

## 1. Выбор статьи:

Caldeira J., Moura G. V. Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy //Available at SSRN 2196391. – 2013.

[https://www.researchgate.net/publication/256043956\\_Selection\\_of\\_a\\_Portfolio\\_of\\_Pairs\\_Based\\_on\\_Cointegration\\_A\\_Statistical\\_Arbitrage\\_Strategy](https://www.researchgate.net/publication/256043956_Selection_of_a_Portfolio_of_Pairs_Based_on_Cointegration_A_Statistical_Arbitrage_Strategy)

В рамках статьи я выбрал построение статистического арбитража при выборе пар на основе коинтеграции. В связи с трудностью построения статистического арбитража из-за необходимости больших вычислительных мощностей, рассматривается частный случай статистического арбитража – парный трейдинг.

## 2. Данные:

<https://www.kaggle.com/datasets/franciscofeng/augmented-china-stock-data-with-fundamentals>

## 3. Реализация алгоритма

Первым шагом в рамках реализации модели стал отбор активов, торги которым будут осуществляться в дальнейшем. Одним из параметров изначального отбора стал объём торгов актива как показатель ликвидности, необходимый для гарантирования осуществления сделки проведения статистического арбитража. Соответственно, рассматривались акции, которые имели топ 5% по объёму с 2011 по 2020 год. Другим параметром стало количество торгуемых дней - так как для того, чтобы осуществлять арбитраж необходимо, чтобы актив присутствовал на рынке – активу нужно торговаться хотя бы 95% от максимального количества дней актива.

Вторым основным шагом стал отбор активов относительно коинтеграции. В рамках определения в статье активы коинтегрированы если отдельный из активов нестационарный, а их линейная комбинация стационарна. В рамках проверки стационарности отдельных активов был проведен тест ADF тест (расширенный тест Дики-Фуллера). А после проверки нестационарности для каждого возможного сочетания активов был проведён двухшаговый тест Engle Granger. После этого были найдены параметр (в рамках работы я это называю гаммой), относительно которых линейная комбинация активов стационарна. Относительно стратегии полученная гамма может быть как положительная, так и отрицательная, так что после этого работа представлена только относительно полученной разницы между активами. То есть, другими словами, разница по модулю между активами - теперь новый актив, с которым осуществляется торговля.

В рамках вышеотмеченной статьи предлагалось привести спред коинтегрируемых активов относительно стандартного нормального распределения (в дальнейшем я

это буду называть  $z_t$ ). В дальнейшем стратегия торговли осуществлялась следующая:

- Покупка при  $z_t < -1.85$
- Шорт при  $z_t > 1.85$
- Заккрытие шорта при  $z_t < 0.7$
- Заккрытие длинной позиции  $z_t > -0.7$

В рамках данного алгоритма так же было введено: закрытие позиции если в течение 50 дней позиция находится в убытке, стоп-лосс при позиции -5%, а также в случае, если стандартное распределение спреда активов вернулось в районе  $[-1.4, 1.4]$ , то алгоритм снова начинает работать.

После этого была проведена реализация на тестовой выборке, а также для проверочного датасета относительно тестовой выборки осуществился еще один отбор пар активов относительно их показателей эффективности инвестиционного портфеля, в качестве которых я выбрал  $r^2$  и показатель Шарпа. Первый необходим как проверка, что алгоритм торгует без сильных просадок и равномерно получает прибыль. В то время как второй показывал, насколько прибыль преобладает над риском в реализации стратегии. Относительно того, что показатель шарпа необходим быть более 1, а  $r^2$  более 0.6 для осуществления равномерной торговли. Так же чтобы избежать уменьшения доходности я выбрал 20 лучших активов и реализовал их в проверочном датасете.

В рамках работы я улучшил работу алгоритма относительно отбора активов (изначально осуществлялся на основе объема и торгуемых дней, а также в дальнейшем отбор по тестовой выборке (то есть тестовая выступала валидационной) для проверочной выборки относительно инвестиционной эффективности портфеля. Добавил окно изменение стандартизированного спреда. Помимо этого, реализовал способ возобновления работы торгов в случае достижения стоп-лосса или долгого срока торгов. А также уменьшил диапазон срабатывания до ~90% доверительного интервала попадает активов (В статье указано ~95%, из-за чего торги осуществляются довольно редко).

После поправки коэффициента шарпа в связи с тем, что торговля осуществлялась до 11 мая 2022 года на проверочных данных я добился следующих результатов:

- Доходность: 7.60%,
- Показатель Шарпа: 0.61,
- Отношение накопленной доходности к максимальной просадке: -410

Довольно хорошая доходность и хорошая тенденция получения прибыли, в то время как очень довольно большой риск в рамках применения данной модели. Так же была довольно небольшая просадка на проверочных данных из-за чего получилось такое отношение накопленной доходности к максимальной просадке.