Study on Intraday Momentum of Chinese Stock Market Based on R and Multiple Linear Regression Models

Zejian Deng[1, a] Yuchan Liu[1,b]

[a]e-mail: 1134601752@qq.com, [b]e-mail: 243230184@qq.com

School of Economics and Management, NJUST, Nanjing, Jiangsu, China[1]

Abstract-There has been much research on momentum, but most of them focused on the monthly or weekly frequency. Did the momentum also exist at the intraday level? To answer the question, we used R and multiple linear regression models to analyze the high-frequency trading data of the Chinese stock market and found that the returns of the first and seventh half-hour could significantly predict the returns of the last half-hour both in and out of the sample. Moreover, from the perspective of asset allocation and market timing, this intraday momentum has yielded considerable economic gains. In addition, a series of robustness tests were carried out to prove that intraday momentum was not an accidental phenomenon.

Keywords: R; Multiple Linear Regression; Stock Return Forecast; Investment

1. INTRODUCTION

Jegadeesh and Titman (1993) [1] identified a well-known momentum strategy that buys past winners and sells past losers, generating significant positive returns over the 3-12 month holding period. In contrast to this cross-sectional momentum, Moskowitz (2012) [2] and Neely (2014) [3] show the time series momentum of the monthly frequency of stock returns. Sun (2016) [4] and Renault (2017) [5] found that based on intraday momentum, high-frequency investor sentiment could predict intraday stock returns. In addition, a recent study by Gao (2018) [6] demonstrates that the first half-hour returns of the US stock market are positively predicting the last half-hour returns. Similarly, this paper studies the predictability of intraday momentum returns in Chinese stock markets.

ICEMME 2022, November 18-20, Nanjing, People's Republic of China

DOI 10.4108/eai.18-11-2022.2327114

2. DATA

This paper uses the representative CSI 300 Index Fund (510300) to measure the return of China's stock market. First of all, the CSI 300 Index covers the 300 stocks with the largest market capitalization and the best liquidity on the Shanghai and Shenzhen stock exchanges. As a representative of cross-exchange indexes, it can reflect the overall trend of China's securities market. Secondly, using CSI 300 index fund instead of CSI 300 index itself is because index fund can better represent the actual situation of market trading and has more guiding significance for investors' practice. The intraday data was obtained from RESST high-frequency database, and the sample period was from April 28, 2012, to December 30, 2018.

To study the predictability of intraday return of the Chinese stock market at day t, this paper uses the price at 10:00 am of the trading day and the closing price of the previous day (3:00 pm) to calculate the return of the first half-hour. After that, from 10:00 am to 3:00 pm, the return is calculated every half hour. It is important to note that the fifth half-hour return is calculated by the price of 11:30 am and 1:30 pm since the Chinese stock market has a 90-minute trading break from 11:30 am to 1:00 pm. So we can get eight half-hour returns per day:

$$r_{j,t} = \frac{p_{j,t}}{p_{j-1,t}} - 1, j = 1, \cdots, 8, \tag{1}$$

Where $P_{j,t}$ is the jth half-hour price on day t.Note that $P_{0,t}$ is the closing price of the previous trading day, that is, $P_{0,t}$= $P_{8,t-1}$. In other words, this paper uses the closing price of the previous trading day as the opening price of the current trading day, so the return of the first half-hour contains overnight information from the previous trading day

3. EMPIRICAL ANALYSIS

3.1 In-sample regression

Promoted by Gao (2018) [6], this paper uses the following predictive regression model to test the in-sample predictability of intraday momentum:

$$c = \alpha + \beta_1 r_{1,t} + \varepsilon_t \tag{2}$$

Where, $r_{8,t}$ and $r_{1,t}$ are the eight-half-hour return and the first-half-hour return of trading day t respectively, and ε_t is the error term with the mean value of zero.

The in-sample estimates in Eq (2) are reported in the first column of panel A of Table 1. The positive slope of r_1 was 0.0894, which was statistically significant at the 1% level, and the R^2 in-sample was 1.8%. Such a high R^2 forecast is impressive not only because almost all monthly forecasters have low R^2, but also because the R^2 values for US intraday equity returns are small (see Gao2018) [6].

Gao (2018)[6] also found that returns in the penultimate half-hour can positively predict returns in the last half-hour. With this in mind, further, run the following regression equation:

$$r_{8,t} = \alpha + \beta_7 r_{7,t} + \varepsilon_t \tag{3}$$

The second column of panel A of Table 1 reports the results of the in-sample estimate using the seventh half-hour return. The seventh half-hour return r_7 predicted the last half-hour return r_8 with a positive slope of 0.178, which was statistically significant at the 5% level, with an in-sample R^2 of 3.0%. And the coefficient slope and R^2 of the prediction using the seventh half-hour is much higher than using the first half-hour. In other words, intraday momentum in the seventh half-hour returns is stronger than in the first half-hour returns, contrary to what is found in US equities (see Gao 2018) [6].

Since r_1 and r_7 can independently predict r_8, how can they together predict r_8? To this end, the regression was carried out as follows:

$$r_{8,t} = \alpha + \beta_1 r_{1,t} + \beta_7 r_{7,t} + \varepsilon_t \tag{4}$$

The third column 3 of panel A of Table 1 reports the results of the in-sample estimates using the first and seventh half-hour returns. Interestingly, their regression values differ from the results when using each of them respectively. With the addition of r_7, the coefficient of r_1 decreased and became less significant, and the R^2 of the combined prediction was only 3.8%, which was close to the R^2 (3.0%) obtained by using r_7 alone. This seems to indicate that r_7 has absorbed the explanatory power of r_1.

TABLE 1 IN-AND OUT- SAMPLE PREDICTION RESULTS

Predictor	r_1	r_7	r_1 *and* r_7
	Panel A: In-sample prediction		
β_1	0.0894***(2.910)		0.080**(2.486)
β_7		0.178**(2.20)	0.165**(2.052)
Intercept	0.00(0.789)	0.00(0.722)	0.00(0.788)
R^2	0.018	0.030	0.038
	Panel B: Out of sample prediction		
R^2_{OS}	1.58***	1.81**	2.68***

3.2 Out-of-sample prediction

So far, our in-sample predictions have yielded excellent results. However, Welch(2008)[8] pointed out that the predictive power of many macroeconomic variables came from over-fitting in the sample, which was unstable and would lead to good performance in the sample and poor predictive power out of the sample. Moreover, out-of-sample forecasts are more useful in practical applications because out-of-sample return forecasts can guide financial market participants to make investment decisions. Thus, out-of-sample forecasting is a more rigorous test of return predictability. Given this, this paper focuses on out-of-sample tests in the next empirical analysis.

To generate out-of-sample forecasts of the half-hour returns, this paper uses recursive (extended) estimation windows such as Rapach(2010)[7], Neely(2014)[3], and Gao(2018)[6]. Specifically, the entire sample containing T observations is divided into an in-sample part containing the first m observations and an out-of-sample part containing the last q observations. For example, the return to the eighth half-hour of the first out-of-sample prediction is obtained from the following equation:

$$\hat{r}_{8,m+1} = \hat{\alpha}_m + \hat{\beta}_{1,m} r_{1,m+1} \tag{5}$$

Where $\hat{\alpha}_m$ and $\hat{\beta}_{1,m}$are the coefficients obtained from the least-squares regression in Eq.(2). The second out-of-sample predicted value is then calculated by the following equation.

$$\hat{r}_{8,m+2} = \hat{\alpha}_{m+1} + \hat{\beta}_{1,m+1} r_{1,m+2} \tag{6}$$

By proceeding in this manner to the end of the out-of-sample, a series of q out-of-sample forecasts of the eighth half-hour gain $\{\hat{r}_{8,t}\}_{t=m+1}^{T}$can be obtained. using a similar approach, a series of out-of-sample forecasts can also be obtained based on Eq.(3) and Eq.(4). The window period for prediction in this paper is April 28, 2012, to July 9, 2014, with 361 observations. This length was chosen by referring to Gao (2018)[6], using 1/4 of the entire sample length as the prediction window period.

Following the conventions of earnings forecasting, this paper uses the R_{OS}^2 statistic to assess the out-of-sample forecasting accuracy of the forecasting model of interest relative to the popular historical mean model, where the historical mean model uses the historical mean for forecasting: $\bar{r}_{8,t+1} = 1/t \sum_{k=1}^{t} r_{8,t}$. Welch(2008)[8] found that many popular forecasting models have difficulty exceeding historical averages. The out-of-sample R_{OS}^2 is calculated from the following equation.

$$R_{OS}^2 = 1 - \frac{\sum_{t=m+1}^{T} \left(r_{8,t} - \hat{r}_{8,t}\right)^2}{\sum_{t=m+1}^{T} \left(r_{8,t} - \bar{r}_{8,t}\right)^2} \tag{7}$$

Where $r_{8,t}$、$\bar{r}_{8,t}$、$\hat{r}_{8,t}$ are the actual return, historical mean return, and forecast return for the last half hour of trading day t, respectively, and m and T are the initial estimation window and the length of the entire sample period, respectively.

R_{OS}^2 measures the reduction in mean squared forecast error (MSFE) relative to the universal historical mean return forecast. To further determine whether the forecasting model produces a statistically significant improvement in MSFE, the Clark statistic is used in this paper:

$$f_t = \left(r_{8,t} - \bar{r}_{8,t}\right)^2 - \left(r_{8,t} - \hat{r}_{8,t}\right)^2 + \left(\bar{r}_{8,t} - \hat{r}_{8,t}\right)^2 \tag{8}$$

By regressing $\{f_s\}_{s=m+1}^{T}$ on a constant, the Clark statistic can be obtained, which is equivalent to the t-statistic corresponding to that constant. In addition, the p-value of the one-sided (upper tail) test is easily obtained under the standard normal distribution.

Panel B of Table 1 reports the out-of-sample forecasting results using the first half-hour return r_1, the seventh half-hour return r_7, and the joint use of both. Similar to the in-sample prediction results, r_7 outperforms r_1 (1.81 > 1.58), but its significance level is lower than that of r_1. Again, the best prediction is achieved when using both jointly, with R_{OS}^2 reaching 2.68 and significant at the 1% level.

Overall, this paper finds that intraday returns in the Chinese stock market are predictable due to the presence of intraday momentum. Compared to the U.S. stock market, the intraday returns of the Chinese stock market are much more predictable. In addition, the penultimate (i.e., seventh) half-hour returns of the Chinese stock market are more predictable than the first half-hour returns, while the opposite is true for the U.S. stock market (see Gao 2018)[6].

4. Economic Significance

In addition to the above theoretical research, this paper will also examine the economic significance of intraday momentum from a more practical perspective, including asset allocation and market timing.

4.1 Asset allocation

Following Rapach(2010)[7], Neely(2014)[3], and Gao (2018)[6], this paper calculates the deterministic equivalent return of a mean-variance investor with a relative risk aversion coefficient of 3 allocating between equities and risk-free instruments using the most recent half-hourly return forecast (CER). At the end of the penultimate half-hour, investors allocate their stock weights optimally for the last half-hour of trading day t as:

$$w_t = \frac{1}{\gamma}\frac{\hat{r}_{8,t} - r_{f,t}}{\hat{\sigma}_{8,t}^2} \tag{9}$$

Where γ is the relative risk aversion coefficient of the investor, and $\hat{r}_{8,t}$、$r_{f,t}$ denote the predicted return and risk-free return in the last half hour of day t. Similar to Thompson (2008), Neely (2014) [3], and Rapach (2010) [7], this paper uses a one-year moving window of the last half hour to predict the volatility $\hat{\sigma}_{8,t}^2$. In addition, this paper restricts w to between 0 and 1.5 to prevent short selling and allows no more than 50% leverage. In addition, this paper imposes a relatively loose constraint on the portfolio weights, with w being limited to between -0.5 and 1.5, which means that investors can short or borrow 50% of the margin.

Portfolio returns can be realized by investors who allocate their wealth using Eq(9).

$$R_t = w_t r_{8,t} + (1 - w_t) r_{f,t} \tag{10}$$

For the entire out-of-sample cycle, the CERs achieved are:

$$CER = \bar{R}_p - 0.5\gamma\sigma_p^2 \tag{11}$$

Where $\bar{R}_p$ and σ_p^2 are the mean and variance of the portfolio returns for the out-of-sample evaluation period, respectively. The excess CER (CER gain) is the difference between the CER obtained using r_1,r_7 as predictors and the CER obtained from the benchmark model using the historical mean as the predictor.

TABLE 2 ASSET ALLOCATION

Predictor	Avg ret(%)	Std dev(%)	Sratio	CER (%)	CER gain (%)
Panel A: Risk-assets weights are limited to lie between 0 and 1.5					
$\overline{r_8}$	-1.26	9.56	-0.13	-2.63	
r_1	14.67	8.60	1.71	13.56	16.19
r_7	10.82	8.68	1.25	9.69	12.32
r_1 *and* r_7	16.29	8.74	1.86	15.15	17.78
Panel B: Risk-assets weights are limited to lie between -0.5 and 1.5					
$\overline{r_8}$	-2.40	9.69	-0.25	-3.81	
r_1	18.87	9.10	2.07	17.62	21.43
r_7	13.85	9.23	1.50	12.58	16.39
r_1 *and* r_7	20.42	9.29	2.20	19.12	22.93

Table 2 reports the asset allocation results. Panel A shows the economic values when the risky asset weights are restricted to 0 to 1.5. The historical average forecast yields the lowest average portfolio return of -1.26%. In contrast, the first half-hour return r_1 and the seventh half-hour return r_7 alone produce higher average returns of 14.67% and 10.82%, respectively. Unlike the previous calculation of R_{OS}^2, the average return of r_1 is higher than that of r_7. In addition, combining the two predictors r_1 and r_7 yields the highest annual average return of 16.29%.

Of course, it is necessary to consider risk. Surprisingly, the standard deviation is consistently low in the last half-hour return forecasts where the average portfolio return is high. The historical average forecast $\overline{r_8}$ produces the highest standard deviation at 9.56%, and both forecasters produce low standard deviations at around 8.7%. Thus, the ranking of the portfolio performance return forecasts used does not change after accounting for risk. Specifically, the historical average forecasts have the lowest Sharpe ratio and CER, while the r_1 and r_7 portfolio forecasts r_8 yield the largest Sharpe ratio and CER.

Panel B of Table 2 shows the economic value when the risky asset weights are constrained to be between -0.5 and 1.5. Stronger returns are obtained with this relatively loose constraint. More importantly, the findings above prove that our results are robust. Overall, mean-variance investors can achieve substantial economic returns by switching from a random walk model to an intraday momentum model.

4.2 Market timing

In the market timing test, this paper follows Gao (2018) [6] and uses the returns of the first half-hour and the seventh half-hour as timing signals to trade in the last half-hour. More specifically, if the timing signal is positive, a long position is opened in the market at the beginning of the last half hour, otherwise, a short position will be opened in the market. It is important to note that long or short positions should be closed at the end of the last half hour of each trading day.

First, the first half-hour return r_1 is used as the trading signal. Thus, a market timing strategy based on the timing factor r_1 on trading day t will realize the following return in the last half hour.

$$\eta(r_1) = \begin{cases} r_8, & \text{if } r_1 > 0 \\ -r_8, & \text{if } r_1 \leq 0 \end{cases} \tag{12}$$

Second, using the seventh half-hour return r_7 of day t as a trading signal, the last half-hour return can be realized as:

$$\eta(r_7) = \begin{cases} r_8, & \text{if } r_7 > 0 \\ -r_8, & \text{if } r_7 \leq 0 \end{cases} \tag{13}$$

Third, combine r_1 and r_7 and go long only when both returns are positive and short only when both returns are negative. Otherwise, the short position is chosen. Thus, the realized return can be calculated as.

$$\eta(r_1, r_7) = \begin{cases} r_8, & \text{if } r_1 > 0 \text{ and } r_7 > 0 \\ -r_8, & \text{if } r_1 \leq 0 \text{ and } r_7 \leq 0 \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

Finally, two benchmark strategies are also considered in this paper. One of the benchmark strategies is the 'Always Long' strategy, which means to always hold a long position at the beginning of the last half hour and close it at the end. Another one is the popular "buy and hold" strategy, which buys stocks at the beginning of the whole sample and holds until the end of the entire sample.

As shown in table5, the economic value of these two benchmark strategies is relatively low. In particular, the always long strategy has a mean return of only 2.12%, resulting in the lowest Sharpe ratio of 0.27. In contrast, the timing strategy that uses both r_1 and r_7, $\eta(r_1, r_7)$, has the lowest standard deviation of 6.21%, resulting in the highest Sharpe ratio of 2.18 and a position win ratio of 59.25%, the highest of all strategies. which is the highest among all strategies. Overall, all three timing strategies outperform the benchmark strategy, suggesting that, like asset allocation, intraday momentum-based timing strategies can also deliver economic returns.

TABLE 3 TIMING STRATEGY PERFORMANCE

Timing	Avg ret(%)	Std dev (%)	SRatio	Skewness	Kurtosis	Success(%)
Panel A: Market timing						
$\eta(r_1)$	15.43	7.77	1.98	1.47	22.58	52.56
$\eta(r_7)$	11.18	7.80	1.43	1.20	22.57	48.34
$\eta(r_1, r_7)$	13.53	6.21	2.18	2.85	47.15	59.25
Panel B: Benchmarks						

Always long	2.12	7.84	0.27	-1.29	22.78	49.93
Buy-and-hold	10.71	24.35	0.44	-0.48	12.10	49.79

5. ROBUSTNESS TESTS

Some robustness tests are performed in this section to further validate the previous main findings.

5.1 Replacement of estimation window

Rossi (2012) [9] argues that in practice, an arbitrary choice of estimation window size may lead to very different out-of-sample results. Therefore, the selection of window period has a great influence on the out-of-sample prediction results. Conserdering this, the window period is replaced next. In the previous study, the window period for the initial regression was 2012.5.28-2013.02.03. The window period is now replaced with 2012.5.28-2015.12.30.

Table 4 reports the out-of-sample results for the new prediction window, where the stock weights are restricted to be between -0.5 - 1.5. The first column in Table 4, R^2_{OS}, are all positive and significant, indicating that r_1 and r_7 still have predictive power during the new window period. Similarly, the results in the second and third columns of Table 4 also show that the predictions of r_1 and r_7 are still economically significant during the new window period.

In short, this paper finds a robust result in intraday momentum, which yields substantial forecasting gains from both statistical and economic perspectives. In particular, the 7th half-hour return is more powerful in predicting the last half-hour return than the first half-hour return. (Out-of-sample R^2_{OS} is 1.64% and 5.95%, respectively)

TABLE 4 REPLACEMENT WINDOW RESULTS

Predictor	R^2_{OS} (%)	CER gain(%)	Sharp ratio
r_1	1.64***	8.56	1.78
r_7	5.95***	6.46	1.27
r_1 *and* r_7	3.71***	7.53	1.47

5.2 Other half-hourly gains

The previous analysis leads to the conclusion that r_1 and r_7 can predict r_8. However, in addition to r_1 and r_7, can other half-hourly returns also predict r_8? To answer this question, the same research method as before is used to predict r_8 using r_2-r_6, respectively, and the statistically significant R^2_{OS} as well as the economically significant CER gain and Sharp ratio are reported. second, the mean of r_1-r_7 is also used to predict r_8. Finally, for comparison, the prediction results of r_1 and r_7 are also put on the table.

As shown in Table 5, the out-of-sample prediction results R^2_{OS} for the second half-hour to the sixth half-hour are all negative and insignificant, which means all of them are poor predictors. What's more, the mean combination shows great prediction power for its R^2_{OS} is 2.19% and significant at 1% level.As we have known that r_2-r_6 are all poor predictors, the prediction power of the mean combination is come from r_1 and r_7.Also, we can draw the same conclusion from CER gain and Sharp ratio.

TABLE 5 OTHER HALF-HOURLY FORECAST RESULTS

Predictor	R^2_{OS}(%)	CER gain(%)	Sharp ratio
r_1	1.58^{***}	21.44	2.07
r_2	-0.11	3.20	0.02
r_3	-0.24	2.82	-0.01
r_4	-0.78	-3.70	-0.81
r_5	-0.47	3.18	0.08
r_6	-0.10	11.22	0.95
r_7	1.81^{**}	16.39	1.50
r_1 *and* r_7	2.68^{***}	22.93	2.20
Mean combination	2.19^{***}	24.75	2.67

6. CONCLUSION

This paper proves the existence of intraday momentum in the Chinese stock market. The returns of the first and seventh half-hour could predict the returns of the last half-hour, which is statistically significant both in- and out-of-sample. Compared to the U.S. stock market, the intraday returns of the Chinese stock market are much more predictable. In addition, the seventh half-hour return has stronger power than the first half-hour return in predicting the last half-hour return while the opposite is true for the U.S. stock market (see Gao 2018)[6]. What's more, from the perspective of asset allocation and market timing, using intraday momentum as a trading basis can easily achieve excess returns. Finally, we perform three robustness tests, including changing the out-of-sample window period and verifying whether the remaining half-hour returns(r_2-r_6) also have the predictive power, which all show that intraday momentum was not an accidental phenomenon.

REFERENCES

[1]Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. J. Finance 48, 65–91

[2]Moskowitz, T.J., Ooi, Y.H., Pedersen, L.H., 2012. Time-series momentum. J. Financ. Econ. 104, 228–250

[3]Neely, C.J., Rapach, D.E., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: the role of technical indicators. Manag. Sci. 60, 1772–1791.

[4]Sun, L., Najand, M., Shen, J., 2016. Stock return predictability and investor sentiment: a high-frequency perspective. J. Bank. Finance 73, 147–164.

[5]Renault, T., 2017. Intraday online investor sentiment and return patterns in the US stock market. J. Bank. Finance 84, 25–40.

[6]Gao, L., Han, Y., Li, S.Z., Zhu, G., 2018. Market intraday momentum. J. Financ. Econ. 129, 394–414

[7]Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev. Financ. Stud. 23, 821–862.
[8]Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21, 1455–1508.
[9]Rossi, B., Inoue, A., 2012. Out-of-sample forecast tests robust to the choice of window size. J. Bus. Econ. Stat. 30, 432–453.