A Study of Cross-Industry Return Predictability in the Chinese Stock Market

May 31, 2022

Abstract

We investigate cross-industry return predictability for the Shanghai and Shenzhen stock exchanges, by constructing 6- and 26- industry portfolios. The dominance of retail investors in these markets, in conjunction with the gradual diffusion of information hypothesis provide the theoretical background that allows us to employ machine learning methods to test for cross-industry predictability. We find that Oil, Telecommunications and Finance industry portfolio returns are significant predictors of other industries. Our out-of-sample forecasting exercise shows that the OLS post-LASSO estimation outperforms a variety of benchmarks and a long-short trading strategy generates an average annual excess return of 13%.

JEL Classifications: C22, C52, C53, C55, G14

Keywords: Return Predictability; Shrinkage; LASSO; Model selection; Industry Portfolio

1 Introduction

Equity markets in China over the past 20 years have experienced substantial growth with a market of $7.2tn in March 2019, making it the second largest in the world[1]. The Shanghai and Shenzhen stock markets have also undergone considerable institutional changes in their relatively short history (Huang et al., 2018). The Chinese stock market exhibits distinctive characteristics, as China prohibits most foreign investors from acquiring shares of its companies as well as the Chinese investors from participating in foreign markets, due to Chinese government regulations[2]. Moreover, Chinese stock markets are dominated by retail investors (i.e. non-professionals working through brokers), in contrast to developed markets (such as the US and the UK) that are dominated by institutional investors. For example, in 2017 retail investors in China accounted for 82% of turnover, with the same figure being less than 20% in the US[3].

These distinctive market characteristics have a number of implications for both the time series properties of stock returns as well as their potential predictability. Firstly, the domination of momentum driven retail investors suggests that these markets exhibit high volatility that one would consider difficult to explain during normal periods on other major exchanges across the globe[4]. Secondly, the high concentration of retail investors

[1] In February 2019 the index provider MSCI made the decision to increase the weight associated to Chinese stocks in its Emerging Markets index.

[2] Limits on foreign investors could slow the diffusion of information.

[3] In 2017 retail investors in China accounted for 82% of turnover and estimated to hold a third of the market's total common shares outstanding see https://www.ft.com/content/175c3afc-44c6-11e9-a965-23d669740bfb, in the US this figure is below 20% for the same year, and has increased above 20% in 2021, see, https://www.ft.com/content/7a91e3ea-b9ec-4611-9a03-a8dd3b8bddb5. Adding to this, the strong performance of the major stocks listed on the Shanghai and Shenzhen indexes in late 2019 was attributable to the majority view of retail investors that shares were undervalued - According to Margaret Yang of CMC Markets Singapore, available from https://www.ft.com/content/291f4e82-fa06-11e9-98fd-4d6c20050229. Given that the retail investors' role has increased in other markets (such as the US) studying a market which is already dominated my retail investors (China) may provide insight about the developments in other markets that move towards this direction (i.e. increased turnover by retail investors).

[4] Foucault et al. (2011), show empirically that retail trading positively impacts the volatility of stock

suggests that information spread on these markets is gradual which implies that the time taken for prices to adjust to fair values is longer. This is further exacerbated by the fact that short selling stocks is prohibited on these markets.

As is documented in Cakici et al. (2017), when attempting to forecast stock market returns in China studies have produced conflicting results that are not always in-line with expectations, and predictability is generally weak. Despite this, they find that returns in China are predictable using a number of factors such as the book-market ratio. Liu et al. (2019) construct size and value factors in China, and find that these are able to explain most of the reported anomalies in China. The authors also suggest that replicating a US model in China is questionable due to the difference in economic and financial systems. Further studies, such as, Chen et al. (2017) have found that Chinese stock returns are predictable, showing that international volatility can forecast the subsequent days' stock market returns. Interesting research from Jordan et al. (2014) reports that the returns of China's 15 largest trading partners can forecast the returns of China's A-share index in a similar vein to the research of Rapach et al. (2013). From the contrasting results that researchers have found when attempting to forecast stock returns in China, further analysis is needed in order to (i) assess whether returns are predictable in China and (ii) uncover which industry portfolios are most important when forecasting the returns of other industry portfolio's. We attempt to contribute to the literature by demonstrating return predictability in China using machine learning methods.

Hong and Stein (1999) suggest that market agents are only able to process partial information, not the entire news that hits the stock market. Especially for retail investors, investing and pricing assets would more than likely not be their core employment. Predictability may occur if investors observe and therefore processes information at different

market returns.

points in time, where the asset prices can remain incorrect until the relevant information diffuses. For industry portfolios of the Chinese stock market in our case, Hong et al. (2007) explain that investors tend to focus on segments of the market such as industry portfolios. They present evidence that supports the predictability of the market returns using industry portfolio returns. When an investor specialises in a particular industry and receives an incorporates relevant information, this may have a knock on effect for other industries which is not reflected in prices. It creates the potential for cross-industry return predictability. Hong and Stein (1999) offer an appealing theoretical framework to investigate cross-industry predictability in China due to the prevalence of retail investors in the local markets. We thus proceed in this study to employ machine learning methods that allows us to investigate stock return predictability in such a setup.

In this paper, we use monthly data from July 1997 to December 2017 to explore cross-industry return predictability both in- and out-of-sample. We use data for all Chinese A-shares listed on the Shanghai and Shenzhen exchanges[5]. To the best of our knowledge, this is the first paper to construct industry portfolios using all A-shares listed on the Shanghai and Shenzhen stock exchanges using Standard Industry Classification (SIC). This allows us to sort stocks into aggregated 6 industry portfolios and disaggregated 26 industry portfolios[6].

We employ statistical learning methods to investigate the possibility of cross-industry predictability. This introduces a flexible framework for the selection of predictors for each industry portfolio. Such a data-driven approach is naturally motivated by the high number of potential predictors (Zhang et al., 2020). Another reason is that the evident cross-industry portfolio return correlations may lead to imprecise estimates in the con-

[5]A-shares are denominated in Chinese yuan and are restricted to Chinese citizens.

[6]Industry portfolio data is currently limited in the CSMAR database post December 2017, which restricts our sample.

text of an OLS regression including all possible predictors. For the purpose of selecting predictors, we use the Least Absolute Shrinkage and Selection Operator (LASSO) of Tibshirani (1996). The use of machine learning methods in the context of portfolio return predictability models accounts for a growing body of literature (e.g. Huck, 2019; Rapach et al., 2019). We follow Belloni and Chernozhukov (2013) and Belloni et al. (2017), and use the LASSO to select the predictors for each industry portfolio return and then estimate OLS post-LASSO regressions including only the predictors chosen by the LASSO. This two-step procedure aims to alleviate any attenuation bias stemming from penalised regression methods.

In recent studies, Huck (2019) and Rapach et al. (2019) use machine learning methods and find cross-industry return predictability which is consistent with the gradual diffusion of information over economically linked industries. Another strand of the literature looks into firms specific predictability. Hou (2007) attributes evidence of return predictability to the assumption that not all firms react at the same time to common information, which conforms to the gradual information diffusion hypothesis. Specifically for China, Jiang et al. (2011) demonstrate that industry portfolios are predictable using certain macroeconomic factors.

We begin by constructing Chinese industry portfolios providing a gateway for future research[7]. Second, we reveal patterns in cross-industry predictability in China's stock market. Finally, we demonstrate the economic significance of our findings in the context of an out-of-sample forecasting exercise.

Our study provides robust statistical evidence in favour of adopting the two-step OLS post-LASSO estimation in predictive regressions for Chinese industry portfolio returns.

[7]We make data available for further use at the following link `https://sites.google.com/view/yawenzheng/data`.

Our in-sample predictive regressions uncover that Oil, Telecommunications and Finance industries are the most frequently selected predictors. Using the arguments outlined in Menzly and Ozbas (2010), we postulate that these three industries possess strong economic links with the rest and therefore intuitively possess predictive power. For example, lagged industry portfolio returns associated to the oil industry portfolio bear negative and statistically significant relationships with other industries. We argue that the predictability discovered is due to vertical supply chain links, as strong returns in the oil industry signal price rises in oil that are borne by industries further down the production process. This is consistent with Huang and Mollick (2020) and Nandha and Faff (2008) who document this finding for other countries.

Moreover, we showcase the substantial benefits of the OLS post-LASSO estimation for forecasting purposes. In particular, forecasts from these models outperform forecasts obtained from regressions containing all lagged industry return portfolios. The improvement, in terms of forecasting, is more prominent for the disaggregated 26 industry classification. We proceed to demonstrate that the results are robust in the presence of macroeconomic factors and the Elastic Net of Zou and Hastie (2005) as a variable selection procedure[8].

In our effort to investigate the economic significance of cross-industry predictability, we construct a long-short industry rotation portfolio using out-of-sample forecasts from recursive OLS post-LASSO models. We show that this portfolio generates a statistically significant 22.10% average annual excess return. This is larger than our alternative methods and also yields the highest Sharpe ratio. We then examine whether risk factors in the

[8]When weekly and daily data are used, our models far outperform a variety of benchmarks that include: i) forecasts from AR(1) models; and ii) the historical average. These results are available upon request. It is worth noting that the historical average has considerable coverage in the Finance literature with many studies documenting that it is difficult to beat (Welch and Goyal, 2008; Campbell and Thompson, 2008).

literature (e.g. Fama and French, 2015) are able to explain the variation in the long-short industry rotation portfolio. These results show that the commonly used factors cannot explain significant variation in the long-short industry rotation portfolio. They also reveal that this portfolio generates an average annual risk-adjusted return of 13.14%. We note that this cannot be explained by cross-industry momentum (Moskowitz and Grinblatt, 1999).

Our study relates both to the stock return predictability literature and the literature using statistical learning in Finance applications (e.g. Rapach et al., 2019; Huck, 2019). We rely on gradual information diffusion as an economic explanation as to why we observe predictability on these markets (Hong and Stein, 1999; Menzly and Ozbas, 2010; Hou, 2007)[9]. We provide novel insights and findings on cross-industry return predictability from Shanghai and Shenzhen stock exchanges by showing that economically related firms, in terms of the vertical supply chain, hold predictive power for other industries.

The remainder of this paper is as follows. Section 2 provides details on the construction of the industry portfolios. Section 3 discusses the econometric methodology followed in this study and Section 4 reports the empirical findings. Section 5 concludes the paper.

2 Industry Portfolio Construction

Our data is downloaded from the China Stock Market and Accounting Research (CSMAR) database. We use monthly stock data for all Chinese A-shares listed on the Shanghai and Shenzhen stock exchanges from July 1997 to December 2017. Using each firm's SIC, we sort stocks into both 6 broad industry portfolios, and more granular 26

[9]The gradual diffusion of information hypothesis is now a well established mechanism when analyzing financial markets, for example, Chen and Lu (2017) find that it plays a decisive role in explaining momentum in options markets.

industry portfolios in the spirit of Fama and French (1997)[10]. Specifically, we build value weighted portfolios at the end of June in each year and then rebalance them at the end of June the following year. We omit stocks from the portfolio if they have missing data in June of year i, and we check for data at the next time we re-balance in year $i+1$[11].

Table 1 reports the average, the standard deviation and the Sharpe ratio for the excess industry portfolio returns[12]. Panel A presents the twenty six industries classification, in which the average ranges from 0.54% of the Utilities portfolio to 1.44% of the Machinery portfolio. The standard deviations vary from 8.57% to 13.54% and the Sharpe ratio from 0.05 (Arts, Fishing, Hunting and Tourism) to 0.16 (Beverages). Panel B presents statistics for the six industries classification returns where the average ranges from 0.72% of the Public Utility portfolio to 0.87% of the Finance portfolio. The standard deviations vary from 8.44% to 10.08% and the Sharpe ratio from 0.07 (Conglomerates) to 0.10 (Finance). In this case, the range of the statistics is lower (as expected) in comparison to the twenty-six industries classification.

3 Cross-Industry Return Predictability

We examine cross-industry return predictability within a predictive regression framework. The ith industry portfolio excess return is a function of (potentially) all other lagged

[10] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data˙library.html. We initially follow directly the 30 industry portfolio definitions as in Fama and French (1997), however, Smoke, Clothes, Carry and Coal contain no constituents from the CSMAR database. Whilst the CSMAR database provides broadly defined industry portfolios (16 industries) for comparability to the research in developed stock markets and to link more closely to the industry portfolio - analyst relationship as in Hong et al. (2007), we generate industry portfolios in a more narrowly defined 26 industry portfolios.

[11] Re-balancing annually avoids considerable transaction costs.

[12] With reference to the time-series properties of the industry portfolios, we run unit root tests and reject the null hypothesis of a unit root for each of the industry portfolios. We use the Augment Dickey Fuller test, with the relevant number of lags being selected by the Akaike Information Criterion.

Table 1: Summary statistics, Industry portfolio excess returns, July 1997–December 2017 The table reports summary statistics for value-weighted monthly excess returns for each industry portfolio using stocks listed on the Shanghai and Shenzhen stock exchanges. Portfolios are constructed using the US industry portfolio definitions from the data library of Kenneth French. Returns are calculated in excess of the 1-month risk-free rate. Panel A reports summary statistics for the China A-share value-weighted 26 industry portfolios. Panel B reports summary statistics for the China A-share 6 value-weighted industry portfolios. $E(R_p) - R_f$ denotes the expected excess portfolio return in %, σ_p denotes the portfolio standard deviation in % and $\frac{E(R_p)-R_f}{\sigma_p}$ is the Sharpe Ratio of the portfolio.

Panel A: 26 Industry portfolio excess returns

ID	Definition	$E(R_p) - R_f$	σ_p	$\frac{E(R_p)-R_f}{\sigma_p}$
Food	Food Products and Agriculture	0.92	9.00	0.10
Beer	Beverages	1.36	8.67	0.16
Games	Arts, Fishing, Hunting and Tourism	0.55	10.8	0.05
Books	Printing and Publishing	1.00	10.34	0.10
Hshld	Furs, Leather, Feather, Related Products	1.27	13.54	0.09
Hlth	Heath Care, Nursing Care, Medicine, Bio Products	1.01	8.57	0.12
Chems	Chemical Products	0.75	9.31	0.08
Txtlds	Textile and wool textile industry	0.92	11.12	0.08
Cnstr	Construction and Related Products	0.60	9.29	0.06
Steel	Steel Works	1.02	11.55	0.09
FabPr	Machinery	1.44	9.84	0.15
ElcEq	Electrical Machinery and Equipment	1.09	9.31	0.12
Autos	Transportation Equipment and Fabric Products	0.90	9.77	0.09
Mines	Metal Mining	0.82	9.60	0.09
Oil	Oil and Gas Extraction	1.01	9.36	0.11
Util	Utilities	0.54	8.53	0.06
Telcm	Communication Service	0.88	10.63	0.08
Servs	Personal, Public and Business Services	0.81	9.8	0.08
BusEq	Communication and Related Equipment	0.98	9.63	0.10
Paper	Paper and Allied Products	0.81	9.59	0.08
Trans	Transportation and Support Services	0.63	8.64	0.07
Whlsl	Wholesale	0.79	9.51	0.08
Rtail	Retail Trade	0.82	9.34	0.09
Meals	Hotels	0.76	10.26	0.07
Fin	Banking, Cong., Ins., Real Est., Trading	0.81	8.54	0.09
Other	Everything Else	0.75	10.36	0.07

Panel B: 6 Industry portfolio excess returns

ID	Definition	$E(R_p) - R_f$	σ_p	$\frac{E(R_p)-R_f}{\sigma_p}$
Fin	Finance	0.87	8.78	0.10
Pub	Public Utility	0.72	8.63	0.08
Prop	Properties	0.78	9.64	0.08
Cong	Conglomerates	0.74	10.08	0.07
Ind	Industrials	0.73	8.44	0.09
Com	Commerce	0.79	9.35	0.08

industry portfolios.

$$IND_{i,t} = \delta_i + \sum_{j=1}^{N} \beta_{j,N} IND_{j,t-1} + \varepsilon_{i,t}, \tag{1}$$

where $IND_{i,t}$ is the ith industry portfolio return in excess of the risk-free rate at time t, $\beta_{j,N}$ is the coefficient associated to the jth lagged industry portfolio return. Note that in the case of N=26, 27 parameters would have to be estimated for each of the 26 regressions including all predictors. With the objective of estimating fewer parameters, we use LASSO as a regularisation method that selects only relevant predictors in each model (Tibshirani, 1996).

The objective function of LASSO for model (1) is:

$$\hat{\beta}_{LASSO} = \underset{\delta_i, \beta_i}{\operatorname{argmin}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} ||IND_{i,t} - \delta_i - \sum_{i=1}^{N} \beta_{i,N} IND_{i,t-1}||^2 + \lambda_i ||\beta_i||_1 \right\}, \tag{2}$$

with $\lambda_i \geq 0$ as the regularisation parameter. When $\lambda_i = 0$ the model reduces to an OLS regression as no penalisation is taking place. LASSO permits model coefficients to shrink to zero and yields sparse solutions in a data-based manner. We determine λ_i using a 10-fold cross-validation technique[13]. This splits the sample into 10 disjoint random subsamples using the first 9 for training and the tenth for evaluation. This process repeats 10 times and we choose λ_i with the minimum mean squared prediction error.

Notably, LASSO estimates suffer from downward bias in the magnitude which means that the penalty term tends to "overshrink" the coefficients of relevant predictors selected by LASSO (Fan and Li, 2001). Belloni and Chernozhukov (2013) and others recommend re-estimating model coefficients for the LASSO-selected predictors using OLS post-LASSO[14]. Furthermore, Belloni et al. (2017) postulate that penalised regression

[13] The main conclusions (in terms of which industry portfolios are selected as useful predictors) are robust when we switch to the 5-fold cross-validation technique.

[14] Note that inference on OLS post-LASSO coefficients constitutes post-selection inference. However,

methods cause an attenuation bias that be resolved by applying OLS to predictors selected using a variable selection technique in the first stage[15].

4 Results

4.1 In-Sample Analysis

Table 2 presents cross-industry return predictive regressions using the 6 industry classification. The predictors for each industry are selected by the LASSO and then OLS estimation is used to obtain the coefficients for each regression. There are two predictors that are significant for all industries: Fin (Finance) and Ind (Industrials). In all cases (regressions), the slope coefficient corresponding to Fin is positive, whereas the slope coefficient for Ind is negative. Table 3 presents cross-industry return predictive regressions using the 26 industry classification. Three predictors appear to be significant for most industry portfolios: Oil (Oil and Gas Extraction), Telcm (Communication Service) and Fin (Banking, Cong., Ins., Real Est., Trading) industry portfolio returns. In all cases (regressions), the slope coefficient corresponding to Telcm and Fin is positive, whereas the slope coefficient for Oil is negative.

Although the $\bar{R}^2$ values are low, Campbell and Thompson (2008) and Rapach et al. (2013) show that a monthly R^2 close to 0.5% can still provide significant predictability benefits and economically meaningful results. It is noteworthy to mention that the regression fit of our method is better than the ones appearing in the relevant literature (e.g.

as discussed at length in Rapach et al. (2019), conventional t-statistics are valid asymptotically as show by Zhao et al. (2017). We refer the interested reader to these papers.

[15] We considered analysis that allow for four lagged dependent variables in each regression in the LASSO variable selection procedure. Results are robust to those we show in the main text. We also conducted comparable analysis to those in the main text for data observed at weekly and daily frequencies. In general our results remain consistent.

Rapach et al., 2013).

The predictive ability of the financial sector is a finding of Rapach et al. (2019). Levine et al. (2003) argue that a liquid and relatively unconstrained financial system is important for the profitability of firms. Therefore strong performance of the financial industry signals positive performance of other industries within the economy. Intuitively, this makes sense as a strong financial sector signals favourable future economic conditions.

Regarding the Oil industry, Wang et al. (2019) find predictive power for stock market returns. Our results show that the Oil industry portfolio lagged returns are negatively related to the Steel, Fabric Production, Electrical Equipment, Autos, Mines, Utilities, Retail and Transport industry portfolio returns. We conjecture that the significant negative link stems from the production process and the vertical supply chain. A strong performing Oil industry suggests rising demand (and price) for Oil. Therefore, industries heavily reliant on oil for production, such as Steel or Autos incur higher costs that may not necessarily be pushed on to the other industries or firms at a later stage in the production process/tertiary sector. For retail, rising prices in Oil can raise distribution costs. Nandha and Faff (2008) and Huang and Mollick (2020) provide consistent results to ours and show that il price rises yield a negative impact on stock returns for other countries[16].

Turning our attention to the telecommunications industry portfolio[17], there are a number of reasons why this industry can predict the returns of other industries. Firstly, the telecommunications industry has strong links to economic growth. Roller and Waver-

[16]Kilian and Park (2009) note that the picture is slightly more complicated and the source of the change in the Oil industry will have different effects across sectors, an Oil-demand shock will suppress stock returns whereas an Oil-price shocks caused by positive global economic performance has a positive impact.

[17]The largest constituent of Telcm industry in China is China United Network Communications Group (China Unicom) and contributes to a significant proportion of the entire industry throughout our sample. For example, it made up 47% of the industry in December 2017, the final month of our sample.

man (2001) show that telecommunication infrastructure influences many firms. This is because telecommunication infrastructure increases communication within and between firms, hence encouraging further productive activities and generate lower transaction costs of information acquisition. Therefore, development in the telecommunications industry may lead benefits across other industries. Secondly, positive signals from the telecommunications industry suggest that firms and consumers are engaging with the industry, and this is likely to positively impact on other firms across other industries. As it is noted by Norton (1992), telecommunication expansion encourages specific industries to grow, particularly when these industries can benefit from greater telecommunication infrastructure.

In the specific case of China, the telecommunications is a strategic industry that Chinese policy makers target with measures including deregulation and allowing foreign firms into the market. As a result, direct investment results in improvements to infrastructure and facilitates economic growth (Wu, 2004).

The importance of the telecommunications industry is also reflected in the Chinese government investments, with it being part owner of many large telecommunication corporations such as China Unicom. We note that to a certain degree, each and every industry will rely on telecommunication infrastructure, whether it be for producing a product or providing payment services. Coherent with the findings of Menzly and Ozbas (2010) that economically related industries yield predictive power, we uncover predictability that may stem from Telcm industry as a result of their supplier role to other industries.

We proceed to assess the robustness of the LASSO variable selection procedure, by running a joint test of significance on the predictors deemed as insignificant in each regression. Specifically, for each industry we run an F-test of joint significance in the context of the OLS regression for the predictors that are found insignificant by the LASSO procedure. The results are presented in Table 4. For the 26 industry classification, we see that only in the case of the Beer, Autos, Mines and BusEq we have a marginal (at the 5% level) rejection of the null of insignificance of the predictors dropped by LASSO. In the 6 industry case classification, the results fully support the LASSO procedure, as the null hypothesis is not rejected in any of the cases[18].

We also apply the Elastic Net of Zou and Hastie (2005) as a variable selection procedure[19]. This is in order to account for the concerns expressed in Huck (2019) that the LASSO procedure arbitrarily selects one from a group of correlated predictors. Essentially, the Elastic Net regression is a mixture regression. The penalty term of the Elastic Net combines two types of regression penalties l_1 (LASSO), which shrinks coefficients to zero, and l_2 (ridge regression) which promotes diversity by not allowing any coefficients to be exactly zero. Therefore, Elastic Net is less parsimonious compared with the LASSO and on many occasions will select more lagged industry returns as predictors. On the whole, OLS post-Elastic Net estimation results show that similar industries are selected as relevant predictors in the sample, alleviating any concerns.

Overall, our in-sample analysis suggests that OLS post-LASSO procedure works well in finding the most relevant predictors for industry portfolio returns. The telecommunications, finance, and oil sectors appear to be the most commonly selected industries for predictive regressions. This links well with the rationale in Cohen and Frazzini (2008) and Menzly and Ozbas (2010). The novelty in our results is that telecommunications

[18] We also conduct a simulation analysis in the Appendix where we generate independent data with the same mean and standard deviation as our industry portfolio returns. We then run LASSO to check whether the procedure is driven by false positives. These results show that the procedure seldom selects any simulated industry portfolio return within predictive regressions attempting to explain our simulated data.

[19] The results are reported in the Appendix.

appears to be a strong leading indicator for industry portfolio returns. We postulate that this may be due to their prominence as a supplier to other firms in the form of payment systems and communication services.

Our findings here also support the gradual diffusion of information hypothesis in Hong and Stein (1999) corresponding with the high concentration of retail investors in China[20]. Our results also build on Huck (2019) and Rapach et al. (2019) who find evidence of gradual information diffusion across returns with economic links[21]. The implication of our in-sample analysis suggests that retail investors should monitor current conditions in telecommunications, finance and oil industry portfolios as they have predictive power for other sectors on the Shanghai and Shenzhen markets.

[20]The importance of retail investors in conjunction with the gradual diffusion of information hypothesis, is elaborated upon in Ben-Rephael et al. (2017) who find that institutional or professional attention responds more quickly to news and also lead retail attention. For the purpose of markets with a higher degree of retail investors, this suggests that the speed of information diffusion may be slower.

[21]When considering pairwise predictive regressions as in Rapach et al. (2013), we find that Oil, Telcm and Fin are frequently found to lead other industry portfolios at a higher level when compared to other pairwise connections, these results are reported in the Appendix, see Table A1. Diebold and Yilmaz (2009) measures of connectedness also portray an important role of these three industry portfolios, see table A2, the net directional connections of the industries are reported in Table A3.

Table 2: OLS Post-LASSO Regression Estimation Results: 6 Industry Classification; July 1997–December 2017 The table reports the OLS post-LASSO slope coefficient estimates. The predicted variable is the excess return for the industry portfolio in the column heading. The predictors are selected from the complete set of lagged industry excess returns. Each predictive regression model includes an intercept term. Newey-West standard errors are used for the calculation of the OLS post-LASSO t-test. – indicates that the corresponding predictor was not selected by LASSO. $\bar{R}^2$ is the adjusted R-squared in % from each regression. LM is the Godfrey (1978) test for autocorrelation. ARCH is the Engle (1982) test of time varying conditional variance. Italic font indicates rejection of the null hypothesis at the 10% significance level, bold font indicates rejection of the null hypothesis at the 5% level. No formatting indicates no rejection of the null.

Regressor	Finance	Public Utility	Properties	Conglomerates	Industrials	Commerce
Finance	0.21	*0.26*	**0.28**	*0.29*	**0.31**	*0.01*
Public Utility	-	*0.62*	0.21	0.56	0.38	0.44
Properties	**0.33**	0.2	0.81	0.11	0.2	0.23
Conglomerates	-0.04	-0.21	-	-0.15	-0.17	-0.18
Industrials	-0.42	**-0.91**	**-0.89**	**-1.01**	**-0.86**	**-1**
Commerce	-	0.21	0.35	0.37	0.3	0.41
$\bar{R}^2$ (%)	6.49	10.22	4.22	7.84	9.98	9.15
ARCH	**19.88**	**14.92**	**21.88**	**30.88**	**12.95**	**16.98**
LM	5.67	5.13	*12.24*	3.21	*11.03*	5.48

Table 3: OLS Post-LASSO Regression Estimation Results: 26 Industry Classification; July 1997–December 2017 The table reports the OLS post-LASSO slope coefficient estimates. The predicted variable is the excess return for the industry portfolio in the column heading. The predictors are selected from the complete set of lagged industry excess returns. Each predictive regression model includes an intercept term. Newey-West standard errors are used for the calculation of the OLS post-LASSO t-test. – indicates that the corresponding predictor was not selected by LASSO. $\bar{R}^2$ is the adjusted R-squared in % from each regression. LM is the Godfrey (1978) test for autocorrelation. ARCH is the Engle (1982) test of time varying conditional variance. Italic font indicates rejection of the null hypothesis at the 10% significance level, bold font indicates rejection of the null hypothesis at the 5% level. No formatting indicates no rejection of the null.

Regressor	Food	Beer	Games	Books	Hshld	Hlth	Chems	Txlds	Cnstr	Steel	Fabpr	ElcEq	Autos
Food	-	-	-	-	-	-	-	-	-	-	-	-	-
Beer	-	-	-	-	-	-	-	-	-	-	-	-	-
Games	-	-	-	-	-	-	-	-	-	-	-	-0.19	-
Books	-	-	-	-	-	-	-	-	-	-	-	-0.05	-
Hshld	-	-	-	-	-	-	-	-	-	-	-	-	-
Hlth	-	-	-	-	-	-	-	-	-	-	-	-0.15	-
Chems	-	-	-	-	-	-	-	-	-	-	-	-	-
Txtlds	-	-	-	-	-	-	-	-	-	-	-	-0.04	-
Cnstr	-	-	-	-	-	-	-	-	-	-	-	-	-
Steel	-	-	-	-	-	-	-	-	-	-	-	*0.25*	-
FabPr	-	-	-	-	-	-	-	-	-	0.08	-	**0.24**	-
ElcEq	-	-	-	-	-	-	-	-	-	-	-	-	-
Autos	-	-	-	-	0.12	-	-	-	-	-	-	**0.34**	-
Mines	-	-	-	-	-	-	-	-	-	-	-	**-0.48**	-
Oil	-	-	-	-	-	-	-	-	-	-	**-0.38**	**-0.35**	-
Util	-	-	-	-	-	-	-	-	-	-	-	-	-
Telcm	**0.2**	-	**0.3**	**0.25**	**0.27**	**0.18**	**0.22**	**0.24**	0.12	-	0.22	**0.39**	0.09
Servs	-	-	-	-	-	-	-	-	-	-	-	-	-
BusEq	-	-	-	-	-	-	-	-	-	-	-	-0.18	-
Paper	-	-	-	-	-	-	-	-	-	-	-	-	-
Trans	-	-	-	-	-	-	-	-	-	-	-	-	-
Whlsl	-	-	-	-	-	-	-	-	-	-	-	-	-
Rtail	-	-	-	-	-	-	-	-	-	-	-	-	-
Meals	-	-	-	-	-	-	-	-	-	-	-	-	-
Fin	-	-	-	-	-	-	-	-	-	0.16	**0.31**	**0.32**	0.18
Other	-	-	-	-	-	-	-	-	-	-	-	-	-
$\bar{R}^2$ (%)	5.62	-	8.82	6.39	7.45	4.93	6.19	5.23	1.93	2.8	9.86	21.3	5.1
ARCH	**28.7**	-	**26.95**	**11.4**	*5.79*	**30.44**	**28.06**	**16.75**	**23.38**	**20.43**	8.7	**14.1**	**29.72**
LM	*11.8*	-	5.42	**12.85**	0.64	**13.21**	*12.03*	8.38	*11.47*	*10.98*	1.39	8.13	*11.02*

Table 3 continued: OLS Post-LASSO Regression Estimation Results: 26 Industry Classification; July 1997–December 2017

Regressor	Mines	Oil	Util	Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Fin	Other
Food	-	-	-	-	-	-	-	-	-	-	-	-	-
Beer	-	-	-	-	-	-	-	-	-	-	-	-	-0.12
Games	-	-	-	-	-	-	-	-	-	-	-	-	-0.18
Books	-	-	-	-	-	-	-	-	-	-	-	-	-
Hshld	-	-	-	**-0.17**	-	-	-	-	-	-	-	-	-
Hlth	-	-	-	-	-	-	-	-	-	-	-	-	-0.17
Chems	-	-	-	-	-	-	-	-	-	-	-	-	-0.13
Txtlds	-	-	-	-	-	-	-	-	-	-	-	-	-
Cnstr	-	-	-	-	-	-	-	-	-	-	-	-	-
Steel	-	-	-	-	-	-	-	-	-	-	-	-	*0.26*
FabPr	-	-	-	-	-	-	-		-	-	-	-	**0.24**
ElcEq	-	-	-	-	-	-	-	-	-	-	-	-	-
Autos	-	-	-	-	-	-	0.07	-	-	-	-	-	**0.44**
Mines	-	-	-	-	-	-	-	-	-	-	-	-	**-0.55**
Oil	-	-	**-0.33**	-	-	-	-	-	-	-	-	-	**-0.35**
Util	-	-	-	-	-	-	-	-	-	-	-	-	-
Telcm	-	-	0.12	-	**0.23**	-	0.12	0.1	*0.18*	**0.18**	0.13	0.06	**0.3**
Servs	-	-	-	-	-	-	-	-	-	-	-	-	-
BusEq	-	-	-	-	-	-	-	-	-	-	-	-	-
Paper	-	-	-	-	-	-	-	-	-	-	-	-	-
Trans	-	-	-	0.19	-	-	-	-	-	-	-	-	-
Whlsl	-	-	-	-	-	-	-	-	-	-	-	-	-
Rtail	-	-	-	-	-	-	-	-	-	-	-	-	-
Meals	-	-	-	-	-	-	-	-	-	-	-	-	-
Fin	-	**0.16**	**0.31**	0.09	-	-	0.06	0.08	0.07	-	0.11	0.11	*0.33*
Other	-	-	-	-	-	-	-	-	-	-	-	-	-
$\bar{R}^2$ (%)	-	2.07	8.1	4.37	6.25	-	5.16	3.45	6.12	4.11	4.19	3.01	17.8
ARCH	-	**33.4**	**18.6**	1.9	**22.9**	-	**39.6**	**15.6**	**33.5**	**27**	**32.2**	**14.9**	**25.3**
LM	-	**19.5**	**15**	4.91	8.73	-	**13.5**	9.7	**13.6**	*10.9*	5.4	*11.4*	3.3

4.2 Out-of-Sample Results

We proceed to explore the out of sample performance of the OLS post-LASSO estimation procedure. Stock return predictability is examined in the context of a voluminous literature, with many papers documenting poor out of sample performance of predictive regression models compared to simple AR(1) models, or the historical average (Welch and Goyal, 2008; Rapach et al., 2013). Furthermore, Welch and Goyal (2008) use models that contain many factors (the so called "kitchen sink"regression) to forecast market returns as a benchmark. Rapach et al. (2010) show that these models tend to perform poorly out-of-sample.

Our first out-of-sample exercise compares our OLS post-LASSO estimation procedure against a model using all lagged industry returns as predictors. More specifically, we conduct OLS post-LASSO estimation recursively on expanding monthly windows using the first 60 months as our initial window. We forecast each industry return one-month ahead and then add the prevailing month's data in the next estimation. Note that we perform LASSO as variable selection at each recursion so the predictors have the potential to change throughout time.

Table 5 reports, for each industry portfolio return, the relative root mean squared error (RRMSE) statistic for the regression that contains all lagged industry portfolio returns relative to the OLS post-LASSO model. Panel A reports results from industry portfolios sorted according to the 26 industry classification and Panel B shows results from industry portfolios sorted according to the 6 industry classification. RRMSE values greater than 1 indicate that the OLS post-LASSO estimation procedure renders lower root mean squared error (RMSE) than the one corresponding to the OLS post-LASSO estimation procedure.

Table 4: Joint Significance Tests of Regression Coefficients of Predictive Regressions July 1997–December 2017 The table reports joint significance tests of regression coefficients associated to the lagged industry return variables in each predictive regression. Panels A and B refer to industry portfolios sorted according to the 26 and 6 industry classification definitions respectively. The first column reports the F-test with the null hypothesis being that the predictors dropped by LASSO are jointly insignificant. Bold font indicates rejection of the null at the 5% level. The second column reports the number of lagged industry returns dropped by LASSO.

Panel A: 26 industry classification		
Model	**Post-Non LASSO F-test**	**Number of Variables dropped by LASSO**
Food	1.48	25
Beer	**1.60**	26
Games	1.37	25
Books	1.49	25
Hshld	0.74	24
Hlth	1.40	25
Chems	1.29	25
Txtlds	1.43	25
Cnstr	1.24	25
Steel	1.43	24
FabPr	0.91	23
ElcEq	0.90	14
Autos	**1.58**	24
Mines	**1.61**	26
Oil	1.24	25
Util	0.86	23
Telcm	1.12	23
Servs	1.33	25
BusEq	**1.65**	25
Paper	0.97	23
Trans	1.20	24
Whlsl	1.30	24
Rtail	1.40	25
Meals	1.00	24
Fin	0.98	24
Other	0.54	15
Panel B: 6 industry classification		
Model	**Post-Non LASSO F-test**	**Number of Variables dropped by LASSO**
Finance	0.71	2
Public Utility	1.87	0
Properties	1.99	1
Conglomerates	1.49	0
Industrials	1.96	0
Commerce	1.87	0

Table 5: Out-of-sample Forecasting Results The table reports out-of-sample forecasting results from rolling OLS post-LASSO regressions relative to its corresponding OLS regression including all predictors. Panel A reports results from industry portfolio returns with the 26 industry classification and Panel B reports results from the 6 industry classification. Out-of-sample forecasts are formed recursively starting with 60 data points and forecast 1-period ahead return with LASSO performed at each recursion. We report the relative root mean squared error, RRMSE. RRMSE > 1 indicates that the OLS post-LASSO model outperforms its corresponding OLS regression; bold font highlights these cases. ***, **, * indicates that the LASSO forecast outperforms alternatives at the 1%, 5% and 10% significance level respectively using the Giacomini and White (2006) test of statistical significance.

Panel A: 26 industry return classification					
	RRMSE	p-value		RRMSE	p-value
Food	**1.08**	0.15	Mines	**1.06**	0.08**
Beer	**1.09**	0.03**	Oil	**1.14**	0.01***
Games	**1.11**	0.00***	Util	**1.11**	0.06**
Books	**1.14**	0.01***	Telcm	**1.24**	0.00***
Hshld	**1.20**	0.00***	Servs	**1.10**	0.02**
Hlth	**1.07**	0.09*	BusEq	**1.08**	0.07*
Chems	**1.10**	0.02**	Paper	**1.09**	0.06*
Txtlds	**1.13**	0.02**	Trans	**1.08**	0.03**
Cnstr	**1.08**	0.03**	Whlsl	**1.08**	0.04**
Steel	**1.09**	0.01***	Rtail	**1.09**	0.06*
FabPr	**1.15**	0.00***	Meals	**1.16**	0.00***
ElcEq	**1.07**	0.02**	Fin	**1.12**	0.00***
Autos	**1.04**	0.36	Other	**1.08**	0.05*
Panel B: 6 industry return classification					
	RRMSE	p-value		RRMSE	p-value
Finance	**1.03**	0.15	Conglomerates	0.99	0.55
Public Utility	0.97	0.16	Industrials	0.98	0.41
Properties	**1.02**	0.33	Commerce	0.96	0.04**

As can be seen from Panel A forecasts from the OLS post-LASSO procedure outperforms the models using all lagged industry portfolio returns as predictors in all cases. In particular, 12 out of 26 models generate RRMSEs greater than 1.10 with 96% of models generating RRMSEs greater than 1.05. Turning to Panel B, the results are not conclusive regarding the dominance of one of the two competing models in the out of sample context. In two out of six cases the OLS post-LASSO estimation generates a lower RMSE than the one of the "kitchen sink"regression. Note however that the RRMSE values for all models are higher than or equal to 0.96[22].

Out of a possible 4732 times an industry could be selected by our OLS post-LASSO procedure, oil, telecommunications and finance are selected 1407, 3717, and 1136 times respectively. Therefore, it is clear that these industries are leading indicators of other industry portfolio returns out-of-sample. This finding provides further substance that industry portfolio returns with economic linkages predict future industry returns out-of-sample. Our analysis also shows investors are able to benefit from conducting variable selections procedures in order to obtain more reliable out-of-sample forecasts for industry portfolio returns.

4.3 Industry Rotation Portfolios

Our second out-of-sample exercise constructs long-short industry rotations portfolios using forecasts from OLS post-LASSO regressions. This sheds light on the economic significance of cross-industry linkages, and whether they contain useful information in generating positive and significant returns. We construct our long-short industry rotation

[22] We assess OLS post-LASSO models with weekly and daily data. The results are very similar with those we report in Table 5. Forecasts generated by OLS post-LASSO models on weekly and daily data are far superior to AR(1) forecasts and also the historical average; something we do not find in monthly data. Note also that monthly, weekly and daily out-of-sample results are robust to changing the size of the initial estimation window to 8 years, and 10 years of data.

portfolio in the following manner. First we use the first 60 months of data and conduct OLS post-LASSO estimation for every industry portfolio. We then generate N 1-step ahead forecasts of each industry portfolio return and then sort the industries in ascending order according to these forecasts. We then take those industries in the 70^{th} and 30^{th} percentiles of the distribution and form equal weighted portfolios. We then go long the top percentile portfolio and short the bottom percentile portfolio. We then repeat this procedure until the end of the sample which means rebalancing occurs each month. Thus, we have a long-short industry rotation portfolio based on information in real-time and a forecast made using this information.

Table 6 reports the average annual return of the industry rotation portfolio (and its corresponding t-statistic) constructed using the OLS post-LASSO forecasts, as well as industry rotation portfolios constructed using forecasts from: i) an AR(1) model; and ii) forecasts of the historical average. We also report each portfolio's Sharpe ratio. Note that the average annual return of the industry rotation portfolio constructed using forecasts from our OLS post-LASSO models is more than twice the historical average and 5.35% larger than the rotation portfolio formed based on AR(1) forecasts. All average annual excess returns are significant at the 1% level. Further note that the industry rotation portfolio formed on OLS post-LASSO forecasts possesses a larger Sharpe ratio than the other benchmarks indicating larger excess returns for a given level of risk.

In Table 7, we now test whether risk factor exposures are able to explain the variation in our industry rotation portfolio sorted on OLS post-LASSO forecasts. We add the emerging market momentum (MOM) factor to the emerging market 5-factors of Fama and French (2015), all of which are available from Kenneth French's data library. We also present factor model results for long-short industry rotation portfolios sorted on AR(1) forecasts and the historical average. We can see that the long-short industry rotation portfolio based on OLS post-LASSO forecasts delivers an annualised risk adjusted return of 13.14% which is significant at the 1% level. Note that this is higher than the benchmark long-short rotation portfolios considered. Adding to this, the β associated to the market is 0.04 and statistically insignificant[23] which suggests that the long-short industry rotation portfolio is insulated from market movements. The only significant factor is 'high minus low' (HML) for the long-short industry rotation portfolio formed on OLS post-LASSO forecasts and also for the historical average portfolio. Note that for the AR(1) and historical average long-short industry rotation portfolios, CMA factor is also significant.

Table 6: Long-Short Industry Rotation Portfolio Descriptive Statistics from 2002–2017 The table reports the descriptive statistics from three industry rotation portfolios that are constructed in month t using different models to forecast industry returns in month $t+1$. Forecasts are made recursively using the initial 60 months to construct our first forecasts. We then add 1 month of realised data for each recursion until the end of our sample in December 2017. At each recursion, all forecasts are sorted and then we form equal weighted portfolios of those in the 70$^{\text{th}}$ and 30$^{\text{th}}$ percentiles of the distribution and construct a long-short portfolio; with rebalancing occurring monthly. The three industry rotation portfolios are constructed from: i) models using OLS post-LASSO estimation; ii) AR(1) models; and iii) historical averages. We report the average annual return, its corresponding t-statistic and each industry rotation's Sharpe ratio. In order to compute the significance of the Sharpe ratio differences across strategies, we compute rolling regressions in a framework similar to our out-of-sample period starting at a base of 60 months and storing the rolling Sharpe ratio from each of run and each strategy. We complete a t-test of equal means to compare the OLS post LASSO versus the AR(1) and historical average, we reject the null hypothesis at the 1% significance level, suggesting that the Sharpe ratio is superior for the OLS post LASSO strategy.

Portfolio Sorts on:	**OLS post-LASSO**	**AR(1)**	**Historical Ave.**
Average Annual Excess Return (%)	22.10	12.30	7.43
t-stat	[5.66]	[2.96]	[5.61]
Sharpe Ratio	0.49	0.29	0.19
Sharpe Ratio Difference	-	0.2	0.3
t-stat	-	[17.27]	[44.27]

We also compute a cross-sectional industry momentum portfolio in the spirit of

[23]Note the betas of the long portfolio and short portfolio are close to one.

Moskowitz and Grinblatt (1999) to investigate whether our long-short industry rotation portfolio formed on OLS post-LASSO forecasts is capturing cross-sectional industry momentum. This portfolio, for each month t, is constructed by sorted the previous 12 month's cumulative excess returns in ascending order and we construct a long-short portfolio based on the top 70% and bottom 30% percentiles of the cross-sectional distribution of the sorted cumulative returns over the previous year. Note regressing this portfolio on the risk factors as in Table 7 generates a statistically insignificant annualised α of 2.88%[24].

Overall, it is clear that these risk factors explain very little of the variation in our long-short industry rotation portfolio. We can also see that the long-short industry rotation portfolios formed on AR(1) forecasts and historical averages generate significant risk adjusted returns[25]. While these estimates are significant, which suggests even these portfolios generate meaningful risk adjusted returns, they are considerably lower than the risk adjusted return from the long-short portfolio formed on OLS post-LASSO forecasts. Therefore, these results highlight that signals from lagged industry portfolio returns are informative for generating risk adjusted average returns that are not captured by cross-sectional industry momentum.

In general, our out-of-sample analysis highlights the benefits of utilising OLS post-LASSO shrinkage methods when predicting future Chinese industry portfolio returns. It demonstrates how one is able to use machine learning methods to identify the most relevant predictors. While this approach is atheoretical, we provide substantial evidence that this method chooses leading indicators with theoretical underpinning (e.g. Cohen and Frazzini, 2008). We also show that trading on cross industry portfolio return forecasts

[24]These results are available on request.

[25]It is important to note that investors are willing to pay to exploit information even for marginal gains (Campbell and Thompson, 2008).

from OLS post-LASSO estimates has the ability to generate statistically significant and economically meaningful positive returns out of sample.

Taken together, these exercises indicate that gradual diffusion of information is present when examining Chinese industry portfolio returns, and that one can benefit from utilising machine learning methods out-of-sample. Interestingly, our analysis shows that our strategy generates abnormal returns even after accounting for momentum that arguably captures the behaviour of retail investors as they obtain information gradually. This shows that there is a role for shrinkage methods in order to identify economically linked industries and benefit from trading on forecasts using key leading indicators.

4.4 Inclusion of Macroeconomic Predictors

4.4.1 In-Sample Estimation

As a robustness check of our results we proceed to estimate the OLS post-LASSO models of industry portfolio returns including macroeconomic variables as predictors. This allows us to uncover the importance of economic state variables for the prediction of industry portfolio returns. The macroeconomic factors considered are the total index of industrial production, consumer price index and the M1 money supply. All variables are from the FRED economic database[26]. and we convert them into monthly % growth rates.

Tables 8 and 9 report results for industry portfolio return predictive regressions including our macroeconomic factors for the 6 and 26 industry portfolio classifications respectively. Note that macroeconomic factors are selected by the LASSO procedure in 5 out of the 6 industry portfolio regressions and 12 out of 26 industry portfolio predictive regressions. We can see that Oil, Telecommunications and Finance industries are

[26] https://fred.stlouisfed.org.

Table 7: Factor Models for Long-Short Industry Rotation Portfolios from 2002 to 2017 The table presents results for factor model on long-short industry rotation portfolios from July 2002 to December 2017. Industry rotation portfolios are constructed in month t using different models to forecast industry returns in month $t+1$. Forecasts are made recursively using the initial 60 months to construct our first forecasts. We then add 1 month of realised data for each recursion until the end of our sample in December 2017. At each recursion, all forecasts are sorted and then we form equal weighted portfolios of those in the 70$^{\text{th}}$ and 30$^{\text{th}}$ percentiles of the distribution and construct a long-short portfolio; with rebalancing occurring monthly. The three industry rotation portfolios are constructed from: i) models using OLS post-LASSO estimation; ii) AR(1) models; and iii) historical averages. We use the emerging market factors from Kenneth French's data library. MKT is the market risk premium; SMB and HML are the 'small minus big' and 'high minus low' factors; RMW and CMA are the 'robust minus weak' and conservative minus aggressive' factors; and MOM is the momentum factor. t-ratios are in square brackets and calculated using heteroscedasticity robust standard errors. In order to compute the significance of alpha differences across strategies, we compute rolling regressions in a framework similar to our out-of-sample period starting at a base of 60 months and storing the alphas (α) from each of the regressions and strategies, completing a t-test of equal means, we reject the null hypothesis at the 1% significance level for each test suggesting significantly larger alphas for the OLS post LASSO strategy.

Portfolio	**Ann. α**	**Ann. α Difference**	**MKT**	**SMB**	**HML**
OLS post-LASSO	13.14%	-	0.04	-0.02	0.4
	[4.48]	-	[0.72]	[-0.14]	[1.88]
AR(1)	11.18%	1.96%	0.03	0.02	-0.08
	[3.53]	[5.62]	[0.51]	[0.12]	[-0.47]
Historical Ave.	5.18%	7.96%	0.01	0.08	0.35
	[2.10]	[38.96]	[0.13]	[0.60]	[2.07]
Portfolio	**RMW**	**CMA**	**MOM**	R^2	
OLS post-LASSO	0.22	0.19	-0.13	5.75%	
	[0.92]	[1.05]	[-1.41]		
AR(1)	0.00	0.35	0.04	1.52%	
	[0.01]	[1.65]	[0.44]		
Historical Ave.	0.14	-0.42	0.06	4.62%	
	[0.45]	[-1.96]	[0.93]		

still the most frequently selected predictors with similar levels of significance; also signs remain consistent with our main results. Interestingly, where LASSO selects economic factors within these regressions, the adjusted R-squared statistics are relatively larger than analogous regressions we report in Table 3.

In Table 10 we report F-tests on the joint significance of regression coefficients that are comparable and conducted in the same manner as those in Table 4. Again, these results remain consistent with our baseline results suggesting that LASSO is selecting only those predictors relevant for each industry portfolio[27].

Table 8: OLS Post-LASSO Regression Estimation Results: 6 Industry Classification with Macroeconomic Factors; July 1997–December 2017 The table reports the OLS post-LASSO slope coefficient estimates. The predicted variable is the excess return for the industry portfolio in the column heading. The predictors are selected from the complete set of lagged industry excess returns. Each predictive regression model includes an intercept term. Newey-West standard errors are used for the calculation of the OLS post-LASSO t-test. – indicates that the corresponding predictor was not selected by LASSO. $\bar{R}^2$ is the adjusted R-squared in % from each regression. LM is the Godfrey (1978) test for autocorrelation. ARCH is the Engle (1982) test of time varying conditional variance. Italic font indicates rejection of the null hypothesis at the 10% significance level, bold font indicates rejection of the null hypothesis at the 5% level. No formatting indicates no rejection of the null.

Regressor	Finance	Public Utility	Properties	Conglomerates	Industrials	Commerce
Finance	0.20	**0.25**	*0.26*	**0.19**	**0.31**	0.27
Public Utility	-	*0.59*	-	-	0.38	0.42
Properties	**0.34**	0.19	0.18	-	*0.20*	0.22
Conglomerates	-	-0.11	-	-	-0.07	-0.07
Industrials	**-0.52**	**-0.98**	**-0.80**	-	**-0.96**	**-1.08**
Commerce	-	0.15	*0.43*	-	0.24	0.34
Industrial production growth	**0.17**	**0.14**	*0.12*	-	**0.16**	*0.15*
Money growth	-0.01	-0.01	*-0.01*	-	-0.01	-0.01
Inflation	0.00	*0.00*	*-0.01*	-	0.00	0.00
$\bar{R}^2$ (%)	11.28	14.04	9.65	2.64	14.44	12.65
ARCH	**19.82**	*12.31*	**24.28**	**30.45**	9.89	**15.05**
LM	2.99	4.44	10.39	3.14	7.68	4.01

[27] We run two further robustness checks. First, we include the industry portfolio returns along with the macroeconomic variables and one additional variable, the market risk premium which is the excess return of the emerging market risk factor as in Kenneth French's data library and re-complete in our analysis. Finance, Oil and Telecm remain important predictors regardless; We also include the industry portfolios alongside the emerging market risk factors (MKT, SMB, HML, RMW, MOM and CMA) from Kenneth French's data library as predictors. Again, the results that Finance, Oil and Telcm are important industries and frequently selected by LASSO when forecasting the industry portfolios are persistent.

Table 9: OLS Post-LASSO Regression Estimation Results: 26 Industry Classification with Macroeconomic Factors; July 1997–December 2017 The table reports the OLS post-LASSO slope coefficient estimates. The predicted variable is the excess return for the industry portfolio in the column heading. The predictors are selected from the complete set of lagged industry excess returns along with: Industrial production growth; Money growth; and Inflation. Each predictive regression model includes an intercept term. Newey-West standard errors are used for the calculation of the OLS post-LASSO t-test. – indicates that the corresponding predictor was not selected by LASSO. $\bar{R}^2$ is the adjusted R-squared in % from each regression. LM is the Godfrey (1978) test for autocorrelation. ARCH is the Engle (1982) test of time varying conditional variance. Italic font indicates rejection of the null hypothesis at the 10% significance level, bold font indicates rejection of the null hypothesis at the 5% level. No formatting indicates no rejection of the null.

Regressor	Food	Beer	Games	Books	Hshld	Hlth	Chems	Txlds	Cnstr	Steel	Fabpr	ElcEq	Autos
Food	-	-	-	-	-	-	-	-	-	-	-	-	-
Beer	-	-	-	-	-	-	-	-	-	-	-	-	-
Games	-	-	-	-	-	-	-	-	-	-	-	-0.19	-
Books	-	-	-	-	-	-	-	-	-	-	-	-0.07	-
Hshld	-	-	-	-	-	-	-	-	-	-	-	-	-
Hlth	-	-	-	-	-	-	-	-	-	-	-	-0.15	-
Chems	-	-	-	-	-	-	-	-	-	-	-	-0.08	**-0.37**
Txtlds	-	-	-	-	-	-	-	-	-	-	-	-	-
Cnstr	-	-	-	-	-	-	-	-	-	-	-	-	-
Steel	-	-	-	-	-	-	-	-	-	-	-	*0.24*	*0.29*
FabPr	-	-	-	-	-	-	-	-	-	0.05	-	**0.24**	-
ElcEq	-	-	-	-	-	-	-	-	-	-	-	-	-
Autos	-	-	-	-	0.12	-	-	-	-	-	-	*0.23*	-
Mines	-	-	-	-	-	-	-	-	-	-	-	**-0.5**	-
Oil	-	-	-	-	-	-	-	-	-	**-0.38**	**-0.4**	**-0.36**	**-0.35**
Util	-	-	-	-	-	-	-	-	-	-	-	-	-0.17
Telcm	**0.2**	-	**0.29**	**0.25**	**0.27**	**0.18**	**0.2**	**0.24**	**0.12**	0.1	0.21	**0.4**	**0.21**
Servs	-	-	-	-	-	-	-	-	-	-	-	-	-
BusEq	-	-	-	-	-	-	-	-	-	-	-	-	-
Paper	-	-	-	-	-	-	-	-	-	-	-	-	-
Trans	-	-	-	-	-	-	-	-	-	-	-	-	-
Whlsl	-	-	-	-	-	-	-	-	-	-	-	-	-
Rtail	-	-	-	-	-	-	-	-	-	-	-	-	-
Meals	-	-	-	-	-	-	-	-	-	-	-	-	-
Fin	-	-	-	-	-	-	-	-	-	*0.34*	**0.28**	**0.31**	**0.43**
Other	-	-	-	-	-	-	-	-	-	-	-	-	-
Industrial production growth	-	-	0.1	-	-	-	**0.14**	-	-	**0.18**	0.09	0.11	0.13
Money growth	-	-	-	-	-	-	-	-	-	-	-	-0.01	*-0.01*
Inflation	-	-	-	-	-	-	-	-	-	0	**-0.01**	0	**-0.01**
$\bar{R}^2$ (%)	5.62	-	9.75	6.39	7.45	4.93	8.85	5.23	1.93	10.9	12.53	22.66	19.67
ARCH	**28.70**	-	**23.55**	*11.40*	5.79	**30.44**	**24.27**	**16.75**	**23.38**	*10.92*	8.54	*10.66*	8.91
LM	*11.80*	-	5.78	**12.85**	0.64	**13.21**	*12.14*	8.38	*11.47*	*11.05*	1.47	8.01	7.24

Table 9 continued: OLS Post-LASSO Regression Estimation Results: 26 Industry Classification with Macroeconomic Factors; July 1997–December 2017

Regressor	Mines	Oil	Util	Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Fin	Other
Food	-	-	-	-	-	-	-	-	-	-	-	-	-
Beer	-	-	-0.09	-	-	-	-	-	-	-	-	-	-
Games	-	-	-	-	-	-	-	-	-	-	-	-	-
Books	-	-	-	-	-	-	-	-	-	-	-	-	-
Hshld	-0.02	-	-	-	-	-	-	-	-	-	-	-	-
Hlth	-	-	-	-	-	-	-	**-0.24**	-	-	-	-	-
Chems	**-0.28**	-	-0.18	-	-	-	-	-	-	-	-	-	-
Txtlds	-	-	-	-	-	-	-	-	-	-	-	-	-
Cnstr	-	-	-	-	-	-	-	-	-	-	-	-	-
Steel	-	-	-	-	-	-	-	0.13	-	-	-	-	-
FabPr	*0.19*	-	-	-	-	-	-	-	-	0.10	-	-	-
ElcEq	-	-	-	-	-	-	-	-	-	-	-	-	-
Autos	-	-	-	-	-	-	0.05	-	-	-	-	-	-
Mines	-	-	-	-	-	-	-	-	-	-	-	-	-
Oil	*-0.29*	-	**-0.33**	-	-	-	-	**-0.29**	-	**-0.23**	-	-	-
Util	-0.18	-	-	-	-	-	-	-	-	-	-	-	-
Telcm	**0.22**	-	**0.17**	-	**0.23**	-	0.14	*0.17*	**0.22**	**0.21**	**0.19**	0.05	**0.21**
Servs	-	-	-	-	-	-	-	-	-	-	-	-	-
BusEq	-	-	-	-	-	-	-	-	-	-	-	-	-
Paper	-	-	-	-	-	-	-	-	-	-	-	-	-
Trans	-	-	-	-	-	-	-	-	-	-	-	-	-
Whlsl	-	-	-	-	-	-	-	-	-	-	-	-	-
Rtail	-	-	-	-	-	-	-	-	-	-	-	-	-
Meals	-	-	-	-	-	-	-	-	-	-	-	-	-
Fin	**0.40**	-	**0.43**	-	-	-	-	*0.23*	-	-	-	0.07	-
Other	-	-	-	-	-	-	-	-	-	-	-	-	-
Industrial production growth	**0.21**	-	*0.12*	-	-	-	*0.13*	**0.13**	-	*0.11*	-	**0.11**	-
Money growth	-0.01	-	*-0.01*	-	-	-	-	**-0.01**	-	-	-	-	-
Inflation	0.00	-	*-0.01*	*-0.01*	-	-	0.00	**-0.01**	-	*0.00*	-	0.00	-
$\bar{R}^2$	16.18	-	13.74	1.76	6.25	-	7.94	14.10	5.88	10.22	3.71	6.44	4.60
ARCH	**22.39**	-	**12.44**	2.44	**22.88**	-	**37.58**	6.56	**34.82**	**21.63**	**32.07**	*10.74*	**34.45**
LM	8.57	-	9.53	3.68	8.73	-	**13.89**	7.30	**13.39**	7.85	4.97	8.69	4.83

4.4.2 Out-of-Sample Estimation

In Table 11 we report the RRMSE statistic corresponding to the RMSE comparison of the OLS post-LASSO procedure and the ”kitchen sink” regression in the presence three macroeconomic factors. The forecasting exercise setup is the same as the one in Section 4.2. We estimate the models recursively and perform 1-step ahead forecasts at each recursion. RRMSE statistics greater than 1 indicate that the OLS post-LASSO outperforms the regression containing all lagged industry return portfolios. Again, these results are consistent with those in Table 5. In particular, using the 26 industry return classification the OLS post-LASSO forecasting model outperforms the forecast regression containing all lagged industry portfolio returns. Note also that 25/26 of RRMSE are greater than 1.05 which suggests that the OLS post-LASSO forecasting model is strictly better to the chosen benchmark. In Panel B which presents the results for the 6 industry classification, the two models are indistinguishable in terms of forecasting performance, with the RRMSE ranging between 0.98 and 1.02.

On the whole, it is clear that our robustness exercises do not contradict our main findings. Therefore our conclusions hold even in the face of including macroeconomic factors into the information set and extend on Jiang et al. (2011). This provides further evidence in cross-industry return predictability both in- and out-of-sample. Even in the face of additional variables capturing economic conditions, our results show that economically linked industries act as leading indicators for other industry portfolio returns (see e.g. Cohen and Frazzini, 2008; Menzly and Ozbas, 2010).

Table 10: Joint Significance Tests of Regression Coefficients of Predictive Regressions with Macroeconomic Factors July 1997–December 2017 The table reports joint significance tests of regression coefficients associated to the lagged industry return variables and macroeconomic factors in each predictive regression. Panels A and B refer to industry portfolios sorted according to the 26 and 6 industry classification definitions respectively. The first two columns report the the F-test and corresponding p-value for those lagged industry portfolio returns and macroeconomic factors that were not selected by OLS. The second column reports the number of lagged industry returns and macroeconomic factors dropped by LASSO. – shows that no industry was selected during estimation. Bold font indicates significance at the 5% level respectively.

Panel A: 26 industry classification

Model	**Post-Non LASSO F-test**	**Number of Variables dropped by LASSO**
Food	**1.51**	28
Beer	**1.76**	29
Games	1.27	27
Books	1.49	28
Hshld	0.72	27
Hlth	1.43	28
Chems	1.21	27
Txtlds	1.40	28
Cnstr	1.26	28
Steel	1.14	23
FabPr	0.89	24
ElcEq	0.84	15
Autos	1.21	20
Mines	1.02	19
Oil	**1.59**	29
Util	0.74	21
Telcm	1.28	28
Servs	1.35	28
BusEq	**1.79**	29
Paper	1.05	25
Trans	0.95	21
Whlsl	1.51	28
Rtail	1.24	24
Meals	1.24	28
Fin	0.94	25
Other	**1.63**	28

Panel B: 6 industry classification

Model	**Post-Non LASSO F-test**	**Number of Variables dropped by LASSO**
Finance	1.24	3
Public Utility	1.87	0
Properties	1.68	2
Conglomerates	**2.53**	8
Industrials	1.96	0
Commerce	1.87	0

Table 11: Out-of-sample Forecasting Results with Macroeconomic Factors The table reports out-of-sample forecasting results from rolling OLS post-LASSO regressions relative to its corresponding OLS regression including all predictors. The macroeconomic factors we include in the LASSO selection procedure are the index of industrial production; M1 money growth; and consumer price inflation. Panel A reports results from industry portfolio returns with the 26 industry classification and Panel B reports results from the 6 industry classification. Out-of-sample forecasts are formed recursively starting with 60 data points and forecast 1-period ahead return with LASSO performed at each recursion. We report the relative root mean squared error, RRMSE. RRMSE > 1 indicates that the OLS post-LASSO model outperforms its corresponding OLS regression; bold font highlights these cases. ***, **, * indicates that the LASSO forecast outperforms alternatives at the 1%, 5% and 10% significance level respectively using the Giacomini and White (2006) test of statistical significance.

Panel A: 26 industry return classification					
	RRMSE	P-value		RRMSE	P-value
Food	**1.12**	0.06**	Mines	**1.07**	0.05**
Beer	**1.13**	0.01***	Oil	**1.18**	0.00***
Games	**1.10**	0.00***	Util	**1.15**	0.01***
Books	**1.10**	0.01**	Telcm	**1.30**	0.00***
Hshld	**1.26**	0.00***	Servs	**1.12**	0.01***
Hlth	**1.09**	0.02**	BusEq	**1.10**	0.02**
Chems	**1.12**	0.1***	Paper	**1.12**	0.01**
Txtlds	**1.13**	0.02**	Trans	**1.10**	0.01**
Cnstr	**1.10**	0.02**	Whlsl	**1.09**	0.02**
Steel	**1.07**	0.02**	Rtail	**1.08**	0.07*
FabPr	**1.16**	0.00***	Meals	**1.17**	0.00***
ElcEq	**1.09**	0.03**	Fin	**1.11**	0.01***
Autos	**1.04**	0.26	Other	**1.10**	0.03**
Panel B: 6 industry return classification					
	RRMSE	P-value		RRMSE	P-value
Finance	**1.02**	0.69	Conglomerates	0.98	0.50
Public Utility	0.98	0.52	Industrials	1.00	0.94
Properties	1.00	0.94	Commerce	1.00	0.90

5 Conclusion

This paper constructs industry portfolios using all A-stocks listed on the Shanghai and Shenzhen stock exchanges. Based on the unique characteristics of investors in these markets and the gradual diffusion of information hypothesis (Hong and Stein, 1999), we conduct an empirical investigation of cross-industry return predictability both in- and out-of-sample. We use the LASSO as a variable selection procedure and then proceed to OLS post-LASSO estimation of the parameters following recommendations in Belloni and Chernozhukov (2013) and Belloni et al. (2017).

Our results suggest that the telecommunications, finance and oil industries lead other industry portfolio returns. This links well with the economic linkages argument in Cohen and Frazzini (2008) and supplier-customer linkages in Menzly and Ozbas (2010). Out-of-sample forecasting models uncover that our OLS post-LASSO estimation procedure outperforms forecasts obtained from predictive regressions using all lagged industry portfolio returns. These results are robust not only in the presence of macroeconomic factors, but also when conducting variable selection using the Elastic Net procedure in Zou and Hastie (2005).

We highlight the economic significance of cross-industry return predictability by constructing a long-short industry rotation portfolio formed on forecasts from OLS post-LASSO models and show that this portfolio earns an average annual excess return of 13% per annum. This portfolio outperforms long-short industry rotation portfolios formed on forecasts from: i) an AR(1) process; and ii) historical averages. Notably, we document that the information content in our long-short industry rotation portfolio is unrelated to cross-industry momentum.

Our analysis has important implications for academic research and practical investing alike. We provide substantial evidence supporting the gradual diffusion of information hypothesis for Chinese industry portfolio returns. This paper shows that one can use statistical learning methods to identify predictors with linkages underpinned by theory or economic rationale to obtain more accurate point forecasts of industry portfolio returns, and earn significant risk adjusted returns by trading on forecasts from automated variable selection procedures.

Appendix

Table A1: Pairwise Granger causality test results: 26 Industry Classification; July 1997–December 2017 The table reports the least square, β_2 coefficients for the pairwise predictive regression following Rapach et al. (2013) We test the null hypothesis that the coefficient associated to industry j is equal to zero versus the alternative that it does not. ***,**,* represent rejection at the 1%, 5% and 10% levels respectively. In the rows, we report ind_j the independent variable, and in the columns we report, ind_i the dependent variable.

Regressor	Food	Beer	Games	Books	Hshld	Hlth	Chems	Txlds	Cnstr	Steel	Fabpr	ElcEq	Autos
Food	-	0.02	0.29*	0.17	0.22*	0.4***	0.28*	0.36**	0.02	-0.07	0.09	0.03	-0.16
Beer	-0.03	-	0.17	0.09	0.17	0.13	0.06	0.1	-0.12	-0.03	0.05	-0.02	-0.07
Games	-0.14	-0.02	-	-0.06	0.12	0.04	0.03	0.03	-0.01	-0.08	-0.05	-0.12	-0.17*
Books	0.05	0.02	0.22*	-	0.16	0.17	0.08	0.17	0.01	-0.05	0.06	-0.02	-0.11
Hshld	-0.05	-0.04	-0.01	-0.07	-	0.01	-0.03	0	-0.02	-0.07	-0.05	-0.06	-0.07
Hlth	-0.4***	-0.15	0.08	-0.08	0.15	-	-0.08	0.08	-0.15	-0.1	-0.03	-0.27	-0.33***
Chems	-0.19	-0.02	0.14	-0.02	0.23*	0.13	-	0.08	-0.12	-0.19	0.01	-0.22	-0.44***
Txtls	-0.1	-0.01	0.09	-0.08	0.12	0.08	0.02	-	-0.01	-0.06	0.05	-0.15	-0.22**
Cnstr	0.01	-0.02	0.23**	0.11	0.24**	0.22**	0.21*	0.21*	-	-0.07	0.09	0.1	-0.08
Steel	0.07	0.13**	0.17**	0.1	0.14*	0.17***	0.23***	0.17*	0.07	-	0.07	0.09	0.1
FabPr	0.15	0.11	0.22**	0.16	0.2*	0.27***	0.23**	0.2*	0.08	0.1	-	0.19	-0.02
ElcEq	-0.01	0.01	0.23	0.18	0.2	0.38***	0.16	0.25*	-0.05	-0.08	0.06	-	-0.22
Autos	0.17	0.08	0.37***	0.15	0.3***	0.37***	0.43***	0.35***	0.08	0.07	0.17	0.28**	-
Mines	0	0.08	0.14	0	0.16	0.14	0.16	0.03	-0.06	-0.02	-0.02	-0.04	-0.16
Oil	-0.11	-0.06	0.02	-0.03	0.05	-0.02	0	-0.03	-0.14*	-0.22**	-0.12	-0.11	-0.11
Util	-0.07	-0.05	0.14	0.08	0.19	0.18	0.12	0.06	-0.18	-0.2	0	-0.01	-0.28*
Telcm	0.23***	0.13**	0.35***	0.34***	0.31***	0.24***	0.27***	0.26***	0.12*	0.09	0.19***	0.3***	0.12
Servs	0.04	0.02	0.31**	0.18	0.17	0.3**	0.18	0.24*	-0.01	-0.03	0.08	0.04	-0.19
BusEq	-0.01	0.04	0.21	0.06	0.21*	0.26**	0.19	0.22	-0.07	-0.11	0.02	0	-0.31*
Paper	-0.05	0.04	0.11	0.02	0.18	0.17	0.29*	0.13	0	-0.07	0.04	0.04	-0.19
Trans	0.09	-0.02	0.24**	0.19*	0.26**	0.25***	0.26**	0.22	0.09	-0.14	0.1	0.13	-0.06
Whlsl	-0.07	0.02	0.14	0.02	0.18	0.28**	0.22	0.2	-0.05	-0.06	0.03	-0.04	-0.38***
Rtail	0.11	0.03	0.38***	0.19	0.27**	0.46***	0.38**	0.37**	0.06	0.04	0.15	0.13	-0.1
Meals	0.03	0.02	0.31***	0.16	0.19*	0.15	0.18	0.28**	0	0.02	0.1	0.02	-0.1
Fin	0.12	0.18*	0.25***	0.18*	0.24**	0.24***	0.26***	0.19*	0.19	0.17	0.18*	0.15	0.21*
Other	-0.05	0.02	0.26*	0.05	0.16	0.23**	0.21	0.2	-0.02	-0.01	0.04	-0.02	-0.23*

Table A1 continued: Pairwise Granger causality test results: 26 Industry Classification; July 1997–December 2017

Regressor	Mines	Oil	Util	Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Fin	Other
Food	0.03	0.02	0.1	-0.05	0.09	0.16	0.15	0.04	0.13	-0.02	0.1	-0.09	0.02
Beer	0.03	0	-0.04	-0.09	0.01	0.02	0.01	-0.08	0.01	-0.08	0.01	-0.18*	-0.09
Games	0	-0.02	0.04	-0.04	-0.19	-0.02	-0.02	0.04	-0.05	-0.14	-0.05	-0.09	-0.13
Books	-0.01	0.01	0.08	-0.06	-0.05	0.03	0.05	0.06	0.01	-0.03	0.08	-0.05	0.1
Hshld	-0.05	-0.03	0	-0.11**	-0.08	-0.05	-0.03	0	-0.03	-0.06	-0.05	-0.05	-0.02
Hlth	-0.09	-0.03	-0.03	-0.16	-0.29*	-0.12	-0.05	-0.11	-0.11	-0.5***	-0.09	-0.15*	-0.21
Chems	-0.13	-0.01	-0.03	-0.12	-0.21	-0.22	-0.05	-0.07	-0.1	-0.37**	-0.15	-0.13	-0.29
Txtls	-0.02	0	0.05	-0.07	-0.13	0	0.05	0.01	-0.06	-0.09	-0.03	-0.05	-0.02
Cnstr	-0.04	0.05	0.14	-0.06	0.09	0.14	0.13	0	0.22*	0.02	0.06	-0.14	0.11
Steel	0.14	0.12*	0.11	0.06	0.09	0.22**	0.15	0.11	0.17*	0.08	0.1	-0.02	0.09
FabPr	0.11	0.06	0.12	-0.03	0.15	0.18*	0.18*	0.09	0.16	0.16	0.16	-0.01	0.2*
ElcEq	0.02	0.02	0.09	-0.05	0.04	0.11	0.06	-0.01	0.16	-0.06	0.07	-0.07	0.16
Autos	0.15	0.07	0.26***	0.03	0.25**	0.47***	0.31***	0.14	0.46***	0.23	0.17	-0.04	0.34***
Mines	-	0.15*	0.09	0.03	0.01	0.1	0.07	0.01	0.11	-0.07	0.02	-0.17	-0.07
Oil	-0.11	-	-0.12	-0.05	-0.05	-0.04	-0.02	-0.09	-0.07	-0.1	-0.02	-0.18**	-0.13
Util	-0.14	0.07	-	-0.04	0.05	-0.04	0.05	-0.15	0.09	-0.14	-0.03	-0.16	-0.01
Telcm	0.17**	0.16***	0.18***	-	0.31***	0.21***	0.19***	0.16**	0.25***	0.19***	0.16**	0.06	0.23***
Servs	0.03	0.02	0.11	-0.08	-	0.08	0.11	0.08	0.17	-0.09	0.09	-0.05	0.15
BusEq	-0.07	0.01	0.08	-0.11	0	-	0.06	-0.01	0.12	-0.13	0.01	-0.16*	0.02
Paper	0.07	0.08	0.12	0	-0.02	0.05	-	0.09	0.08	-0.07	0.05	-0.04	0.04
Trans	0	0.12	0.18	0.12	0.13	0.15	0.19	-	0.18	0.07	0.04	-0.13	0.1
Whlsl	-0.03	0.04	0.1	-0.08	-0.03	0.12	0.09	-0.03	-	-0.13	0.02	-0.07	0.08
Rtail	0.11	0.06	0.18	-0.04	0.06	0.28	0.21	0.11	0.28*	-	0.16	-0.05	0.22
Meals	0.03	0.02	0.04	-0.06	0.02	0.12	0.15	0.05	0.14	0.08	-	-0.02	0.1
Fin	0.32***	0.34***	0.29***	0.11	0.15	0.27***	0.2**	0.21*	0.25***	0.14	0.18*	-	0.11
Other	0.04	0.07	0.09	-0.05	-0.1	0.12	0.11	0.03	0.15	-0.07	0.06	-0.04	-

Table A2: Spillover Table and Net Directional Connections, Chinese Stock Market Industry Returns, 07/1997-12/2017
The table reports the variance decomposition of a VAR(2) model with monthly returns of the industry portfolios following the Diebold and Yilmaz (2009) measures of connectedness. Total spillover reported contribution from each of the industry portfolios and contribution to each of the industry portfolios.

Regressor	Food	Beer	Games	Books	Hshld	Hlth	Chems	Txlds	Cnstr	Steel	Fabpr	ElcEq	Autos
Food	7.98	7.15	13.33	2.31	0.06	4.30	1.09	0.47	0.58	0.73	2.70	1.84	0.53
Beer	0.26	31.98	2.21	0.20	0.06	2.64	0.52	0.15	0.31	1.21	1.44	1.28	0.26
Games	0.16	7.00	35.56	0.20	0.13	0.74	0.72	0.07	0.60	0.31	0.91	1.03	1.03
Books	0.65	3.83	15.15	18.49	0.15	0.80	0.70	0.10	0.63	0.65	2.26	1.29	0.71
Hshld	0.20	5.85	15.97	2.31	38.67	1.56	0.32	0.05	0.33	0.20	0.50	0.59	0.72
Hlth	0.42	10.60	17.00	2.96	0.21	12.51	0.54	0.08	0.94	0.74	2.49	1.58	0.62
Chems	0.66	6.75	16.28	2.22	0.19	2.80	9.10	0.26	0.56	0.93	2.26	1.40	0.90
Txtlds	1.27	3.86	18.00	3.96	0.26	2.83	4.04	10.52	0.96	0.89	2.06	1.41	0.60
Cnstr	0.45	4.37	3.69	1.03	0.36	3.83	1.99	1.21	12.85	0.78	1.29	0.78	0.84
Steel	0.25	1.25	5.43	1.32	0.36	1.53	5.22	0.99	0.56	16.15	2.57	3.14	1.18
FabPr	2.45	6.64	10.35	1.77	0.55	2.43	1.52	0.45	1.44	0.88	17.70	1.18	0.20
ElcEq	0.30	7.18	12.51	2.64	0.33	3.48	2.07	0.32	1.17	1.27	3.70	6.28	0.39
Autos	0.12	3.90	9.49	2.50	0.20	3.43	2.51	0.57	1.32	2.17	3.27	1.65	8.92
Mines	0.18	1.84	4.37	2.20	0.25	1.90	4.66	0.66	0.62	2.48	2.43	2.83	1.64
Oil	0.22	3.07	2.71	0.40	0.36	1.87	0.19	0.24	0.42	1.22	0.76	0.57	1.12
Util	0.12	4.09	5.65	1.38	0.38	5.54	1.83	1.58	2.10	0.54	1.46	1.28	1.76
Telcm	0.57	2.20	2.01	0.34	1.81	0.94	0.40	0.32	0.39	0.75	0.43	0.78	0.46
Servs	0.71	6.65	17.64	2.87	0.44	2.17	1.38	0.28	0.94	0.68	2.66	2.30	0.83
BusEq	0.49	5.28	11.62	2.63	0.36	2.63	3.24	0.32	1.62	1.86	2.34	1.67	0.86
Paper	0.74	4.26	14.68	1.98	0.09	3.38	4.81	0.56	1.54	2.12	2.20	2.03	2.19
Trans	0.25	3.81	3.97	1.39	0.19	3.11	2.26	1.27	0.96	1.03	1.35	1.30	2.13
Whlsl	0.32	5.10	14.54	3.41	0.44	3.40	3.04	0.83	1.65	1.18	1.99	2.45	1.67
Rtail	0.70	6.66	12.94	3.75	0.15	3.67	1.58	0.47	0.70	0.89	2.66	1.07	0.46
Meals	0.56	4.72	17.89	1.95	0.40	2.18	3.44	0.50	0.68	0.69	2.15	1.44	0.94
Fin	0.16	1.98	1.92	0.72	0.11	3.42	0.47	0.40	0.40	0.54	2.03	0.91	0.62
Other	0.44	4.08	16.75	1.62	0.49	3.02	3.65	0.78	1.74	1.02	2.86	1.10	1.18
Contribution To Others	12.64	122.15	266.10	48.06	8.35	67.58	52.20	12.92	23.16	25.76	50.79	36.90	23.84
Contribution Including Own	20.62	154.10	301.66	66.55	47.00	80.11	61.29	23.45	36.01	41.91	68.47	43.18	32.76

Table A2 continued: Spillover Table, Chinese Stock Market Industry Returns, 07/1997-12/2017

Regressor	Mines	Oil	Util	Telcm	Servs	BusEq	Paper	Trans	Whlsl	Rtail	Meals	Fin	Other	Contribution From Others
Food	0.92	17.97	0.45	24.65	0.46	1.15	1.16	0.24	0.68	2.14	0.25	6.03	0.86	92.02
Beer	0.94	26.59	0.62	10.88	0.22	0.37	2.91	0.97	1.95	2.31	0.39	8.45	0.89	68.02
Games	1.46	11.42	1.54	23.05	0.38	1.86	1.24	0.26	1.04	2.68	0.51	5.51	0.58	64.44
Books	2.81	14.32	0.35	27.32	0.58	0.73	0.54	0.75	0.26	0.93	1.51	3.89	0.62	81.51
Hshld	0.99	11.85	0.31	8.99	1.44	0.41	0.20	0.11	0.13	1.19	0.40	6.16	0.56	61.33
Hlth	0.86	16.92	0.38	18.73	0.73	0.54	1.53	0.43	0.55	1.23	0.17	6.25	0.98	87.49
Chems	1.55	20.57	0.78	18.59	0.76	1.36	0.99	0.11	0.24	2.57	0.23	7.09	0.85	90.90
Txtlds	3.21	13.78	0.70	16.96	0.55	0.79	0.79	0.24	0.44	2.54	0.80	8.34	0.22	89.48
Cnstr	1.27	27.93	0.94	14.88	0.83	1.26	1.50	0.20	0.56	2.81	0.23	13.27	0.85	87.15
Steel	0.48	24.46	1.36	11.01	0.15	1.51	2.23	0.22	2.48	4.30	1.67	9.62	0.58	83.85
FabPr	1.52	17.72	0.98	15.40	0.51	2.51	1.79	0.41	1.08	2.10	0.49	7.60	0.34	82.30
ElcEq	1.62	19.31	0.66	23.93	0.77	1.17	1.62	0.20	0.41	1.56	0.21	6.09	0.80	93.72
Autos	1.53	20.72	1.02	16.55	0.62	1.94	1.70	0.28	1.04	3.91	0.55	9.53	0.54	91.08
Mines	8.00	30.00	1.07	12.28	0.18	1.94	2.54	0.09	0.70	2.67	1.30	12.76	0.45	92.00
Oil	0.52	71.96	0.76	3.57	0.33	1.47	1.87	0.40	0.39	1.43	0.74	2.57	0.84	28.04
Util	1.09	28.03	9.44	16.99	0.11	1.07	2.34	0.34	0.64	1.67	0.81	8.90	0.86	90.56
Telcm	0.19	23.99	0.61	54.75	0.55	1.63	1.32	1.68	0.75	0.81	0.71	0.71	0.90	45.25
Servs	1.57	15.94	0.37	28.62	5.24	0.98	0.95	0.09	0.57	1.43	0.12	3.62	0.94	94.76
BusEq	0.70	24.55	1.70	15.58	0.64	6.88	1.96	0.25	0.89	3.65	0.32	7.00	0.97	93.12
Paper	1.10	16.01	1.14	17.49	0.56	1.57	9.65	0.30	0.94	2.55	0.43	7.15	0.53	90.35
Trans	0.77	26.83	1.74	19.56	0.49	1.27	2.63	9.46	0.83	2.53	0.69	9.31	0.88	90.54
Whlsl	1.02	20.61	0.81	18.47	0.40	1.51	2.04	0.12	5.51	2.35	0.31	6.27	0.56	94.49
Rtail	1.39	20.92	0.70	22.45	1.01	1.06	1.26	0.29	0.70	5.39	0.39	8.32	0.43	94.61
Meals	1.23	12.42	1.13	17.51	1.27	1.56	1.75	1.06	0.76	2.49	11.68	9.27	0.31	88.32
Fin	0.50	36.61	0.77	9.54	0.23	1.70	1.84	0.18	0.70	2.45	1.51	29.68	0.62	70.32
Other	1.66	18.73	0.27	21.84	0.88	1.50	1.06	0.11	0.48	1.50	0.41	8.45	4.37	95.63
Contribution To Others	30.90	518.18	21.14	434.83	14.63	32.85	39.76	9.35	19.18	55.79	15.13	182.15	16.94	2141.29
Contribution Including Own	38.90	590.14	30.58	489.58	19.87	39.73	49.41	18.81	24.69	61.18	26.81	211.83	21.31	Spillover index =82.36%

Table A3: Net Directional Connections Table, Chinese Stock Market Industry Returns, 07/1997-12/2017 The table reports the net directional connections following the Diebold and Yilmaz (2009) measures of connectedness.

	Contribution From Others	Contribution To Others	Net Directional Connections
Food	92.02	12.64	-79.38
Beer	68.02	122.15	54.13
Games	64.44	266.10	201.67
Books	81.51	48.06	-33.45
Hshld	61.33	8.35	-52.97
Hlth	87.49	67.58	-19.91
Chems	90.90	52.20	-38.69
Txtlds	89.48	12.92	-76.56
Cnstr	87.15	23.16	-63.99
Steel	83.85	25.76	-58.09
FabPr	82.30	50.79	-31.51
ElcEq	93.72	36.90	-56.81
Autos	91.08	23.84	-67.25
Mines	92.00	30.90	-61.10
Oil	28.04	518.18	490.14
Util	90.56	21.14	-69.42
Telcm	45.25	434.83	389.58
Servs	94.76	14.63	-80.13
BusEq	93.12	32.85	-60.27
Paper	90.35	39.76	-50.59
Trans	90.54	9.35	-81.19
Whlsl	94.49	19.18	-75.31
Rtail	94.61	55.79	-38.82
Meals	88.32	15.13	-73.19
Fin	70.32	182.15	111.83
Other	95.63	16.94	-78.69

Simulation Analysis checking for False Positive

Here we conduct a simulation analysis to check our findings that Oil, Telcm and Fin lead other industries is not a result of data mining. In order to do so, we generate 1,000 simulations of data with the same mean and standard deviations of our industry portfolio returns. Then, for each simulation, we conduct LASSO for each of our simulated industry portfolio returns and count the proportion of times LASSO selects each industry portfolio returns in the predictive regressions.

Tables A4 and A5 show results for the 26 and 6 simulated industry portfolio returns respectively. As can be seen here, LASSO selects Oil, Telcm and Fin around 1% of the time for our 26 simulated industry portfolio returns and Fin no more than 17.2% of the time for 6 industry portfolio returns. This suggests that our procedure and results are not driven by false positives or data mining. On the whole, we can see that our simulation analysis shows LASSO rarely selects any simulated industry portfolio return within the predictive regressions.

Table A4: Simulation Analysis 26 Simulated Industry Portfolio Returns The table reports the proportion of times LASSO selects our simulated industry portfolio return data in a predictive regression that explains one of our simulated industry portfolio returns. The simulated data are independent and have the same mean and standard deviation as one of our industry portfolio returns. We run 1,000 simulations for each dataset and conduct LASSO for each simulation.

Regressor	Food	Beer	Games	Books	Hshld	Hlth	Chems	Txlds	Cnstr	Steel	Fabpr	ElcEq	Autos
Food	6.90%	5.50%	7.20%	7.70%	6.00%	7.20%	7.30%	6.40%	6.50%	7.90%	7.60%	8.00%	5.30%
Beer	5.70%	5.60%	3.90%	5.30%	4.50%	5.40%	4.70%	5.30%	3.50%	5.50%	4.80%	4.30%	7.40%
Games	3.10%	5.30%	4.10%	4.00%	5.10%	3.80%	3.70%	4.40%	4.40%	5.00%	4.70%	4.10%	3.90%
Books	2.80%	4.30%	3.90%	3.50%	4.00%	3.40%	3.40%	3.40%	3.20%	3.00%	3.20%	2.80%	3.10%
Hshld	2.80%	2.90%	3.70%	2.30%	3.20%	3.90%	3.10%	2.70%	3.60%	2.40%	1.70%	3.00%	3.30%
Hlth	2.60%	2.00%	2.80%	3.40%	2.10%	1.80%	2.80%	2.40%	2.40%	1.70%	2.50%	2.10%	2.10%
Chems	1.70%	2.10%	2.00%	1.70%	2.50%	1.80%	2.80%	1.80%	2.80%	1.80%	2.70%	1.70%	2.30%
Txtls	1.60%	3.00%	1.00%	2.90%	2.70%	1.70%	1.30%	1.50%	1.70%	2.90%	2.60%	2.70%	2.10%
Cnstr	1.60%	1.20%	2.20%	0.90%	1.80%	1.80%	1.30%	2.20%	1.90%	2.40%	1.50%	1.20%	1.20%
Steel	2.90%	2.00%	1.00%	1.90%	1.80%	1.30%	2.00%	1.20%	2.50%	1.20%	1.30%	2.00%	1.80%
FabPr	1.60%	1.90%	1.60%	1.40%	2.20%	1.20%	1.30%	1.50%	1.70%	1.00%	2.20%	1.30%	1.90%
ElcEq	1.70%	1.60%	1.10%	2.10%	1.30%	1.60%	1.60%	1.30%	0.60%	1.20%	1.40%	1.50%	1.40%
Autos	1.40%	1.50%	0.80%	1.50%	1.00%	1.20%	1.10%	0.80%	0.80%	1.10%	1.60%	1.20%	0.90%
Mines	1.40%	0.80%	1.40%	1.20%	1.30%	1.30%	1.70%	1.40%	1.40%	1.20%	1.60%	1.00%	1.30%
Oil	1.50%	0.80%	2.00%	1.00%	1.50%	1.60%	1.20%	2.00%	1.10%	0.90%	1.60%	0.50%	1.90%
Util	0.80%	0.90%	0.80%	1.70%	0.80%	0.90%	1.00%	0.50%	0.80%	0.60%	1.10%	1.00%	0.50%
Telcm	1.00%	0.90%	1.00%	1.00%	1.20%	1.10%	0.60%	0.70%	1.00%	1.00%	1.10%	1.40%	1.60%
Servs	1.30%	1.60%	1.20%	0.80%	1.00%	1.00%	0.70%	1.50%	0.80%	0.70%	1.00%	0.70%	0.50%
BusEq	1.80%	0.70%	0.90%	1.20%	0.90%	0.90%	0.90%	0.80%	0.80%	0.90%	1.60%	1.50%	0.50%
Paper	0.60%	1.20%	1.00%	1.00%	1.00%	0.80%	0.90%	1.20%	1.50%	2.00%	0.60%	0.80%	1.30%
Trans	0.60%	0.90%	0.80%	0.70%	1.10%	1.10%	0.90%	0.70%	0.90%	0.80%	0.90%	1.60%	0.50%
Whlsl	1.50%	1.90%	0.60%	1.20%	1.20%	0.80%	1.50%	0.70%	1.10%	1.00%	1.00%	0.60%	0.90%
Rtail	0.60%	1.30%	1.20%	0.90%	0.80%	1.10%	0.90%	1.60%	0.90%	0.40%	1.10%	1.60%	0.80%
Meals	0.90%	0.50%	1.00%	1.10%	1.10%	1.40%	0.80%	0.90%	1.30%	0.70%	0.80%	1.00%	1.60%
Fin	0.60%	0.90%	0.90%	0.70%	0.70%	0.70%	0.60%	0.50%	0.50%	1.10%	0.40%	1.00%	0.50%
Other	0.20%	1.00%	0.90%	1.20%	0.40%	1.00%	0.40%	0.60%	1.30%	0.80%	0.80%	1.50%	1.70%

Table A4 Continued: Simulation Analysis 26 Simulated Industry Portfolio Returns The table reports the proportion of times LASSO selects our simulated industry portfolio return data in a predictive regression that explains one of our simulated industry portfolio returns. The simulated data are independent and have the same mean and standard deviation as one of our industry portfolio returns. We run 1,000 simulations for each dataset and conduct LASSO for each simulation.

Regressor	Food	Beer	Games	Books	Hshld	Hlth	Chems	Txlds	Cnstr	Steel	Fabpr	ElcEq	Autos
Food	7.10%	6.70%	6.80%	6.50%	6.30%	5.90%	6.30%	5.90%	7.60%	6.00%	5.00%	7.30%	6.20%
Beer	5.30%	6.50%	5.20%	5.20%	4.30%	5.20%	4.60%	4.70%	4.70%	5.40%	4.70%	4.80%	6.10%
Games	3.30%	3.60%	3.50%	4.20%	4.10%	4.20%	3.50%	4.70%	4.50%	4.60%	4.00%	4.90%	5.50%
Books	2.40%	3.70%	3.60%	3.80%	2.60%	2.50%	2.80%	4.90%	4.10%	3.60%	4.10%	3.20%	4.50%
Hshld	2.50%	2.50%	2.90%	4.10%	1.70%	3.40%	3.50%	3.70%	3.10%	3.20%	3.10%	2.30%	2.90%
Hlth	2.80%	2.40%	2.00%	3.70%	2.60%	3.90%	2.60%	3.20%	2.40%	1.90%	3.10%	2.10%	1.90%
Chems	2.80%	2.70%	2.30%	2.30%	2.30%	1.80%	2.30%	2.60%	2.20%	3.60%	2.80%	2.90%	2.20%
Txtls	1.90%	2.30%	2.20%	2.40%	2.10%	1.80%	2.40%	2.00%	2.70%	2.70%	1.90%	2.20%	2.20%
Cnstr	2.20%	1.60%	2.00%	1.60%	2.10%	2.00%	1.50%	1.90%	1.40%	2.00%	2.20%	2.20%	2.10%
Steel	1.60%	0.90%	1.50%	1.00%	1.80%	1.40%	1.70%	1.90%	1.30%	1.80%	1.90%	1.20%	2.40%
FabPr	1.60%	1.00%	1.50%	1.40%	1.50%	1.40%	1.60%	1.10%	1.30%	1.90%	1.90%	1.00%	1.10%
ElcEq	1.50%	1.50%	2.00%	1.00%	1.90%	1.10%	1.70%	1.00%	1.70%	1.50%	1.00%	1.30%	1.40%
Autos	1.90%	0.80%	1.60%	1.00%	1.30%	1.00%	1.10%	1.90%	1.30%	2.00%	1.60%	1.00%	1.30%
Mines	1.00%	1.10%	0.90%	2.10%	0.80%	1.40%	1.10%	1.00%	1.30%	1.90%	1.60%	1.40%	1.00%
Oil	1.80%	1.60%	1.60%	1.00%	1.20%	1.50%	0.80%	0.90%	0.80%	1.20%	1.10%	1.00%	0.90%
Util	1.40%	0.90%	1.00%	1.00%	1.10%	0.80%	1.60%	0.50%	1.10%	1.60%	1.70%	1.30%	1.40%
Telcm	1.40%	1.00%	0.90%	1.50%	1.80%	1.60%	1.40%	0.70%	1.00%	1.10%	1.30%	1.40%	1.00%
Servs	1.10%	0.90%	1.10%	1.00%	1.10%	0.60%	1.00%	0.90%	1.00%	1.00%	1.30%	1.10%	1.00%
BusEq	1.20%	1.40%	0.80%	1.60%	1.00%	1.10%	1.40%	1.40%	0.80%	1.00%	1.30%	1.50%	1.10%
Paper	0.70%	1.10%	0.70%	0.60%	1.20%	1.00%	0.90%	0.70%	0.60%	0.90%	1.00%	1.20%	1.20%
Trans	0.90%	0.80%	1.10%	1.10%	0.70%	1.00%	0.90%	0.80%	0.50%	0.70%	1.00%	1.00%	1.10%
Whlsl	0.70%	0.70%	0.90%	0.50%	1.40%	1.00%	0.70%	0.70%	0.80%	0.70%	1.20%	0.80%	0.70%
Rtail	1.40%	0.60%	0.70%	0.40%	1.00%	1.30%	0.80%	0.80%	1.20%	1.10%	1.10%	1.40%	0.70%
Meals	0.50%	1.10%	1.10%	1.10%	0.80%	0.90%	0.70%	0.90%	0.60%	0.70%	1.70%	0.50%	0.60%
Fin	0.70%	1.50%	1.00%	1.30%	0.50%	0.70%	0.40%	1.50%	1.20%	0.80%	0.60%	0.70%	0.60%
Other	0.80%	1.20%	1.50%	1.10%	0.40%	0.80%	1.40%	1.00%	0.50%	0.70%	0.50%	1.00%	0.60%

Table A5: Simulation Analysis 6 Simulated Industry Portfolio Returns The table reports the proportion of times LASSO selects our simulated industry portfolio return data in a predictive regression that explains one of our simulated industry portfolio returns. The simulated data are independent and have the same mean and standard deviation as one of our industry portfolio returns. We run 1,000 simulations for each dataset and conduct LASSO for each simulation.

Independent Simulated Variables with Same Mean and Standard deviation of Industry Portfolio Returns						
	Finance	**Public Utility**	**Properties**	**Conglomerates**	**Industrials**	**Commerce**
Finance	16.70%	15.20%	13.80%	17.20%	16.40%	14.20%
Public Utility	9.10%	8.70%	8.00%	10.70%	10.30%	9.40%
Properties	7.00%	7.50%	9.10%	6.40%	8.10%	7.40%
Conglomerates	7.20%	6.80%	4.40%	6.60%	5.00%	6.00%
Industrials	4.90%	6.00%	5.40%	4.70%	5.20%	6.30%
Commerce	4.80%	4.60%	5.50%	4.40%	4.50%	5.30%

References

Belloni, A., Chernozhukov, V., 2013. Least squares after model selection in high-dimensional sparse models. Bernoulli 19, 521–547.

Belloni, A., Chernozhukov, V., Fernández-Val, I., Hansen, C., 2017. Program evaluation and causal inference with high-dimensional data. Econometrica 85, 233–298.

Ben-Rephael, A., Da, Z., Israelsen, R.D., 2017. It depends on where you search: Institutional investor attention and underreaction to news. The Review of Financial Studies 30, 3009–3047.

Cakici, N., Chan, K., Topyan, K., 2017. Cross-sectional stock return predictability in china. European Journal of Finance 23, 581–605.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Chen, J., Jiang, F., Liu, Y., Tu, J., 2017. International volatility risk and chinese stock return predictability. Journal of International Money and Finance 70, 183–203.

Chen, Z., Lu, A., 2017. Slow diffusion of information and price momentum in stocks: Evidence from options markets. Journal of Banking and Finance 75, 98–108.

Cohen, L., Frazzini, A., 2008. Economic links and predictable returns. Journal of Finance 63, 1977–2011.

Diebold, F.X., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. The Economic Journal 119, 158–171.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of Econometric Society , 987–1007.

Fama, E.F., French, K.R., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of American Statistical Association 96, 1348–1360.

Foucault, T., Sraer, D., Thesmar, D.J., 2011. Individual investors and volatility. Journal of Finance 66, 1369–1406.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.

Godfrey, L.G., 1978. Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. Econometrica: Journal of Econometric Society , 1303–1310.

Hong, H., Stein, J.C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. Journal of Finance 54, 2143–2184.

Hong, H., Torous, W., Valkanov, R., 2007. Do industries lead stock markets? Journal of Financial Economics 83, 367–396.

Hou, K., 2007. Industry information diffusion and the lead-lag effect in stock returns. Review of Financial Studies 20, 1113–1138.

Huang, W., Lai, P.C., Bessler, D.A., 2018. On the changing structure among Chinese equity markets: Hong Kong, Shanghai, and Shenzhen. European Journal of Operational Research 264, 1020–1032.

Huang, W., Mollick, A.V., 2020. Tight oil, real WTI prices and US stock returns. Energy Economics 85, 104574.

Huck, N., 2019. Large data sets and machine learning: Applications to statistical arbitrage. European Journal of Operational Research 278, 330–342.

Jiang, F., Tu, J., Rapach, D., Strauss, J.K., Zhou, G., 2011. How predictable is the chinese stock market? Journal of Financial Research , 107–121.

Jordan, S.J., Vivian, A., Wohar, M.E., 2014. Sticky prices or economically-linked economies: The case of forecasting the chinese stock market. Journal of International Money and Finance 41, 95–109.

Kilian, L., Park, C., 2009. The impact of oil price shocks on the US stock market. International Economic Review 50, 1267–1287.

Levine, R., et al., 2003. More on finance and growth: more finance, more growth? Review-Federal Reserve Bank of Saint Louis 85, 31–46.

Liu, J., Stambaugh, R.F., Yuan, Y., 2019. Size and value in china. Journal of Financial Economics 134, 48–69.

Menzly, L., Ozbas, O., 2010. Market segmentation and cross-predictability of returns. Journal of Finance 65, 1555–1580.

Moskowitz, T.J., Grinblatt, M., 1999. Do industries explain momentum? Journal of Finance 54, 1249–1290.

Nandha, M., Faff, R., 2008. Does oil move equity prices? A global view. Energy Economics 30, 986–997.

Norton, S.W., 1992. Transaction costs, telecommunications, and the microeconomics of macroeconomic growth. Economic Development and Cultural Change 41, 175–196.

Rapach, D.E., Strauss, J.K., Tu, J., Zhou, G., 2019. Industry return predictability: A machine learning approach. Journal of Financial Data Science 1, 9–28.

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.

Rapach, D.E., Strauss, J.K., Zhou, G., 2013. International stock return predictability: What is the role of the United States? Journal of Finance 68, 1633–1662.

Roller, L.H., Waverman, L., 2001. Telecommunications infrastructure and economic development: A simultaneous approach. American Economic Review 91, 909–923.

Tibshirani, R., 1996. Regression shrinkage and selection via the LASSO. Journal of Royal Statistical Society: Series B (Methodological) 58, 267–288.

Wang, Y., Pan, Z., Liu, L., Wu, C., 2019. Oil price increases and the predictability of equity premium. Journal of Banking and Finance 102, 43–58.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

Wu, Y., 2004. China's economic growth: A miracle with Chinese characteristics. volume 6. Routledge.

Zhang, H., He, Q., Jacobsen, B., Jiang, F., 2020. Forecasting stock returns with model uncertainty and parameter instability. Journal of Applied Econometrics .

Zhao, S., Shojaie, A., Witten, D., 2017. In defense of the indefensible: A very naive approach to high-dimensional inference. arXiv Available from: https://arxiv.org/pdf/1705.05543.pdf .

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of Royal Statistical Society: Series B (Methodology) 67, 301–320.