

Influential Article Review – Applying Statistics in Stock Market Decision

Evelyn Franklin

Elliott Valentine

Seb Villarreal

This paper examines the stock market. We present insights from a highly influential paper. Here are the highlights from this paper: Forecasting stock returns is extremely challenging in general, and this task becomes even more difficult given the turbulent nature of the Chinese stock market. We address the stock selection process as a statistical learning problem and build cross-sectional forecast models to select individual stocks in the Shanghai Composite Index. Decile portfolios are formed according to rankings of the forecasted future cumulative returns. The equity market's neutral portfolio—formed by buying the top decile portfolio and selling short the bottom decile portfolio—exhibits superior performance to, and a low correlation with, the Shanghai Composite Index. To make our strategy more useful to practitioners, we evaluate the proposed stock selection strategy's performance by allowing only long positions, and by investing only in A-share stocks to incorporate the restrictions in the Chinese stock market. The long-only strategies still generate robust and superior performance compared to the Shanghai Composite Index. A close examination of the coefficients of the features provides more insights into the changes in market dynamics from period to period. For our overseas readers, we then present the insights from this paper in Spanish, French, Portuguese, and German.

Keywords: Stock selection, Stock return prediction, Statistical learning, Lasso, Elastic net

SUMMARY

- We follow the steps described in Section 2.2 and set the look-back rolling window B as 12 months, such that the actual performance evaluation period is from January 2003 to December 2016. In setting the look-forward window F to one month, we evaluate three trading strategies based on the forecasted one-month return from the elastic net model. First, we apply our strategy to buy the top-decile stocks while selling short the bottom-decile stocks. We do this because asset pricing literature often uses a long-short dollar-neutral strategy to test market efficiency and examine certain features' usefulness. We denote the portfolio constructed using this strategy as the Equity-Market-Neutral portfolio, as it seeks to exploit investment opportunities unique to some specific group of stocks while maintaining a neutral exposure to broad groups of stocks. We then evaluate our strategy's performance by only buying the top-decile stocks that adapt to the regulations on short-sales in the Chinese stock market. The EMN-ENET portfolio has the smoothest overall cumulative return curve, and especially during the 2008 global financial crisis, because the

proposed forecast model effectively detects changes in the market environment and successfully selects underperforming stocks to benefit from selling short. Additionally, we observe that the stock selection strategies still significantly outperform the index where the OLS estimator is used instead of the ENET estimator but are not as effective as with the ENET approach. This demonstrates the value of considering a more sophisticated ENET estimator over a simple OLS estimator.

- We delineate each strategy's downside risk by comparing the drawdowns of each portfolio from their high watermarks. At each time t , each strategy's high watermark M is the highest value of historical cumulative returns, and can be calculated as $M_t = \max_{k=1, \dots, t} \{R_{\sim 1, k}\}$ where $R_{\sim 1, k} = \prod_{j=1}^k R_j - 1$ and R_j is the portfolio's return at time j . The portfolio's drawdown at time t is defined as D_t/M_t . Figure 3 compares the percentage of drawdowns for different strategies. In the plot, the point in time at which the time series lines last touch the 0-reference line is the time when each portfolio reaches its historical highest cumulative return. The success in identifying winning and losing stocks also generates a portfolio that negatively correlates with the Shanghai Composite Index. In contrast to the EMN-ENET strategy, the two long-only strategies are less stable with slightly larger annualized standard deviations than the Shanghai Composite Index. However, they still deliver higher excess returns than the index. Table 3 presents the annual returns of each strategy from 2003 to 2015. In 2007, the highest annual return among the proposed strategies was 216%, which is nearly twice as large as the index's highest annual return during the entire evaluation period. Hence, the proposed strategies achieve much higher Sharpe and Calmar ratios than the index, with similar risks but much greater upside potential.
- We also observe whether B-share stocks are included in the top-performing portfolios makes a small difference.

HIGHLY INFLUENTIAL ARTICLE

We used the following article as a basis of our evaluation:

Wu, W., Chen, J., Xu, L., He, Q., & Tindall, M. L. (2019). A statistical learning approach for stock selection in the Chinese stock market. *Financial Innovation*, 5(1), 1–18.

This is the link to the publisher's website:

<https://jfin-swufe.springeropen.com/articles/10.1186/s40854-019-0137-1>

INTRODUCTION

China's annual GDP growth has averaged more than 9% since 2000. Meanwhile, the Chinese stock market has experienced substantial fluctuations during this period. The Shanghai Composite Index reached a historical high of over 6000 in October 2007, only to fall to approximately half that level at present. This disparity between rapidly growing GDP and lackluster stock market performance poses a dilemma for investors. How can one benefit from the burgeoning Chinese economy by investing in the Chinese stock market? As Rapach and Zhou (2013) have indicated, forecasting stock returns is extremely challenging in general; this task is even more difficult when it involves the turbulent Chinese stock market. Kang et al. (2002) studied contrarian and momentum strategies for the Chinese stock market in earlier years, while Li et al. (2017) examined the performance of trends following strategies in Chinese commodity futures markets. The recent introduction of advanced statistical approaches to portfolio management, such as machine learning, has been highly successful (e.g. Li and Hoi 2015; Shen and Wang 2017; Wu et al. 2018; Gu et al. 2019). This paper presents a statistical learning approach to select top-performing Chinese stocks that can potentially generate superior returns and significantly outperform the market.

Some researchers have attempted to identify the underlying drivers for Chinese stock returns. Wang and Xu (2004) discovered that the value factor does not explain the cross-sectional differences in the Chinese stock market due to the market's speculative nature and low-quality accounting information. On

the other hand, the size factor does carry certain explanatory power. Kling and Gao (2008) explored the positive feedback process between Chinese stock share prices and institutional investors' sentiment. Further, Yuan et al. (2008) found that the equity ownership of mutual funds positively effects performance. These types of information can all contribute to the stock selection process. However, these methods benefit institutional rather than individual investors, as the latter cannot easily access this information.

Given the Chinese stock market's poor performance, an increasing number of institutions and investors are searching for stocks with high excess returns compared to the Shanghai Composite Index. Empirical finance often ranks stocks according to a risk factor over a look-back window, and long and short stock positions are subsequently taken according to this ranking. Literature has also proposed numerous factors (Fama and French 1992; Ang et al. 2006; Frazzini and Pedersen 2014), with Harvey et al. (2016) identifying more than 200 such factors. Factor-ranking procedures are simple and effective, but have some inherent problems. For example, a particular factor's risk premium may be unstable over time. Fama and French (1992) found that their value factor exhibits a positive risk premium over the long-term in the U.S. market; however, we discovered that the value premium—estimated as the annual return difference between the Russell 1000 Value Index and Russell 1000 Growth Index—has been negative in all five-year rolling windows from 2009 to 2015.

The simple factor-ranking approach can be extended to include multiple risk factors in the portfolio-formation process, but this creates another problem. If we follow a traditional approach, we would select stocks ranked in the top group for all risk factors. However, a high-dimensional problem occurs with many factors, in that the stocks would be scattered sparsely in an N-dimensional space, with N being the number of factors that we would incorporate. It may be impossible to find enough stocks in the top rankings of all factors to form a diversified portfolio. One could use an ad hoc method by applying a composite score to combine the factors into a ranking. Mohanram (2005) combined traditional fundamental factors to create an index-GSCORE. While a long-short strategy based on the GSCORE earns significant excess returns, this method lacks theoretical guidance about how to assign weights among factors, and subjective factors always influence the method's precision in forming a composite score. These problems make the method unstable in practice. As previously mentioned, both simple and multiple factor-ranking approaches cannot effectively construct a portfolio with the highest future return, and thus, we must consider other ways to improve performance.

Cross-sectional regression plays an important role in finance to explain variations in stock prices (Sharpe 1964; Fama and French 1992; Carhart 1997). For example, Fama and French (1992) indicate that the cross-section regression provides a good description of returns on portfolios formed based on size, BE/ME, and term-structure risk factors. This leads to the gradual development of a method for determining factors' weight through regressions. While some subsequent works (Hou et al. 2015; Fama and French 2015) attempted to discover more factors to explain stock returns using cross-sectional regression models, the regressions' accuracy is undermined by the effects of multicollinearity and overfitting using a large number of factors.

In this paper, we propose an advanced statistical model, the elastic net (Zou and Hastie 2005), to resolve this issue when regressing stock returns over a large number of factors. Our method differs from most existing methods by building cross-sectional forecast models for stock returns, and selects stocks based on these models' predicted returns. Our approach is rooted in the Fama-French-Carhart model (Carhart 1997), hereafter "the FFC four-factor model," but greatly expands its scope to include more statistical factors. Given the low quality of accounting information in the Chinese stock market, we focus on forecasting future returns using only the statistical factors derived from historical stock prices. We handle this as a supervised statistical learning problem in which portfolios are formed according to the forecast returns' rankings. We find that the highest ranked portfolios generate robust and superior performance. Our forecast methods deviate from traditional econometric approaches and are more in line with the approach taken by Varian (2014), who argues that data analysis in statistics and econometrics can be divided into four categories: 1) prediction, 2) summarization, 3) estimation, and 4) hypothesis testing. The statistical learning methods that we employ fall into the first category as our goal is to identify the stocks with the highest future returns.

This paper contributes to existing literature in the following ways. First, the proposed statistical learning approaches effectively rank, and hence select, the top performing stocks relying on the predictability of the forecast model. This is different from traditional cross-sectional regression models (Fama and French 1992; Carhart 1997). Second, building a statistical learning model provides a data-adaptive guidance on how to combine information from different factors (which we treat as features) in the proposed forecast model. Third, the use of the elastic net estimator (Zou and Hastie 2005) improves the estimation precision (accuracy) of the forecast model when multicollinearity is present and also avoids the overfitting that occurs when a large number of features are considered. Another advantage of the proposed method is its interpretability and feature selection capability. Lastly, the proposed stock selection framework is flexible and can be modified by investors to include user-specific features. The remainder of our paper is organized as follows: Section 2 discusses our motivation and explains the proposed methodology. Subsequently, Section 3 presents the data and features used in this study. Section 4 presents the empirical results, while the final section concludes.

CONCLUSION

This study constructed a cross-sectional statistical forecast model for stock selection in the Chinese stock market. Based on the forecast of future cumulative returns, the proposed approach allows investors to identify stocks that are likely to perform well, and to construct corresponding portfolios. Although the Chinese stock market collectively cannot generate satisfactory results, our empirical results indicate that it is still possible to generate significant excess returns through an active, quantitative stock selection process. Additionally, and regarding the features used in this study, far fewer useful features were observed during the 2008 financial crisis than in other periods of time. Meanwhile, the frequent occurrence and negative signs in the one-month return (R1) feature strongly indicate a short-term mean-reversion in the Chinese stock market. Further, despite the fact that the forecast models are constructed using both A- and B-share stocks, solely investing in A-share stocks from the top-decile portfolio can still yield incredibly good performance.

APPENDIX

**TABLE 1
FEATURES USED IN THE STUDY**

Features	Symbol
Return based:	
1. One-month return	R1
2. Two-month cumulative return	R2
3. Three-month cumulative return	R3
4. Four-month cumulative return	R4
5. Five-month cumulative return	R5
6. Six-month cumulative return	R6
7. Seven-month cumulative return	R7
8. Eight-month cumulative return	R8
9. Nine-month cumulative return	R9
10. Ten-month cumulative return	R10
11. Eleven-month cumulative return	R11
12. Twelve-month cumulative return	R12
Statistics based:	
13. Standard deviation calculated from the past 12 months of returns	SD
14. Skewness calculated from the past 12 months of returns	Skew
15. Kurtosis calculated from the past 12 months of returns	Kurt
16. Sharpe ratio calculated from the past 12 months of returns	Sharpe
17. Alpha with respect to the market calculated from the past 12 months of returns	Alpha
18. Market beta calculated from the past 12 months of returns	Beta
19. Correlation with index calculated from the past 12 months of returns	Corr
20. Short-term volatility calculated from the daily returns in the past 20 days	Vol
21. Distance/drawdown from the high watermark in the past 12 months	DD

**TABLE 2
PERFORMANCE MEASURES DURING 2003–2015**

Statistics	Index	EMN-ENET	Long-ENET	LongA-ENET	EMN-OLS	Long-OLS	LongA-OLS
AR (%)	6.08	20.59	22.03	22.52	13.12	16.75	17.34
Std (%)	28.88	19.08	35.27	35.58	16.80	34.81	35.40
Best MR (%)	27.45	25.16	35.24	41.14	20.92	29.38	38.31
Worst MR (%)	-24.63	-20.38	-26.46	-26.46	-21.20	-27.18	-28.35
ShR	0.21	1.08	0.67	0.63	0.78	0.48	0.49
MD (%)	70.97	35.75	66.58	64.74	37.41	67.39	66.46
CarR	0.09	0.58	0.33	0.34	0.35	0.25	0.26
Skewness	-0.21	-0.06	0.15	0.20	-0.17	-0.01	0.09
Kurtosis	1.10	3.39	1.08	1.34	3.77	0.59	1.01
Correlation	1.00	-0.10	0.82	0.83	-0.12	0.84	0.83

TABLE 3

ANNUAL RETURNS FOR 2003–2015

Year	Index	EMN-ENET	Long-ENET	LongA-ENET	EMN-OLS	Long-OLS	LongA-OLS
2003	-0.19	17.47	-9.95	-11.40	19.50	-11.61	-10.29
2004	-15.40	35.52	-3.63	-2.91	34.94	-3.62	-2.54
2005	-8.33	48.73	8.62	9.60	22.32	-5.01	-5.11
2006	130.43	26.92	117.34	142.38	15.73	93.60	109.85
2007	96.66	18.04	216.39	201.58	0.68	205.97	209.02
2008	-65.39	13.81	-56.19	-53.52	19.15	-57.98	-56.57
2009	79.98	60.94	172.12	177.43	40.01	165.45	166.35
2010	-14.31	-2.63	10.52	10.42	7.83	12.35	12.52
2011	-21.68	0.53	-29.17	-30.12	-7.91	-32.86	-33.09
2012	3.17	36.69	23.79	21.43	22.00	12.83	8.84
2013	-6.75	23.70	19.94	21.00	12.14	11.88	12.91
2014	52.87	-19.72	33.62	33.69	-26.26	29.72	31.02
2015	9.41	27.18	29.86	22.85	21.17	32.95	27.07
2016	-12.31	25.04	-3.84	-4.16	22.07	-7.81	-9.33

TABLE 4
AVERAGE PERFORMANCE MEASURES DURING 2003–2015 FOR 5-YEAR ROLLING WINDOWS

Statistics	Index	EMN-ENET	Long-ENET	LongA-ENET
AR (%)	11.20	22.15	33.08	34.07
Std (%)	29.56	20.36	36.98	37.32
Best MR (%)	22.28	19.46	29.97	31.48
Worst MR (%)	-20.72	-16.04	-19.47	-19.83
ShR	0.38	1.10	0.89	0.89
MD (%)	51.37	28.60	46.51	45.37
CalR	0.25	0.84	0.80	0.85
Skewness	-0.08	-0.25	0.26	0.23
Kurtosis	0.69	2.85	0.46	0.62
Correlation	1.00	-0.07	0.81	0.82

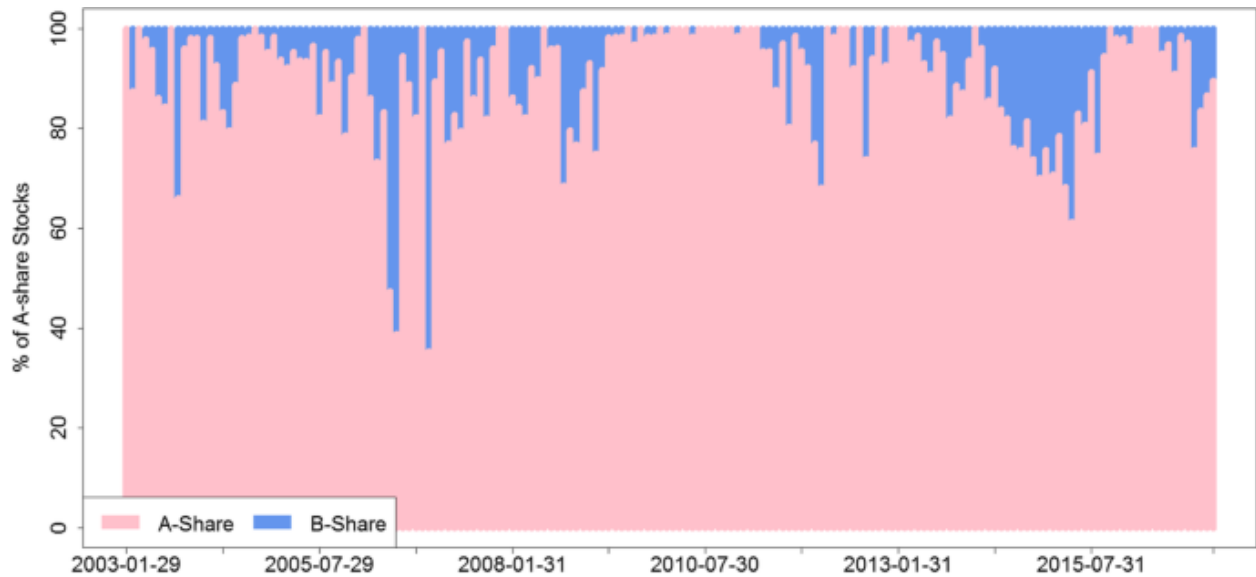
TABLE 5
AVERAGE ANNUALIZED RETURN OVER A 5-YEAR ROLLING WINDOW

Period	Index	EMN-ENET	Long-ENET	LongA-ENET
2003–2007	31.12	28.90	47.67	49.49
2005–2009	20.94	32.48	54.85	59.52
2007–2011	-3.84	16.13	24.18	24.58
2009–2013	3.05	21.60	25.90	25.76
2011–2015	4.74	11.66	12.79	11.01

TABLE 6
FEATURE IMPORTANCE SCORES IN DIFFERENT TIME PERIODS

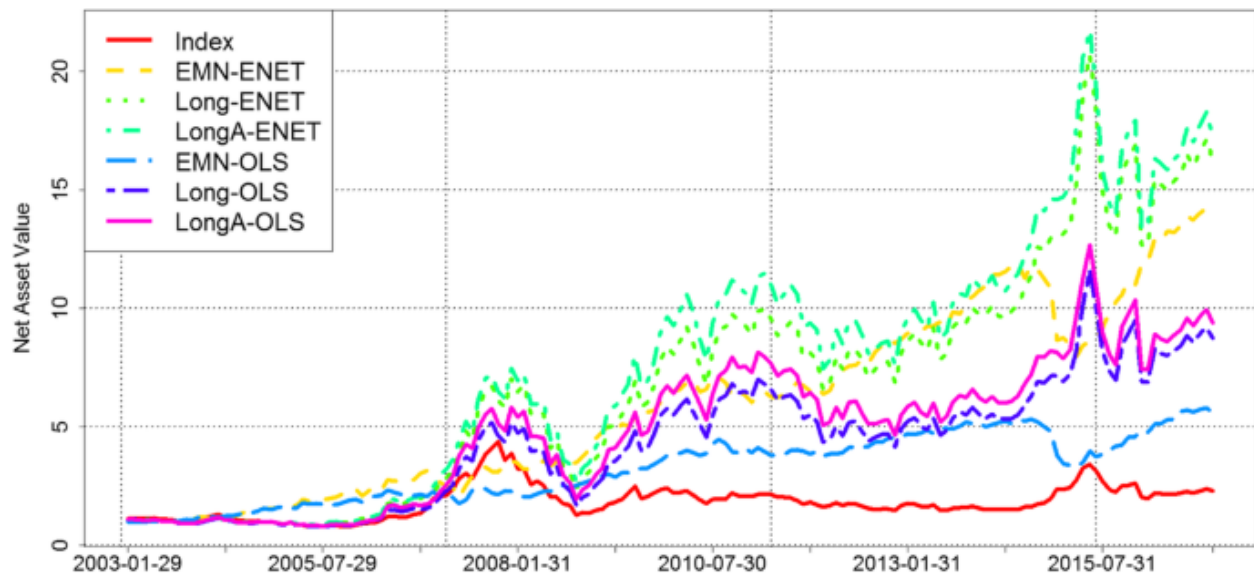
Factors	2003-2004	2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	2015-2016	2003-2016
R1	0.71	1.00	1.00	0.96	0.50	1.00	1.00	0.88
R2	1.00	1.00	1.00	0.46	1.00	0.88	0.46	0.83
R3	0.75	0.96	0.50	0.17	0.08	0.58	1.00	0.58
R4	0.67	0.63	0.46	0.29	0.17	0.04	0.38	0.38
R5	1.00	1.00	0.50	0.21	0.46	0.67	0.67	0.64
R6	0.58	1.00	0.54	0.33	0.54	0.92	0.83	0.68
R7	0.25	0.63	0.92	0.42	0.00	0.21	0.67	0.44
R8	0.08	0.42	0.00	0.13	0.13	0.00	0.42	0.17
R9	0.92	0.88	0.67	0.67	0.04	0.13	0.42	0.53
R10	0.17	0.75	0.54	0.00	0.46	0.71	0.17	0.40
R11	0.54	0.54	0.17	0.83	0.75	0.33	0.63	0.54
R12	1.00	1.00	0.79	0.29	0.13	0.21	0.92	0.62
SD	0.79	0.46	0.42	0.33	0.54	0.79	0.83	0.60
Skew	0.50	0.50	0.38	1.00	0.92	1.00	0.50	0.68
Kurt	0.54	1.00	0.50	0.88	0.79	0.54	0.96	0.74
Sharpe	0.13	0.58	0.00	0.42	0.13	0.83	0.50	0.37
Alpha	0.42	0.83	0.04	0.17	0.54	0.25	0.58	0.40
Beta	0.13	0.33	0.38	0.13	0.50	0.96	0.33	0.39
Corr	0.75	0.79	0.21	0.58	0.58	0.33	0.67	0.56
Vol	0.58	0.79	0.25	0.83	0.42	0.79	0.79	0.64
DD	0.71	0.38	0.46	0.58	0.42	0.08	0.13	0.39
Average FI	0.58	0.74	0.46	0.46	0.43	0.54	0.61	0.55

FIGURE 1
PERCENTAGE OF A-SHARE STOCKS IN THE TOP-DECILE PORTFOLIO



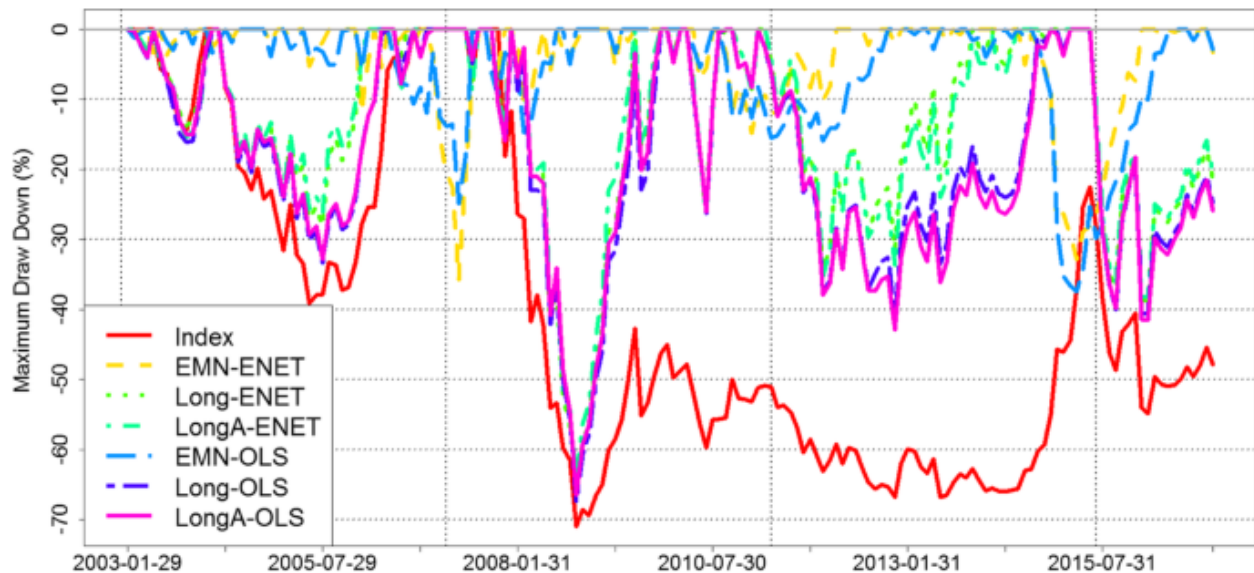
This plot represents the percentage of A-share stocks (the vertical distance of the pink bar) at each time point in the study period

FIGURE 2
CUMULATIVE RETURNS OF PORTFOLIOS



This plot contains the evolution of investing \$1 at the starting point over the entire study period by different strategies (EMN-ENET, Long-ENET, LongA-ENET, EMN-OLS, Long-OLS, and LongA-OLS). The performance of Shanghai Composite Index over the same time period is also included (red solid line) as a reference

FIGURE 3
MAXIMUM DRAWDOWN PERCENTAGE OF EACH PORTFOLIO



This plots displays the maximum drawdowns of each strategy from its previous high-watermark by different strategies (EMN-ENET, Long-ENET, LongA-ENET, EMN-OLS, Long-OLS, and LongA-OLS). The maximum drawdowns of Shanghai Composite Index over the same time period is also included (red solid line) as a reference

FIGURE 4
TIME SERIES OF THE COEFFICIENTS FOR RETURN BASED FEATURES

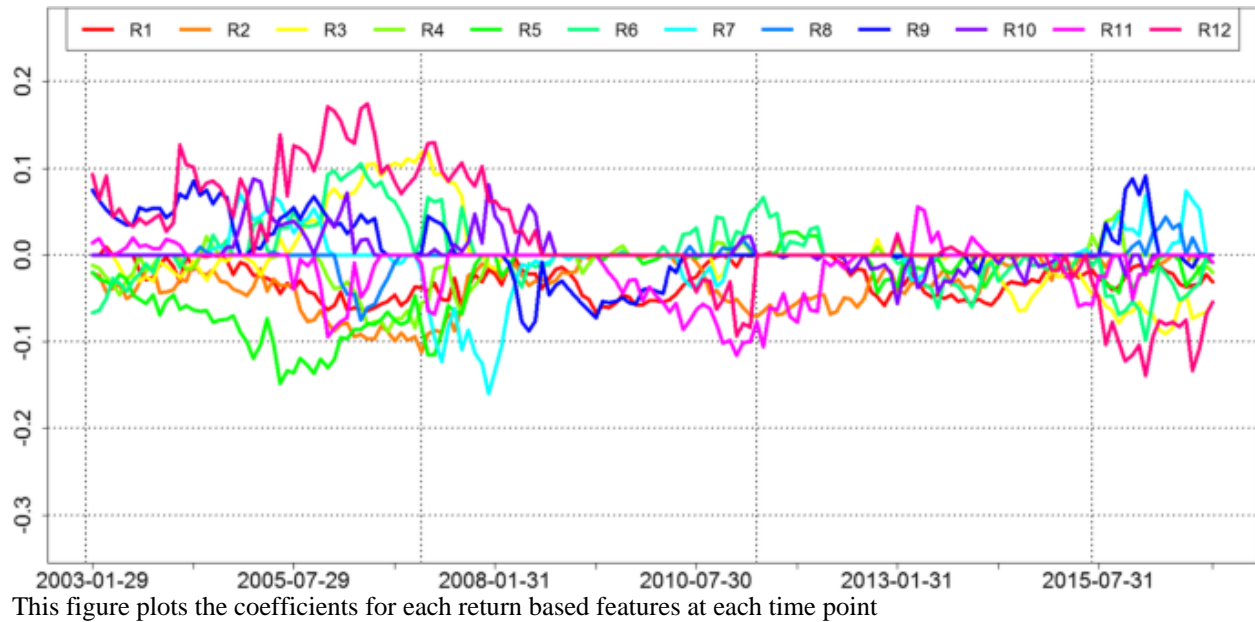
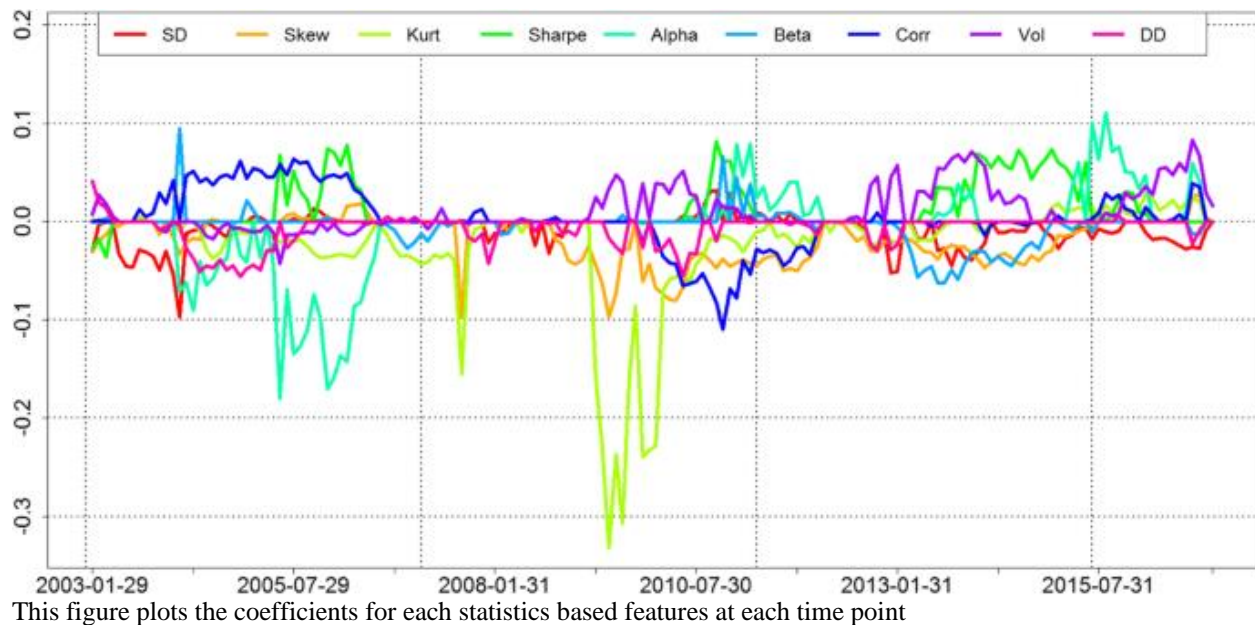


FIGURE 5
TIME SERIES OF THE COEFFICIENTS FOR STATISTICS BASED FEATURES



REFERENCES

Allen, F., Qian, J., Shan, S. C., and Zhu, J. L. (2017). Explaining the disconnection between china’s economic growth and stock market performance. Wharton and SAIF Working Paper.

Ang A, Hodrick RJ, Xing Y, Zhang X (2006) The cross-section of volatility and expected returns. J Financ 61:259–299.

Bai J, Ng S (2008) Forecasting economic time series using targeted predictors. J Econ 146(2):304–317.

Blitz D, van Vliet P (2007) The volatility effect: lower risk without lower return. J Portf Manag 34(1):102–113.

- Carhart MM (1997) On persistence in mutual fund performance. *J Financ* 7(1):57–82.
- Chen L-W, Yu H-Y (2016) Nearness to the 52-Week High and Low Prices, Past Returns, and Average Stock Returns (June 17, 2016). 29th Australasian Finance and Banking Conference 2016. Available at SSRN: <https://ssrn.com/abstract=2797149> or <http://dx.doi.org/10.2139/ssrn.2797149>.
- Efron B, Hastie T, Johnstone I (2004) Least angle regression. *Ann Stat* 10:407–499.
- Fama EF, French KR (1992) The cross-section of expected stock returns. *J Financ* 47(2):427–465.
- Fama EF, French KR (2015) A five-factor asset pricing model. *J Financ Econ* 116:1–22.
- Fama EF, MacBeth JD (1973) Commonality in the determinants of expected stock returns. *J Financ Econ* 81(3):607–636.
- Frazzini A, Pedersen LH (2014) Betting against beta. *J Financ Econ*:171–125.
- Gu, S., Kelly, B., and Xiu, D. (2019). Empirical asset pricing via machine learning. working paper.
- Harvey CR, Liu Y, Zhu H (2016) and the cross-section of expected returns. *Rev Financ Stud* 29(1):5–68.
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning: data mining*, 2nd edn. Springer, New York.
- Haugen RA, Baker NL (1996) Commonality in the determinants of expected stock returns. *J Financ Econ* 41(3):401–439.
- Hoerl AE, Kennard RW (1970) Ridge regression: applications to nonorthogonal problems. *Technometrics* 12(1):69–82.
- Hou K, Xue C, Zhang L (2015), Digesting Anomalies: An Investment Approach. *Rev Financ Stud* 28(3):650–705
- Jegadeesh N, Titman S (1990) Evidence of predictable behavior of security returns. *J Financ* 45(3):881–898.
- Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: implications for stock market efficiency. *J Financ* 48(1):65–91.
- Kang J, Liu M-H, Ni SX (2002) Contrarian and momentum strategies in the China stock market: 1993–2000. *Pac Basin Financ J* 10(3):243–265.
- Kling G, Gao L (2008) Chinese institutional investors' sentiment. *J Int Financ Mark Inst Money* 18(4):374–387.
- Li B, Hoi SCH (2015) *Online portfolio selection: principles and algorithms*. Crc Press.
- Li B, Zhang D, Zhou Y (2017) Do trend following strategies work in chinese futures markets? *J Futur Mark* 37(12):1226–1254.
- Lintner J (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev Econ Stat* 47(1):13–37.
- Mohanram PS (2005) Separating winners from losers among lowbook-to-market stocks using financial statement analysis. *Rev Acc Stud* 10:133–170.
- Montanari A, Nguyen P-M (2017) Universality of the elastic net error. In: *Information theory (ISIT)*. 2017 IEEE International Symposium, pp 2338–2342 IEEE.
- Moskowitz TJ, Ooi YH, Pedersen LH (2010) Time series momentum. *J Financ Econ* 104:228–250.
- Rapach D, Zhou G (2013) Forecasting stock returns. *Handbook of economic forecasting*:39328–39383.
- Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. *J Financ* 19(3):425–442.
- Shen W, Wang J (2017) Portfolio selection via subset resampling. *AAAI*, pp 1517–1523.
- Shen W, Wang J, Ma S (2014) Doubly regularized portfolio with risk minimization. *AAAI*, pp 1286–1292.
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Methodol* 58(1):267–288.
- Varian HR (2014) Big data: new tricks for econometrics. *J Econ Perspect* 28(2):3–28.
- Wang F, Xu Y (2004) What determines chinese stock returns? *Financ Anal J* 60(6):65–77.
- Wang L, Zhu J (2010) Financial market forecasting using a two-step kernel learning method for the support vector regression. *Ann Oper Res* 174(1):103–120.

- Welsch RE, Zhou X (2007) Application of robust statistics to asset allocation models. *Statistical J* 5(1):97–114.
- Wu W, Chen J, Yang Z(B), Tindall ML (2008) A Cross-Sectional Machine Learning Approach for Hedge Fund Return Prediction and Fund Selection (August 16, 2018). Available at SSRN: <https://ssrn.com/abstract=3238466> or <http://dx.doi.org/10.2139/ssrn.3238466>.
- Yuan R, Xiao JZ, Zou H (2008) Mutual funds' ownership and firm performance: evidence from China. *J Bank Financ* 32(8):1552–1565.
- Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Ser B* 67(2):301–320.

TRANSLATED VERSION: SPANISH

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

VERSION TRADUCIDA: ESPAÑOL

A continuación se muestra una traducción aproximada de las ideas presentadas anteriormente. Esto se hizo para dar una comprensión general de las ideas presentadas en el documento. Por favor, disculpe cualquier error gramatical y no responsabilite a los autores originales de estos errores.

INTRODUCCIÓN

El crecimiento anual del PIB de China ha promediado más del 9 por ciento desde 2000. Mientras tanto, el mercado de valores chino ha experimentado fluctuaciones sustanciales durante este período. El Shanghai Composite Index alcanzó un máximo histórico de más de 6000 en octubre de 2007, sólo hasta aproximadamente la mitad de ese nivel en la actualidad. Esta disparidad entre el rápido crecimiento del PIB y el desempeño de la bolsa de bajo nivel plantea un dilema para los inversores. ¿Cómo se puede beneficiarse de la floreciente economía china invirtiendo en el mercado de valores chino? Como han indicado Rapach y Zhou (2013), la previsión de rendimientos de las acciones es extremadamente difícil en general; esta tarea es aún más difícil cuando se trata del turbulento mercado de valores chino. (2002) estudiaron estrategias de contrario e impulso para el mercado de valores chino en años anteriores, mientras que Li et al. (2017) examinaron el desempeño de las tendencias siguiendo estrategias en los mercados chinos de futuros de materias primas. La reciente introducción de enfoques estadísticos avanzados para la gestión de carteras, como el aprendizaje automático, ha sido muy exitosa (por ejemplo, Li y Hoi 2015; Shen y Wang 2017; 2018; 2019). Este documento presenta un enfoque de aprendizaje estadístico para seleccionar acciones chinas de alto rendimiento que potencialmente pueden generar rendimientos superiores y superar significativamente al mercado.

Algunos investigadores han intentado identificar los factores subyacentes para los rendimientos de las acciones chinas. Wang y Xu (2004) descubrieron que el factor de valor no explica las diferencias transversales en el mercado bursátil chino debido a la naturaleza especulativa del mercado y la información contable de baja calidad. Por otro lado, el factor de tamaño tiene cierto poder explicativo. Kling y Gao (2008) exploraron el proceso de retroalimentación positiva entre los precios de las acciones chinas y el sentimiento de los inversores institucionales. Además, Yuan y otros (2008) encontraron que la propiedad de capital de los fondos de inversión afecta positivamente el desempeño. Todos estos tipos de información pueden contribuir al proceso de selección de stocks. Sin embargo, estos métodos benefician a inversores institucionales y no individuales, ya que estos últimos no pueden acceder fácilmente a esta información.

Dado el bajo rendimiento del mercado bursátil chino, un número cada vez mayor de instituciones e inversores están buscando acciones con altos rendimientos excesivos en comparación con el Shanghai Composite Index. Las finanzas empíricas a menudo clasifican las acciones de acuerdo a un factor de riesgo

sobre una ventana de retroceso, y las posiciones de acciones largas y cortas se toman posteriormente de acuerdo con este ranking. La literatura también ha propuesto numerosos factores (Fama y francés 1992; 2006; Frazzini y Pedersen 2014), con Harvey et al. (2016) identificando más de 200 factores de este tipo. Los procedimientos de clasificación de factores son simples y eficaces, pero tienen algunos problemas inherentes. Por ejemplo, la prima de riesgo de un factor en particular puede ser inestable con el tiempo. Fama y French (1992) constatan que su factor de valor presenta una prima de riesgo positiva a largo plazo en el mercado estadounidense; sin embargo, descubrimos que la prima de valor, estimada como la diferencia de retorno anual entre el índice de valor Russell 1000 y el índice de crecimiento Russell 1000, ha sido negativa en todas las ventanas rodantes de cinco años de 2009 a 2015.

El enfoque simple de clasificación de factores se puede ampliar para incluir múltiples factores de riesgo en el proceso de formación de cartera, pero esto crea otro problema. Si seguimos un enfoque tradicional, seleccionaríamos acciones clasificadas en el grupo superior para todos los factores de riesgo. Sin embargo, un problema de alta dimensión ocurre con muchos factores, en el que las poblaciones estarían dispersas en un espacio de la dimensión N , siendo N el número de factores que incorporaríamos. Puede ser imposible encontrar suficientes acciones en los primeros puestos de todos los factores para formar una cartera diversificada. Se podría utilizar un método ad hoc aplicando una puntuación compuesta para combinar los factores en una clasificación. Mohanram (2005) combinó factores fundamentales tradicionales para crear un índice-GSCORE. Mientras que una estrategia larga y corta basada en el GSCORE obtiene rendimientos excesivos significativos, este método carece de orientación teórica sobre cómo asignar ponderaciones entre factores, y los factores subjetivos siempre influyen en la precisión del método en la formación de una puntuación compuesta. Estos problemas hacen que el método sea inestable en la práctica. Como se mencionó anteriormente, tanto los enfoques simples como los de clasificación de múltiples factores no pueden construir eficazmente una cartera con el mayor rendimiento futuro y, por lo tanto, debemos considerar otras maneras de mejorar el rendimiento.

La regresión transversal desempeña un papel importante en las finanzas para explicar las variaciones en los precios de las acciones (Sharpe 1964; Fama y Francia 1992; Carhart 1997). Por ejemplo, Fama y Francés (1992) indican que la regresión transversal proporciona una buena descripción de los rendimientos de las carteras formadas en función de los factores de riesgo de tamaño, BE/ME y estructura a término. Esto conduce al desarrollo gradual de un método para determinar el peso de los factores a través de regresiones. Mientras que algunas obras posteriores (Hou et al. 2015; Fama y French 2015) intentaron descubrir más factores para explicar los rendimientos de las acciones utilizando modelos de regresión transversal, la precisión de las regresiones se ve socavada por los efectos de la multicolinealidad y el sobreajuste utilizando un gran número de factores.

En este documento, proponemos un modelo estadístico avanzado, la red elástica (Zou y Hastie 2005), para resolver este problema al retroceder las ganancias de las acciones sobre un gran número de factores. Nuestro método difiere de la mayoría de los métodos existentes mediante la creación de modelos de previsión transversal para las devoluciones de stock y selecciona las existencias en función de los rendimientos previstos de estos modelos. Nuestro enfoque se basa en el modelo Fama-French-Carhart (Carhart 1997), en adelante "el modelo de cuatro factores ffc", pero amplía en gran medida su alcance para incluir más factores estadísticos. Dada la baja calidad de la información contable en el mercado bursátil chino, nos centramos en la previsión de rendimientos futuros utilizando sólo los factores estadísticos derivados de los precios históricos de las acciones. Manejamos esto como un problema de aprendizaje estadístico supervisado en el que se forman portafolios de acuerdo con las clasificaciones de los rendimientos de previsión. Encontramos que las carteras mejor clasificadas generan un rendimiento robusto y superior. Nuestros métodos de previsión se desvían de los enfoques econométricos tradicionales y están más en línea con el enfoque adoptado por Varian (2014), quien argumenta que el análisis de datos en estadística y econometría se puede dividir en cuatro categorías: 1) predicción, 2) resumen, 3) estimación y 4) pruebas de hipótesis. Los métodos de aprendizaje estadístico que empleamos entran en la primera categoría, ya que nuestro objetivo es identificar las acciones con los mayores rendimientos futuros.

Este documento contribuye a la literatura existente de las siguientes maneras. En primer lugar, los enfoques de aprendizaje estadístico propuestos clasifican efectivamente y, por lo tanto, seleccionan, las

acciones de mayor rendimiento que dependen de la previsibilidad del modelo de previsión. Esto es diferente de los modelos tradicionales de regresión transversal (Fama y francés 1992; Carhart 1997). En segundo lugar, la creación de un modelo de aprendizaje estadístico proporciona una guía adaptativa de datos sobre cómo combinar información de diferentes factores (que tratamos como características) en el modelo de previsión propuesto. En tercer lugar, el uso del estimador de red elástico (Zou y Hastie 2005) mejora la precisión de estimación (precisión) del modelo de previsión cuando existe multicolinealidad y también evita el sobreajuste que se produce cuando se considera un gran número de entidades. Otra ventaja del método propuesto es su interpretabilidad y capacidad de selección de características. Por último, el marco de selección de acciones propuesto es flexible y puede ser modificado por los inversores para incluir características específicas del usuario. El resto de nuestro documento está organizado de la siguiente manera: la Sección 2 analiza nuestra motivación y explica la metodología propuesta. Posteriormente, la Sección 3 presenta los datos y características utilizados en este estudio. La Sección 4 presenta los resultados empíricos, mientras que la sección final concluye.

CONCLUSIÓN

Este estudio construyó un modelo de previsión estadística transversal para la selección de acciones en el mercado de valores chino. Sobre la base de la previsión de rendimientos acumulados futuros, el enfoque propuesto permite a los inversores identificar acciones que probablemente funcionen bien y construir carteras correspondientes. Aunque el mercado bursátil chino colectivamente no puede generar resultados satisfactorios, nuestros resultados empíricos indican que todavía es posible generar rendimientos excesivos significativos a través de un proceso activo y cuantitativo de selección de acciones. Además, y con respecto a las características utilizadas en este estudio, se observaron muchas menos características útiles durante la crisis financiera de 2008 que en otros períodos de tiempo. Mientras tanto, la ocurrencia frecuente y los signos negativos en la característica de rendimiento de un mes (R1) indican fuertemente una reversión media a corto plazo en el mercado de valores chino. Además, a pesar del hecho de que los modelos de previsión se construyen utilizando acciones A y B, invertir únicamente en acciones A de la cartera de decil superior todavía puede producir un rendimiento increíblemente bueno.

TRANSLATED VERSION: FRENCH

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

VERSION TRADUITE: FRANÇAIS

Voici une traduction approximative des idées présentées ci-dessus. Cela a été fait pour donner une compréhension générale des idées présentées dans le document. Veuillez excuser toutes les erreurs grammaticales et ne pas tenir les auteurs originaux responsables de ces erreurs.

INTRODUCTION

La croissance annuelle du PIB de la Chine a été en moyenne supérieure à 9 % depuis 2000. Pendant ce temps, le marché boursier chinois a connu des fluctuations importantes au cours de cette période. L'indice composite de Shanghai a atteint un sommet historique de plus de 6000 en octobre 2007, pour tomber à environ la moitié de ce niveau à l'heure actuelle. Cette disparité entre une croissance rapide du PIB et des performances boursières ternes pose un dilemme aux investisseurs. Comment peut-on bénéficier de l'essor de l'économie chinoise en investissant dans le marché boursier chinois? Comme Rapach et Zhou (2013) l'ont indiqué, il est extrêmement difficile en général de prévoir les rendements des actions; cette tâche est encore plus difficile lorsqu'elle implique le marché boursier chinois turbulent. Kang et coll. (2002) ont

étudié les stratégies contraires et dynamiques pour le marché boursier chinois au cours des années précédentes, tandis que Li et coll. (2017) ont examiné la performance des tendances suivant les stratégies sur les marchés chinois à terme des produits de base. L'introduction récente d'approches statistiques avancées en matière de gestion de portefeuille, comme l'apprentissage automatique, a été couronnée de succès (p. Ex. Li et Hoi, 2015; Shen et Wang 2017; Wu et coll. 2018; Gu et coll. 2019). Cet article présente une approche d'apprentissage statistique pour sélectionner les actions chinoises les plus performantes qui peuvent potentiellement générer des rendements supérieurs et surpasser considérablement le marché.

Certains chercheurs ont tenté d'identifier les facteurs sous-jacents pour les rendements boursiers chinois. Wang et Xu (2004) ont découvert que le facteur de valeur n'explique pas les différences transversales sur le marché boursier chinois en raison de la nature spéculative du marché et de l'information comptable de mauvaise qualité. D'autre part, le facteur de taille a une certaine puissance explicative. Kling et Gao (2008) ont exploré le processus de rétroaction positive entre le cours des actions chinoises et le sentiment des investisseurs institutionnels. De plus, Yuan et coll. (2008) ont constaté que la propriété en actions de fonds communs de placement a un effet positif sur le rendement. Ces types d'information peuvent tous contribuer au processus de sélection des actions. Toutefois, ces méthodes profitent aux investisseurs institutionnels plutôt qu'individuels, car ces derniers ne peuvent pas facilement accéder à ces informations.

Compte tenu de la piètre performance du marché boursier chinois, un nombre croissant d'institutions et d'investisseurs sont à la recherche d'actions à rendement excédentaire élevé par rapport à l'indice composite de Shanghai. La finance empirique classe souvent les actions en fonction d'un facteur de risque au-dessus d'une fenêtre de recul, et des positions d'actions longues et courtes sont ensuite prises selon ce classement. La littérature a également proposé de nombreux facteurs (Fama et Français 1992; Ang et coll. 2006; Frazzini et Pedersen 2014), harvey et coll. (2016) identifiant plus de 200 facteurs de ce genre. Les procédures de classement des facteurs sont simples et efficaces, mais ont des problèmes inhérents. Par exemple, la prime de risque d'un facteur particulier peut être instable au fil du temps. Fama et Français (1992) ont constaté que leur facteur de valeur présente une prime de risque positive à long terme sur le marché américain; toutefois, nous avons découvert que la prime de valeur — estimée comme la différence de rendement annuelle entre l'indice de valeur Russell 1000 et l'indice de croissance Russell 1000 — a été négative dans toutes les fenêtres roulantes sur cinq ans de 2009 à 2015.

L'approche simple de classement des facteurs peut être étendue pour inclure plusieurs facteurs de risque dans le processus de formation du portefeuille, mais cela crée un autre problème. Si nous suivons une approche traditionnelle, nous sélectionnerons les actions classées dans le groupe supérieur pour tous les facteurs de risque. Cependant, un problème de haute dimension se produit avec de nombreux facteurs, en ce que les stocks seraient dispersés peu dans un espace dimensionnel N , avec N étant le nombre de facteurs que nous incorporerions. Il peut être impossible de trouver suffisamment d'actions dans le classement supérieur de tous les facteurs pour former un portefeuille diversifié. On pourrait utiliser une méthode ad hoc en appliquant un score composite pour combiner les facteurs dans un classement. Mohanram (2005) a combiné les facteurs fondamentaux traditionnels pour créer un indice-GSCORE. Bien qu'une stratégie à long terme basée sur le GSCORE rapporte des rendements excédentaires significatifs, cette méthode manque de conseils théoriques sur la façon d'attribuer des poids entre les facteurs, et les facteurs subjectifs influencent toujours la précision de la méthode dans la formation d'un score composite. Ces problèmes rendent la méthode instable dans la pratique. Comme nous l'avons déjà mentionné, les approches simples et multiples de classement des facteurs ne peuvent pas construire efficacement un portefeuille avec le rendement futur le plus élevé, et nous devons donc envisager d'autres moyens d'améliorer le rendement.

La régression transversale joue un rôle important dans la finance pour expliquer les variations du cours des actions (Sharpe, 1964; Fama et Français 1992; Carhart, 1997). Par exemple, Fama et Français (1992) indiquent que la régression transversale fournit une bonne description des rendements des portefeuilles constitués en fonction de la taille, du BE/ME et des facteurs de risque liés à la structure à terme. Cela conduit au développement graduel d'une méthode pour déterminer le poids des facteurs par des régressions. Alors que quelques œuvres ultérieures (Hou et coll. 2015; Fama et Français 2015) ont tenté de découvrir plus de facteurs pour expliquer les rendements des actions à l'aide de modèles de régression transversale,

l'exactitude des régressions est minée par les effets de la multicollinéarité et de la suraresse en utilisant un grand nombre de facteurs.

Dans cet article, nous proposons un modèle statistique avancé, le filet élastique (Zou et Hastie, 2005), pour résoudre ce problème lors de la régression des rendements des stocks sur un grand nombre de facteurs. Notre méthode diffère de la plupart des méthodes existantes en construisant des modèles de prévision transversale pour les rendements des actions, et sélectionne les stocks en fonction des rendements prévus de ces modèles. Notre approche est enracinée dans le modèle Fama-Français-Carhart (Carhart, 1997), ci-après « le modèle à quatre facteurs FFC », mais élargit considérablement sa portée pour inclure davantage de facteurs statistiques. Compte tenu de la faible qualité de l'information comptable sur le marché boursier chinois, nous nous concentrons sur la prévision des rendements futurs en utilisant uniquement les facteurs statistiques dérivés des cours boursiers historiques. Nous traitons cela comme un problème d'apprentissage statistique supervisé dans lequel les portefeuilles sont formés en fonction du classement des rendements prévisionaux. Nous constatons que les portefeuilles les mieux classés génèrent des performances robustes et supérieures. Nos méthodes de prévision s'écartent des approches économétriques traditionnelles et sont plus conformes à l'approche adoptée par Varian (2014), qui fait valoir que l'analyse des données dans les statistiques et l'économétrie peut être divisée en quatre catégories : 1) prédiction, 2) synthèse, 3) estimation et 4) tests d'hypothèses. Les méthodes d'apprentissage statistique que nous employons entrent dans la première catégorie car notre objectif est d'identifier les stocks ayant les rendements futurs les plus élevés.

Cet article contribue à la littérature existante de la façon suivante. Premièrement, les approches d'apprentissage statistique proposées classent et sélectionnent donc efficacement les stocks les plus performants en fonction de la prévisibilité du modèle de prévision. Cela est différent des modèles traditionnels de régression transversale (Fama et Français 1992; Carhart, 1997). Deuxièmement, l'établissement d'un modèle d'apprentissage statistique fournit une orientation adaptative des données sur la façon de combiner l'information provenant de différents facteurs (que nous traitons comme des caractéristiques) dans le modèle de prévision proposé. Troisièmement, l'utilisation de l'estimateur élastique net (Zou et Hastie, 2005) améliore la précision d'estimation (précision) du modèle de prévision lorsque la multicollinéarité est présente et évite également le suraressement qui se produit lorsqu'un grand nombre de caractéristiques sont prises en compte. Un autre avantage de la méthode proposée est sa capacité d'interprétation et de sélection des fonctionnalités. Enfin, le cadre de sélection des actions proposé est flexible et peut être modifié par les investisseurs pour inclure des fonctionnalités spécifiques à l'utilisateur. Le reste de notre document est organisé comme suit : la section 2 traite de notre motivation et explique la méthodologie proposée. Par la suite, la section 3 présente les données et les caractéristiques utilisées dans cette étude. La section 4 présente les résultats empiriques, tandis que la dernière section se termine.

CONCLUSION

Cette étude a construit un modèle de prévision statistique transversale pour la sélection des actions sur le marché boursier chinois. Sur la base des prévisions de rendements cumulatifs futurs, l'approche proposée permet aux investisseurs d'identifier les actions susceptibles de bien performer et de construire des portefeuilles correspondants. Bien que le marché boursier chinois ne puisse collectivement pas générer de résultats satisfaisants, nos résultats empiriques indiquent qu'il est encore possible de générer des rendements excédentaires significatifs grâce à un processus actif et quantitatif de sélection des actions. En outre, et en ce qui concerne les caractéristiques utilisées dans cette étude, beaucoup moins de caractéristiques utiles ont été observées au cours de la crise financière de 2008 que dans d'autres périodes de temps. Pendant ce temps, l'occurrence fréquente et les signes négatifs dans la fonction de retour d'un mois (R1) indiquent fortement un retour moyen à court terme sur le marché boursier chinois. En outre, malgré le fait que les modèles de prévision sont construits à l'aide d'actions a- et b-actions, uniquement investir dans des actions A du portefeuille de déciles haut peut encore donner des performances incroyablement bonnes.

TRANSLATED VERSION: GERMAN

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

ÜBERSETZTE VERSION: DEUTSCH

Hier ist eine ungefähre Übersetzung der oben vorgestellten Ideen. Dies wurde getan, um ein allgemeines Verständnis der in dem Dokument vorgestellten Ideen zu vermitteln. Bitte entschuldigen Sie alle grammatikalischen Fehler und machen Sie die ursprünglichen Autoren nicht für diese Fehler verantwortlich.

EINLEITUNG

Chinas jährliches BIP-Wachstum hat seit dem Jahr 2000 durchschnittlich mehr als 9 % betragen. Unterdessen hat der chinesische Aktienmarkt in dieser Zeit erhebliche Schwankungen erlebt. Der Shanghai Composite Index erreichte im Oktober 2007 ein historisches Hoch von über 6000, um dann auf etwa die Hälfte dieses Niveaus zu fallen. Diese Diskrepanz zwischen dem schnell wachsenden BIP und der glanzlosen Börsenperformance stellt die Anleger vor ein Dilemma. Wie kann man von der aufkeimenden chinesischen Wirtschaft profitieren, wenn man in den chinesischen Aktienmarkt investiert? Wie Rapach und Zhou (2013) angedeutet haben, ist die Prognose der Aktienrenditen im Allgemeinen äußerst anspruchsvoll; Diese Aufgabe ist noch schwieriger, wenn es um den turbulenten chinesischen Aktienmarkt geht. Kang et al. (2002) untersuchten in früheren Jahren kontrarianische und impulsgebende Strategien für den chinesischen Aktienmarkt, während Li et al. (2017) die Entwicklung von Trends nach Strategien auf den chinesischen Rohstoffterminkmärkten untersuchten. Die jüngste Einführung fortgeschrittener statistischer Ansätze für das Portfoliomanagement, wie z. B. Maschinelles Lernen, war sehr erfolgreich (z. B. Li und Hoi 2015; Shen und Wang 2017; Wu et al. 2018; Gu et al. 2019). Dieses Papier stellt einen statistischen Lernansatz zur Auswahl der leistungsstärksten chinesischen Aktien vor, die potenziell höhere Renditen generieren und den Markt deutlich übertreffen können.

Einige Forscher haben versucht, die zugrunde liegenden Treiber für chinesische Aktienrenditen zu identifizieren. Wang und Xu (2004) entdeckten, dass der Wertfaktor die Querschnittsunterschiede am chinesischen Aktienmarkt aufgrund der spekulativen Natur des Marktes und der minderwertigen Buchhaltungsinformationen nicht erklärt. Andererseits hat der Größenfaktor eine gewisse Erklärungskraft. Kling und Gao (2008) untersuchten den positiven Feedback-Prozess zwischen chinesischen Aktienkursen und der Stimmung institutioneller Investoren. Darüber hinaus stellte Yuan et al. (2008) fest, dass sich die Beteiligung an Investmentfonds positiv auf die Performance ausgewirkt hat. Diese Arten von Informationen können alle zum Bestandsauswahlprozess beitragen. Diese Methoden kommen jedoch eher institutionellen als einzelnen Anlegern zugute, da diese nicht ohne weiteres auf diese Informationen zugreifen können.

Angesichts der schwachen Entwicklung des chinesischen Aktienmarktes suchen immer mehr Institute und Investoren nach Aktien mit hohen Überschüssen im Vergleich zum Shanghai Composite Index. Empirische Finanzierungen ordnen Aktien oft nach einem Risikofaktor über ein Blick-Back-Fenster, und Long- und Short-Aktienpositionen werden anschließend nach diesem Ranking genommen. Die Literatur hat auch zahlreiche Faktoren vorgeschlagen (Fama und Französisch 1992; Ang et al. 2006; Frazzini und Pedersen 2014), mit Harvey et al. (2016) identifiziert mehr als 200 solcher Faktoren. Factor-Ranking-Verfahren sind einfach und effektiv, haben aber einige inhärente Probleme. Beispielsweise kann die Risikoprämie eines bestimmten Faktors im Laufe der Zeit instabil sein. Fama und French (1992) stellten fest, dass ihr Wertfaktor langfristig einen positiven Risikoaufschlag auf dem US-Markt aufweist; Wir stellten jedoch fest, dass die Wertprämie – geschätzt als jährliche Renditedifferenz zwischen dem Russell 1000 Value Index und dem Russell 1000 Growth Index – in allen fünfjährigen Rollfenstern von 2009 bis 2015 negativ war.

Der einfache Faktor-Ranking-Ansatz kann erweitert werden, um mehrere Risikofaktoren in den Portfolio-Bildungsprozess einzubeziehen, aber dies schafft ein weiteres Problem. Wenn wir einem traditionellen Ansatz folgen, würden wir Aktien auswählen, die in der Top-Gruppe für alle Risikofaktoren eingestuft werden. Ein hochdimensionales Problem tritt jedoch mit vielen Faktoren auf, da die Bestände nur spärlich in einem N-dimensionalen Raum verstreut würden, wobei N die Anzahl der Faktoren ist, die wir einbeziehen würden. Es ist möglicherweise unmöglich, genügend Aktien in den Top-Rankings aller Faktoren zu finden, um ein diversifiziertes Portfolio zu bilden. Man könnte eine Ad-hoc-Methode verwenden, indem man eine zusammengesetzte Punktzahl anwendet, um die Faktoren in einem Ranking zu kombinieren. Mohanram (2005) kombinierte traditionelle fundamentale Faktoren zu einem Index-GSCORE. Während eine lange kurze Strategie, die auf dem GSCORE basiert, signifikante Überschüsse erzielt, fehlt es dieser Methode an theoretischer Anleitung, wie Gewichte zwischen Faktoren zugewiesen werden können, und subjektive Faktoren beeinflussen immer die Genauigkeit der Methode bei der Bildung einer zusammengesetzten Punktzahl. Diese Probleme machen die Methode in der Praxis instabil. Wie bereits erwähnt, können sowohl einfache als auch mehrere Faktor-Ranking-Ansätze nicht effektiv ein Portfolio mit der höchsten zukünftigen Rendite erstellen, und daher müssen wir andere Möglichkeiten zur Leistungssteigerung in Betracht ziehen.

Die Querschnittsregression spielt im Finanzwesen eine wichtige Rolle, um die Schwankungen der Aktienkurse zu erklären (Sharpe 1964; Fama und French 1992; Carhart 1997). Beispielsweise weisen Fama und French (1992) darauf hin, dass die Querschnittsregression eine gute Beschreibung der Renditen von Portfolios liefert, die auf Größen-, BE/ME- und Termstrukturrisikofaktoren basieren. Dies führt zur schrittweisen Entwicklung einer Methode zur Bestimmung des Gewichts von Faktoren durch Regressionen. Während einige nachfolgende Werke (Hou et al. 2015; Fama und French 2015) versuchten, weitere Faktoren zu entdecken, um Aktienrenditen mithilfe von Querschnittsregressionsmodellen zu erklären, die Genauigkeit der Regressionen wird durch die Auswirkungen von Multikollinearität und Überanpassung unter Verwendung einer großen Anzahl von Faktoren untergraben.

In diesem Papier schlagen wir ein erweitertes statistisches Modell vor, das elastische Netz (Zou und Hastie 2005), um dieses Problem zu lösen, wenn die Aktienrenditen über eine große Anzahl von Faktoren zurückgehen. Unsere Methode unterscheidet sich von den meisten vorhandenen Methoden durch die Erstellung von Querschnittsprognosemodellen für Aktienrenditen und wählt Aktien basierend auf den prognostizierten Renditen dieser Modelle aus. Unser Ansatz wurzelt im Fama-French-Carhart-Modell (Carhart 1997), im Folgenden "das FFC-Vier-Faktor-Modell", erweitert aber seinen Anwendungsbereich stark um weitere statistische Faktoren. Angesichts der geringen Qualität der Buchhaltungsinformationen am chinesischen Aktienmarkt konzentrieren wir uns auf die Prognose zukünftiger Renditen, wobei wir nur die statistischen Faktoren verwenden, die aus historischen Aktienkursen abgeleitet werden. Wir behandeln dies als ein beaufsichtigtes statistisches Lernproblem, bei dem Portfolios nach den Rankings der prognostizierten gebildet werden. Wir stellen fest, dass die am höchsten bewerteten Portfolios robuste und überlegene Performance generieren. Unsere Prognosemethoden weichen von traditionellen ökonometrischen Ansätzen ab und entsprechen eher dem Ansatz von Varian (2014), der argumentiert, dass Datenanalysen in Statistiken und Ökonometrie in vier Kategorien unterteilt werden können: 1) Vorhersage, 2) Zusammenfassung, 3) Schätzung und 4) Hypothesentests. Die statistischen Lernmethoden, die wir anwenden, fallen in die erste Kategorie, da unser Ziel darin besteht, die Bestände mit den höchsten zukünftigen Renditen zu identifizieren.

Dieses Papier trägt auf folgende Weise zur vorhandenen Literatur bei. Erstens ordnen die vorgeschlagenen statistischen Lernansätze effektiv die leistungsstärksten Aktien, die sich auf die Vorhersehbarkeit des Prognosemodells stützen, effektiv ein und wählen daher diese aus. Dies unterscheidet sich von herkömmlichen Querschnittsregressionsmodellen (Fama und French 1992; Carhart 1997). Zweitens bietet die Erstellung eines statistischen Lernmodells eine datenadaptive Anleitung zum Kombinieren von Informationen aus verschiedenen Faktoren (die wir als Features behandeln) im vorgeschlagenen Prognosemodell. Drittens verbessert die Verwendung des elastischen Netzschätzers (Zou und Hastie 2005) die Schätzgenauigkeit (Genauigkeit) des Prognosemodells, wenn Multikollinearität vorhanden ist, und vermeidet auch die Überanpassung, die auftritt, wenn eine große Anzahl von Features

berücksichtigt wird. Ein weiterer Vorteil der vorgeschlagenen Methode ist ihre Auslegungsfähigkeit und Featureauswahlfähigkeit. Schließlich ist der vorgeschlagene Rahmen für die Aktienausswahl flexibel und kann von Anlegern so geändert werden, dass er nutzerspezifische Merkmale enthält. Der Rest unseres Papiers ist wie folgt organisiert: Abschnitt 2 erläutert unsere Motivation und erläutert die vorgeschlagene Methodik. Anschließend werden in Abschnitt 3 die in dieser Studie verwendeten Daten und Merkmale dargestellt. Abschnitt 4 stellt die empirischen Ergebnisse dar, während der letzte Abschnitt abgeschlossen wird.

SCHLUSSFOLGERUNG

Diese Studie erstellte ein querschnittsübergreifendes statistisches Prognosemodell für die Aktienausswahl am chinesischen Aktienmarkt. Basierend auf der Prognose zukünftiger kumulativer Renditen ermöglicht der vorgeschlagene Ansatz Anlegern, Aktien zu identifizieren, die sich wahrscheinlich gut entwickeln, und entsprechende Portfolios zu konstruieren. Obwohl der chinesische Aktienmarkt insgesamt keine zufriedenstellenden Ergebnisse erzielen kann, deuten unsere empirischen Ergebnisse darauf hin, dass es immer noch möglich ist, durch einen aktiven, quantitativen Aktienausswahlprozess signifikante Überschüsse zu erzielen. Darüber hinaus wurden in Bezug auf die in dieser Studie verwendeten Merkmale während der Finanzkrise 2008 weit weniger nützliche Merkmale beobachtet als in anderen Zeiträumen. Unterdessen deuten die häufigen Auftretene und negativen Anzeichen in der einmonatigen Rendite (R1) stark auf eine kurzfristige mittelmäßige Umkehr am chinesischen Aktienmarkt hin. Trotz der Tatsache, dass die Prognosemodelle sowohl mit A- als auch mit B-Aktien erstellt werden, kann allein die Investition in A-Aktien aus dem Top-Dezil-Portfolio immer noch eine unglaublich gute Performance abwerfen.

TRANSLATED VERSION: PORTUGUESE

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

VERSÃO TRADUZIDA: PORTUGUÊS

Aqui está uma tradução aproximada das ideias acima apresentadas. Isto foi feito para dar uma compreensão geral das ideias apresentadas no documento. Por favor, desculpe todos os erros gramaticais e não responsabilize os autores originais responsáveis por estes erros.

INTRODUÇÃO

O crescimento anual do PIB da China foi de mais de 9% desde 2000. Enquanto isso, o mercado acionário chinês sofreu flutuações substanciais durante este período. O Índice Composto de Xangai atingiu uma alta histórica de mais de 6000 em outubro de 2007, apenas para cair para aproximadamente metade desse nível no momento. Essa disparidade entre o rápido crescimento do PIB e o fraco desempenho do mercado de ações coloca um dilema para os investidores. Como se pode beneficiar da crescente economia chinesa investindo no mercado de ações chinês? Como Rapach e Zhou (2013) indicaram, a previsão de retorno das ações é extremamente desafiadora em geral; essa tarefa é ainda mais difícil quando envolve o turbulento mercado de ações chinês. Kang et al. (2002) estudaram estratégias contrarianas e de impulso para o mercado de ações chinês em anos anteriores, enquanto Li et al. (2017) examinaram o desempenho das tendências seguindo estratégias nos mercados futuros de commodities chineses. A recente introdução de abordagens estatísticas avançadas para a gestão de portfólio, como o machine learning, tem sido altamente bem sucedida (por exemplo, Li e Hoi 2015; Shen e Wang 2017; Wu et al. 2018; Gu et al. 2019).

Este artigo apresenta uma abordagem de aprendizagem estatística para selecionar ações chinesas de alto desempenho que podem potencialmente gerar retornos superiores e superar significativamente o mercado.

Alguns pesquisadores tentaram identificar os drivers subjacentes para o retorno das ações chinesas. Wang e Xu (2004) descobriram que o fator valor não explica as diferenças transversais no mercado de ações chinês devido à natureza especulativa do mercado e informações contábeis de baixa qualidade. Por outro lado, o fator tamanho carrega certo poder explicativo. Kling e Gao (2008) exploraram o processo de feedback positivo entre os preços das ações chinesas e o sentimento dos investidores institucionais. Além disso, Yuan et al. (2008) descobriram que a propriedade dos fundos mútuos afeta positivamente o desempenho. Esses tipos de informações podem contribuir para o processo de seleção de ações. No entanto, esses métodos beneficiam investidores institucionais e não individuais, pois estes não podem acessar facilmente essas informações.

Dado o fraco desempenho do mercado acionário chinês, um número crescente de instituições e investidores estão procurando ações com altos retornos excedentes em comparação com o Índice Composto de Xangai. As finanças empíricas geralmente classificam as ações de acordo com um fator de risco sobre uma janela de olhar para trás, e posições de ações longas e curtas são posteriormente tomadas de acordo com este ranking. A literatura também propôs inúmeros fatores (Fama e francês de 1992; Ang et al. 2006; Frazzini e Pedersen 2014), com Harvey et al. (2016) identificando mais de 200 fatores desse tipo. Os procedimentos de classificação de fatores são simples e eficazes, mas têm alguns problemas inerentes. Por exemplo, o prêmio de risco de um fator específico pode ser instável ao longo do tempo. Fama e French (1992) descobriram que seu fator de valor apresenta um prêmio de risco positivo no longo prazo no mercado americano; no entanto, descobrimos que o prêmio de valor — estimado como a diferença de retorno anual entre o Índice de Valor Russell 1000 e o Índice de Crescimento Russell 1000 — tem sido negativo em todas as janelas de cinco anos de 2009 a 2015.

A abordagem simples de classificação de fatores pode ser estendida para incluir múltiplos fatores de risco no processo de formação de portfólio, mas isso cria outro problema. Se seguirmos uma abordagem tradicional, selecionaremos ações classificadas no grupo principal para todos os fatores de risco. No entanto, um problema de alta dimensão ocorre com muitos fatores, na medida em que os estoques seriam dispersos escassamente em um espaço N-dimensional, sendo N o número de fatores que incorporaríamos. Pode ser impossível encontrar ações suficientes no topo do ranking de todos os fatores para formar um portfólio diversificado. Pode-se usar um método ad hoc aplicando uma pontuação composta para combinar os fatores em um ranking. Mohanram (2005) combinou fatores fundamentais tradicionais para criar um índice-GSCORE. Embora uma estratégia de longo prazo baseada no GSCORE ganhe retornos excessivos significativos, este método carece de orientação teórica sobre como atribuir pesos entre fatores, e fatores subjetivos sempre influenciam a precisão do método na formação de uma pontuação composta. Esses problemas tornam o método instável na prática. Como mencionado anteriormente, tanto abordagens simples quanto múltiplas de classificação de fatores não podem efetivamente construir um portfólio com o maior retorno futuro e, portanto, devemos considerar outras maneiras de melhorar o desempenho.

A regressão transversal desempenha um papel importante nas finanças para explicar as variações nos preços das ações (Sharpe 1964; Fama e francês 1992; Carhart 1997). Por exemplo, Fama e francês (1992) indicam que a regressão transversal fornece uma boa descrição dos retornos em carteiras formadas com base no tamanho, BE/ME e fatores de risco de estrutura de prazo. Isso leva ao desenvolvimento gradual de um método para determinar o peso dos fatores através de regressões. Enquanto alguns trabalhos subsequentes (Hou et al. 2015; Fama e french 2015) tentaram descobrir mais fatores para explicar o retorno das ações usando modelos de regressão transversal, a precisão das regressões é prejudicada pelos efeitos da multicolinearidade e do excesso de adaptação usando um grande número de fatores.

Neste artigo, propomos um modelo estatístico avançado, a rede elástica (Zou e Hastie 2005), para resolver esse problema ao regressir o retorno das ações sobre um grande número de fatores. Nosso método difere da maioria dos métodos existentes, construindo modelos de previsão transversal para retornos de ações, e seleciona ações com base nos retornos previstos desses modelos. Nossa abordagem está enraizada no modelo Fama-French-Carhart (Carhart 1997), a partir de agora "o modelo de quatro fatores da FFC", mas expande consideravelmente seu escopo para incluir mais fatores estatísticos. Dada a baixa qualidade

das informações contábeis no mercado acionário chinês, focamos na previsão de retornos futuros utilizando apenas os fatores estatísticos derivados dos preços históricos das ações. Tratamos disso como um problema de aprendizagem estatística supervisionada no qual as carteiras são formadas de acordo com os rankings de retornos previstos. Descobrimos que as carteiras mais bem classificadas geram desempenho robusto e superior. Nossos métodos de previsão se desviam das abordagens econométricas tradicionais e estão mais alinhados com a abordagem tomada por Varian (2014), que argumenta que a análise de dados em estatísticas e econometria pode ser dividida em quatro categorias: 1) previsão, 2) resumição, 3) estimativa e 4) teste de hipóteses. Os métodos estatísticos de aprendizagem que empregamos se enquadram na primeira categoria, pois nosso objetivo é identificar os estoques com maiores retornos futuros.

Este artigo contribui para a literatura existente das seguintes formas. Em primeiro lugar, a aprendizagem estatística proposta se aproxima efetivamente classificar e, portanto, selecionar, as ações de melhor desempenho que dependem da previsibilidade do modelo de previsão. Isso é diferente dos modelos tradicionais de regressão transversal (Fama e francês de 1992; Carhart 1997). Em segundo lugar, a construção de um modelo de aprendizagem estatística fornece uma orientação adaptativa de dados sobre como combinar informações de diferentes fatores (que tratamos como características) no modelo de previsão proposto. Em terceiro lugar, o uso do estimador líquido elástico (Zou e Hastie 2005) melhora a precisão de estimativa (precisão) do modelo de previsão quando a multicolinearidade está presente e também evita o excesso de adaptação que ocorre quando um grande número de características são consideradas. Outra vantagem do método proposto é sua capacidade de interpretação e seleção de recursos. Por fim, o quadro de seleção de ações proposto é flexível e pode ser modificado pelos investidores para incluir recursos específicos do usuário. O restante do nosso trabalho é organizado da seguinte forma: a Seção 2 discute nossa motivação e explica a metodologia proposta. Posteriormente, a Seção 3 apresenta os dados e características utilizados neste estudo. A seção 4 apresenta os resultados empíricos, enquanto a seção final termina.

CONCLUSÃO

Este estudo construiu um modelo de previsão estatística transversal para a seleção de ações no mercado de ações chinês. Com base na previsão de retornos acumulados futuros, a abordagem proposta permite aos investidores identificar ações que provavelmente terão um bom desempenho e construir carteiras correspondentes. Embora o mercado acionário chinês coletivamente não possa gerar resultados satisfatórios, nossos resultados empíricos indicam que ainda é possível gerar retornos excessivos significativos através de um processo ativo e quantitativo de seleção de ações. Além disso, e em relação aos recursos utilizados neste estudo, muito menos recursos úteis foram observados durante a crise financeira de 2008 do que em outros períodos de tempo. Enquanto isso, a ocorrência frequente e os sinais negativos no recurso de retorno de um mês (R1) indicam fortemente uma reversão média de curto prazo no mercado de ações chinês. Além disso, apesar do fato de que os modelos de previsão são construídos usando ações A e B, investir apenas em ações A da carteira superior ainda pode produzir um desempenho incrivelmente bom.