ридж-регрессии говорится о возможной медленной сходимости алгоритма. Достоинством метода является получение истинных оптимальных значений структурных параметров модели, при котором достигается максимум обоснованности и соответственно апостериорной вероятности структурных параметров при полученных наблюдениях в выборке.

4 Экспериментальное исследование

4.1 Структура экспериментальных данных

Данные для экспериментальных исследований были предоставлены компанией "Markov Processe Int.". В них присутствуют индексы $n = 11$ активов за каждый день в период с 30 октября 2015 года по 1 января 2016 года (всего $N = 57$ дней). Для этих же дней предоставлены данные доходностей 10 хедж-фондов. Все величины даны в процентах.

Также для каждого фонд известен показатель оборачиваемости инвестиционного портфеля инвестиционного фонда за прошлый год, используемый далее в результатах эксперимента 4.3 и описанный в следующей секции 4.2.

4.2 Критерий проверки качества оцененной модели динамики портфеля

Для оценки работы методов *LOO, LHO, AIC, ME* был выбран внешний критерий под названием оборачиваемость инвестиционного портфеля инвестиционного фонда. *Turn over* или *оборачиваемость* ν - показатель торговой активности фонда в течение года, выраженный в процентах от средних суммарных активов фонда. Он выражается как отношение:

$$\nu = \frac{\text{показатель торговой активности фонда в течении года}}{\text{средние суммарные активы фонда}}, \text{ где}$$

$$\text{показатель торговой активности фонда в течении года} =$$
$$= \sum_{t=2}^{N} \sum_{i=1}^{n} |z_t^i m_t^i - z_t^i m_{t-1}^i|,$$
$$\text{средние суммарные активы фонда} = \frac{1}{N} \sum_{t=1}^{N} \sum_{i=1}^{n} z_t^i m_t^i.$$

Распишем оборачивамость в обозначениях из постновки задачи :

$$\nu = \frac{\sum_{t=2}^{N} \sum_{i=1}^{n} |z_t^i m_t^i - z_t^i m_{t-1}^i|}{\frac{1}{N} \sum_{t=1}^{N} \sum_{i=1}^{n} z_t^i m_t^i} = N \frac{\sum_{t=2}^{N} \frac{\sum_{i=1}^{n} |z_t^i m_t^i - z_t^i m_{t-1}^i|}{\sum_{i=1}^{n} z_t^i m_t^i} \sum_{i=1}^{n} z_t^i m_t^i}{\sum_{t=1}^{N} \sum_{i=1}^{n} z_t^i m_t^i} =$$
$$= N \frac{\sum_{t=2}^{N} \sum_{i=1}^{n} |\frac{z_t^i m_t^i}{\sum_{i=1}^{n} z_t^i m_t^i} - \frac{z_t^i m_{t-1}^i}{\sum_{i=1}^{n} z_t^i m_t^i}| \sum_{i=1}^{n} z_t^i m_t^i}{\sum_{t=1}^{N} \sum_{i=1}^{n} z_t^i m_t^i} =$$
$$= \left\{\beta_t^i = \frac{z_t^i m_t^i}{\sum_{i=1}^{n} z_t^i m_t^i};\ z_t^i = (x_t^i + 1) z_{t-1}^i\right\} =$$
$$\nu = N \frac{\sum_{t=2}^{N} \sum_{i=1}^{n} |\beta_t^i m - (x_t^i + 1) \frac{z_{t-1}^i m_{t-1}^i}{\sum_{i=1}^{n} z_t^i m_t^i}| \sum_{i=1}^{n} z_t^i m_t^i}{\sum_{t=1}^{N} \sum_{i=1}^{n} z_t^i m_t^i}$$

Выразим стоимость портфеля через его доходность: $\sum_{i=1}^{n} z_t^i m_t^i = (y_t + 1) \sum_{i=1}^{n} z_{t-1}^i m_{t-1}^i$.

$$\nu = N \frac{\sum_{t=2}^{N} \sum_{i=1}^{n} |\beta_t^i m - (x_t^i + 1) \frac{z_{t-1}^i m_{t-1}^i}{(y_t+1) \sum_{i=1}^{n} z_{t-1}^i m_{t-1}^i}| \sum_{i=1}^{n} z_t^i m_t^i}{\sum_{t=1}^{N} \sum_{i=1}^{n} z_t^i m_t^i} =$$
$$= N \frac{\sum_{t=2}^{N} \sum_{i=1}^{n} |\beta_t^i - \frac{(x_t^i+1)}{(y_t+1)} \beta_{t-1}^i| \sum_{i=1}^{n} z_t^i m_t^i}{\sum_{t=1}^{N} \sum_{i=1}^{n} z_t^i m_t^i}$$

Запишем рекуррентная формулу для стоимости портфеля в момент t через стоимость портфеля в некоторый начальный момент времени: $\sum_{i=1}^{n} z_t^i m_t^i = \prod_{s=1}^{t} (1 + y_s) \sum_{i=1}^{n} z_0^i m_0^i$. Тогда:

$$\nu = N \frac{\sum_{t=2}^{N} \sum_{i=1}^{n} |\beta_t^i - \frac{(x_t^i+1)}{(y_t+1)} \beta_{t-1}^i| \prod_{s=1}^{t} (1 + y_s)}{\sum_{t=1}^{N} \prod_{s=1}^{t} (1 + y_s)} \tag{47}$$

В дальнейших исследованиях используется информация о показателях оборачиваемости портфелей 10 фондов, описанных выше. За внешний критерием наилучшего метода оценки структурного параметра λ возьмем сумму средних квадратичных отклонений истиной оборачиваемости от ее оценки для каждого фонда, полученной с помощью найденных векторов коэффициентов регрессии $\boldsymbol{\beta}$.

4.3 Результаты экспериментального исследования

Для первых трех методов эксперименты проводились на сетке значений $\lambda \in [0.01; 6000]$. Область была взята из примеров работы итерационного алгоритма для нескольких фондов и обозначения возможных значений параметра λ.

Результаты алгоритма максимизации обоснованности 1, а также результаты работы методов проиллюстрированы на рисунках и графиках 4.3, 4.3, 4.3. В соответствии с критерием останова, алгоритм покоординатного подъема вычислял максимум логарифма обоснованности с точностью до $\epsilon = 1e - 6$.

Стоит заметить, что оборачиваемость инвестиционного портфеля инвестиционного фонда дается за прошедший год, поэтому предварительно эта величина была поделена на 6, чтобы она могла быть сравнима с показателем 2х-месячной активности портфеля, оцененной нашей моделью через коэффициенты регрессии.

4.4 Выводы

Главным критерием оценки метода подбора структурных параметров для нас является полученная после его настройки оборачиваемость.

Как и ожидалось, лучший результат был получен методом максимальной обоснованности, так как параметр λ подбирался на всей возможной области значений и соответствует близкому к оптимальному зна-

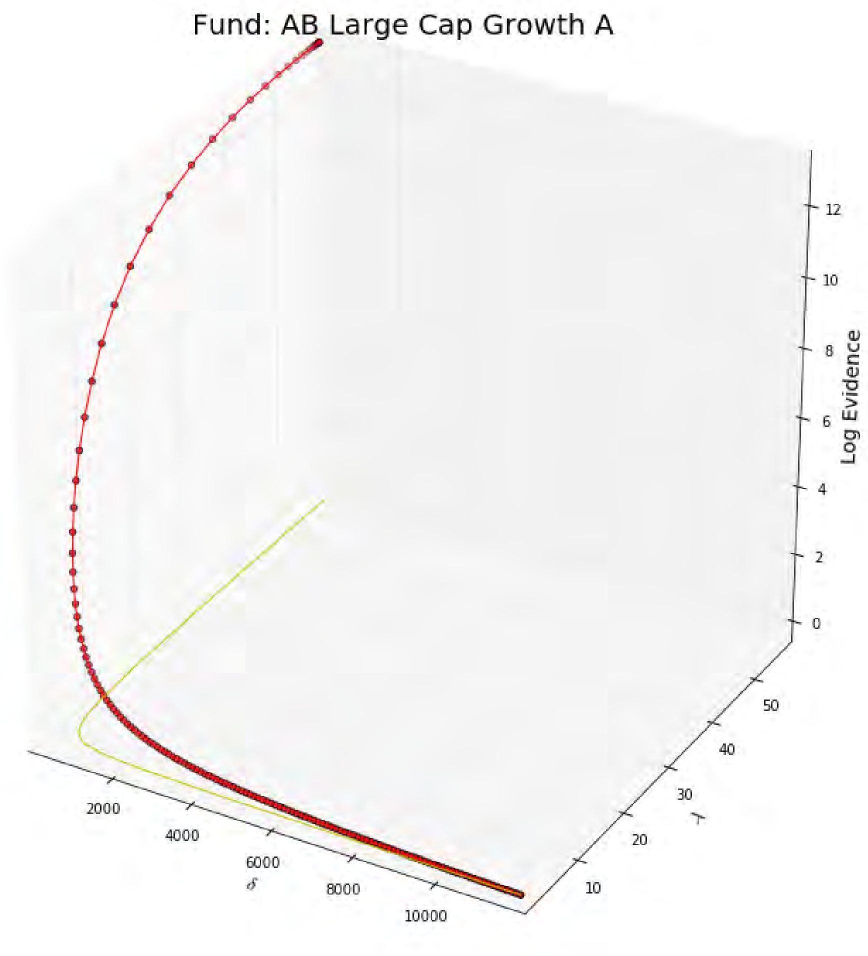
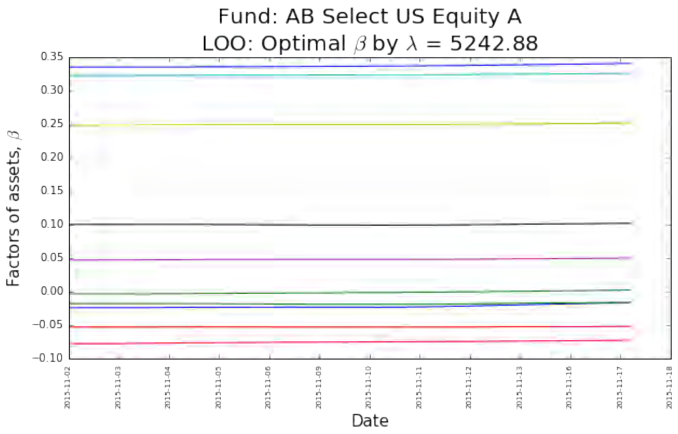
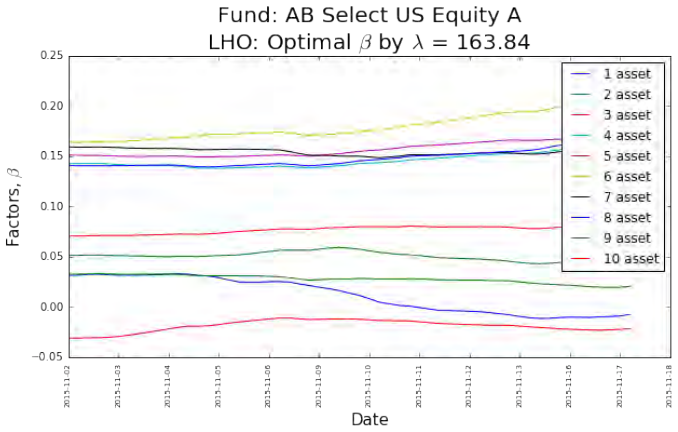
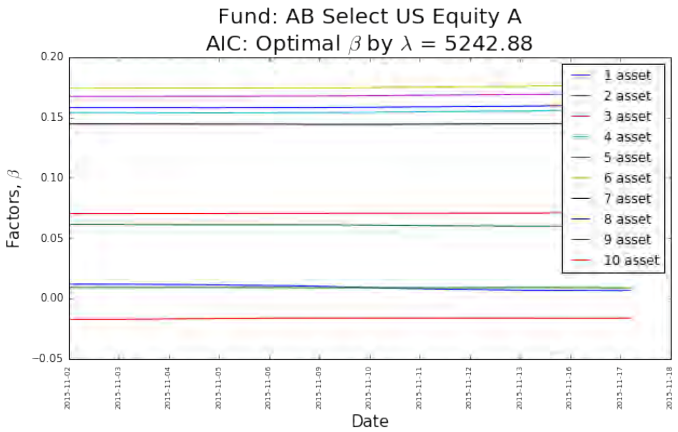
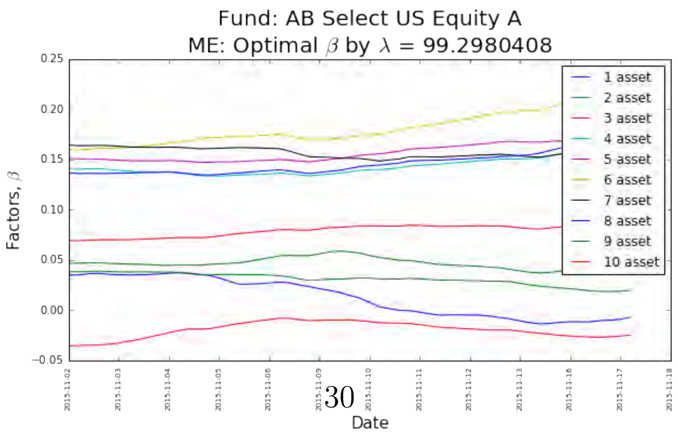


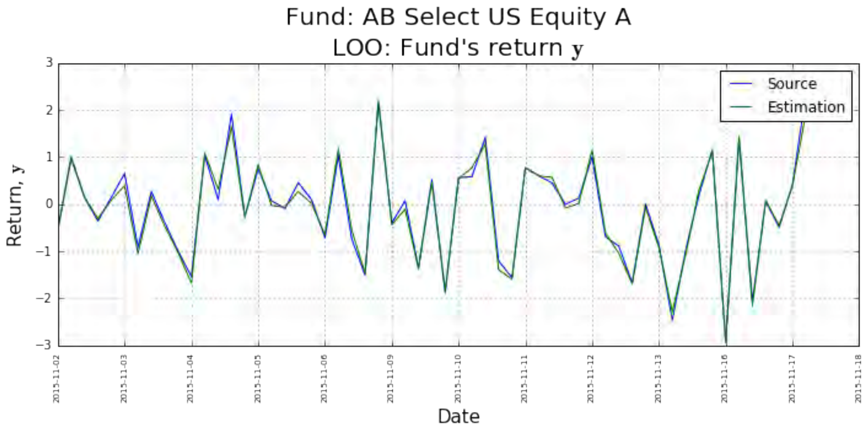
Рис. 1: 185 итераций метода покоординатного подъема для максимизации обоснованности и соответствующие точки логарифма обоснованности

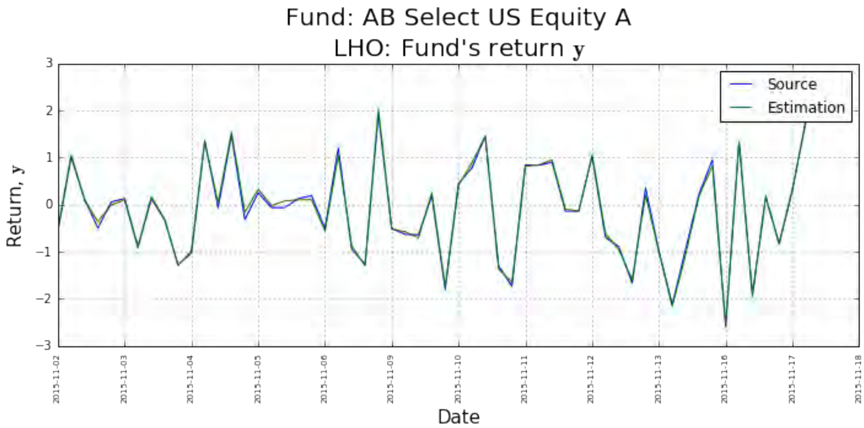

Fund: AB Select US Equity A
LOO: Optimal β by λ = 5242.88
Factors of assets, β
Date

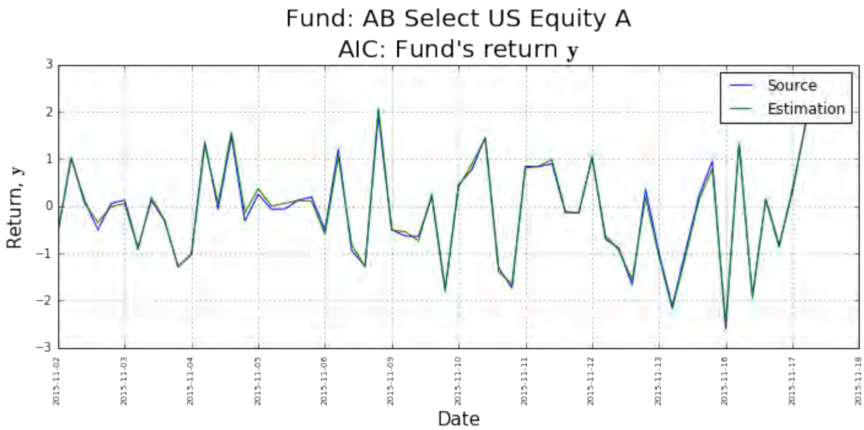

Fund: AB Select US Equity A
LHO: Optimal β by λ = 163.84
Factors, β
Date
1 asset
2 asset
3 asset
4 asset
5 asset
6 asset
7 asset
8 asset
9 asset
10 asset


Fund: AB Select US Equity A
AIC: Optimal β by λ = 5242.88
Factors, β
Date
1 asset
2 asset
3 asset
4 asset
5 asset
6 asset
7 asset
8 asset
9 asset
10 asset


Fund: AB Select US Equity A
ME: Optimal β by λ = 99.2980408
Factors, β
Date
1 asset
2 asset
3 asset
4 asset
5 asset
6 asset
7 asset
8 asset
9 asset
10 asset


Fund: AB Select US Equity A
LOO: Fund's return y
Source
Estimation
Return, y
Date


Fund: AB Select US Equity A
LHO: Fund's return y
Source
Estimation
Return, y
Date


Fund: AB Select US Equity A
AIC: Fund's return y
Source
Estimation
Return, y
Date

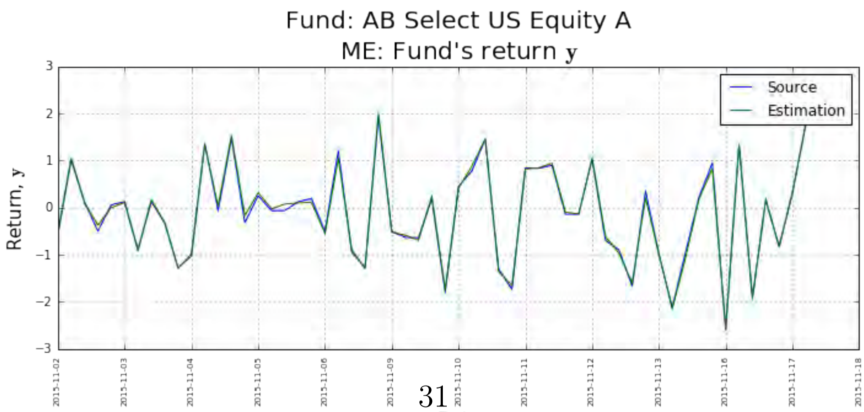

Fund: AB Select US Equity A
ME: Fund's return y
Source
Estimation
Return, y
Date

Таблица 2: Усредненные оборачиваемости (turn over) фондов и их оценки, полученные при настройке структурного параметра λ соответствующими методами

	Фонд 1	Фонд 2	Фонд 3	Фонд 4	Фонд 5	Фонд 6
Turn Over (TO)	29.33	12.5	16.83	9.83	12.33	58.
LOO TO	53.90	40.71	24.28	30.18	29.71	26.11
LHO TO	1319.90	254.56	36.38	114.11	154.52	39.91
AIC TO	1690.17	34.56	23.17	28.80	46.51	21.21
ME TO	87.52	68.81	54.85	79.41	97.10	50.76

	Фонд 7	Фонд 8	Фонд 9	Фонд 10	Ср. кв. отклонение
Turn Over (TO)	9.33	19.	77.5	10.33	—
LOO TO	30.80	63.30	284.94	30.87	221.50
LHO TO	315.17	382.39	652.93	88.96	1522.78
AIC TO	315.17	47.82	652.93	63.16	1786.08
ME TO	62.03	83.38	123.73	60.41	177.90

Таблица 3: Полученные в эксперименте значения гиперпараметра гладкости λ

	Фонд 1	Фонд 2	Фонд 3	Фонд 4	Фонд 5	Фонд 6
LOO TO	5242.88	327.68	5242.88	5242.88	5242.88	655.36
LHO TO	2.56	5.12	327.68	40.96	20.48	163.84
AIC TO	0.32	655.36	5242.88	5242.88	327.68	5242.88
ME TO	873.43	92.60	137.53	86.22	57.70	99.29

	Фонд 7	Фонд 8	Фонд 9	Фонд 10
LOO TO	5242.88	163.84	40.96	5242.8
LHO TO	10.24	0.02	2.56	81.92
AIC TO	10.24	327.68	2.56	163.84
ME TO	266.20	86.78	262.18	180.53

чению максимума обоснованности. При этом параметр λ оказывается не таким большим, чтобы найденные коэффициенты регрессии - доли активов фондов были сильно сглажены 4.3, но и не таким маленьким, чтобы не происходило согласования между их значениями во времени.

Второй лучший показатель для суммы отклонений от истинных значений оборачиваемости фондов принадлежит методу скользящего контроля, но при этом стоит отметить постоянные завышенные значения λ 3, при которых видно сильное сглаживание коэффициентов регрессии $\boldsymbol{\beta}_t$ 4.3.

Остальные методы показывали не постоянные результаты из которых сложно сделать окончательные выводы, но полученные в данных методах значения оценок оборачиваемости для некоторых фондов сильно превышают истинный показатель.

Несмотря на различие в показателях оборачиваемости данные методы отлично аппроксимируют доходности инвестиционных портфелей 4.3. Таким образом без помощи настройки гиперпараметра сглаживания λ нельзя получить доли активов портфеля, близкие к реальной ситуации.

5 Заключение

В работе рассматривается задача оценивания структурных параметров модели нестационарной регрессии. В качестве прикладного примера, приводящего к необходимости оценивания нестационарной регрессии, в работе обсуждается задача восстановления скрытой стратегии управления инвестиционным портфелем.

Рассматривались три способа подбора структурных параметров: кросс-валиадционные методы, в том числе Leave One Out и его модификация Leave Half Out, критерий Акаике и критерий обоснованности модели. Для последнего метода была поставлена вероятностная постановка и проведен математический вывод обоснованности вероятностной модели, с последующей разработкой алгоритма ее максимизации.

Полученные результаты относительно выбранного внешнего критерия показывают, что при использовании критерия обоснованности полученные коэффициенты нестационарной регрессии более точно описывают реальные данные.

В дальнейшем исследовании будет возможным рассмотреть результаты работы критерия обоснованности на модельных данных, а также провести анализ границ применимости метода оценивания структурного параметра.

Список литературы

[1] Harry M. Markowitz, Portfolio Selection // Journal of Finance, 7, no. 1 (March 1952), pp. 77-91

[2] Костин А. С., Моттль В. В., Красоткина О. В. Алгоритмы динамического программирования для анализа нестационарных сигналов // Журнал вычислительной математики и математической физики. — 2003. — № 2. — С. 103–117

[3] Обобщение информационного критерия Акаике для выбора значений непрерывного параметра в моделях данных / В. В. Моттль, О. В. Красоткина, Е. О. Ежова (Черноусова) // Труды 51 научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук». — 2008. — Т. 3. — С. 58-60.

[4] Bellman R. Dynamic Programming. Princeton University Press, // Princeton, N.J., 1957.

[5] Байесовские методы машинного обучения (курс лекций, Д.П. Ветров, Д.А. Кропотов) / 2015

[6] Ezhova E., Mottl V., Krasotkina O. Estimation of time-varying linear regression with unknown time-volatility via continuous generalization of the akaike information criterion // International Science Index. — 2009. — Vol. 3, no. 3. — P. 132–137.

[7] Bishop, Christopher M. Pattern Recognition and Machine Learning // New York: Springer, 2006.