Machine learning methods combination to solve the problem of predicting the cryptocurrency quotes behavior

Elena V. Alymova, Oleg E. Kudryavtsev

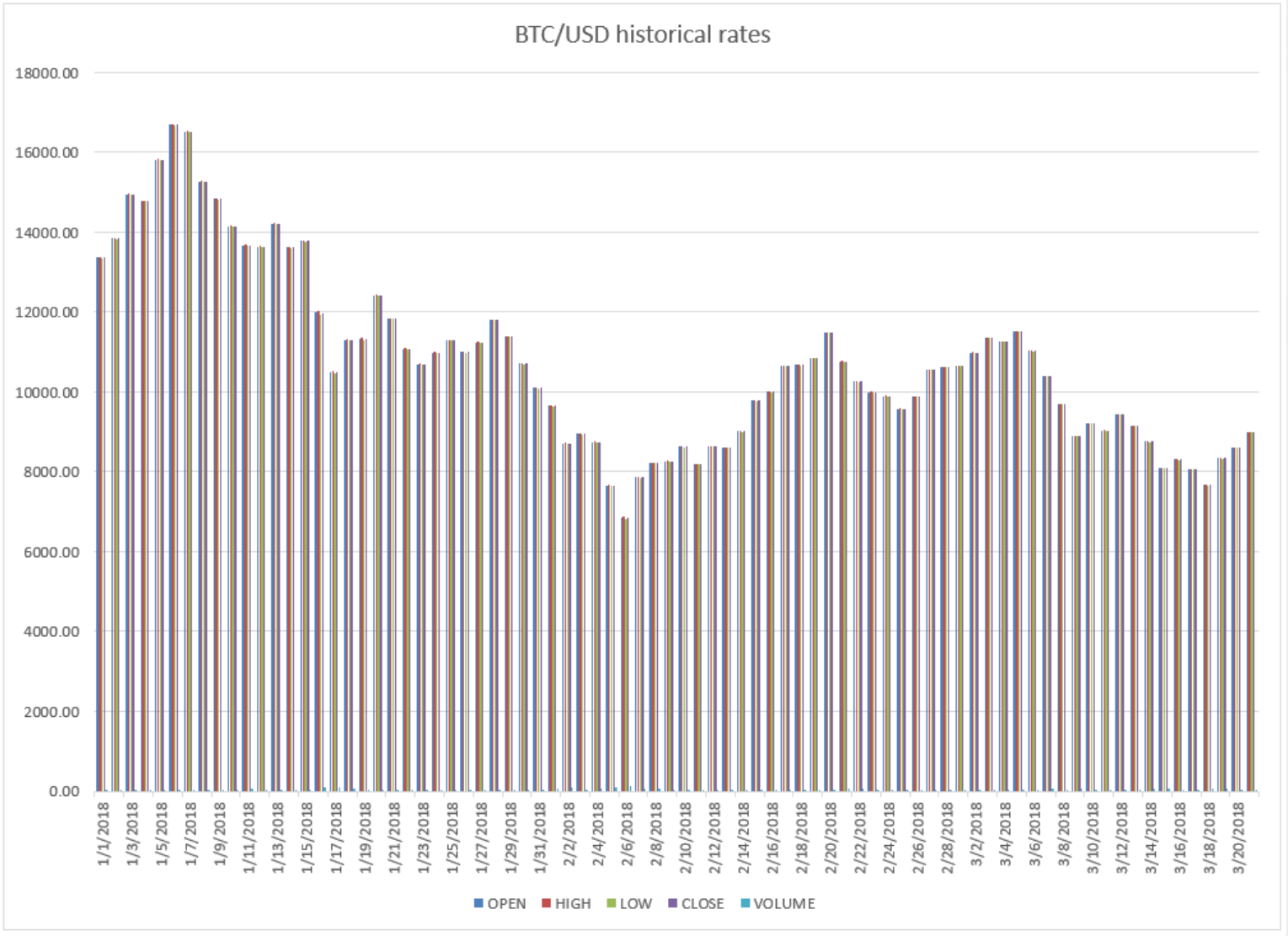
Russian Customs Academy, Rostov-on-Don

5th International Conference on Stochastic Methods
23–27 November 2020, Moscow , Russia

Research goals

- Combine machine learning methods in predicting the behavior of cryptocurrency quotes.
- Develop a methodology for a reasonable choice of a trading strategy on the currency exchange in the next ten-minute period, based on the data of the current twenty-minute trading period.
- Build a multilayer feed-forward neural network as a trend prediction model.
- Evaluate the applicability of the proposed methodology.

Initial data set: the real-time historical trading data of the BTC / USD currency pair with a minute interval



Extending the initial data set with logarithmic returns

logarithmic return at the i-th minute ($i \geq 1$):

$$LOGRET_i = ln\left(\frac{CLOSE_i}{CLOSE_{i-1}}\right)$$

	A	B	C	D	E	F	G
1	OPEN	HIGH	LOW	CLOSE	VOLUME	PERIOD	LOG RETURN
2	966.61	966.61	966.61	966.61	1		
3	966.62	966.63	966.62	966.63	1	1	=LN(D3/D2)
4	967.05	967.06	967.05	967.06	10	2	=LN(D4/D3)
5	967.37	967.4	967.37	967.39	9	3	=LN(D5/D4)
6	967.39	967.39	967.39	967.39	0	4	=LN(D6/D5)
7	967.55	967.79	967.55	967.79	2	5	=LN(D7/D6)
8	967.46	967.46	967.45	967.45	0	6	=LN(D8/D7)
9	967.34	967.34	966.77	966.77	0	7	=LN(D9/D8)
10	965.79	965.79	965.79	965.79	0	8	=LN(D10/D9)
11	965.79	965.79	965.79	965.79	0	9	=LN(D11/D10)
12	965.73	965.73	965.03	965.03	0	10	=LN(D12/D11)
13	965.03	965.03	965.03	965.03	0	11	=LN(D13/D12)
14	965.03	965.03	965.02	965.02	1	12	=LN(D14/D13)
15	965.01	965.36	965	965.36	11	13	=LN(D15/D14)
16	965.3	965.3	965.05	965.05	1	14	=LN(D16/D15)
17	965.05	965.83	965.05	965.83	10	15	=LN(D17/D16)
18	965.11	965.11	965.11	965.11	0	16	=LN(D18/D17)
19	965.11	965.11	965.11	965.11	0	17	=LN(D19/D18)
20	965.12	966.48	965.12	966.47	10	18	=LN(D20/D19)
21	966.46	966.46	966.46	966.46	0	19	=LN(D21/D20)
22	966.41	967.05	966.41	967.05	10	20	=LN(D22/D21)

Assumption

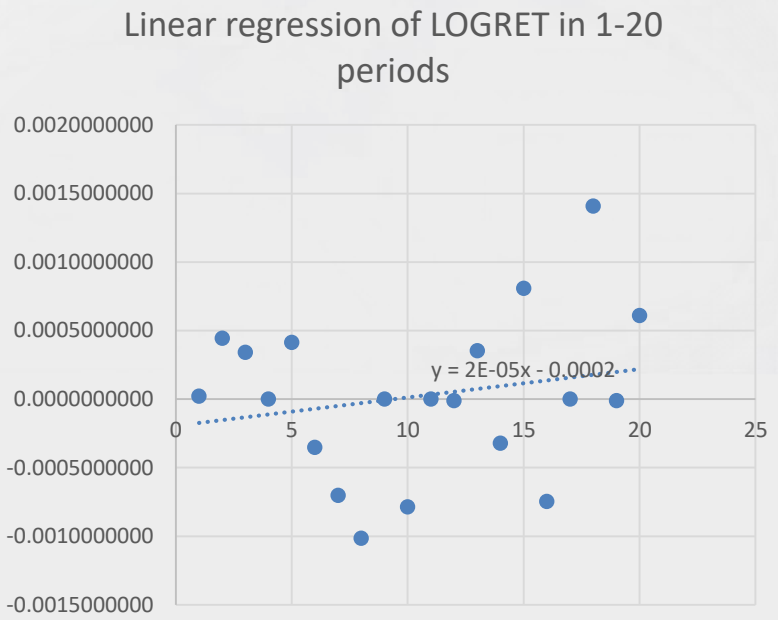
The indicators of the logarithmic return in the current twenty-minute period determine the logarithmic return indicators in the next ten-minute period

Proving the assumption:

For every twenty values of the logarithmic return determine the linear trend direction of the target currency pair quotes behavior in the next ten minutes

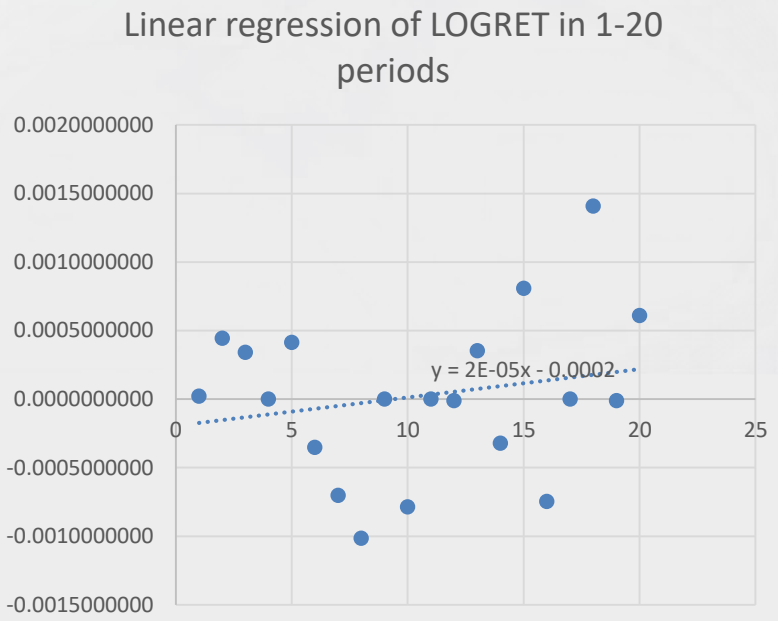
Calculating the linear trend direction

	D	G	H	I
1	CLOSE	LOG RETURN	LIN REG COEF B	B SIGN
2	966.61			
3	966.63	=LN(D3/D2)		
4	967.06	=LN(D4/D3)		
5	967.39	=LN(D5/D4)		
6	967.39	=LN(D6/D5)		
7	967.79	=LN(D7/D6)		
8	967.45	=LN(D8/D7)		
9	966.77	=LN(D9/D8)		
10	965.79	=LN(D10/D9)		
11	965.79	=LN(D11/D10)		
12	965.03	=LN(D12/D11)	=LINEST(G3:G12,,1,1)	=IF(H12 >= 0,1,0)
13	965.03	=LN(D13/D12)	=LINEST(G4:G13,,1,1)	=IF(H13 >= 0,1,0)
14	965.02	=LN(D14/D13)	=LINEST(G5:G14,,1,1)	=IF(H14 >= 0,1,0)
15	965.36	=LN(D15/D14)	=LINEST(G6:G15,,1,1)	=IF(H15 >= 0,1,0)
16	965.05	=LN(D16/D15)	=LINEST(G7:G16,,1,1)	=IF(H16 >= 0,1,0)
17	965.83	=LN(D17/D16)	=LINEST(G8:G17,,1,1)	=IF(H17 >= 0,1,0)
18	965.11	=LN(D18/D17)	=LINEST(G9:G18,,1,1)	=IF(H18 >= 0,1,0)
19	965.11	=LN(D19/D18)	=LINEST(G10:G19,,1,1)	=IF(H19 >= 0,1,0)
20	966.47	=LN(D20/D19)	=LINEST(G11:G20,,1,1)	=IF(H20 >= 0,1,0)
21	966.46	=LN(D21/D20)	=LINEST(G12:G21,,1,1)	=IF(H21 >= 0,1,0)
22	967.05	=LN(D22/D21)	=LINEST(G13:G22,,1,1)	=IF(H22 >= 0,1,0)



Calculating the linear trend direction

	B	C	D	F	G	H	I
1	HIGH	LOW	CLOSE	PERIOD	LOG RETURN	LIN REG COEF B	B SIGN
12	965.73	965.03	965.03	10	-0.0007872303	-0.000121428	0
13	965.03	965.03	965.03	11	0.0000000000	-0.000100219	0
14	965.03	965.02	965.02	12	-0.0000103624	-0.000051054	0
15	965.36	965.00	965.36	13	0.0003522623	0.000016502	1
16	965.30	965.05	965.05	14	-0.0003211753	0.000024445	1
17	965.83	965.05	965.83	15	0.0008079218	0.000125429	1
18	965.11	965.11	965.11	16	-0.0007457508	0.000085900	1
19	965.11	965.11	965.11	17	0.0000000000	0.000068378	1
20	966.48	965.12	966.47	18	0.0014081739	0.000098405	1
21	966.46	966.46	966.46	19	-0.0000103470	0.000089309	1
22	967.05	966.41	967.05	20	0.0006102891	0.000061710	1
23	967.04	967.04	967.04	21	-0.0000103408	0.000035800	1
24	966.81	966.81	966.81	22	-0.0002378675	-0.000003085	0
25	965.43	965.42	965.42	23	-0.0014387524	-0.000080541	0
26	966.80	966.35	966.35	24	0.0009628476	-0.000050187	0
27	967.68	966.01	966.01	25	-0.0003519013	-0.000031838	0
28	966.73	966.03	966.03	26	0.0000207035	-0.000082684	0
29	966.71	966.71	966.71	27	0.0007036643	-0.000055852	0
30	967.00	966.71	967.00	28	0.0002999416	0.000034309	1
31	967.00	967.00	967.00	29	0.0000000000	0.000026974	1
32	967.00	967.00	967.00	30	0.0000000000	0.00006089	1



Trading strategies based on the coefficient B

strategy 1: if at the end of the current twenty-minute period the coefficient $B \geq 0$, then in the first minute of the next ten-minute period the cryptocurrency should be bought at the current price to sell at the first higher price in the next nine minutes;

strategy -1: if at the end of the current twenty-minute period the coefficient $B < 0$, then in the first minute of the next ten-minute period the cryptocurrency should be sold at the current price to buy at the first lower price in the next nine minutes.

Checking the effectiveness of strategies

if the coefficient $B \geq 0$ in the first minute of the next ten-minute period, then the price at the open should be less than the maximum cost in the next ten-minute period. In this case, strategy 1 is considering to be effective;

if the coefficient $B < 0$ in the first minute of the next ten-minute period, then the price at the close should be higher than the minimum cost in the next ten-minute period. In this case, strategy -1 is considering to be effective.

Checking the formulated criteria of effectiveness on historical data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	TIME	OPEN	HIGH	LOW	CLOSE	VOLUME	PERIOD	LOG RETURN	SPAN	LIN REG COEF B	B SIGN	MAX 21-29	MIN 21-29	FACT STRATEGY
22	0:31:00	966.41	967.05	966.41	967.05	10	20	0.0006102891	11 - 20	0.000061710	1	967.68	965.42	1
23	0:32:00	967.04	967.04	967.04	967.04	1	21	-0.0000103408	12 - 21	0.000035800	1	967.68	965.42	1
24	0:35:00	966.81	966.81	966.81	966.81	0	22	-0.0002378675	13 - 22	-0.000003085	1	967.68	965.42	1
25	0:36:00	965.43	965.43	965.42	965.42	0	23	-0.0014387524	14 - 23	-0.000080541	1	967.68	965.42	1
26	0:38:00	966.80	966.80	966.35	966.35	3	24	0.0009628476	15 - 24	-0.000050187	0	967.68	966.01	1
27	0:41:00	966.81	967.68	966.01	966.01	5	25	-0.0003519013	16 - 25	-0.000031838	0	967.68	966.01	1
28	0:50:00	966.63	966.73	966.03	966.03	2	26	0.0000207035	17 - 26	-0.000082684	0	967.04	966.03	1
29	0:54:00	966.71	966.71	966.71	966.71	0	27	0.0007036643	18 - 27	-0.000055852	0	967.20	966.71	0
30	0:57:00	966.71	967.00	966.71	967.00	3	28	0.0002999416	19 - 28	0.000034309	0	967.20	966.71	0
31	0:58:00	967.00	967.00	967.00	967.00	5	29	0.0000000000	20 - 29	0.000026974	1	967.20	967.00	1
32	0:59:00	967.00	967.00	967.00	967.00	1	30	0.0000000000	21 - 30	0.00006089	1	967.25	967.00	1
33	1:00:00	967.00	967.00	967.00	967.00	1	31	0.0000000000	22 - 31	0.000060826	1	967.25	967.00	1
34	1:03:00	967.04	967.04	967.04	967.04	0	32	0.0000413642	23 - 32	0.000047726	0	967.25	967.00	1
35	1:04:00	967.00	967.00	967.00	967.00	0	33	-0.0000413642	24 - 33	-0.000053330	0	967.25	967.00	0
36	1:07:00	967.00	967.00	967.00	967.00	1	34	0.0000000000	25 - 34	-0.000008962	0	967.25	966.10	1
37	1:08:00	967.00	967.00	967.00	967.00	2	35	0.0000000000	26 - 35	-0.000040572	0	967.25	966.10	1
38	1:10:00	967.20	967.20	967.20	967.20	2	36	0.0002068038	27 - 36	-0.000040328	0	967.25	966.01	1
39	1:14:00	967.20	967.20	967.20	967.20	0	37	0.0000000000	28 - 37	-0.000008088	0	967.25	966.01	1
40	1:15:00	967.20	967.20	967.20	967.20	3	38	0.0000000000	29 - 38	0.000005765	0	967.25	966.00	1
41	1:18:00	967.25	967.25	967.25	967.25	0	39	0.0000516943	30 - 39	0.000006078	0	967.25	965.50	1
42	1:19:00	967.21	967.21	967.21	967.21	0	40	-0.0000413552	31 - 40	0.000000689	1	967.21	965.50	0
43	1:22:00	967.21	967.21	967.00	967.00	2	41	-0.0002171429	32 - 41	-0.000013787	1	967.21	965.50	0
44	1:27:00	967.00	967.00	967.00	967.00	0	42	0.0000000000	33 - 42	-0.000011029	1	967.00	965.50	0

On the historical data for every five successful strategies, there is one ineffective one

Strategy 0 (wait strategy) to minimize the number of unprofitable decisions

Strategy 0 is chosen, when the value of the coefficient B in the first minute of the next ten-minute period is insignificant.

Insignificant means that the coefficient B modulo is less than the specified parameter h.

In the current research, the value selection of the parameter h was performed empirically on historical data and is equal to 0.0003, which made it possible to increase the number of profitable strategies.

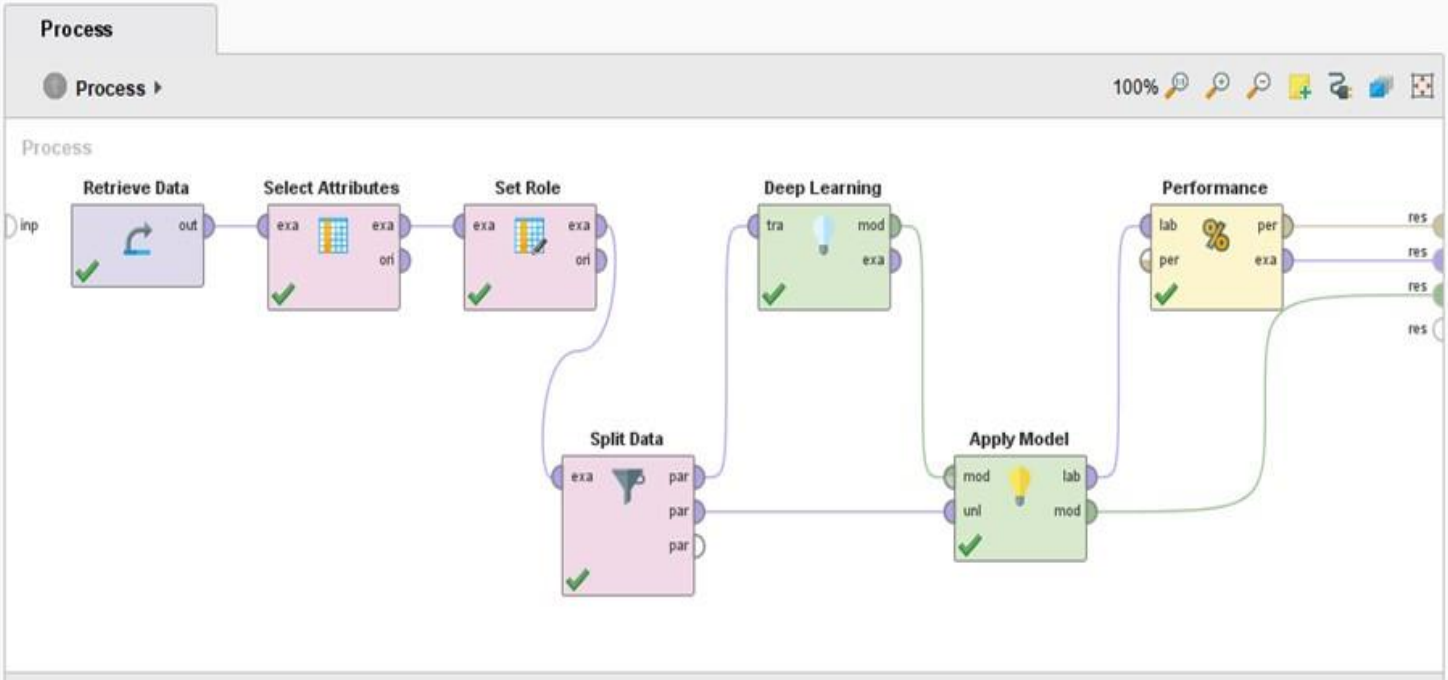
The wrong decision to buy or sell cryptocurrency is making in every eighth case when strategy 0 is using as a third alternative of behavior on the trade market.

Neural network train data

- $X_n = \{LOGRET_{n+1}, LOGRET_{n+2}, ..., LOGRET_{n+20}\}$;
- $Y_n = \{LINREGB_{n+21,n+30}\}$, where
- $LOGRET_i$ – logarithmic return at the i-th minute ($i \geq 1$);
- $LINREGB_{j,k}$ — coefficient B of the linear regression equation based on logarithmic returns $LOGRET_j, ..., LOGRET_k, (k > j)$;
- n — number of the observed period ($n \geq 0$).

If predicted value LINREG$B \geq 0$, then the price increase of the cryptocurrency is predicted (signal to buy), otherwise the price of the cryptocurrency is predicted to fall (signal to sell)

Neural network in Rapid Miner



Deep Learning is based on a multi-layer feed-forward artificial neural network that is trained with stochastic gradient descent using back-propagation.

root_mean_squared_error

```
root_mean_squared_error: 0.000089048 +/- 0.000000000
```

Results of prediction

Checking the effectiveness criteria of strategies on the data predicted by the neural network shows that for every three effective strategies there is one ineffective strategy.

A decrease in the number of effective strategies is a consequence of predicting the linear regression coefficient value accuracy.

The forecasting accuracy of the implemented and trained neural network is 80%.

Conclusion

A methodology for choosing a strategy for trading on the currency exchange in the next ten-minute period, based on the data of the current twenty-minute trading period is proposed.

There are three strategies of trading activities: buying at the beginning of the next period to sell at the first higher price (strategy 1), sell at the beginning of the next period to buy at the first lower price (strategy -1). The third strategy is inaction (strategy 0) when the value of the linear regression coefficient at the beginning of the next ten-minute period is insignificant for the chosen criterion.

Conclusion

On historical data every sixth decision in choosing a strategy leads to losses without introducing strategy 0, and when strategy 0 is applying, every eighth decision is unprofitable.

When using predicted values of linear regression coefficients, every fourth chosen strategy is unsuccessful.

The proposed approach effectiveness depends on predicting the linear regression coefficient accuracy and the parameter h choice of the coefficient significance.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research under Grant No 18-01-00910