

Арифуллин Артур Исхакович

Эссе к вступительному заданию 2 этапа ЦМФ

27 июня 2022 г.

Реализация алгоритмической стратегии по китайским компаниям

Для реализации выбрана стратегия статистического арбитража, ключевой идеей которой является выбор портфеля пар на основе коинтеграции. Стратегия статистического арбитража — это торговая или инвестиционная стратегия, используемая для извлечения прибыли из финансовых рынков, находящихся вне состояния равновесия [1].

Временной ряд называется интегрированным порядком k и обозначается $y_t \sim I(k)$, если он и его разности до порядка $k - 1$ включительно нестационарны, а его разность порядка k стационарна [2]. Коинтеграция — это статистический признак, при котором два временных ряда, которые интегрированы порядка 1, или $I(1)$, могут быть линейно объединены для получения одного временного ряда, который является стационарным, или $I(0)$. Формальное определение коинтеграции приведено ниже:

Определение. $n \times 1$ вектор временного ряда y_t коинтегрирован при выполнении следующих условий:

- Каждый из его элементов по отдельности является стационарным
- Существует такой ненулевой вектор γ такой, что γy_t стационарен

Используемая в этой стратегии техника торговли парами основана на предположении, что линейная комбинация цен возвращается к долгосрочному равновесию, и торговое правило может быть построено с учетом ожидаемых временных отклонений. Подробнее с идеей можно ознакомиться в [1].

Пусть есть две скорректированные (эджастированные)¹ цены активов P_t^1 и P_t^2 , для их получения можем использовать коэффициент корректировки цены, данный в датасете в качестве столбца `qfq_factor`. Мы не можем использовать цену открытия или закрытия, так как в данной стратегии нам важно работать с доходностью активов, которая может быть учтена не в полной мере из-за таких событий как, например, выплата дивидендов.

¹ в некоторых материалах можно встретить слово «эджастированный» от англ. adjusted

В исходном датасете дано достаточно много активов. Требуется выбрать наиболее подходящие из них. Первый критерий, по которому будут отобраны активы- высокая ликвидность², так как в арбитражных стратегиях очень важно избегать ситуации проскальзывания, из-за которой может быть потеряна немалая часть доходности. Вторым критерием будет являться достаточно большая корреляция³ между первым и вторым активом.

Далее нам требуется подобрать такую γ , чтобы ряд $\epsilon_t = P_t^1 - \gamma P_t^2$ был стационарным, γ можем найти с помощью метода наименьших квадратов. Стационарность можем проверить с помощью теста Дики-Фуллера, уровень значимости которого в данной реализации- 5%. На обучающей выборке для каждой подходящей пары сохранены в датафрейм **result**⁴ значения γ , μ_ϵ и σ_ϵ , где μ_ϵ - выборочное среднее ϵ_t , σ_ϵ - выборочная дисперсия ϵ_t . Все вышеперечисленные вычисления были производятся на обучающей выборке.

На тестовой выборке вводится индикатор $z_t = (\epsilon_t - \mu_\epsilon)/\sigma_\epsilon$ ⁵. Будем говорить, что мы открываем позицию **long** для z_t , если мы покупаем актив 1 (с ценой P_t^1) и продаем актив 2 (с ценой P_t^2), и открываем позицию **short**, если если мы продаем покупаем актив 1 (с ценой P_t^1) и покупаем актив 2 (с ценой P_t^2). Для торговли можем придерживаться следующего правила:

$$\begin{aligned} z_t < -2 & - \text{открыть позицию long для } z_t \\ z_t > 2 & - \text{открыть позицию short для } z_t \\ z_t < 0.75 & - \text{закрыть позицию short для } z_t \\ z_t > -0.5 & - \text{закрыть позицию long, для } z_t \end{aligned}$$

Далее реализуем стратегию на тестовой выборке. Для подсчета доходностей от каждой пары используются следующие формулы:

$$\begin{aligned} r_{it}^{raw} &= \ln\left(\frac{P_t^1}{P_{t-1}^1}\right) - \gamma \ln\left(\frac{P_t^2}{P_{t-1}^2}\right) \\ R_{it} &= e^{r_{it}^{raw}} - 1 \end{aligned}$$

R_{it} - доходность стратегии для i -ой пары

² показателем ликвидности в исходном датасете является столбец turnover

³ в процессе выполнения работы изначально корреляция изначально считалась между логарифмами цен, однако это дало результаты хуже: коэффициент Шарпа не превышал 0.9

⁴ с содержанием result можно ознакомиться в прикрепленном JupyterNotebook-файле

⁵ μ_ϵ и σ_ϵ берутся из датафрейма result

Также стоит учесть транзакционные издержки. Сделать это можно следующим образом: пусть значение C описывает стоимость транзакции, мы можем скорректировать цену в день купли/продажи так, чтобы была верно подсчитана доходность, то есть если в момент времени τ была произведена покупка актива, то цена, поправленная на стоимость транзакции будет равна $P_{\tau}^{new} = (1 + C)P_{\tau}$, аналогично в случае продажи- $P_{\tau}^{new} = (1 - C)P_{\tau}$. В таком случае транзакционные издержки корректно «съедают» часть доходности.

Для каждой пары сохраняем в датафрейм df_return⁶ данные о доходностях для каждой пары, а также рассчитать и сохранить коэффициенты Шарпа SR и максимальную просадку MDD :

$$SR = \frac{\mu_R \sqrt{252}}{\sigma_R},$$

$$MDD = \sup_{t \in [0, T]} [\sup_{s \in [0, t]} R_s - R_t]$$

Далее с помощью итераций, в каждой из которых последовательно ограничиваем показатель Шарпа снизу и максимальную просадку сверху, собираем набор пар активов, состоящий из пар, торговля которыми приносит доходность с минимальной просадкой и максимальным коэффициентом Шарпа. Итоговую доходность стратегии можно рассчитать по формуле:

$$R_t = \sum_{i=1}^N \omega_i R_{it}, \text{ где } \omega_i - \text{вес каждой пары}$$

$$\sum_{i=1}^N \omega_i = 1$$

В данной реализации $\omega_1 = \omega_2 = \dots = \omega_N$

Как итог, получаем набор пар, торговля которыми на тестовой выборке, приносит прибыль с коэффициентом Шарпа - **4.67** и максимальной просадкой **10.6%**.

Финальный расчет результатов на финальной выборке показал следующие результаты:

- Доходность стратегии за указанный период - 36%
- Показатель Шарпа - 2.74
- Максимальная просадка - 12.8%

⁶ df_return представляет из себя датафрейм, где 0-ой столбец соответствует датам, т.е значению t , а i -ый столбец соответствует вектору R_{it} , это уменьшает количество лишних вычислений при подсчете показателей стратегии при торговле набором пар

Список литературы:

- [1] João F. Caldeira, Guilherme V. Moura. Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy
- [2] Идентификация коинтегрированных пар акций на фондовых рынках (<https://habr.com/ru/post/332558/>)
- [3] Использование статистических методов для анализа временных рядов (<https://habr.com/ru/post/540868/>)