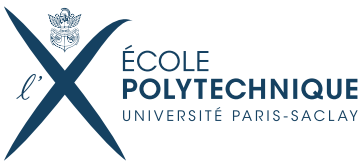
M2 Probabilité et Finance

Allocation d'actifs et arbitrage multi-assets

Project 11 : Cointégration et construction de stratégie actions

Cheng SHAN
2 mai 2020



1 introduction

In this report, we will study the paper about using cointegration relationship between index and stocks' price to build a index tracking portfolio and then extend the tracking portfolio method to build a long-short market neutral portfolio. The report is organized as follows : firstly we will comment on the paper and analyze the motivation to do linear regression under the condition of cointegration relationship and how to use this property to do index tracking in DIJA [1001[1]]. And we will present the main results of the paper. The second part is to present our simulation result in CAC 40 while applying the method presented in the paper and we finish the report by the conclusion.

2 Comments on the article

One main work of financial econometric is to find the model to represent the relationship between different economic variables. Economists have tried to apply conventional statistical methods to the data, however the economic data are rarely to satisfy the fundamental hypothesis that the data are identically independently distributed (IID), which will cause many issues in the model. Essentially, some form of least squares regressions became the well-known methods to be applied to explain economic phenomena.

We all know for using a linear regression model, we need to satisfy several conditions. Before Gauss-Markov, we understood that a linear model should have the IID residuals and the distribution needs to be a Gaussian. According to Gausse-Markov theorem, we don't necessarily need a Guassian distribution of the residuals. We just need that the residual is homoscedastic residual, of zero mean and uncorrelated. However, it's still hard for the realistic economical data to satisfy these conditions. For economists, it's still a problem about when and how to apply linear regression model. Thanks to ergodic theorem, for a long-run, a stationary process will have a similar properties as the IID ones. So it's a convention that when using linear regression model to model two economic time series, they need to be stationnary otherwise, we will have a high possibilities to have a spurious regression. Simulations presented in Granger and Newbold (1974, 1986) show that it is possible to run regressions on unrelated data and find significant relationships where there should be none.

It shows that many economical data aren't stationary and instead they are of integrated order greater than 1. So a common practice is to firstly difference a integrated time series to make it stationary and then use the stationary time series to build linear regression model. The disadvantage of this method is that by differencing a time series, we will lose a long-run information about this time series. Cointegration, concept proposed by Nobel laureate Clive Granger and Paul Newbold, has been widely applied in financial econometric to solve this problem. It has evolved as an extremely powerful statistical technique beacause it allows the application of simple estimation methods to non-stationary variables. The main advantage of cointegration analysis, as compared to the classical but rather limited concept of correlation, is that it enables the use of the entire information set comprised in level financial variables.

The paper has focused on the application of cointegration analysis to asset management. More precisely, the paper presents three applications : a classic index tracking strategy, a long-short equity market neutral strategy and a number of strategies combining index tracking and long-short market neutral.

2.1 Index tracking strategy

Firstly for the index tracking strategy, they have studied the Dow Jones Industrial Average index. They find the index price and the price of a selected bucket of stocks are cointegrated. Based on the linear regression model, they use the regression coefficients as the weights of each

stock in portfolio. They have successfully built a porfolio whose return is fairly close to that DIJA index while considering the trading costs. One important thing that they find is the way to select the stocks to construct the portfolio that will affect the final results. They have used the weights' ranking to select the most important stocks. They find that the more stocks we select to construct the portfolio, the more trading cost will we have. They have confirmed that when doing calibration of linear regression model, when the historical data that are more than one year are already enough for the work. 3 years' calibration is not significantly better than the calibration with one year's data.

We need to pay attention to our model's validity when doing linear regression. According to the cointegration's theory, we need to check is the residual of a linear regression is stationary. In this paper, they have checked the ADF test results for each regression are overall below the 1% critical value. They checked that the tracking portfolios have similar returns and volatility with the DIJA index and are highly correlated with them. And the excess return from the index tracking are uncorrelated with the markets and have low volatility and sligthly leptokurtic distribution (excess kurtosis is not zero) with positive skewness. They provide a standard for constructing a tracking portfolio. From their analysis method, we can learn how to identify a good tracking portfolio.

2.2 Long-short market neutral strategy

From the contruction of a simple index tracking strategy, a natural idea to extend this method is to replicate 'plus' and 'minus' benchmarks. It means that we build two portfolios, one tracks the market's return plus an excess return and another one tracks the market's return minus an excess return. Then by long the 'plus' portfolio and short the 'minus' portfolio, we can build the self-financing long-short strategies. This type of long-short strategy is expected to generate returns according to the spread between the 'plus' and 'minus' portfolio with a fairly low volatility. As the 'plus/minus' porfolios have the correlation coefficients with the market close to 1 which are their beta repectively. So if we combine a long-short portfolio with them, we will have a portfolio of beta zero, and it will be market neutral.

The paper finds that the success of long-short strategy highly depends on the stock selection methods. They find that all long-short strategies based on the daily re-ranking stock selection method will produce negative results after accounting for the transaction costs with one exception. Author has analyzed that the main reason for this failure comes form the transaction costs. If we re-rank the stocks each ten day, we will have a frequent changing of selected universe and so we have more fees for trading costs. On the contrary, if we choose an annual re-ranking or frequency based re-ranking stock selection methods, the total returns, net of transaction costs, tend to be positive.

By the back-test results, we can't increase the spread between the benchmarks as big as we want without a corresponding increase in the volatility and kurtosis of the returns. Another reason is that when we increase the spread, it will weaken the cointegration relationship as the reconstructed benchmarks diverge from the market index.

The number of stocks in each portfolio is also a key element of the success of long-short portfolio. In order to have a cointegration relationship, a minimum number of stocks is required in the tracking portfolio. The most successful strategy found in the back-test is the combination of 30-stocks in the 'plus' portfolio with 20 or 25-stocks in the 'minus' portfolio. Apart from these parameters, the calibration period is also another important one. A minimum of years is required to construct a cointegration relationship. Beyond this number, the difference of final results is small.

2.3 Combinations of index tracking and long-short market neutral strategies

The last section of the paper is to present how to combine different index tracking or long-short market neutral strategies together in order to enhance the properties of the strategies. The first approach is called 'fund of funds'. The idea is to invest equally into a number of similar strategies and it will reduce significantly the volatility of its returns. The second enhancement of the long-short market neutral strategy concerns mainly its performance during significantly market downturns. The last method is useful when we want to exposure to the market return. We can transport the alpha gained in the long-short market neutral framework to an index, through the use of derivatives.

To sum up , the paper has shown that using 'fund of fund' and 'portable alpha' technique, can significantly improve the characteristics of the individual index tracking and long-short market neutral strategies.

3 Simulation with CAC40

In this part, I will present my results when applying the method of this paper to CAC 40. I have mainly studied the index tracking and long-short strategies.

3.1 Data

The provided data are available from 2001-01-02 to 2010-03-09, in other words, we have totally 2396 trading days' data. We have 4o stocks' close price plus the each day's CAC40 index value. However we have missing data for several stocks :

Equity name	number of available days	number of missing days
Credit Agricole	247	2149
Electricite de France	1273	1123
Gaz de France	1177	1219
Suez	1970	426

TABLE 1 – The stocks who have missing data

The missing data for these stocks are more than one year. I have tried to find the missing historical data in the internet but unfortunately, the web-sites like Yahoo Finance don't have the data for them. So what i decide is to drop these stocks and use the remaining 36 stocks to do the study.

FIGURE 1 – Cumulative return of CAC40 and the portfolio constructed by the 36 remaining stocks according to their weights in CAC40

From the fig 1, we can see when using only 36 stocks to construct a portfolio, the return will be greater than CAC40. In the following analysis , we will still focus on these 36 stocks and CAC40, not using a reconstructed index, because according to the paper, with a smaller bucket of stocks to replicate a index is still possible and has an cointegration relationship with the index. So we will keep them.

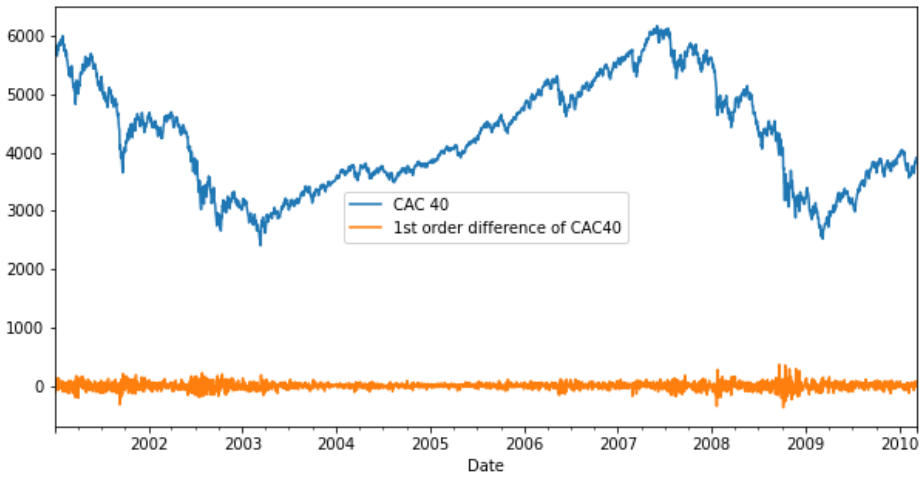


FIGURE 2 – Evolution of CAC40 from 2001 to 2010

From the fig 2, we can clearlt see that the price of CAC 40 is not stationary and it doesn't evolve around a mean. However, we can see that the first order difference of CAC40 evolve always around zero and seems to be stationary, which is the same phenomena as that of DIJA. We can do the ADF test to check our observation.

	ADF of price	ADF of first order difference of CAC40	1% cirtical value	5 % critical value
CAC40	-2.05	-11.04	-3.43	-2.86

TABLE 2 – ADF test results of CAC40 price and its first order difference

According to table2,ADF of price is bigger than 5% critical vaule, so even under 5% tolerance standard, we should reject the null hypothesis, which means price of CAC 40 is not stationary. In the contrast, we can see the ADF of the first order difference of CAC40 is smaller than 1% critical value. So we accept that it doesn't have unit root and it's stationary. So the price of CAC40 is of the integrated order 1 (I(1)). For the other 36 stocks, we observe the same results,and we will not list the detail of the results. So all the stocks' prices are not stationary and they are of integrated order 1 (I(1)). So under the classical theory, we shouldn't use directly the price of stocks to do regression with each other which will led a spurious regression, instead we should use the first order difference of price to do linear regression model. But here with cointegration theory, we can use them directly under the condition that their residual is stationary.

3.2 Index tracking

As mentioned in the paper, the motivation to construct the tracking portfolio based on a co-integration relationship with the index has three different aspects :

1. Tracking error is stationary by construction and hence mean-reverting.
2. The stocks weights in the portfolio are stable and consequently reduced amount of rebalancing and hence will reduce the trading costs.
3. We will profit from the better use of information and in particular the information contained in stock prices.

We have followed the two important stages to do an index tracking : selecting the stocks to be included in the tracking portfolio and then, determining the holding in each stock based on a cointegration optimisation technique. In our back-test procedure, I have chosen to use the simplest stock selection criterion, i.e. the price ranking of the stocks in the index at the moment of the portfolio construction. Firstly we decide how many stock we want to use. Say we want to use 20 stocks, then we order all the stocks descendingly according to their weights in the index, which is the product of the weight parameter and the price of the stock. So the composition of a given portfolio is not constant through time, as the stock ranking is based on prices that are changing.

The second stage of index tracking is to determine the portfolio holdings in each of the stocks selected in the previous stage. Same as that done in the paper, we rebalance our portfolio every ten days. Every ten days, we determine the stocks weights based on the ordinary least square (OLS) coefficients of the cointegration equation that regresses the index log price on the portfolio stocks log prices over a given calibration period prior to the portfolio's construction moment :

$$\log(index_t) = c_1 + \sum_{k=1}^{n} c_{k-1} \log(P_{k-1}) + \epsilon_t \tag{1}$$

For the calibration period, we have tried to use 1,2,3,5 years which will give four different results. The log transformation is applied to produce more homogeneous series based on the fact that if the level variables are cointegrated, so will be their logarithms. After doing the OLS, we need to check if the residual is stationary. Only if the residual is stationary, our OLS is valid. When use the estimated coefficients of OLS in the portfolio construction, we have normalised them to sum up to one and thus have provided the composition of the tracking portfolio. About how we determine the value of our constructed portfolio in the back-test, we have used the method in the paper. Assuming that the portfolio weights $w_{k,T}$ are estimated at time T, the

price of the portfolio at time $T + x, x \leq 10$, can be computed based on the prices $P_{k,T}$ and $P_{k,T+x}$ of the n stocks in the portfolio as follows :

$$\pi_{T+x} = \pi_{T-1} \sum_{k=1}^{n} \frac{w_{k,T}}{P_{k,T}} P_{k,T+x} \tag{2}$$

In the framework of our strategy, the transaction costs were incurred on each portfolio re-balancing, i.e. every 10 trading days. We assume a 20 bps of transaction costs in the back-test, so we have :

$$TC_T = 0.002 \sum_{k=1}^{n} abs(W_{k,T} - w_{k,T-10}) P_{k,T} \tag{3}$$

3.2.1 ADF test statistics

In order to ensure that the tracking portfolios were validly constructed, we need to test the residuals of each OLS regression estimated for stationarity, using the Engle-Granger methodology for testing cointegration relationships as done in the paper. We have chosen 10,20,30 stocks and used 1 year, 2 years, 3 years and 5 years calibration period to build the tracking portfolios. So we have 12 different portfolios. Their ADF tests results are shown in the following graphs.

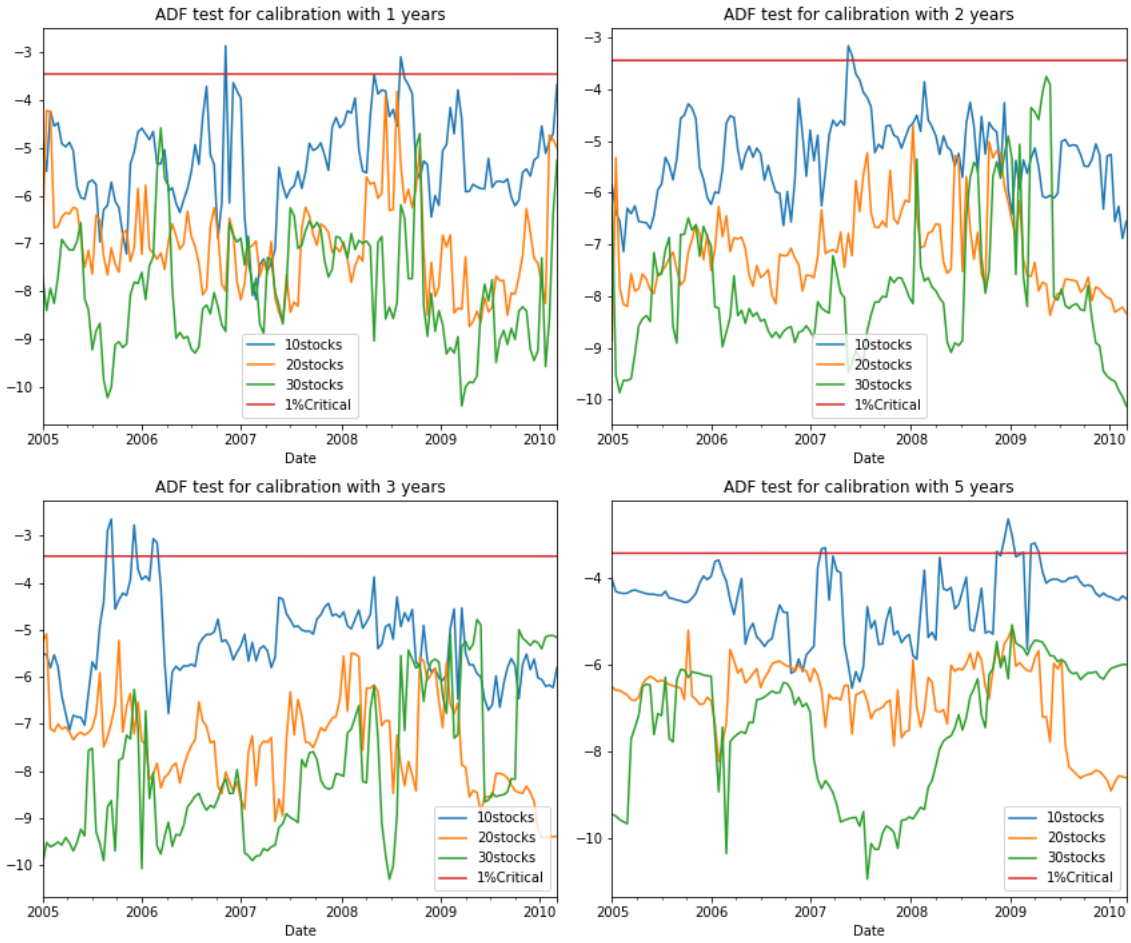


FIGURE 3 – ADF test results of each OLS for different portfolios

We can see from the fig3 that when we construct the tracking portfolio with 20,30 stocks for 1 year or 2 years or 3 years calibration period, we will 100% pass the ADF test while considering 1% critical value. And for constructing with 10 stocks, we can still pass ADF test in more than 92% cases. It means when using 20 or 30 stocks to build the portfolio, we have a strong cointegration between selected stocks and index prices. And when use 10 stocks, we can still have cointegration relationship in more than 92% cases. As expected, the degree of cointegraion increases with the number of stocks in the tracking portfolio and with the calibration period.

3.3 Returns of the tracking portfolios

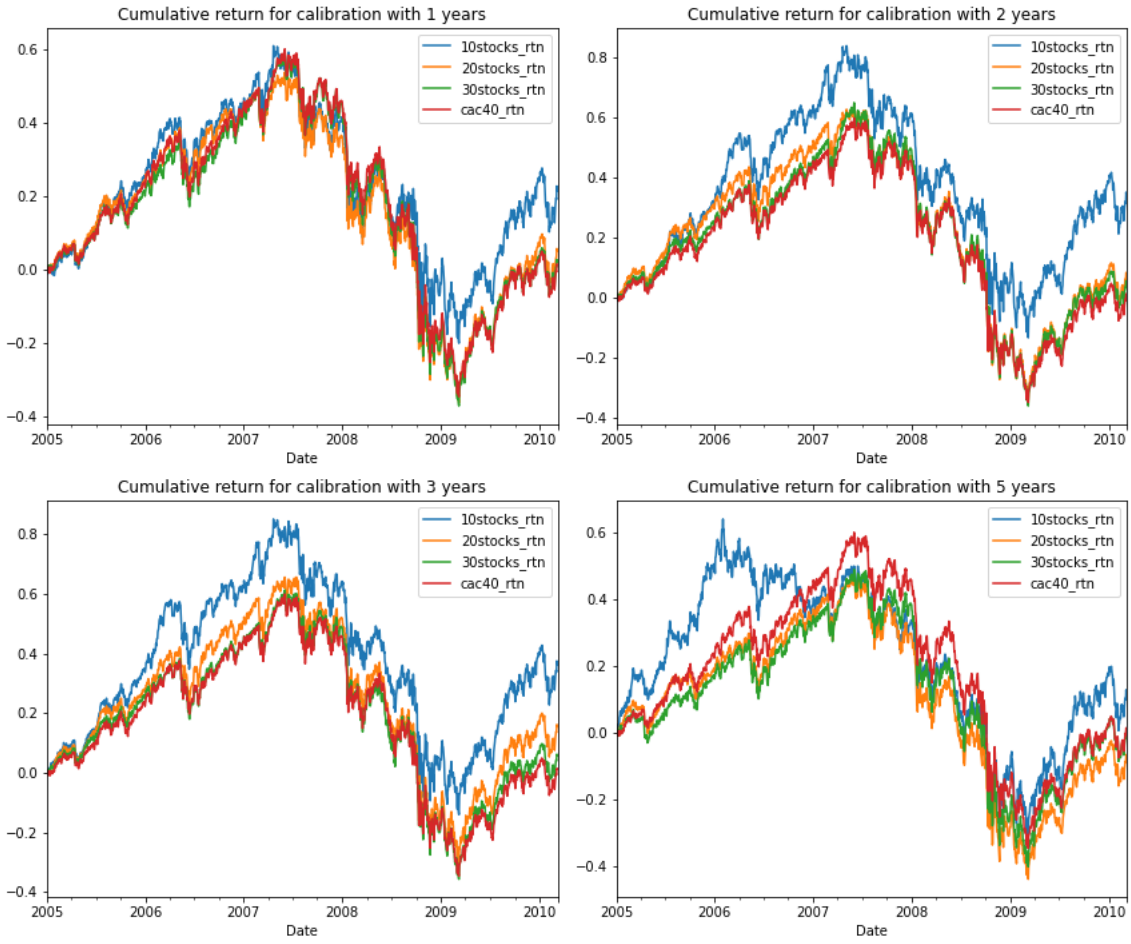


FIGURE 4 – Cumulative returns for different portfolios which have extracted trading costs

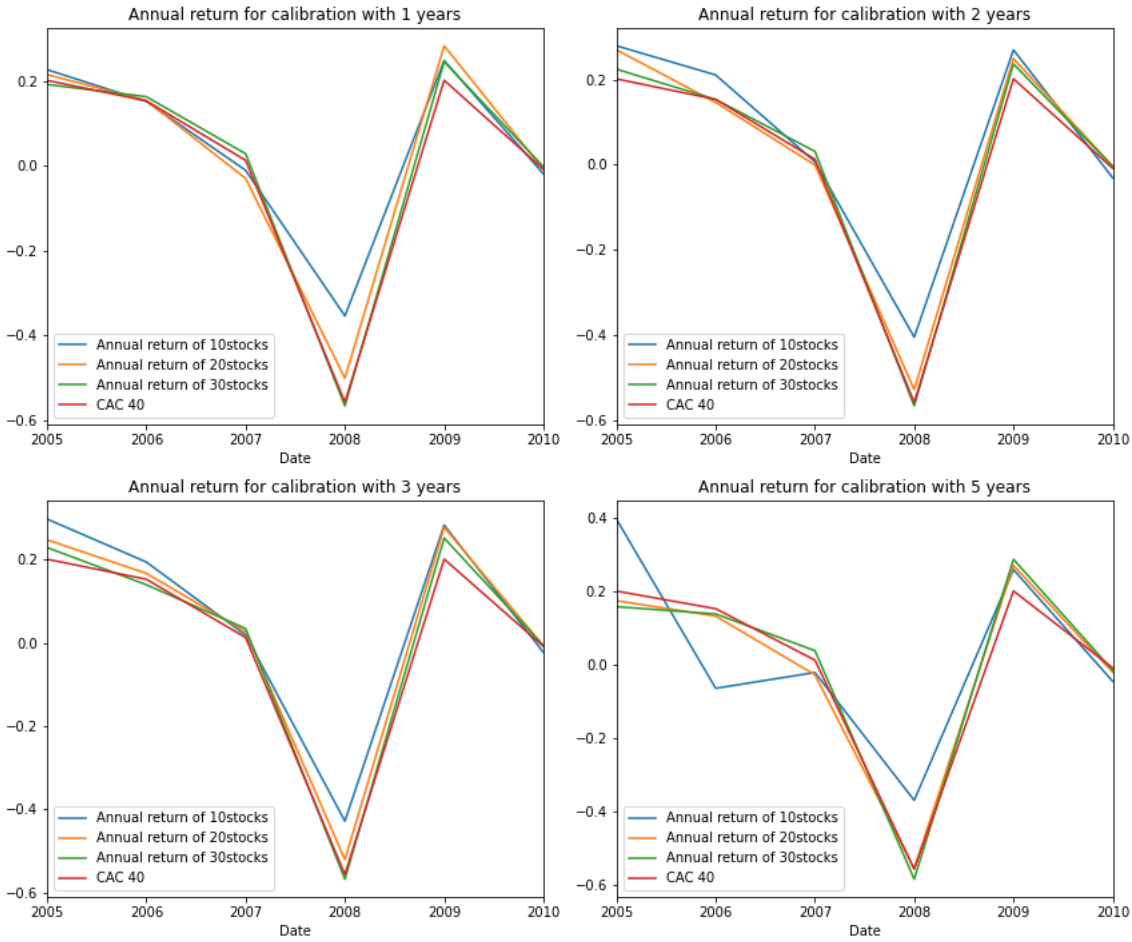


FIGURE 5 – Annual returns for different portfolios

The figure4 is our tracking results for different portfolios with different calibration years and number of selected stocks. As shown in the graph, tracking portfolio with 30 stocks is the best tracking portfolio then the second one is the tracking portfolio with 20 stocks. The portfolio with 10 stocks is far from the index CAC40's return. Another remark is that the portfolios using 1 year to calibrate the model is better to track the index.For example, when we use 5-year data to calibrate the linear model, before 2009, we have a bad tracking of the index return for all three different number of stocks. For 10-stock portfolio, we see that it has badly tracked the index which is consistent with the ADF testing result. We can conclude that, in order to better track the index, we need to have a better cointegration relationship between the index and the selected stocks. We can see the same result from the annual return graph in fig5. Even in our tracking period, the CAC40 is very volatile and CAC40 has a bad performance, it's not interesting to see the sharp ratio or information ratio of the index, we will still list the information ratio of CAC40 and for different portfolios using 1 year data to calibrate. From fig 6, we can see that by using 30 stocks, we have a very close IR as that of CAC40 index.

Date	IR_CAC40	IR_10stocks	IR_20stocks	IR_30stocks
2005	0.112855	0.113864	0.110997	0.097991
2006	0.069010	0.058516	0.061117	0.069235
2007	0.004637	-0.006882	-0.016375	0.007548
2008	-0.094624	-0.061005	-0.082268	-0.096805
2009	0.052695	0.063096	0.070688	0.061654
2010	-0.014207	-0.032892	-0.017006	-0.006331

FIGURE 6 – Information ratios for different portfolios using 1 year to calibrate

3.4 Long-short market neutral strategy

Having constructed the simple tracking strategy, according to the paper, we can construct two types of different tracking portfolios, one is the CAC40 plus one fixed annual return index and anther one is the CAC40 minus a fixed annual return. We have tried to build 10,20,30 stocks' portfolios while trying to track CAC40 plus and minus 1%, 2% or 5%. In our analysis, for simplicity, we have omitted the trading costs. The reason for this is that our portfolio is auto-financing and long-short balanced and we want to see if even we don't have trading cost, we could make a big profit.

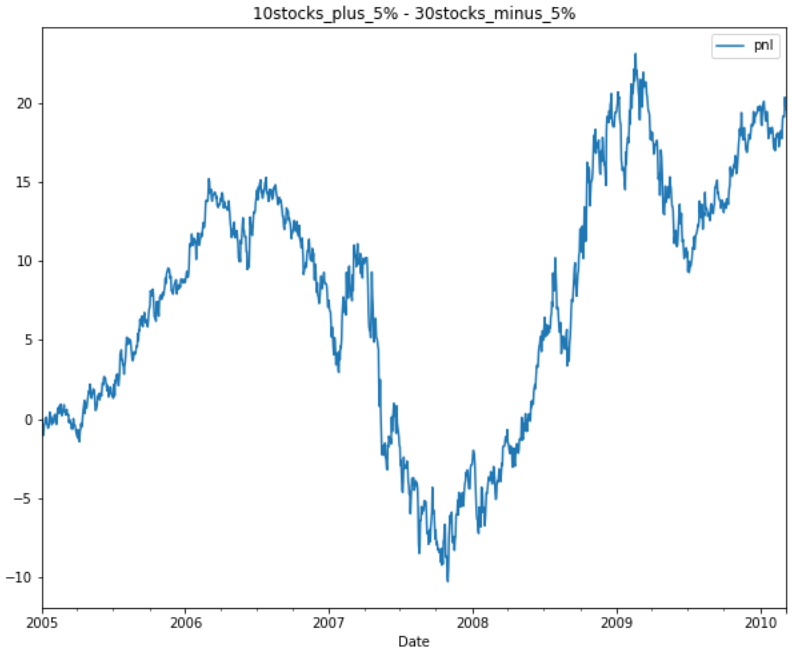


FIGURE 7 – Cumulative pnl for long-short portfolio

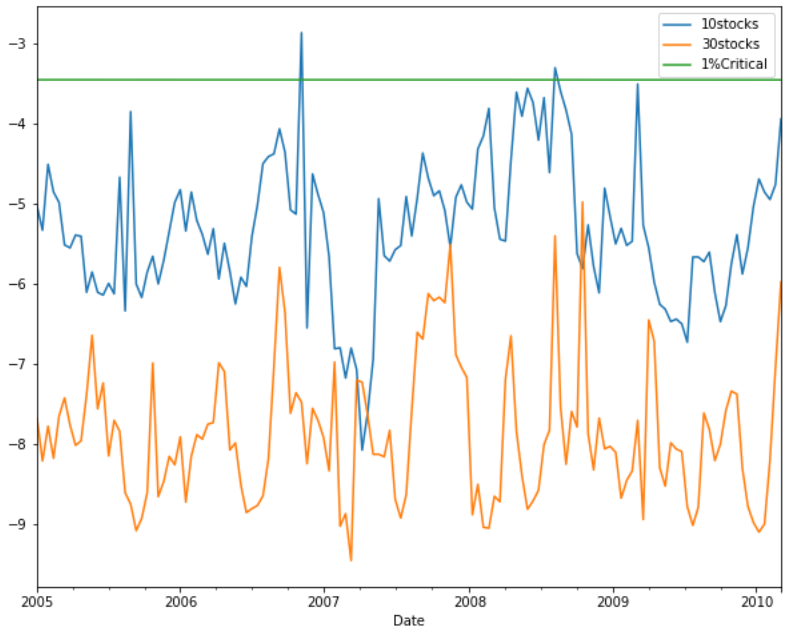


FIGURE 8 – ADF test for long and short portfolio

The combination of 10-stock portfolio to track CAC40 plus (long) 5% and the one of 30-stock to track CAC40 minus 5% (short) is the best long-short portfolio that I have found. As shown in fig8, the two portfolios have a good cointegration relationship with their target index, so it's reasonable for us to do the regression to build our portfolio. We have find the correlation between the pnl of our long-short portfolio and the one of CAC40 is 0.207, which is low and indicates the market-neutral property of our portfolio. The information ratio of this long-short portfolio is 0.023 while the information ratio of CAC40 is 0.0005. Although our portfolio doesn't really seem to be a good strategy in terms of information ratio, it is still much better than market CAC 40 and it's market neutral. The period we are focusing is very volatile and CAC 40 has a very back performance. However our portfolio can still make some profits which prove the efficience to build a long-short strategy based on cointegration strategy.

4 Conclusion

In our report, we have presented our tracking portfolio of CAC 40 by using different number of stocks and different calibration period. We find that the more stocks we use in our portfolio, the more close we will be to the index CAC 40 and the relationship of cointegration will be stronger between the selected stocks and the index. Then we present one possible long-short market neutral portfolio which has a information ratio 0.12 and it's much better than CAC40 in this back-test period. We see that when we increase the spread between the long and short portfolio, it's harder for us to have a strong cointegration relationship between stocks and index CAC40. We have oberseved the correlation of the return between the long-short portfolio and the return of CAC 40 is about 0,2 which proved the market neutral.

Références

[1] Carol Alexander and Anca Dimitriu. The cointegration alpha : Enhanced index tracking and long-short equity market neutral strategies. 2002.