

Pairs Trading

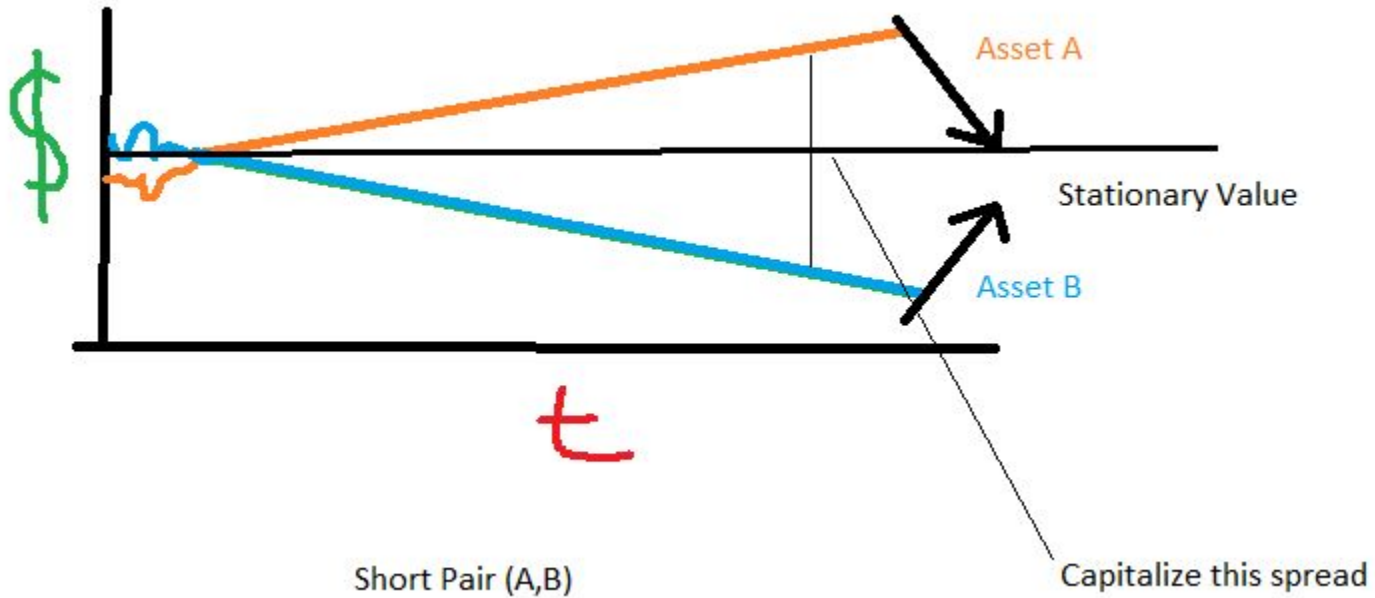
Автор: Максим Шехунов

В чем заключается стратегия?

1. Находим наиболее *связанные между собой пары акций
2. Предполагается, что ряд полученный из их отношения сильно связанных пар стационарен поэтому будет возвращаться к своему среднему
3. В момент отклонения от среднего, мы ожидаем возвращение и поэтому для нас это сигнал

* в первой версии связь будет определяться корреляцией, во второй – коинтеграцией

В чем заключается стратегия?



Данные

Нашу стратегию будем использовать на [датасете](#), в котором представлены дневные торговые данные китайских компаний.

	date	ticker	open	high	low	close	volume	outstanding_share	turnover	pe	pe_ttm	pb	ps	ps_ttm	dv_ratio	dv_ttm	total_mv	qfq_factor
0	2005-01-04	sh600000	0.77	0.77	0.75	0.76	3808939.0	9000000000.0	0.004232	17.1990	14.4219	2.0777	3.1439	2.2097	6.9549	6.9549	2693520.0	8.895254
1	2005-01-05	sh600000	0.76	0.76	0.74	0.75	5225244.0	9000000000.0	0.005806	16.9240	14.1913	2.0445	3.0937	2.1744	7.0679	7.0679	2650455.0	8.895254
2	2005-01-06	sh600000	0.75	0.75	0.73	0.74	4298099.0	9000000000.0	0.004776	16.6991	14.0026	2.0173	3.0525	2.1455	7.1632	7.1632	2615220.0	8.895254
3	2005-01-07	sh600000	0.74	0.75	0.73	0.74	4362864.0	9000000000.0	0.004848	16.7491	14.0446	2.0233	3.0617	2.1519	7.1418	7.1418	2623050.0	8.895254
4	2005-01-10	sh600000	0.75	0.77	0.74	0.77	7115260.0	9000000000.0	0.007906	17.3240	14.0575	2.0082	3.1668	2.2258	6.9048	6.9048	2713095.0	8.895254
5	2005-01-11	sh600000	0.77	0.78	0.76	0.77	6895175.0	9000000000.0	0.007661	17.5240	14.2198	2.0314	3.2033	2.2515	6.8260	6.8260	2744415.0	8.895254
6	2005-01-12	sh600000	0.78	0.78	0.76	0.77	2797449.0	9000000000.0	0.003108	17.4240	14.1386	2.0198	3.1851	2.2386	6.8651	6.8651	2728755.0	8.895254
7	2005-01-13	sh600000	0.77	0.78	0.77	0.77	1284827.0	9000000000.0	0.001428	17.3990	14.1183	2.0169	3.1805	2.2354	6.8750	6.8750	2724840.0	8.895254
8	2005-01-14	sh600000	0.77	0.78	0.76	0.76	1731048.0	9000000000.0	0.001923	17.2990	14.0372	2.0053	3.1622	2.2225	6.9147	6.9147	2709180.0	8.895254
9	2005-01-17	sh600000	0.76	0.76	0.75	0.75	2620773.0	9000000000.0	0.002912	16.9990	13.7938	1.9705	3.1074	2.1840	7.0368	7.0368	2662200.0	8.895254

Первая реализация

Отбираем пары, у которых корреляция > 0.8

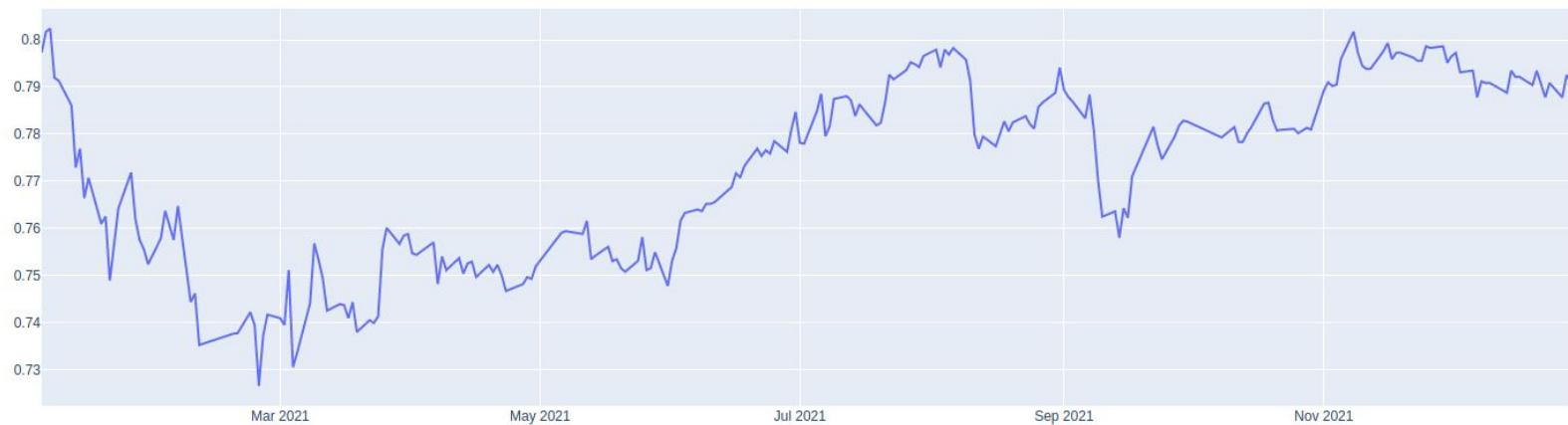


Первая реализация

Стандартизируем (standardized) отношение стоков в каждой из отобранных пар

Первая реализация

Было:



Первая реализация

Стало:



Первая реализация

Определяем уровни, при выходе за которые открываем позиции,
в данной реализации это просто гиперпараметр



Первая реализация

Определяем уровни, при возврате в которые закрываем позиции, тоже гиперпараметр



Первая реализация

Во время бэктеста будем считать стандартизированное отношение в окне

Первая реализация

Во время бэктеста будем считать стандартизированное отношение в окне.

Допустим, рассматриваем отношение стоков A/B , тогда при превышении **верхней** границы ожидаем возврат к среднему, то есть **уменьшение** отношения, поэтому **продаем A и покупаем B**

Первая реализация

Во время бэктеста будем считать стандартизированное отношение в окне.

Допустим, рассматриваем отношение стоков A/B , тогда при превышении **верхней** границы ожидаем возврат к среднему, то есть **уменьшение** отношения, поэтому **продаем A и покупаем B**.

Если отношение ниже **нижней** границы, то соответственно **покупаем A и продаем B**

Первая реализация

В каком количестве покупать?

В данной реализации я покупаю акции таким образом, чтобы выполнялись эти условия:

$$P_A |Q_A| + P_B |Q_B| = I$$

$$P_A Q_A + P_B Q_B = 0$$

Где Q_A и Q_B количество купленных акций А и В соответственно.

Первая реализация

Для большей устойчивости модели будем торговать не одной парой, а параллельно сразу несколькими.

Первая реализация

Результаты	Валидационная выборка	Тестовая выборка
Показатель Шарпа	0.080	0.068
Накоп. доходность / макс. просадка	-0.033	0.163



Выводы

- Каждый день, когда пересечен порог оставлять заявки не лучшая затея, поэтому нужно добавить анализ, чтобы более точно выбирать точку входа
- Корреляция иногда ошибается в желаемой связанности рядов, поэтому нужно попробовать другие методы, например, коинтеграцию
- Нужно протестировать использование спреда, вместо отношения двух рядов для поиска расхождений
- Подумать о других способах обработки NaN значений, вместо текущего отбрасывания рядов, где их больше 20%

Вторая реализация

В данной реализации всё то же самое, кроме того, что корреляция была дополнена коинтеграцией для лучшей фильтрации пар.

Почему коинтеграция?

Если 2 наших ряда – цены на акции коинтегрированы, то их линейная комбинация $a_t - \beta \cdot b_t$ будет стационарным рядом, то есть будет сохранять свои параметры распределения, а значит вращаться около своего среднего. Раз эта линейная комбинация стационарный ряд, то и их отношение будет возвращаться к своему среднему. Такие ряды нам и нужны.

Вторая реализация

Результаты	Валидационная выборка	Тестовая выборка
Показатель Шарпа	0.080	0.077
Накоп. доходность / макс. просадка	-0.007	0.118

