ISSN 2616-5902 Vol. 3, Issue 11: 31-36, DOI: 10.25236/AJBM.2021.031106

Empirical Research on pairs trading -- Based on ICBC and Agricultural Bank of China

Ganyin Cai

University of Nottingham Ningbo China, Ningbo, Zhejiang, China

Abstract: *Based on the theory of paired trading, this paper deals with the actual transaction data of A-shares of Industrial and Commercial Bank of China and Agricultural Bank of China through empirical methods to verify the possibility of paired trading. The main conclusions of this paper are as follows: Although the time series data of ABC and ICBC contain unit root after DF test, the Engle-Grange test proves that the sequence regression of the two is co-integrated, and the relative price relationship with statistical significance can still be generated, and finally the pair trading can be realized.*

Keywords: *Pair trading; time series; Dickey-Fuller test; unit root*

1. Introduction

Pair trading is an investment strategy widely used in quantitative trading. Its core logic is that two stocks with similar fundamentals and long-term stable correlation of price fluctuation trend. Once the price of one of them deviates from the value of its long-term relationship at a certain moment, it can make profits by buying the stocks with relatively low price and selling the stocks with relatively high price. Through the empirical analysis of the time series data of ICBC and ABC, the leading bank stocks in China's A-share market, this paper verifies the applicability of the paired trading investment strategy in Chinese bank stocks. Specifically, this paper selects the daily closing prices of ICBC and Agricultural Bank of China from January 4, 2016 to December 31, 2020.Then, the Dickey-Fuller test was used to test whether the time series data had unit root, and the Engle-Granger test was used to test whether the regression of the two groups of data was co- integrated in the case of unit root. Finally, the relative value relationship of the two stocks was obtained, and the pairs trading strategy was implemented.

The test and auto regression models used in this paper are as follows. The distributed lag regression can be put into a more general model called Auto Regression (AR) model, the standard Dickey-Fuller test (1979) can be extended to test unit root in an AR (p) model,

$$y_t = c_t + \beta y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + \varepsilon_t,$$

In practice, we don't know if two unit root series has a common root., we use the Engle-Granger test for this purpose. For two unit root series, Engle and Granger proposed to run a linear regression of these two serious, then test unit root in the residuals of the regression using the ADF test statistic. The idea is that, if a common unit root exists, it will be cancelled and the residuals will have no unit root.

2. Data description

We chose the leading banks in the A-share market, Industrial and Commercial Bank of China and Agricultural Bank of China. This paper first describes the price trend (closed price) of Agricultural Bank of China and ICBC after January 4, 2016(Non-trading day automatically fills, so date to May 2019, does not affect price trends).Then take the logarithm of each stock price and make it even. Therefore, the regression coefficient of the two after taking the logarithm reflects the percentage of the impact of a unit change in the price of one stock on that of another stock. The result is shown in Figure 1.

ISSN 2616-5902 Vol. 3, Issue 11: 31-36, DOI: 10.25236/AJBM.2021.031106

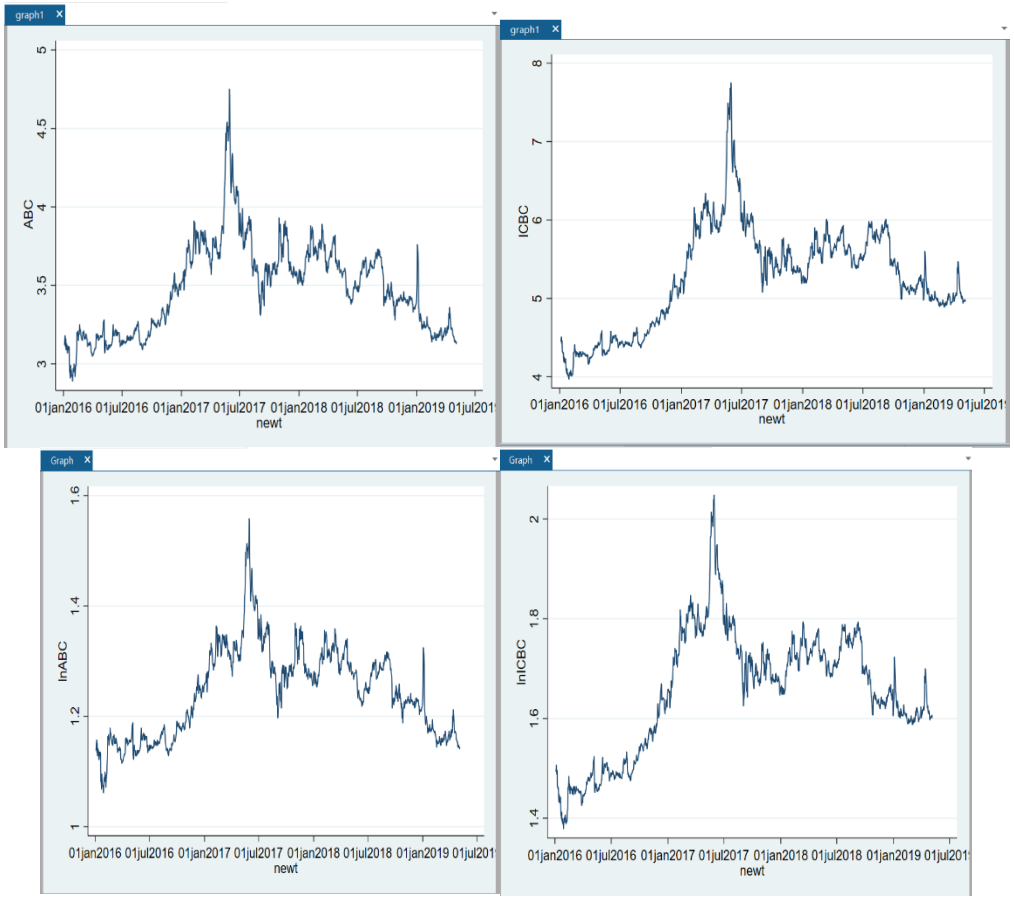

Figure 1: The price trend

As can be see, the price fluctuation trend of ICBC and Agricultural Bank of China is roughly the same, and after taking the logarithm, the price fluctuation is more stable (the ordinate price change is smaller).

3. Empirical Results and Discussion

3.1 Selection of lag order

Before performing the unit root test on the stock price time series data to verify whether it has a stochastic trend, we first select the optimal lag term. The lag order p can be selected using information criteria (IC) in the Dickey-Fuller test through the command varsoc. The code results are shown in Figure 2

```
varsoc D1.lnABC , maxlag(5)

Selection-order criteria
Sample:  10jan2016 - 05may2019                  Number of obs      =      1212
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	3766.13				.000117	-6.21308	-6.2115*	-6.20887*
1	3766.3	.34253	1	0.558	.000117	-6.21171	-6.20855	-6.2033
2	3766.34	.0746	1	0.785	.000118	-6.21013	-6.20537	-6.1975
3	3766.34	.01287	1	0.910	.000118	-6.20849	-6.20215	-6.19165
4	3772.88	13.074	1	0.000	.000117	-6.21762	-6.2097	-6.19658
5	3775.05	4.3387*	1	0.037	.000117*	-6.21955*	-6.21005	-6.19431

```
Endogenous:  D.lnABC
 Exogenous:  _cons
```

ISSN 2616-5902 Vol. 3, Issue 11: 31-36, DOI: 10.25236/AJBM.2021.031106

```
. varsoc D1.lnICBC , maxlag(5)

Selection-order criteria
Sample:  10jan2016 - 05may2019                  Number of obs      =      1212
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	3649.61				.000142	-6.02082	-6.01923	-6.01661*
1	3650.08	.93841	1	0.333	.000142	-6.01994	-6.01677	-6.01152
2	3650.51	.85018	1	0.357	.000142	-6.01899	-6.01424	-6.00637
3	3650.91	.79892	1	0.371	.000143	-6.018	-6.01166	-6.00117
4	3661.56	21.308	1	0.000	.00014	-6.03393	-6.02601	-6.01289
5	3666.19	9.2513*	1	0.002	.000139*	-6.03991*	-6.03041*	-6.01467

```
Endogenous:  D.lnICBC
 Exogenous:  _cons
```

Figure 2: Optimal lag order

According to the above measurement results, we select order 5 for the lag term of unit root test of two groups of time series data, that is, Dickey-Fuller is used to test whether there is a stochastic trend in the T-period price and T-5 period price regression in the autoregressive model.

3.2 Unit root test

After selecting the optimal lag term order, Dickey-Fuller test is used to verify whether the sequence has a unit root. The code results are shown in Figure 3.

In Dickey-Fuller test, H0 is assumed to be the existence of unit root of time series data. So in this whole analysis, we're taking a 95% confidence level. In the first table, the test statistic is equal to -1.841, which is larger than the critical value of -3.410 with a significance level of 5%, and the null hypothesis H0 cannot be rejected. In other words, the time series data has a unit root, and the price fluctuation of Agricultural Bank of China has a stochastic trend. In the second table, the test statistic is equal to -1.470, which is larger than the critical value of -3.410 with a significance level of 5%, and the null hypothesis H0 cannot be rejected. That is, the time series data has a unit root, and the price fluctuation of ICBC also has a stochastic trend.

After verifying that the time series data of the two groups of stock prices contain unit root at the same time, we used Engle-Granger test to verify whether the residual term after linear regression of the two groups of data contains unit root. If the residual term after regression does not contain unit root, then it is confirmed that the two sets of time series data contain common unit root. Such regression is co-integrated and regression parameters are valid. Firstly, we perform linear regression on the stock prices of Agricultural Bank of China and ICBC after taking logarithm, taking the stock price of Agricultural Bank of China as the dependent variable and the stock price of Industrial and Commercial Bank of China as the independent variable. The result is shown in Figure 4.

```
. dfuller lnABC , trend lags(5) regress

Augmented Dickey-Fuller test for unit root        Number of obs   =      1212
```

	Test Statistic	Interpolated Dickey-Fuller: 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-1.841	-3.960	-3.410	-3.120

```
MacKinnon approximate p-value for Z(t) = 0.6844
```

D.lnABC	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnABC						
L1.	-.0072599	.0039426	-1.84	0.066	-.014995	.0004751
LD.	.0133988	.0287768	0.47	0.642	-.0430595	.0698572
L2D.	.0120957	.0285754	0.42	0.672	-.0439675	.0681589
L3D.	.0017963	.0285286	0.06	0.950	-.054175	.0577677
L4D.	-.0992429	.0285242	-3.48	0.001	-.1552056	-.0432801
L5D.	-.0567227	.0286585	-1.98	0.048	-.1129488	-.0004966
_trend	-5.74e-07	9.15e-07	-0.63	0.531	-2.37e-06	1.22e-06
_cons	.0093986	.0048041	1.96	0.051	-.0000268	.018824

ISSN 2616-5902 Vol. 3, Issue 11: 31-36, DOI: 10.25236/AJBM.2021.031106

```
. dfuller lnICBC, trend lags(5) regress

Augmented Dickey-Fuller test for unit root        Number of obs   =      1212

                                 ------------ Interpolated Dickey-Fuller ------------
                    Test         1% Critical       5% Critical      10% Critical
                 Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(t)             -1.470            -3.960            -3.410           -3.120
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.8392

------------------------------------------------------------------------------
    D.lnICBC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lnICBC |
         L1. |  -.0045745   .0031122    -1.47   0.142    -.0106805    .0015315
         LD. |   .0197005   .0286572     0.69   0.492    -.0365232    .0759242
        L2D. |   .0336505   .0283996     1.18   0.236    -.0220677    .0893688
        L3D. |   .0333822   .0283821     1.18   0.240    -.0223017     .089066
        L4D. |  -.1271877   .0283764    -4.48   0.000    -.1828604   -.0715151
        L5D. |  -.0852511   .0286118    -2.98   0.003    -.1413856   -.0291166
      _trend |  -3.33e-07   1.10e-06    -0.30   0.762    -2.49e-06    1.82e-06
       _cons |    .007903   .0048798     1.62   0.106     -.001671    .0174769
------------------------------------------------------------------------------
```

Figure 3: Dickey-Fuller test

```
. regress lnABC lnICBC

      Source |       SS           df       MS      Number of obs   =     1,218
-------------+----------------------------------   F(1, 1216)      =   9834.80
       Model |  7.29538961         1  7.29538961   Prob > F        =    0.0000
    Residual |  .902020571     1,216  .000741793   R-squared       =    0.8900
-------------+----------------------------------   Adj R-squared   =    0.8899
       Total |  8.19741018     1,217  .006735752   Root MSE        =    .02724

------------------------------------------------------------------------------
       lnABC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lnICBC |   .6243635   .0062959    99.17   0.000     .6120115    .6367154
       _cons |   .2112739   .0104424    20.23   0.000     .1907869     .231761
------------------------------------------------------------------------------
```

Figure 4: Regression results between ICBC and ABC

We use the varsoc command again to select the optimal lag term for the unit root test of the residual term after linear regression. The result is shown in Figure 5.

```
. varsoc D1.e1, maxlag(5)

Selection-order criteria
Sample:  10jan2016 - 05may2019                  Number of obs      =      1212

+---------------------------------------------------------------------------+
| lag |    LL      LR      df    p      FPE       AIC       HQIC      SBIC  |
|-----+---------------------------------------------------------------------|
|  0  |  4332.04                        .000046  -7.14693  -7.14535  -7.14272* |
|  1  |  4333.72  3.3526    1  0.067   .000046  -7.14805  -7.14488  -7.13963 |
|  2  |  4337.89  8.3559    1  0.004   .000046  -7.15329  -7.14854* -7.14067 |
|  3  |  4338.73  1.6724    1  0.196   .000046  -7.15302  -7.14668  -7.13619 |
|  4  |  4341.24  5.0117*   1  0.025   .000046   -7.1555  -7.14758  -7.13446 |
|  5  |  4343.06  3.6418    1  0.056   .000046* -7.15686* -7.14735  -7.13161 |
+---------------------------------------------------------------------------+
Endogenous:  D.e1
 Exogenous:  _cons
```

Figure 5: Optimal lag order

Based on the above information, 0 order lag term is selected for Dickey-Fuller test. Then we

ISSN 2616-5902 Vol. 3, Issue 11: 31-36, DOI: 10.25236/AJBM.2021.031106

conduct DF test for the residual term of the linear regression of stock prices of Agricultural Bank of China and Industrial and Commercial Bank of China after taking logarithms. Results and standards are shown in Figure 6 and Figure 7 respectively.

```
. dfuller e1, regress lags(0)

Dickey-Fuller test for unit root                   Number of obs   =      1217

                                 ------------ Interpolated Dickey-Fuller ------------
                 Test         1% Critical       5% Critical      10% Critical
               Statistic         Value             Value            Value
------------------------------------------------------------------------------
 Z(t)            -4.000           -3.430            -2.860           -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0014

------------------------------------------------------------------------------
        D.e1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        e1 |
         L1. |   -.0285244   .0071304    -4.00   0.000    -.0425136   -.0145351
             |
       _cons |   -.0000519   .0001936    -0.27   0.789    -.0004316    .0003279
------------------------------------------------------------------------------
```

Figure 6: Dickey-Fuller test

Significance level	1%	2.5%	5%	10%
Critical value	-3.90	-3.59	-3.34	-3.04

Asymptotic Critical Values for Cointegration Test: No Time Trend

Figure 7: Critical Values

The test statistic of the residual term after linear regression for unit root test is equal to -4.000.At this point, when we choose whether to reject the null hypothesis, the critical value to be compared is that of the Engle-Granger test. Since there is no time trend item in the autoregressive model above, the key value table used does not contain time trend item. In the Engle-Granger test, we still take the 95% confidence level, that is, the key value under the significance degree of 5% is equal to -2.860.-4.000 is less than -2.860. Therefore, we can reject the null hypothesis, that is, the residual term after the regression of two groups of time series data does not contain unit root, so the regression is co-integrated. Based on this conclusion, the regression parameters of price fluctuation trend of Agricultural Bank of China and Industrial and Commercial Bank of China are meaningful. Once again, we perform linear regression on the stock prices of Agricultural Bank of China and Industrial and Commercial Bank of China after taking the logarithm to obtain the specific relationship between their relative value fluctuations. The regression results are shown in Figure 8

It can be seen from the above table that R squared =0.89 and the adjusted R squared =0.8899. The regression model selects independent variables with a high explanatory power, that is, the price changes of ICBC can explain the price changes of Agricultural Bank of China to a large extent. The test statistic for the slope coefficient is 99.17, and the test statistic for the intercept coefficient is 20.13, far exceeding the critical value at the 95% confidence level. In other words, the parameter is valid instead of 0. Where, the slope coefficient is approximately 0.6244. Since both the dependent variable and the independent variable are the stock price after taking the logarithm, the absolute price level is converted into the growth rate. This means that in the relative price relationship between AGRICULTURAL Bank of China and ICBC, the price of Agricultural Bank of China rises by 0.6244% for every 1% increase in ICBC. Based on this relative price relationship, we can theoretically realize the corresponding

ISSN 2616-5902 Vol. 3, Issue 11: 31-36, DOI: 10.25236/AJBM.2021.031106

matching transaction. The specific operation is as follows: after the price of ICBC rises, THE price of ABC does not rise to the corresponding extent. At this time, it is believed that the price of ABC is undervalued compared with THAT of ICBC, but in the long run, the relative price of the two will return to a stable price relationship. At this point long agricultural Bank while selling ICBC, theoretically can make a profit; When ICBC declines, the decline of AGRICULTURAL Bank exceeds the extent of its regression relationship. Therefore, it is believed that Agricultural Bank is undervalued compared with ICBC, so it can also buy Agricultural Bank and sell ICBC and make a profit. Both of the above situations can achieve the success of the paired trading strategy.

```
. regress lnABC lnICBC
```

Source	SS	df	MS	Number of obs	=	1,218
				F(1, 1216)	=	9834.80
Model	7.29538961	1	7.29538961	Prob > F	=	0.0000
Residual	.902020571	1,216	.000741793	R-squared	=	0.8900
				Adj R-squared	=	0.8899
Total	8.19741018	1,217	.006735752	Root MSE	=	.02724

lnABC	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnICBC	.6243635	.0062959	99.17	0.000	.6120115	.6367154
_cons	.2112739	.0104424	20.23	0.000	.1907869	.231761

Figure 8: Regression results

4. Conclusion

Although the time series data of Agricultural Bank of China and Industrial and Commercial Bank of China contain unit root after DF test, that is, there is a stochastic trend. However, the Engle-Grange test proves that the sequence regression of the two is co-integrated, and the relative price relationship with statistical significance can still be generated, and finally the pair trading can be realized.

References

[1] Friedrich W(2020). Köster, Bastian Huwer, Gerd Kraus et al. Egg production methods applied to Eastern Baltic cod provide indices of spawning stock dynamics[J] Fisheries Research, p227
[2] Steven X(2020). Cadrin Spatial Structure: Theory, estimation and application in stock assessment models[J] Fisheries Research, p229
[3] Sai Liang, Yafei Wang, Tianzhu Zhang et al(2017). Structural analysis of material flows in China based on physical and monetary input-output models[J] Journal of Cleaner Production, p158
[4] Distribution of the estimators for autoregressive time series with a unit root. Dickey DA, Fuller WA. Journal of the American Statistical Association. 1979
[5] Cointegration and Error Correction: Representation, Estimation and Testing. Engle RF, Granger CWJ. Econometrica . 1987