Information Glitters on China's Connected Market*

Keqi Chen Yuehan Wang Xiaoquan Zhu

First version: February 2019

This draft: November, 2019

* The authors are with PBC School of Finance, Tsinghua University. We are grateful to Li An, Xin Chen, Zhuo Chen, Lauren Cohen, Zhiguo He, Guoming Alan Huang, Sida Li, Bibo Liu, Clark Liu, Dong Lou, Hongfeng Peng, Jiaguo George Wang, Neng Wang, Xian Wang, Zumian Xiao, Jianfeng Yu, Hong Zhang, Tongbin Zhang, Hao Zhou, Haoxiang Zhu, Ning Zhu and seminar and conference participants at PBCSF, Tsinghua University and SDUFE for valuable suggestions and comments. Xiaoquan Zhu acknowledges the Major Program of National Social Science of China (grant no. 16ZDA032) for the research grant. We are responsible for all remaining errors. Send correspondence to Xiaoquan Zhu, PBC School of Finance, Tsinghua University, 43 Chengfu Road, Haidian, Beijing 100083, China. Email addresses: chenkq.16@pbcsf.tsinghua.edu.cn (Keqi Chen), wangyh.16@pbcsf.tsinghua.edu.cn (Yuehan Wang), zhuxq.16@pbcsf.tsinghua.edu.cn (Xiaoquan Zhu).

Information Glitters on China's Connected Market

ABSTRACT

This paper explores a persistent mechanism through which investors profit from the potential information contained in capital flows from China's connected market. Using a complete history of daily filings about stock-level holdings of all Hong Kong and international (northbound) investors, we document that weekly changes in northbound shareholdings (HPC) have a positive cross-sectional predictability for future stock returns. A long-short portfolio that exploits this differentiating preference earns abnormal returns of up to 61 basis points per week, which cannot be explained by common factors. Three sets of evidence suggest the information edge and trading impact of northbound investors: (1) HPC has stronger predictability around earnings announcements in firms with more international exposure and global accessibility (*fundamental information edge hypothesis*); (2) Combined with the similar trading behavior with opportunistic insiders, the sign of decay in return predictability after the "penetrating" supervision on investor identity also suggests that *some* northbound investors are probably mainland insiders with private information advantage that tend to make round-trip investment on A shares to hide identity (*round-trip trading hypothesis*); (3) The imitative herding of mainland investors following informed northbound flows would help boost the future price of specific stocks suffering excessive attention in the short run (*copycat herding hypothesis*). Time-series analysis reveals that northbound capital is a short-term predictor of market returns both in and out of sample. Besides, we also observe similar cross-sectional predictability for H-share returns in southbound institutional investors' trading behavior.

JEL Classification: G10; G12; G14

Key words: Copycat herding, Informed trading, Connected markets, Northbound flows, Round-trip trading

1 Introduction

The stake of China’s stock market was almost zero for international institutional investors before the policies of China's Qualified Foreign Institutional Investor (QFII) and RMB QFII (RQFII) were made. However, China has taken a number of steps over the recent years to encourage international use of its currency, most notably with the launch of the Shanghai/Shenzhen-Hong Kong Stock Connect Scheme, a cross-border equity trading link denominated in RMB. The Shanghai-Hong Kong Stock Connect Scheme first launched on 17 November 2014. Based on this successful pilot project, China Securities Regulatory Commission (CSRC) and Hong Kong Securities and Futures Commission (HKSFC) jointly announced that Shenzhen-Hong Kong Stock Connect Scheme would be officially launched on December 5, 2016. Bypassing regulation on positions on spot foreign exchange, it allows Hong Kong and overseas as well as mainland investors to trade and settle an eligible list of stocks listed on the other market via the exchange and clearing house in their home market. All Hong Kong and international (northbound) investors are allowed to trade eligible A-shares listed on the mainland exchanges while only mainland (southbound) investors who have at least RMB 500,000 in their cash-and-investment accounts are eligible to trade Hong Kong-listed shares[1]. By the middle of 2019, northbound investors own more than 4% of the total free-float A-share market capitalization through the Stock Connect Scheme.

With the liberalization of financial market, the benefits of accessing China’s A-share Stock Market become increasingly attractive to both international investors and academic researchers. In summary, there are five streams of research on China’s stock market: the privatization of SOEs, political connections of firms, regulatory environment, A-H cross-listed stocks and the role of China’s stock market around the globe (Shan et al., 2018). Since Chinese government has launched many new policies to improve market access in the last two decades, the connected market is a good laboratory for the intersection of several strands above including regulation effectiveness, cross listing and market integration. Most previous studies use the pilot project on Shanghai Stock Exchange as a quasi-natural experiment and argue that the Stock Connect Scheme introduces a demand shock for A-share markets and result in revaluation and corporate governance optimization (See, for example, Zhong and Lu,

[1] Capital flows from trading accounts in Hong Kong to stocks listed on the mainland exchanges are called northbound flows, because in China, Hong Kong is farther south than Shanghai and Shenzhen. Along the same line, capital flows from mainland institutional investors to Hong Kong stocks are called southbound flows.

2018; Liu, Wang and Wei, 2018, among several others). Although the related literature is large, there is little direct evidence with regard to the following questions: How quickly is new information incorporated into stock prices in developing economies under market segmentation? Are there any unintended consequences during the two-way opening-up of the capital markets on the mainland and Hong Kong? Does the Stock Connect Scheme really improve the international investors' access to A-share market? The purpose of this paper is to show how answers to questions like these can be obtained one by one by decomposing the source from which northbound investors have information edge and return predictability in A shares. Specifically, the persistent predictability of northbound investors' trading behavior suggests that the information contained in northbound smart money is not yet fully incorporated into prices. Round-trip trading of opportunistic insiders posing as northbound flows and copycat herding of mainland investors are probably two of unintended consequences of the Stock Connect Scheme. Regulation arbitrage with regard to hiding investor identity may preemptively exploit the choice set in A-share market and crowd out international capital.

In this paper, we propose a measure of weekly changes in shareholdings of all northbound investors scaled by shares outstanding (HPC) at the stock level. Intuitively, the measure provides a view about the informed trading of northbound investors involving a relatively higher frequency of variation in preferences. Our empirical analysis provides three sets of results. First, we show that HPC significantly predicts future returns of connected stocks on the A-share market. Stocks in the highest HPC decile outperform those in the lowest decile by 0.61% per week (*t*-value = 2.99) with no sign of reversal in the subsequent eight weeks. It suggests that northbound investors may be informed about firm fundamentals and if any, stock mispricing beyond demand pressure. The return predictability holds in a bunch of robustness checks, including subsamples across exchanges and over time, Fama-Macbeth regressions and heterogeneity tests. Besides, the cross-sectional return predictability is also observed in HPC of southbound institutional investors, however, it comes most from mainland firms listed on the Hong Kong Stock Exchange. Since southbound institutional investors are more likely informed about firms with local connection and Chinese capital, the return predictability in this subsample can be rationalized. This finding further supports our hypothesis with regard to exploiting information advantage on the connected market.

Second, we show evidence of *at least* three channels through which gains of northbound investors from informed trading on A shares are realized: (1) fundamental information advantage of northbound investors due to global accessibility, (2) round-trip trading to hide identity of insiders, and (3) copycat herding following informed northbound capital[2]. First, HPC predicts cumulative abnormal returns (CARs) around quarterly earnings announcements (QEAs) and the pattern is stronger in firms with more overseas business income or dual-listed on the mainland and Hong Kong exchange. It can be rationalized because northbound investors are more likely informed about the stocks with more international business exposure or global investor accessibility. Second, some of northbound investors are probably mainland insiders with offshore accounts that make round-trip investment on A shares to hide identity and avoid the unwanted attention from regulators. They may masquerade as northbound investors who can take advantage of lack of oversight on northbound investor identity on the Hong Kong Exchanges to gain access to invisible shareholdings and prevent private information from being disclosed to the public through their unusual trading. We show that this is the case because after the "penetrating" supervision on northbound investor identity was implemented from September 2018, the cross-sectional return predictability of HPC shows sign of decay. In sharp contrast, the return predictability of southbound HPC actually grows after September 2018, which further supports our hypothesis of round-trip trading. Furthermore, we document a striking similarity of trading behaviors between opportunistic insiders and northbound investors. Third, if mainland investors believe that northbound capital has information edge, then the copycat herding would be expected. Moreover, the flow pressure driven by copycat trading would intensify when investors place excessive attention on the most actively traded connected securities. We show that portfolios sorted by HPC using a subsample of stocks on the daily list of the most actively traded securities deliver return reversal, though an attractive return predictability is observed in the first week. Besides, net purchases of the mainland investors are significantly driven by increases in HPC immediately prior to that day. Overall, our path analysis is generally consistent with the information advantage explanation whereby northbound capital delivers strong and sustainable return on the connected markets.

Third, in the time-series regression, we document that the net inflow of northbound capital through

[2] For the term "round-trip investment", we borrow the language from the economic geography research where round-trip investment is defined as capital from emerging economies to offshore financial centers and back as foreign direct investment (See, for example, Fung et al, 2011; Ledyaeva et al., 2015). Based on the previous work, the round-tripping is more likely driven by the regulatory arbitrage and secrecy arbitrage, namely, hiding their true identities from authorities in the home location. Here rather than revisit the link between demand regulatory arbitrage and "foreign" direct investment on the real economy—the focus of prior work in the economic geography research—our focus is "foreign" stock investment on the emerging capital market. We define the "round-trip investment" as the mainland investors that pretend to be northbound capital and trade A shares indirectly through Hong Kong accounts.

the Stock Connect Scheme is a reliable signal on a weekly basis. The market return predictability is statistically and economically significant both in and out of sample. The market timing strategy based on northbound capital prediction even enhances portfolio performance. A plausible explanation for this finding is that information advantage of northbound investors postulates a momentum effect on future returns along with short-term demand pressure driven by noise traders engaging in the copycat strategy.

The remainder of the paper is organized as follows. Section 2 describes the development of China's connected market. Section 3 outlines channels through which changes in shareholdings of northbound investors predict future stock returns and develops our main hypotheses. Section 4 describes the data and presents the empirical analysis. Section 5 concludes.

Literature review

Since the laboratory we exploit is that of the Chinese connected stock market, our paper is relevant to a recently emerging literature studying the impact of market openness on asset prices. One strand of literature aims to explain the price difference of dual-listed AH shares from the perspective of investor learning about the prices (Zhang and Zhang, 2019). Another strand of literature explores whether interconnection between mainland and Hong Kong stock market can improve the efficiency of information flow (Liu, Wang and Wei, 2018; Shan et al., 2018; Zhong and Lu, 2018; Zhong et al., 2018). There is mixed evidence on the change of pricing efficiency after the launch of the Stock Connect Scheme. For example, Liu, Wang and Wei (2018) find that the connected stocks experience significant higher returns in anticipation of positive demand shocks from Hong Kong investors and the interaction between demand shocks and speculative trading results in stock mispricing. However, Zhong and Lu (2018) and Zhong et al. (2018) argue that the program mitigates the spillover of market volatility and enhances corporate governance. Besides, Shan et al. (2018) show that the connected stocks are not vulnerable to financial contagion on the global market and China's stock market can provide diversification benefits for international investors. Overall, our understanding about information edge of northbound investors and the pricing efficiency of connected market is still limited. Rather than revisit demand shocks induced by market openness—the focus of prior work in this area—our key contribution in this paper is to analyze the pricing and policy implication of informed trading on the emerging market due to the involvement of international investors. Our results show that northbound investors can exploit the fundamental information advantage gathered through improved market openness and the private information derived from regulation arbitrage. The information edge allows northbound investors to benefit from the market reaction persistently, which in turn shows evidence that the Chinese stock market is still far from strong efficiency.

The long-standing literature is consistent with a common story in which domestic and foreign investors have different preferences for stock characteristics and different reactions to information (Froot and Teo, 2008; Jia, Wang and Xiong, 2017). In recent years, a prevailing view in finance is that on average, domestic investors tend to outperform foreigners due to local information edge (e.g., Agarwal et al. (2009) in Indonesia; Choe et al. (2005) in Korea; Dvořák (2005) in Indonesia; Ferreira et al. (2017) in 32 countries; Hau (2001) in Germany; Shukla and Van Inwegen (1995) in the United States; Teo (2009) in Asia). Chan, Menkveld and Yang (2008) demonstrate that foreign investors who trade B-shares, have an informational disadvantage relative to domestic investors who trade A-shares. In this terminology, our paper shows that northbound investors who trade A-shares on the Chinese connected market, have an informational advantage relative to mainland investors who also trade connected A-shares. Several studies also hold the opposing view that sophisticated foreign investors may have an advantage over domestic ones through global private information that they have acquired in their own market (e.g., Bailey et al. (2007) in Singapore and Thailand; Brennan and Cao (1997) in the U.S.; Chen et al. (2009) in Taiwan; Choe, Kho and Stulz (1999) in Korea; Froot and Ramadorai (2008) in 25 countries; Froot et al. (2001) in emerging markets; Grinblatt and Keloharju (2000) in Finland; Huang and Shiu (2009) in Taiwan; Maffett (2012) in the cross-country setting; Seasholes (2000) in Thailand and Taiwan). Most of these studies conclude that the degree of investors' sophistications matters in determines which type of participants play a leading role in disclosing information to the market and counterparties[3]. In this way, our paper adds to the large literature by distinguishing between informed trading of northbound investors and copycat trading of domestic investors on the China's connected market.

Our paper also contributes to an extensive literature examining how information is incorporated into asset prices (e.g., Ben-Rephael et al., 2017; Edelen, Ince and Kadlec, 2016; Grossman and Stiglitz, 1976; Engelberg, McLean and Pontiff, 2018). Engelberg, McLean and Pontiff (2018) find that anomaly returns are driven by biased expectations of sentimental investors and become much higher around

[3] See Ferreira et al. (2017) for a more comprehensive literature review about the comparison of information edge between domestic and foreign investors.

earnings announcements when information is released concentrically. Ben-Rephael et al. (2017) argue that institutional demand for information is associated with a risk premium as well as the correction of mispricing. One strand of literature investigates this question from an insider standpoint (e.g., Ali and Hirshleifer, 2017; Chakravarty and Mcconnell, 1999; Kyle, 1985). Another line of literature provides insight into pricing efficiency by exploring whether arbitrageurs are effective in detecting mispricing (Chen, Da and Huang, 2018; Shleifer and Vishny, 1997; Sias, Turtle and Zykaj, 2015; Jiao, Massa and Zhang, 2016). Consistent with Chen, Da and Huang (2018), we compare the trading profits during the entire sample period and specific information events (such as earnings announcements) and show that northbound investors can forecast earnings surprises around announcement dates, which allows them to benefit from the market reaction. We provide new insights relative to examining the informed trading and price impact of a new breed of investors, i.e., northbound capital, and enhance our understanding about the interaction between information edge, copycat herding, and potential regulation arbitrage.

Prior research has relied on quarterly QFII holdings to measure the trading of international investors on the Chinese stock market. To the best of our knowledge, our paper is the first to track northbound trading using daily shareholdings of northbound investors through the Stock Connect Scheme at the stock level. Our measure, changes in northbound shareholdings not only have cross-sectional predictability for stock returns both in the short and long run, but also facilitate the time-series predictability of market return. Empirically, we distinguish between the price effect of informed trading, imitative herding and round-trip investment. Specifically, we find no return reverse in the long run in the entire sample of connected stocks, while the decay of return predictability is observed in a subsample of stocks facing copycat herding caused by limited attention. Besides, we document the similarity of trading behaviors between mainland opportunistic insiders and northbound investors. In this way, our tests shed light on how northbound investors trade on their information edge on the connected stock market and how their trading behaviors affect asset prices through news, attention and regulation arbitrage.

2 Institutional background

We briefly describe the development of Shanghai / Shenzhen-Hong Kong Stock Connect Scheme from 2014, and the background of the Northbound Investor Identification Model.

2.1 Shanghai / Shenzhen-Hong Kong Stock Connect Scheme

The Shanghai-Hong Kong Stock connect scheme is a pilot program established by the Chinese government in order to link the stock markets in Shanghai and Hong Kong. On April 10, 2014, the program was formally announced by Chinese Premier Li, Keqiang at the Boao Forum in Hainan Province of China. The program was finally approved and announced on November 10, 2014 and officially launched on November 17, 2014. During a visit to Shenzhen on January 5, 2015, Chinese Premier Li Keqiang stated that a stock connect program was necessary for Shenzhen and Hong Kong markets based on the establishment of the Shanghai-Hong Kong Stock Connect. On August 16, 2016, CSRC and HKSFC made a joint announcement that preparations for the Shenzhen-Hong Kong Stock Connect started. Finally, it was officially launched on December 5, 2016. The Stock Connect Scheme allows investors to trade through local brokers, and orders are routed through the subsidiaries set up by local exchanges to the opposite markets.

Before the launch of the Stock Connect Scheme, Chinese regulators imposed a bunch of restrictions on foreign investments into the Chinese financial markets. One potential channel to access Chinese stock market is to participate in the QFII and RQFII program. The aim of the two programs is to provide a pilot for relaxing foreign exchange controls in a limited way, as well as to leverage the investment skills of foreign institutions to raise the standards of the Chinese market. However, QFII and RQFII are only accessible to selected and government-approved foreign institutions and have only recently lifted caps on investment quota since September 2019. On the contrary, all Hong Kong investors are allowed to trade eligible shares listed on the Shanghai and Shenzhen Stock Exchange, while mainland investors need to have at least 500,000 RMB in their cash and trading accounts to be qualified for trading Hong Kong shares on the connected market.

Eligible stocks on the Shanghai Stock Exchange include all constituent stocks of the Shanghai Stock Exchange 180 Index and 380 Index, and stocks that are dual-listed in Hong Kong, excluding stocks that are either not traded in RMB or that are included in the exchange's "risk alert board" stocks in the process of delisting or at risk of being delisted. Eligible stocks on the Shenzhen Stock Exchange cover constituents of the Shenzhen Stock Exchange Component Index and Small/Mid Cap Innovation indexes with a market capitalization of at least RMB 6 billion, and A-H shares dual-listed on Shanghai and Hong Kong Stock Exchange. ChiNext stocks open to institutional professional Hong Kong investors initially. Eligible stocks in the Hong Kong Stock Exchange through the Shanghai-Hong Kong Stock Connect include the constituent stocks of the Hang Seng Composite Large Cap Index and the Hang Seng Composite Mid Cap Index and stocks that dual-listed in Shanghai, excluding stocks that are not traded in Hong Kong dollar. Compared with eligible stocks for Shanghai southbound trading, Shenzhen Stock Exchange adds constituents of Hang Seng Small Cap Index with a market capitalization of at least HK$ 5 billion and A-H shares dual-listed on the Shenzhen and Hong Kong Stock Exchange to the existing eligible stocks[4].

Initially, the aggregate quota was 300 billion RMB for Shanghai- and Shenzhen-listed shares, respectively and 250 billion for Hong Kong-listed shares through both Shanghai- and Shenzhen-Hong Kong Stock Connect Scheme, respectively. The two-way aggregate Quota has abolished since August 2016. However, trading through the Stock Connect Scheme is still subject to a daily quota. Before May 2018, the daily quota for the net buying value of cross-border trades was 13 billion RMB for both Shanghai- and Shenzhen-listed shares, and 10.5 billion RMB for Hong Kong-listed shares through both Shanghai- and Shenzhen-Hong Kong Stock Connect. Chinese regulators have quadrupled the two-way daily quota since May 2, 2018, which represents approximately 30% of the daily trading volume in each market.

2.2 Northbound Investor Identification Model

The mainland markets adopt a see-through model for trading and clearing. All orders submitted to the mainland exchanges must bear the respective securities account numbers which will be further carried to the clearing end. However, the arrangement was quite different from that in the Hong Kong securities market as market-wide unique investor identifiers did not exist before the launch of the Northbound Investor Identification Model. Only identifiers of the executing brokers would be submitted to the mainland exchanges when northbound trading is conducted through the connected market. For example, if mainland funds masquerade as northbound investors to gain access to large shareholdings, then only the Hong Kong Securities Clearing Company (HKSCC), rather than the name of the specific client, would appear on the list of ten largest shareholders which must be disclosed to investors[5]. Lack of surveillance may give mainland capitals which are subject to scrutiny by regulators

---

[4] For the detailed list and regular adjustment of eligible stocks, please refer to: http://www.hkex.com.hk/eng/market/sec_tradinfra/chinaconnect/Eligiblestock.htm.

[5] HKSCC is a participant of China Clear (CSDC), and clears and settles cross-border trades with China Clear for Hong

the chance to make round-trip investment on A shares to conceal their true motives. For example, two individual investors, Hanbo Tang and Tao Wang, were verified by CSRC to involved in manipulating share prices in 2016. Specifically, they colluded to use three trading accounts on the Hong Kong exchange and one account on the mainland exchange to create artificial prices and trading volumes of the underlying stock of Zhejiang China Commodities City Group (600415.SH) through Shanghai-Hong Kong Stock Connect and intentionally misled other investors[6]. Similarly, two private funds were verified to be involved in manipulating the prices of four stocks listed on the Shanghai exchange by exploiting the funding advantage of six asset management products through connected accounts from December 2015 to August 2016[7].

To facilitate the market surveillance in accordance with the home market principle, CSRC announced that the establishment of an investor identification system for northbound trading should be actively moved forward in September 2016. In November 2017, Hong Kong Exchanges and Clearing Limited (HKEX) issued a regulation document that introduced the motivation, potential impact on surveillance, tentative timeline and technical setup for the implementation of investor identification. It stated that:

> *"In Hong Kong, the Securities and Futures Commission (SFC) is working closely with HKEX to look into how best to implement an Investor ID regime which will identify orders at the client level rather than the broker level to better detect potential misconduct."*

Finally, on August 24th 2018, CSRC published a press release which stated that consensus proposed by mainland and Hong Kong exchanges had been reached by CSRC and HKFSC that the Northbound Investor Identification Model would come into force on 17 September 2018. Since then, exchange brokers that offer northbound trading services are required to assign a unique number in a standard format, known as the Broker-to-Client Assigned Number (BCAN), to each of their northbound trading clients and provide Client Identification Data (CIA) mapping to HKEX by 3 pm on T-1 day, which will forward the information to Mainland exchanges as the file processor directly or through China

---

Kong and international investors. See, for example, in the 2019 semi-annual report, the second largest shareholder of Shanghai International Airport Co., Ltd (600009.SH) is the Hong Kong Securities Clearing Company (HKSCC) which aggregates all shareholdings of northbound investors.

[6] For more details about the case, please refer to: http://www.csrc.gov.cn/pub/zjhpublic/G00306212/201703/t20170310_313478.htm.

[7] For more details about the case, please refer to: http://www.csrc.gov.cn/pub/newsite/jcj/aqfb/201806/t20180607_339507.html.

Securities Depository and Clearing Corporation (CSDCC).

3 **Hypothesis development**

The section raises the question of why net purchases by northbound investors are associated with higher future returns. One possibility is that the cross-sectional predictability may reflect superior information of Hong Kong and foreign investors that is gradually incorporated into prices through informed trading. This argument warrants further research in our setting, but seems not unlikely, given the findings in previous studies that foreigners are expected to have information advantage over domestic investors in emerging markets (e.g., Bailey et al. (2007); Froot et al. (2001); Huang and Shiu (2009)). Given that the Chinese stock market is dominated by retail investors, northbound investors may be sufficiently sophisticated to uncover firm fundamentals, which enable them to identify the mispriced securities as well as predict the market reaction to news before noise traders (Chen, Da, and Huang (2018); Jiao, Massa, and Zhang (2016)). If the trading decisions of northbound investors are purely speculative, we would not observe a cross-sectional variation in net inflows from international investors considering the revaluation effect induced by demand shocks for all connected stocks after the launch of the Stock Connect Scheme. In this case, there would be not any incremental relation between changes in northbound shareholdings and future returns, controlling for flow pressure driven by contemporaneous net inflows. We may even observe a reverse effect if northbound investors are trade-imbalance-sensitive in response to market openness. If we do observe a persistent return predictability of changes in shareholdings of northbound investors, this would support the notion that information advantage drives the trading decisions. We thus form our first hypothesis.

Hypothesis 1 (fundamental information edge hypothesis): *Changes in shareholdings of northbound investors should positively predict future stock return beyond flow pressure due to the fundamental information advantage.*

We have shown anecdotal evidence in the section of institutional background that mainland capital tends to be channeled to Hong Kong and make round-trip investment on A shares. There might be a mixture of motives, such as low transaction fees through the Stock Connect Scheme, low commission of Hong Kong brokers and especially, lax screening of investor identity. Lack of see-through surveillance might obfuscate the true stakeholders and prevent the potential information contained in "quiet" changes in northbound shareholdings from being incorporated into market prices. In this way, opportunistic insiders who have private information about their own firm, are especially likely to be involved into round-trip investment, because their trading activities through mainland exchanges are subject to scrutiny by regulators. We thus form our second hypothesis.

Hypothesis 2: *Changes in shareholdings of northbound investors may positively predict A-share return partially due to the private information of mainland insiders that make round-trip investment.*

Testing Hypothesis 1 and 2 requires to justify the presence of information edge among northbound investors. It is empirically difficult, however, to distinguish between price movement driven by copycat trading of noise traders and active trading of informed investors. In our paper, rather than rule out the prevalence of flow pressure, we aim to justify the contribution of imitative herding to cross-sectional return predictability. It is well documented that foreign investors are frequently viewed as being closely watched and having significant influence on stock prices in emerging economies with poor liquidity (e.g. Richards, 2005). Imitative trading would help boost the market price of specific securities in which northbound investors increase their holdings most if noise trades believe that northbound capital has information edge. Along the same line, lower expected return of stocks in which northbound investors reduce their holdings most may be partially attributed to fire sales driven by imitative liquidating following informed selling. The rationale leads to our third hypothesis.

Hypothesis 3 (copycat herding hypothesis): *Copycat herding by noise traders can also contribute to the cross-sectional return predictability of changes in northbound shareholdings in the short run*.

4 Empirical Results

4.1 Data and Summary Statistics

We draw from a range of data sources to construct the sample in this paper. We obtain the complete history of daily A-shares held by all northbound investors from HKEX website. The southbound shareholding data are obtained from CHOICE Database. We collect data of aggregate northbound (southbound) capital flows and the daily lists of the ten most actively traded stocks (i.e., Top 10 stocks with the highest daily trading volumes in RMB yuan) from WIND database. We obtain daily stock returns, risk-free rate, analyst coverage, insider trading records and firm-level financial indicators from China Stock Market Trading Research (CSMAR) and RESSET database.

Since the eligible stocks in the connect scheme follows the adjustment of constituents in specific indices, we track all the eligible A shares no less than once in history to address the potential concern on the survivor bias. We drop stock-week observations with purchasing constraints due to the index reconstitution based on the record of sample stocks in the CSMAR database. The final sample includes 1795 A-share stocks, spanning from March 17, 2017 to December 31, 2018. Considering the transaction asynchronism on the connected market, we keep only stock-week positions when a specific security is tradable for northbound investors. To address the potential noise in the weekly return due to share suspension, we delete the weeks with less than two trading days. Note that at the beginning of the available data, the Stock Connect Scheme has already opened up both Shanghai and Shenzhen stock market to international investors and removed the aggregate quotas for limiting the cumulative net purchasing value of northbound capital, which alleviates possible regulatory friction in our analysis

We propose a measure to capture the average trading behavior of northbound investors on individual A shares through the Stock Connect Scheme. Specifically, weekly changes in northbound shareholdings for stock *i* in week *t*, namely *HPC*, is defined as the difference in the ratio of northbound holding to total shares outstanding between the end of week *t* and week *t-1*:

$$HPC_{it} = \frac{Shares\ of\ Stock\ i\ held\ by\ northbound\ investors\ in\ week\ t}{Total\ shares\ outstanding\ of\ Stock\ i\ in\ week\ t} - \frac{Shares\ of\ Stock\ i\ held\ by\ northbound\ investors\ in\ week\ t-1}{Total\ shares\ outstanding\ of\ Stock\ i\ in\ week\ t-1} \tag{1}$$

For example, if the reported holding percent of stock *i* at the end of week *t-1* and week *t* is 10% and 20%, respectively, then HPC_{it} is 10%. We would argue that this measure contains more information than the growth rate of shareholdings of northbound investors. Taking again the numeric example above, if another stock, j, experiences an increase in northbound shareholdings from 1% to 2%, then both stock *i* and *j* have a growth rate of 100% in terms of holding percent. However, our measure, HPC for stock i, 10%, is much larger than that for stock j, 1%. Intuitively, weekly changes in stock *i*' s position contain more information related to northbound investors' preference.

[Table 1 about here]

Table 1 presents the summary statistics of stock-level variables. Panel A focuses on key variable of interest, *HPC* and control variables including the level of northbound shareholdings (HP), log of market capitalization (Size, in thousand RMB yuan), the book-to-market ratio (BM), the gross profits-to-assets ratio (ROA), return volatility computed as the standard deviation of daily excess market return over the past 3 months (VOL), idiosyncratic volatility computed as the standard deviation of residual from 3-month Fama-French three-factor regressions on daily data (IVOL), turnover (TURN), stock return reversal computed as the cumulative return from week t −4 to week t – 1 (REVERSAL) and stock return run-up computed from week t-52 as of the end of week t-5 (RUNUP)[8]. There are 1795 unique A-share stocks and 117,598 stock-week observations in the final sample. Several striking features in summary statistics are worth discussing. First, the average ratio of northbound shareholdings to total shares outstanding is about 0.61%, suggesting that the northbound capital only accounts for a minority of A-share investors. Second, note that the median size of all eligible A shares is up to about 16.52 billion RMB yuan on the China's connected stock market, while the median size of all A shares is just 4.1 billion RMB at the end of 2018. The gap means that the majority of connected stocks are large firms, which is more likely to be selected as eligible assets for foreign investors. Besides, the average autocorrelation of degree one of weekly HPC is quite low about -0.134 in our sample, which validates this measure as a legitimate candidate for timely response to information, rather than stale preference of northbound investors. The non-persistence of HPC is also consistent with the empirical implication of learning model suggested by Zhang and Zhang (2019) in explaining AH premium. Specifically, northbound investors may care more about future capital gains in the short run than dividend policies and voting rights in the long run.

Panel B in Table 1 reports the summary statistics of decile portfolios sorted by weekly HPC. At first glance, the average HPCs in different groups have a wide range from -0.21% to 0.26%. However, the number does not change much over Group 2 to Group 9. In other words, much of the cross-sectional variation comes from the difference between extreme deciles. Furthermore, level of northbound shareholdings (HP) is not evenly distributed among ten decile portfolios. More specifically, both the extreme groups have higher HPs than other groups. Hence, the cross-sectional return predictability is, intuitively, more of an information story, rather than demand pressure. Even though the average firm size in the two extreme groups is slightly larger, there is not much between-group difference for other characteristics such as IVOL, VOL and BM ratio.

4.2 Cross-Sectional Return Predictability of HPC

In this section we present empirical results linking the trading strategy based on weekly HPC of

[8] Details of the sample and variable construction are provided in the Appendix.

northbound investors to future stock performance. First, we conduct standard univariate portfolio sorts as well as Fama-MacBeth (1973) regressions. The results show that the cross-sectional return predictability, which cannot be explained by common factors and DGTW adjustment, is robust across exchanges and over time. Second, we confirm that the cross-sectional predictability of HPC is robust to a variety of settings with different formation and holding period. Especially, the return spread between extreme groups sorted by HPC does not reverse in the long run. Third, much of the variance in HPC as well as the cross-sectional return predictability cannot be completely explained by lagged returns and a bunch of common characteristics.

4.2.1 Univariate Portfolio Sorts

We sort all eligible stocks into decile portfolios based on HPC at the end of each week and then calculate the value weighted returns in each decile for the next week. Following the literature, we also use the Carhart (1997) four-factor model to assess the benchmark adjusted performance of stock portfolios. In order to gauge the economic magnitude of predictability, we estimate the return spread between the top- and bottom-decile portfolios and conduct statistical significance tests using the heteroscedasticity and autocorrelation consistent (HAC) GMM estimator for each group.

[Table 2 about here]

Panel A of Table 2 report the value-weighted returns, in percent per week, of stock portfolios. At the end of each week all stocks are sorted into ten deciles, five quintiles or two halves based on their HPC in the previous week. The portfolios are held for one week. A high-minus-low strategy using the extreme deciles, 1 and 10, with a long position in the high-HPC decile and a short position in the low-HPC decile, is then constructed. As shown in the first row of Panel A, a long-short portfolio that exploits northbound investors' preference earns abnormal returns of up to 61 basis points per week, or almost 31.72 percent per year. The striking pattern with respect to the significant return spread remains in the quintile portfolios as well as the halves. We also examine the variation of cross-sectional predictability across market and over time. At the top of Panel B, we repeat univariate sorts for subsamples of firms listed on Shanghai and Shenzhen Stock Exchange, respectively. All qualitative inferences from Panel A remain unchanged, although the predictability is slightly weaker on the Shenzhen Stock Exchange. The bottom of Panel B reports the results for the two sub-periods. The result for the more recent sub-period as of the end of 2018 shows that the return spread is still significantly large, which is consistent with investor learning over time. One possibility is that northbound investors appear to learn about firm fundamentals or private information gradually to differentiate between "good" and "bad" stocks in the recent period. Besides, mainland investors might learn about the preferences of northbound capital and then profit from copycat trading that coincides with the popularity of such a trading strategy following smart money. Panel C presents Carhart (1997) four-factor alphas and DGTW (1997) benchmark-adjusted returns. The return difference cannot be explained by common factors in the Carhart (1997) four-factor models. Specifically, the four-factor alpha spread between extreme deciles is about 51 basis points per week, which is economically and statistically significant at the 1% level. There is just a slight decrease in return spread when using characteristic-based benchmarks. At the bottom of Panel C, our results remain quantitatively similar if one week is skipped after sorting.

One drawback of the univariate portfolio sorts is that they do not allow for a multivariate analysis. However, it is well documented by the long-standing literature that many firm-level characteristics can successfully predict stock returns, such as size, Book-to-Market ratio and past stock returns. Hence, we run Fama-MacBeth (1973) regressions to confirm that the lagged return as well as a variety of firm characteristics cannot subsume the cross-sectional return predictability of HPC. Table 3 reports results of the second stage of Fama-MacBeth (1973) regressions. Model (1) only includes our key variable of interest and Model (2) contains firm-level characteristics and returns within different horizons. One potential concern is that northbound investors purely position themselves following MSCI indices including selected A shares. Hence, we add to Model (2), MSCI, a dummy variable which identifies the constituents of MSCI indices. Model (3) also controls contemporaneous HPC to exclude the effect of flow pressure, if any. In all model specifications, our coefficient of interest remains significantly positive. The results from Fama-MacBeth regressions provide further evidence that the positive relation between changes in northbound shareholdings and future stock return is likely derived from the information advantage of northbound investors.

[Table 3 about here]

4.2.2 Does Return Predictability Reverse in The Long Run?

In this section, we confirm that the cross-sectional return predictability of HPC does not reverse in the long run. Though the stock-level data in our paper say nothing about the investor-level trading of foreigners, we show evidence that cumulative positive returns of the long-short portfolio are beyond flow pressure even in the most extreme hypothetical scenarios.

[Table 4 about here]

Panel A of Table 4 reports the cumulative returns of hedge portfolios in different lengths of formation and holding period. For example, when the number of formation weeks equals one and the number of holding weeks equals eight, the underlying assumption is that northbound investors trade based on the HPC in one week and do not actively position themselves in the subsequent eight weeks. At the end of the holding period, the average cumulative return spread between extreme quintile portfolios over eight weeks following formation is up to about 1.58% and statistically significant at the 1% level. If we take another extreme case where northbound investors rebalance their portfolios based on HPC each week, then both the formation and holding period is one week. The top left corner in Table 4 show that the most active strategy would deliver an average weekly return of up to 50 basis points, or almost 4% per eight weeks. Following Jegadeesh and Titman (1993), to more accurately simulate real-world trading behaviors, we also adopt a similar design to test for the long-run performance of portfolios with overlapping holding periods based on HPC[9]. Panel B of Table 4 reports the results of portfolios that are rebalanced weekly to maintain equal weights. Overall, although we cannot figure out whether northbound investors tend to hold their portfolios long enough, the return predictability is remarkable in the short run and shows no reversal in the long run.

[Figure 1 about here]

To construct a single time series spanning the entire sample period, we concatenate the portfolio returns of top and bottom quintiles as well as hedge portfolios across the holding week. Figure 1 plots the cumulative excess return spanning from March 2017 to December 2018 if northbound investors adopt the J=1 / K=1 strategy. The downward slope of cumulative excess returns in the bottom quintile as well as the upward slope in the top quintile generates a striking pattern in the long-short strategy, which provides support for the sustainable return predictability as well as potential investor learning.

[9] More specifically, a strategy that selects stocks based on HPC over the past J weeks and holds them for K weeks is constructed as follows: At the beginning of each week t, the eligible stocks are ranked in ascending order on the basis of their HPCs in the past J weeks. Based on these rankings, five quintile portfolios are formed. In each week t, the trading strategy buys the top quintile portfolio and sells the bottom quintile portfolio, holding this position for K weeks. In addition, the strategy closes out the position initiated in month t - K. Hence, under this trading strategy we revise the weights on 1/K of the stocks in the entire portfolio in any given week and carry over the rest from the previous week.

4.2.3 Determinants of changes in northbound shareholdings

Before exploring the information channel through which changes in northbound shareholdings have price implications on the cross section, we would like to understand the characteristics of securities with different HPCs. We run a variety of panel regressions to investigate which of determinants plays a critical role in explaining the cross-sectional variation of HPC.

[Table 5 about here]

Panel A and B of Table 5 report results of panel regressions of HP and HPC on firm characteristics, respectively. All regressions include year-month and industry fixed effects and cluster standard errors by industry. In Panel A, we find a significantly positive relation between stock return run-up, gross profitability, the addition to the MSCI China index, analyst covering and HP as well as a negative relation between the Book-to-Market ratio, turnover and HP. The result suggests that northbound investors prefer past "winners" with high-quality profits, growth potential and rich information environment, rather than assets with more noise trading. However, the R^2 statistics across all specifications are relatively small, no more than 25%, suggesting that in addition to firm fundamentals, there may be other information involved with northbound trading. Panel B presents the determinants of HPC. Although the determinants are only slightly different between HP and HPC, the R^2 statistics for HPC are much smaller than that for HP. Informally, this is suggestive of an explanation where northbound investors respond more to information advantage than to lagged returns related to market inefficiency as well as common factors to detect profit margins. Since one concern is that northbound capital might be purely value-oriented investors tracking assets with these firm characteristics like "White Horse Unit" ("Baima" in Chinese), our analysis of portfolio sorts is repeated by using HPC's residual of specification (1) in Panel B of Table 5. The results, reported in the Appendix, are quantitatively similar.

4.3 Hypothesis 1: Fundamental Information Advantage

In this section, we take three steps to show evidence of *fundamental information advantage hypothesis*. First, if northbound investors trade on firm fundamental information, they would detect mispricing and profit from correction to mispricing in the long run. Accordingly, we track the return spread between extreme groups sorted by HPC and examine whether northbound investors profit from information release. Second, we show that northbound investors could forecast earnings surprises

(SUEs) and cumulative abnormal returns (CARs) around quarterly earnings announcements so as to benefit from stock market reaction. Third, taking as given that northbound investors are more likely informed about firms with more overseas business income or dual-listed on the China mainland and Hong Kong stock market, a stronger pattern driven by information edge is expected in those firms with more international exposure and global accessibility.

4.3.1 Cross-Sectional Returns around Information Release

Fundamental information release is a crucial channel through which public news would affect stock prices, especially on the Chinese stock market where limits to arbitrage lead to mispricing[10]. In such an investment environment without effective price discovery, investors who trade in the direction of information by advance could earn substantial returns when information is released to the public gradually. In Table 6, we track the return spread between extreme quintiles sorted by HPC. In Column 1, we examine the cross-sectional stock returns over a five-day window around quarterly earnings announcements. To be comparable with main results in Table 4, we also compute the cumulative returns from one to eight weeks following announcement dates. The weekly return spread between extreme groups reaches 0.52% (t-value=3.30) over the five-day window, slightly higher than the average stock return spread of 0.50% (t-value=3.31) per week over the entire sample (as reported in the top left corner of Panel A in Table 4). Besides, the cumulative return of long-short strategy increases with the release of fundamental information. More specifically, the cumulative excess return is up to 2.31% (t-value=5.10) as of the end of the eighth week after announcements, much larger than 1.58% (t-value=3.81) in the entire sample (as reported in the top right corner of Panel A in Table 4). The absence of return reversal in the long run further demonstrates that the trading pattern of northbound investors is consistent with the released information about firms' future cash flows.

[Table 6 about here]

Note that, return spread between extreme quintile portfolios sorted by HPC mainly comes more from the bottom group, rather than the top one in Table 6. It is different from the return pattern in periods without earnings news as reported in Table 4. It appears to suggest that the disclosure of firm

[10] On the mainland market, only a group of eligible stocks are approved in margin trading and short selling. Although northbound investors are allowed to short selling in the Shanghai/Shenzhen-Hong Kong Stock Connect Scheme, few eligible stocks have ever been shorted by northbound investors after the market crash in 2015, as reported in HKEX.

earnings is more likely to unveil the intrinsic value of stocks that have been overestimated by uninformed investors. Once the fundamental information is available to the public, prices revert to the reasonable level and informed northbound investors succeed to avoid future loss by reducing their shareholdings in mispriced stocks based on their information edge in advance.

4.3.2 HPCs Forecast CARs and SUEs around QEAs

Informed northbound investors would increase their shareholdings on the stocks with the anticipation of positive abnormal returns around earnings announcements[11]. In the spirit of Chen, Da and Huang (2018), we compare the trading profits during the entire sample period and specific information events (such as earnings announcements) and investigate whether northbound investors benefit from the market reaction. In formal tests, short-term market reaction to announcements is measured as the buy-and-hold abnormal return adjusted by market model over a five-day window [-2, +2], where the market benchmark is defined as the value-weighted portfolio of all A-share stocks listed on Shanghai or Shenzhen exchange[12]. Estimation window is required to be one quarter (66 trading days) before the announcement of earnings for each stock, with minimum observations of 22 trading days. Following Livnat and Mendenhall (2006), standard earnings surprises (SUEs) are defined as: $SUE_{i,t} = \frac{X_{i,t}-X_{i,t-4}}{P_{i,t}}$, where $X_{i,t}$ denotes earnings per share for firm i in quarter t. The key variable of interest is a variant of the ranking variable, changes in northbound shareholdings (HPC), calculated as the difference of holding ratio between t-1 and t-6, where t is the first day of event window. Control variables include firm and stock characteristics.

[Table 7 about here]

Table 7 presents a bunch of panel regressions forecasting the BHARs and SUEs around QEAs. The dependent variables are SUE and CAR [-2, +2] across all specifications in Panel A and Panel B, respectively. Column (1) presents results for the base model with industry and year-month fixed effect. Column (2) examines institutional ownership, column (3) considers analyst covering and column (4) estimate the full specification. As shown in the Table 7, the coefficient estimates of HPC are positive

[11] Informally as shown in Figure I.1 in the Appendix, it appears that increases in northbound shareholdings line up well positive abnormal returns around QEAs.

[12] In untabulated tables, we also employ a longer estimation window [-376, -11], and use Fama-French three factor model and constant model as alternative benchmarks to estimate abnormal returns during the announcement window. The main pattern remains quantitatively similar.

and statistically significant at 5% across all specifications. It confirms that northbound investors tend to increase their shareholdings when they anticipate the earning growth or positive abnormal returns around quarterly earnings announcements.

4.3.3 Heterogeneity Tests

If the main arguments about the presence of information advantage holds, then it appears to be a natural inference that northbound investors are more likely informed about firms with more overseas business income which may improve international trade exposure and global investor accessibility. Then, a stronger cross-sectional predictability driven by information edge would be expected in those firms. Many other motivations and effects, including tax incentives (Azémar et al., 2007) and corporate governance and external supervision optimization (Aggarwal et al., 2011; Zhong and Lu, 2018), can potentially lead to the price implication of foreign investors on the connected stock market. In the current study, overseas business income is simply used as a proxy for a source of information, and this paper does not model or investigate possible underlying forces which lead to build overseas connections in the first place. Instead, the paper focuses on the effect of information intensity on the cross-sectional predictability of HPC.

To study whether a HPC characteristic contains unique information on stock future return, we first report the results for double sorts with respect to HP in Panel A of Table 8. We sort the sample stocks by their stakes in northbound investors into two groups and each group is sorted to quintile portfolios by HPC. We find that both return spreads are positive and statistically significant at 1% in both below-median and above-median HP groups suggesting that our key variable of interest, HPC, contains unique information that do not depend on shareholding situation. Hence, the double sorts indicate that the effects of the characteristics can be decoupled and changes in northbound shareholdings are a more significant driver of stock future price. This result is actually consistent with those related to potential determinants of HP and HPC, as reported in Table 5. Again, it indicates that even though northbound investors may have preferences for their investment choice set, for example, specific industries, the variation of HPC is not directly proportional to historical shareholdings driven by passive rebalancing.

The results for double sorts with respect to A-H cross-listing are presented in Panel B of Table 8. If they have information advantage on the connected market, then northbound investors are more likely informed about AH dual-listed firms. This is intuitive, since dual-listed firms release regular reports both in Chinese and English, or northbound investors are at least with knowledge of H-share market. Moreover, to be listed on the H-share market, the business operation and accounting rules have to be more internationalized and standardized, which may attract more analyst coverage and attention of overseas investors. All the underlying forces can help facilitate northbound investors to exploit their information edge on AH dual-listed firms. Hence, it would be expected that HPC has a stronger cross-sectional predictability for cross-listed firms. The results confirm our conjecture with a larger return spread in the subgroup of cross-listed firms, up to about 0.81% (only about 0.43% in the rest group).

Along the same line, the return spread between extreme quintiles is positive and statistically significant in only above-median group sorted by overseas business income, as shown in Panel C of Table 8. The difference of return spreads between below- and above-median groups is positive and statistically significant at 5%. It suggests that trading behaviors of northbound investors within one week have more cross-sectional predictability for firms with larger overseas business income. The results can be rationalized because northbound investors are more likely informed about firms with more exposure to international investors and trades. More specifically, we can interpret it as a trading strategy that ideally, if most of information edge of northbound investors comes from firms with more overseas business income, then they could earn a weekly return up to 0.53% by entering into the long-short portfolio sorted by HPC within the above-median group and lighten a position of hedge portfolio constructed within the below-median group.

[Table 8 about here]

Furthermore, based on the double sorting results related to overseas income, we repeat the analysis about HPCs forecasting SUEs and CARs around QEAs in both below-median and above-median-overseas-income groups. Panel B of Table 7 reports the results. In the above-median group, HPC of northbound investors have a stronger predictability for earning growth and abnormal stock return around earning announcements. The difference of coefficient estimates between above- and below-median group is positive and statistically significant at 10% across most regressions. It suggests that for firms with higher overseas income, northbound investors are more likely to have information edge and increase their shareholdings when they anticipate positive fundamentals before formal earning announcements.

4.3.4 Further Tests Related to Changes in Southbound Shareholdings

We also examine the cross-sectional predictability of southbound capital flows for H-share returns in Table 9. Consistent with our information hypothesis, changes in southbound shareholdings can also positively predict the future stock return. In the entire sample of connected H shares, the return spread between extreme quintile groups sorted by weekly changes in southbound shareholdings on individual H share is up to about 64 basis point per week. Furthermore, as shown in Panel A, most predictability comes from southbound informed trading in firms listed on the H-share market with Chinese capital or mainland background. Considering the information edge of Chinese institutional investors documented in prior studies (Bailey et al., 2009; Hou and Ye, 2008), it is not surprising that southbound investors, most of which are domestic institutions, are more likely informed about firms closely connected with the mainland. We find no long-run reversal in Panel B, suggesting that beyond price pressure, southbound investors are informed about firm fundamentals and if any, stock mispricing.

[Table 9 about here]

4.4 Hypothesis 2: Round-Trip Trading

One might concern that *some* northbound investors are mainland-sourced capital that trades A shares in a round-trip form, rather than foreigners undertaking cross-border transactions. Anecdotal evidence and cases of cross-border market manipulation in the institutional background section suggest that there are doubts about the identity of northbound investors. However, it is important to recognize that our *round-trip trading hypothesis* is also consistent with the argument about information advantage of northbound investors except that *some* northbound investors are not true foreigners. We first discuss the practicability of round-trip trading. The summary statistics for portfolios sorted by HP in a subsample with positive overseas income are reported in Panel C of Table 1. The amount of overseas income increase almost monotonically with HP. This is likely a sign of round-trip trading using foreign-currency business income through the connected market. To our best knowledge, Hong Kong exchange and regulators did not collect details about investor identity until the launch of Northbound Investor Identification Model[13]. The supervision loophole and data limitation makes it empirically

---

[13] Please refer to https://www.hkex.com.hk/-/media/HKEX-Market/Mutual-Market/Stock-Connect/Reference-Materials/Northbound-Investor-ID-Model/NB-Investor-ID-Information-Paper-Chi.pdf?la=zh-HK for more details about regulation on northbound investor identity. To validate the details in the institutional design, we conducted several rounds of interviews with the designers, regulators and technicians that were involved in the major events. In the routine regulation, CSDCC and the mainland exchanges monitor the broker-level trading of northbound investors. After the establishment of

difficult for us to conduct a direct investigation into who is behind the northbound trading. However, we do find some supportive evidence for our *round-trip trading hypothesis* in this section. First, we examine the price effect of regulation change on the cross-sectional predictability of northbound HPC along with a placebo test. Second, we show a striking similarity in trading behavior between northbound investors and opportunistic insiders on the mainland of China.

4.4.1 Price Effect of Regulation Change

Panel A in Table 10 reports the cross-sectional predictability of northbound HPC for the two sub-periods. The first sub-period spans from March 2017 to August 24, 2018 when HKEX scheduled the rollout of the investor identification model for northbound trading in the Stock Connect Scheme, and the second one from September, 2018 to the present[14]. In Panel A, the return spread between extreme groups sorted by northbound HPC is statistically significant at the 1% level before the announcement of investor identification regulation. However, the cross-sectional predictability within the more recent sub-period shows sign of decay. It suggests that a more stringent regulation with respect to northbound investor identification plays an important role in hamstringing round-trip trading. It can be rationalized that the motive of round-trip trading may be gravely diminished because the regulation change would result in more efficient cross-border surveillance on the connected market.

[Table 10 about here]

There might be concern that potential compounding factors cause the decay of return predictability after September 2018, which weakens our argument about the prevalence of round-trip trading. In Panel B of Table 10, we show further evidence with a placebo test examining the effect of regulation change on the return predictability of changes in southbound shareholdings for H share returns. If it is more of a compounding factor related story, for example, changes in the market structure over time, rather than changes in regulation on northbound investor identity, then the decay in return predictability of southbound HPC would be also expected during the more recent sub-period. However, an even larger return spread for the second sub-period in Panel B rejects the conjecture and shows

Northbound Investor ID model, HKEX keeps the account-level data available for examination by the regulators on the connected markets upon request.

[14] Here we divide the sample weeks based on the announcement date of Northbound investor identification model, i.e., August 24, because it would be expected that once the plan for regulation change is announced to the mutual market, northbound investors may respond to it by reducing round-trip investment and insider trading. 2018. In untabulated tables, we also repeat our main analysis using the introduction date, i.e., September 26, 2018.

further support to our regulation arbitrage hypothesis. Because there have been strict regulations governing the identification of southbound investors on the mainland exchanges from the outset, southbound trading is more likely driven by information edge. Hence, it can be rationalized that the regulation change on the identification of northbound investors have no diminished effect on the return predictability of southbound HPC.

[Table 11 about here]

Furthermore, we find a heterogeneous effect of regulation change on state-owned enterprises (SOEs) and non-SOEs. In Table 11, we compare the average weekly excess returns of portfolios sorted by HPC conditional on firm attribute of ownership. Panel A and B presents the results before and after the launch of Northbound Investor Identification Model, respectively. The return spreads between extreme quintiles sorted by HPC are positive and statistically significant at 1% in both SOEs and non-SOEs before regulation change. However, the return spread is statistically positive in only SOEs after the regulation change and the difference in return spreads between the two groups is positive and statistically significant at 5%. It suggests that the cross-sectional predictability of northbound HPC for non-SOEs are more vulnerable to the stringent regulation on investor identity. This is intuitive, because it is well documented in prior studies that in China, SOEs are subject to heavier scrutiny by authorities (See, for example, Chen et al., 2006; Jiang et al., 2010). Given that the round-trip trades may be more common in non-SOEs, the decay in return predictability can be rationalized in the subgroup due to a narrowed arbitrage opportunity after regulation change.

4.4.2 Similarity between Northbound Investors and Insiders

Under what we call *the round-trip trading hypothesis*, it would be expected that *some* mainland investors who are subject to scrutiny by regulators, would be prone to exploit the information edge by disguising themselves as northbound investors. As mentioned above, authorities can scrutinize round-trip trades more heavily over time. Given the risk of enforcement action, we expect round-trip trading most often in a group of opportunistic insiders whose private information is important enough to make the illegitimate profits high on the connected market. Opportunistic insiders who have offshore accounts, may trade their own firm on the connected markets to avoid the unwanted attention from the regulators. For those who do not have offshore accounts, opportunistic insiders would struggle to conceal identity, but they could still exploit the private information within the bounds of regulatory integrity and firm permission. We therefore expect a significant similarity between a set of opportunistic insiders and northbound investors trading on the same connected A-shares.

Ideally, it is supposed to distinguish between the mainland insiders and foreigners among northbound investors exactly and track their trading behaviors, respectively. However, even after the launch of Northbound Investor Identification Model, Broker-to-Client Assigned Numbers and Client Identification Data - the two main components of the regulation change - are for regulators' market surveillance only and not released to the public. Hence, adapted from Ali and Hirshleifer (2017) as well as Cziraki and Gider (2019), opportunistic insiders are identified from all insiders on the mainland market by measuring the past profitability in our paper. Specifically, we measure the dollar profits of the trades insiders make in the four-week trading days (about one calendar month) prior to QEAs within a five-day window centered at each QEA date. In each quarter, we rank insiders into halves based on the profitability. The top 50% are then identified as opportunistic insiders, a counterpart of northbound investors trading on the same stocks through different channels[15]. We track the trades of all identified opportunistic insiders either within or outside of pre-QEA windows.

[Table 12 about here]

In Table 12, we examine whether northbound investors trade similarly with opportunistic insiders using their private information. Panel A reports the summary statistics of weekly changes in shareholdings of northbound investors and opportunistic insiders, respectively. The mean values of weekly changes in the shareholdings in the two groups are both positive. It suggests that, on average, opportunistic insiders and northbound investors tend to increase their shareholdings on the specific firms. The result is also consistent with the previous literature that information advantages are easier to identify for insider purchases and predictability is to some extent limited to the long side. There are 154 unique firms with opportunistic insiders in our sample, 127 of which are non-SOEs, accounting for about 82.47%. Compared with the share of non-SOEs (about 70% as of the middle of 2019) in the entire Chinese A-shares, a higher percentage in this subsample suggests that insiders in non-SOEs may be more likely involved in regulation arbitrage. It shows further support to our findings in Table 11 that the return predictability of northbound HPC are more affected by regulation change on investor

[15] We aggregate the trades of opportunistic insiders at the firm level and extract the data of changes in northbound shareholdings on the same A shares all these opportunistic insiders trade.

identity in the group of non-SOEs. The average number of aggregated opportunistic insider trades per firm is only 1.22. It is consistent with the fact that many listed firms have restrictions on insider trading. Hence, the sparse opportunistic trades and congruent round-trip ones may provide very revealing information about the true identity of northbound investors. As shown at the bottom of Panel A, the correlation of trading behaviors between the two groups is positive and statistically significant. In Panel B, we regress HPC of northbound investors on contemporaneous weekly changes in shareholdings of opportunistic insiders and the coefficient estimate is significantly positive. Overall, the similarity of contemporaneous changes in shareholdings between opportunistic insiders and northbound capital suggests that there is possibility that *some* northbound investors are opportunistic insiders on the mainland market which take detours to exploit private information on the connected markets.

4.5 Hypothesis 3: Copycat Herding Hypothesis

4.5.1 Copycat Trading of Mainland Investors

In the above analysis testing Hypothesis 1, we address the concern that the return predictability is solely driven by demand pressure. It can scarcely be that there is no information advantage at all and northbound investors just anticipate the copycat trading of noise traders. However, we cannot rule out the possibility that other investors, especially the mainland investors, may regard northbound capital as smart money with information edge and tend to imitate its trading behavior. Then imitative flows from other investors would help boost the stock price and contribute to the return predictability in the short run, at least partially. To detect the copycat trading, we first estimate the daily changes in shareholdings of the mainland investors at the stock level, namely $IMHERD_{i,t}$:

$$IMHERD_{i,t} = \frac{net\ \ buyer\text{-}initiated\ \ volume_{i,t}}{Total\ \ shares\ \ outstanding_{i,t}} - HPC_{it} \tag{2}$$

where the first item denotes the ratio of stock i's net buyer-initiated volume on day t to total shares outstanding, and HPC_{it} denotes daily percentage changes in shareholdings of northbound investors[16].

[Table 13 about here]

In Table 13, we regress $IMHERD_{i,t}$ on the lagged terms of northbound HPCs controlling the

[16] Similarly, we also estimate the weekly changes in shareholdings of the mainland investors at the stock level by subtracting weekly HPCs of northbound investors from the weekly ratio of net buying-initiated volumes to the total shares outstanding at the end of each week.

contemporaneous HPC as well as stock returns. In Column 1, the copycat effect of the mainland investors fades moving from HPC on the immediately lagged trading day to five days earlier. Among one week, changes in northbound shareholdings within the most recent three days significantly predict the trading of mainland investors. As shown in Column 2, mainland investors only track closely northbound trading in the last week, suggesting that the imitative herding is not persistent in the long term. In sharp contrast with the baseline results that HPC of northbound investors has strong return predictability that does not reverse in the long run, $IMHERD_{i,t}$ has even no predictability in the short run[17]. These results provide further evidence supporting our *copycat trading hypothesis* that compared with northbound capital, the mainland investors have limited information on the connected stocks. Rather than arbitragers detect mispricing or insiders exploiting the private information, the mainland investors are in the nature of copycats tracking northbound capital.

4.5.2 Attention-Induced Copycat Herding

To figure out the presence of flow pressure driven by imitative herding, we focus on a subgroup of securities which are more likely affected by investor attention, i.e., stocks on the daily most actively traded lists. Intuitively, these stocks tend to attract more media coverage and public attention, because the apps, such as FT Chinese and The Wall Street Chinese, share the news feed of the league tables of the ten most actively traded stocks every trading day[18]. It is natural that these stocks on the daily lists are more likely affected by imitative herding flows from the mainland investors[19]. Column 3 in Table 13 confirms our conjecture that limited attention would intensify the imitative herding among the mainland investors. Specifically, Top10, a dummy variable, takes the value of 1 if the stock is on the daily league table of the ten most actively traded stocks, and 0, otherwise. We extend the setting in Column 1 by including the indicator variable along with the interaction terms with northbound HPCs. Consistent with the main results in Column 1, the imitative herding of the mainland investors concentrates on the most recent three days (excluding HPC on that day because of a time lag). Moreover, the interaction terms are positive and statistically significant during the last three days,

[17] Please refer to the Internet Appendix (Table I.4) for the detailed results about portfolios sorted by $IMHERD_{i,t}$.

[18] Note that most major data providers like WIND, CSMAR and Choice database, also report the league table of the ten most actively traded stocks. Generally, the data related to the ten most actively traded stocks have more media exposure than those related to northbound shareholdings of all connected A-shares.

[19] Given the potential limited attention effect, we also drop the stocks on the daily lists and repeat our main analysis in Table 2 (univariate sorts) and Table 4 (different formation and holding periods). The results are quantitatively similar and available upon request.

suggesting that for the stocks on the ten most actively traded league tables, northbound capital flows are more likely subject to the imitative trading on the Chinese mainland market. For example, a one-standard deviation increase in daily HPC of northbound investors for a stock on the league table, is associated with, on average, more increase in $IMHERD$ of 0.3601 $\times$ 0.394 (as reported in Panel A in Table 13) = 0.142 the following day relative to one security not on the list. Given that the average $IMHERD$ across the entire sample is 0.579, this corresponds to an economically significant increment in the mainland investors' copycat trading of about 24.50% for those connected stocks subject to more investor attention.

One might expect that if the mainland investors pay excessive but transient attention to the most actively traded stocks, then there would be a visible fall in returns when copycat flows shift their focus with northbound capital. How stock returns behave in a subsample that are likely subject to excessive attention within a short window can also provide us with a hint at the copycat-trading channel. Quintile portfolios sorted by daily HPC deliver a quick return reversal within one week in the subsample of stocks on the league table of northbound trading volumes, though we document a significantly positive return spread in the immediate three days after portfolio formation[20]. However, for the entire sample, the return predictability is rather persistent and lasts at least eight weeks. This baseline makes the dramatic reversal in cross-sectional predictability in the subsample even more interesting. It suggests that copycat trading driven by attention is a contributor to return predictability, however, it is by no means the only factor at play. It is also consistent with the finding documented by Barber and Odean (2008) that retail investors' purchases driven by attention-grabbing events can temporarily inflate stock prices. They confirm the hypothesis that preferences determine choices after attention has determined the choice set. In this terminology, we show evidence that attention determines the choice of the mainland investors after northbound investors' preferences have determined the choice set in our setting of China's connected stock market.

4.6 Path Analysis

Northbound capital can predict stock return in general, either through information advantage due to fundamental information edge, the round-trip trading of insiders with private information, or via

[20] Please refer to Table I.5 in the Internet Appendix for daily returns of quintile portfolios sorted by northbound HPC in the subsample of stocks on the ten most actively traded lists. The return spread decreases monotonically from 0.37% immediately after the northbound HPC becomes public to almost zero as of the end of the first week after formation.

copycat trading. To assess the relative importance of information edge versus herding channels through which HPC can predict future return, we perform a path analysis for high-minus-low returns of portfolios sorted by daily HPC. Following Baderstcher, Shanthikumar and Teoh (2018), we therefore estimate the following regressions:

$$ret_{i,t} = \alpha + \beta_1 HPC_{i,t-1} + \beta_2 HPC_{i,t-2} + \beta_3 IMHERD_{i,t-1} + control_{i,t} + \varepsilon_{i,t} \quad (3)$$

$$IMHERD_{i,t} = \gamma_0 + \gamma_1 HPC_{i,t-1} + control_{i,t} + \epsilon_{i,t} \quad (4)$$

where i indexes firms and t denotes days. All regressions include day and firm fixed effects in addition to the control variables in the vectors $control_{i,t}$, including Size, BM, ROA, SOE, VOL, IVOL, TURN, REVERSAL, RUNUP, MSCI and InstHold with standard errors clustered by firm and trading day. The estimated value of β_1 captures the direct effect of HPC on stock future return, and the estimated value of $\beta_2 \times \gamma_1$ captures the indirect effect of HPC on stock return through attracting imitative herding of mainland investors. We interpret the direct effect as coming from either fundamental information due to global accessibility or private information due to insiders' round-trip trading. Intuitively, if the cross-sectional predictability of information advantage to stock return is economically significant regardless of whether this information edge was induced by processing public information or exploiting private information or other information-related channels, the effect of HPC through the information channel is captured by the corresponding coefficient in the first equation, with the direct effect captured by the $HPC_{i,t-1}$ coefficient. Equation (4) gives the coefficient needed to rescale the $HPC_{i,t-2}$ coefficient in Equation (3) to reflect the sensitivity of $IMHERD_{i,t-1}$ to $HPC_{i,t-2}$.

[Table 14 about here]

Panel A of Table 14 reports key coefficient estimates from the two-step regressions. The percentages at the bottom of Panel B summarize the portion of the total effect of HPC that is through the information versus non-information channel[21]. The preponderance of the total effect of northbound HPC on the future return of connected stocks, 90.09%, comes from the information channel. The

[21] In the Internet Appendix, we also perform a path analysis for high-minus-low returns of portfolios sorted by weekly HPC in the Table I.7. The results are qualitatively similar. The preponderance of the total effect comes from the information channel, up to about 98.90%, larger than that on a daily basis. The imitative herding channel contributes the remaining 1.11%. The lower proportion of copycat effect is intuitive on a weekly basis, because the imitative herding is not persistent and concentrated in the short run, almost within one week. The conclusion that information edge is the primary channel through which HPC predicts stock return on a weekly basis shows consistency with that on a daily basis.

imitative herding channel contributes the remaining 9.91%. Thus, we obtain the conclusion bringing out the theme of information glittering that information edge is the primary channel through which HPC predicts stock return both in the short and long term.

4.7 Time-Series Predictability of HPC

In this section, we explore the time-series predictability of northbound capital on the market returns. We document that the net inflow of northbound capital is a reliable signal on a weekly basis. The outstanding predictability is statistically and economically significant both in and out of sample. More importantly, a trading strategy based on the model prediction can bring substantial profits for investors.

4.7.1 In-sample and out-of-sample prediction

We use the aggregate-level variable, the amount of net inflows from northbound investors (denominated in RMB yuan, $NetInflow$) from Shanghai/Shenzhen-Hong Kong Stock Connect Scheme, to predict future stock market returns in Mainland China. Three market indexes are to be predicted: value weighted portfolio of all A shares, value weighted portfolio of A shares in Shanghai stock exchange (Shanghai Composite Index), and value weighted portfolio of A shares in Shenzhen stock exchange (Shenzhen Composite Index). Available data sample is longer for analysis in this section, spanning from November 2014 to December 2018, which provides us with sufficient observations of 210 weeks to estimate the parameters.

[Table 15 about here]

We apply the commonly used univariate predictive regressions for in-sample forecasting as follows:

$$R^e_{t+1} = \alpha + \beta NetInflow_t + \varepsilon_{t+1} \tag{5}$$

where R^e_{t+1} is the market return in excess of the risk-free rate at week t+1, $NetInflow_t$ is the lagged net inflow of northbound capital. Here our interest is that how well the model fits with real values, so we focus on the in-sample R^2. We also examine the out-of-sample predictive power of net inflows from northbound capital. As documented in Goyal and Welch (2008), out-of-sample test is more relevant for evaluating return predictability in real time as it avoids the over-parameterization issue. In line with Goyal and Welch (2008), Han and Li (2017), we adopt a recursive estimation scheme for out-of-sample forecasts using the following single-predictor model:

$$\mathrm{E}_t(R^e_{t+1}) = \alpha_t + \beta_t NetInflow_t \quad (6)$$

where α_t and β_t are estimated recursively using data from the first week to week t. The initial estimation period is from November 2014 to December 2015. To figure out whether the predictability is stable over time, out-of-sample performance is evaluated in three sub-samples: January 2016 to December 2018, January 2017 to December 2018 and January 2018 to December 2018. We compare the model forecasts with historical average ones through two widely used statistics, i.e., R^2_{OS} and annualized certainty equivalent return (CER).

Table 15 presents the forecasting performance. As shown in Column 1, in-sample R^2 is 2.16% when predicting the excess returns of value weighted portfolio of all A shares. Considering that any single predictor hardly beats the historical benchmark on Chinese stock market, our results indicate that $NetInflow$ is a potentially useful predictor for excess market returns. Out-of-sample examinations provide more striking evidences. R^2_{OS} ranges from 3.48% to 6.37%, which is surprisingly high under such a simple model specification. As for economic implications, an investor who balances portfolio between a risk-free asset and the market portfolio according to model forecast rather than historical mean, could achieve annualized profit gain of 19.09% during January 2016 to December 2018, 24.31% during January 2017 to December 2018, and 42.81% during January 2018 to December 2018. The second and third column in Table 15 show that $NetInflow$ also has outstanding predictability for Shanghai Composite Index and Shenzhen Composite Index, with positive R^2_{OS} and substantial CER. Overall, we conclude that net inflows of northbound capital predict market returns in and out of sample. The significant predictive power could stem from its information content.

4.7.2 Market Timing Strategy

To shed more lights on the time-series predictability, referring to Han and Li (2017), we propose a simple market timing strategy. The design of the strategy is as follows: At the end of each week, investors would take a long position of the market return over the next week, if he or she receives a buying signal. Otherwise, investors would liquidate the risky market portfolio and then use the proceeds to invest in the risk-free asset. The trading signal is identified based on the out-of-sample predictions as illustrated in Section 4.7.1. When the predicted excess market return is positive, it is defined as a buying signal. For brevity, we only report the results of predicting value-weighted portfolio of all A shares. Returns of buy-and-hold strategy are set to be the benchmark. Additionally, to investigate the predictability under different horizons, holding periods range from one to eight weeks. We use the methodology suggested by Jegadeesh and Titman (1993) with overlapping portfolios to derive the time series of the weekly returns for all holding periods larger than one week. Out-of-sample evaluation period spans from January 2016 to December 2018.

Panel A in Table 16 reports annualized excess returns (in percent) and annualized Sharpe ratio for the benchmark as well as the timing strategy. As it stands, the simple buy-and-hold strategy incurs a loss, with negative annualized excess return of -5.28% and negative Sharpe ratio of -0.29. In contrast, our proposed strategy yields remarkable profits. In the case of one-week holding period, the annualized excess return is up to 12.83%, associated with large positive Sharpe ratio of 0.82, significantly outperforming the benchmark. Looking across the different holding periods, excess return reaches the highest when holding the original position for three weeks. In addition, it seems that the forecastability of $NetInflow$ mainly exists in the short run from one week ranging to five weeks, which confirms our argument that the northbound capital is an important signal for the short-term market return. Panel B in Table 16 provides more evidence by estimating the risk-adjusted alphas of investment strategies based on Carhart (1997) four-factor model. We report the results of the benchmark and the timing strategy where the holding period is one week. As anticipated, our proposed strategy earns more significant and larger risk-adjusted return of 0.22% per week, relative to the insignificant risk-adjusted return of 0.01% generated by the buy-and-hold strategy.

[Table 16 about here]

To sum up the time-series analysis, net inflows of northbound capital has outstanding predictability for market returns and its forecasting power can be translated into enormous economic gains.

4.8 Robustness Check

We perform a bunch of robustness tests to determine whether the main findings are sensitive to our research design. Some of these tests have been mentioned or footnoted throughout the text; others are discussed here or reported in the Appendix. To address the concern on estimation errors in extreme groups, we repeat univariate portfolio sorts using the back-tested procedure suggested by Mamaysky et al. (2008) in Table I.1 of the Internet Appendix. The cross-sectional predictability of HPC is even more salient in the back-tested sample. In Table I.2, we examine daily HPCs and find that the cross-sectional predictability is robust to such a trading strategy. We also checked whether our results are sensitive to alternative definitions of aggregate trading behavior of northbound investors. We replace HPC with net capital flows and weekly percentage changes in holding volumes (dollar value) at the individual stock level as alternative proxies for the stock selection of northbound investors in our baseline results. Results is consistent with earlier findings related to HPC. To determine whether northbound flows are purely value investors tracking "good" assets without any motivation driven information advantage, our analysis of portfolio sorts is repeated by using HPC residuals after a bunch of firm-level characteristics, as reported in Table I.3. To address the potential concern on the dominance of private information in the information edge of northbound investors, we repeat the heterogeneity tests in Table 8 in two sub-periods based on the launch of Northbound Investor Identification Model. As shown in Table I.6, the cross-sectional predictability slightly decreases in both groups with different levels of global trade accessibility and international investor exposure. However, the return spreads between extreme portfolios sorted by HPC remain positive and statistically significant, suggesting the fundamental information edge at play.

5 Conclusion

Although there is a growing body of research related to the Stock Connect Scheme, there are still many open questions. Rather than examine the difference of pricing efficiency between connected and unconnected stocks in earlier work on this topic, the current study focuses on the implications of information advantage on the pricing of changes in northbound (southbound) shareholdings. Empirically, we show evidence that northbound (southbound institutional) investors perform well in using information advantage and changes in their shareholdings have a relatively persistent predictability for stock future return on each other's market beyond demand pressure. In this way, we show further empirical support to the implication of the theoretical model in Albuquerque et al. (2009). Our main findings related to northbound investors' information edge as well as their outperformance in China's connected stock market are also consistent with those in Froot et al. (2001) focusing on the important role of international investors in price discovery on emerging markets.

Information edge, round-trip trading and copycat herding all play a visible role in the cross-sectional return predictability of changes in northbound investors' shareholdings. On the one hand, international investors may be informed about mispricing and earn the abnormal return along with the release of public information. Unlike typical demand pressure patterns in asset pricing, we find announcements effect associated with changes in northbound shareholdings, suggesting that northbound purchasing and selling are more likely informed trading. Besides, we cannot rule out the prevalence of round-trip trading of mainland insiders posing as northbound investors. The decay of return predictability after the launch of the Northbound Investor Identification Model as well as the high correlation between HPC and insiders' opportunistic trading shows evidence that *some* mainland insiders pretend to be northbound capital, hide their identities to avoid unwanted attention. On the other hand, a subsample of stocks on the ten most actively traded league tables, which are more likely affected by imitative herding of noise traders, show the long-run reversal following a positive return spread immediately after portfolio formation. It suggests that copycat herding driven by attention also contributes to the return predictability of changes in shareholdings of investors from the counterpart stock market, at least partially. The results provide insight into the pricing implication of information contained in newly introduced investors during China's integration with global financial market.

The extant studies highlight the difficulty in distinguishing between informed trading and noise traders' copycat trading. Our findings motivate future research that seeks to isolate the role of international investors' informative preferences from imitative herding as well as insiders' round-trip trading driven by regulation arbitrage. A limitation of our study is that in our stock-level tests, we are unable to measure to which degree northbound investors make use of copycat herding of noise traders to position themselves. Besides, one possibility that has not been examined in our paper is whether the positive association between HPC over a shorter window and future stock price reflect intra-day momentum trading of northbound investors. Future research can overcome the limitations by incorporating account-level data to track the trading motives of participants on the connected market, attributed to information advantage, copycat herding, regulation arbitrage or momentum strategy.

Reference

Agarwal, S., Faircloth, S., Liu, C., & Rhee, S. G. (2009). Why do foreign investors underperform domestic investors in trading activities? Evidence from Indonesia. *Journal of Financial Markets, 12*(1), 32-53.

Aggarwal, R., Erel, I., Ferreira, M., & Matos, P. (2011). Does governance travel around the world? Evidence from institutional investors. *Journal of Financial Economics, 100*(1), 154-181.

Albuquerque, R., Bauer, G. H., & Schneider, M. (2009). Global private information in international equity markets. *Journal of Financial Economics*, *94*(1), 18-46.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259-299.

Azémar, C., Desbordes, R., & Mucchielli, J. L. (2007). Do tax sparing agreements contribute to the attraction of FDI in developing countries?. *International Tax and Public Finance, 14*(5), 543-562.

Badertscher, Brad, Devin Shanthikumar, and Siew Hong Teoh, 2018, Private firm investment public peer misvaluation, Working paper.

Bailey, W., Mao, C. X., & Sirodom, K. (2007). Investment restrictions and the cross-border flow of information: Some empirical evidence. *Journal of International Money and Finance, 26*(1), 1-25.

Bailey, W. , Cai, J. , Cheung, Y. L. , & Wang, F. . (2009). Stock returns, order imbalances, and commonality: evidence on individual, institutional, and proprietary investors in china. *Journal of Banking and Finance, 33*(1), 0-19.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies, 21*(2), 785-818.

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*(9), 3009-3047.

Brennan, M. J., & Cao, H. H. (1997). International portfolio investment flows. *The Journal of Finance, 52*(5), 1851-1880.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance, 52*(1), 57-82.

Chakravarty, S., & Mcconnell, J. J. (1999). Does insider trading really move stock prices?. *Journal of Financial & Quantitative Analysis, 34*(2), 191-209.

Chan, K., Menkveld, A. J., & Yang, Z. (2008). Information asymmetry and asset prices: Evidence from the China foreign share discount. *The Journal of Finance*, *63*(1), 159-196.

Chen, G., Firth, M., Gao, D. N., & Rui, O. M. (2006). Ownership structure, corporate governance, and fraud: evidence from china. Journal of Corporate Finance, 12(3), 0-448.

Chen, L. W., Johnson, S. A., Lin, J. C., & Liu, Y. J. (2009). Information, sophistication, and foreign versus domestic investors' performance. *Journal of Banking & Finance, 33*(9), 1636-1651.

Chen, Y., Da, Z., & Huang, D. (2018). Arbitrage trading: The long and the short of it. *The Review of Financial Studies, 32*(4), 1608-1646.

Choe, H., Kho, B. C., & Stulz, R. M. (1999). Do foreign investors destabilize stock markets? The Korean experience in 1997. *Journal of Financial Economics, 54*(2), 227-264.

Choe, H., Kho, B. C., & Stulz, R. M. (2005). Do domestic investors have an edge? The trading experience of foreign investors in Korea. *The Review of Financial Studies, 18*(3), 795-829.

Cziraki, P. & Gider J. (2019). The dollar profits to insider trading. Working Paper.

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of finance, 52*(3), 1035-1058.

Dvořák, T. (2005). Do domestic investors have an information advantage? Evidence from Indonesia. *The Journal of Finance, 60*(2), 817-839.

Edelen, R. M., Ince, O. S., & Kadlec, G. B. (2016). Institutional investors and stock return anomalies. *Journal of Financial Economics, 119*(3), 472-488.

Engelberg, J., McLean, R. D., & Pontiff, J. (2018). Anomalies and news. *The Journal of Finance, 73*(5), 1971-2001.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy, 81*(3), 607-636.

Ferreira, M. A., Matos, P., Pereira, J. P., & Pires, P. (2017). Do locals know better? A comparison of the performance of local and foreign institutional investors. *Journal of Banking & Finance*, 82, 151-164.

Froot, K. A., O'connell, P. G., & Seasholes, M. S. (2001). The portfolio flows of international investors. *Journal of financial Economics, 59*(2), 151-193.

Froot, K. A., & Ramadorai, T. (2008). Institutional portfolio flows and international investments. *The Review of Financial Studies, 21*(2), 937-971.

Froot, K., & Teo, M. (2008). Style investing and institutional investors. *Journal of Financial and Quantitative Analysis, 43*(4), 883-906.

Fung, H. G. , & Zhang, Y. G. . (2011). Reported trade figure discrepancy, regulatory arbitrage, and round-tripping: evidence from the China—Hong Kong trade data. *Journal of International Business Studies, 42*(1), 152-176.

Grinblatt, M., & Keloharju, M. (2000). The investment behavior and performance of various investor types: a study of Finland's unique data set. *Journal of Financial Economics, 55*(1), 43-67.

Grossman, S. J., & Stiglitz, J. E. (1976). Information and competitive price systems. *The American Economic Review, 66*(2), 246-253.

Han, X., & Li, Y. (2017). Can investor sentiment be a momentum time-series predictor? Evidence from China. *Journal of Empirical Finance, 42*, 212-239.

Hau, H. (2001). Location matters: An examination of trading profits. *The Journal of Finance, 56*(5), 1959-1983.

Hou, Y. and Ye, D. (2008). Institutional investors, informed trading and market efficiency. *Journal of Financial Research*, 334(4), 131-145 (in Chinese).

Huang, R. D., & Shiu, C. Y. (2009). Local effects of foreign ownership in an emerging financial market: Evidence from qualified foreign institutional investors in Taiwan. *Financial Management, 38*(3), 567-602.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance, 48*(1), 65-91.

Jia, C., Wang, Y., & Xiong, W. (2017). Market segmentation and differential reactions of local and foreign investors to analyst recommendations. *The Review of Financial Studies, 30*(9), 2972-3008.

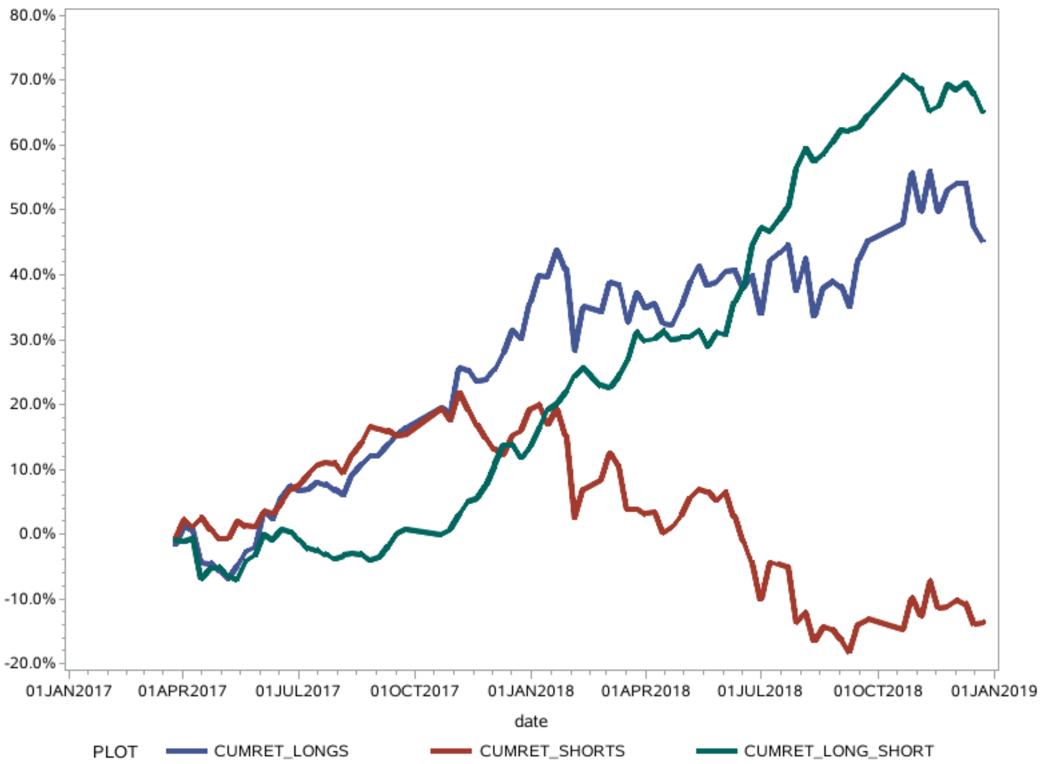
Jiang, G. , Lee, C. M. C. , & Yue, H. . (2010). Tunneling through intercorporate loans: the china experience. *Journal of Financial Economics, 98*(1), 1-20.

Jiao, Y., Massa, M., & Zhang, H. (2016). Short selling meets hedge fund 13F: An anatomy of informed demand. *Journal of Financial Economics, 122*(3), 544-567.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315-1335.
Ledyaeva, S., Karhunen, P., Kosonen, R., & Whalley, J. (2015). Offshore foreign direct investment, capital round-tripping, and corruption: Empirical analysis of Russian regions. *Economic Geography*, *91*(3), 305-341.
Liu, C., Wang, S., & Wei, K. C. (2018). Demand Shock, Speculative Beta, and Asset Prices: Evidence from the Shanghai-Hong Kong Stock Connect Program. Working Paper.
Livnat, J., & Mendenhall, R. R. (2006). Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of accounting research*, *44*(1), 177-205.
Maffett, M. (2012). Financial reporting opacity and informed trading by international institutional investors. *Journal of Accounting and Economics, 54*(2-3), 201-220.
Mamaysky, H., Spiegel, M., & Zhang, H. (2008). Estimating the dynamics of mutual fund alphas and betas. *The Review of Financial Studies*, 21(1), 233-264.
Richards, A. (2005). Big fish in small ponds: The trading behavior and price impact of foreign investors in Asian emerging equity markets. *Journal of Financial and quantitative Analysis, 40*(1), 1-27.
Seasholes, M. S. (2000). Smart Foreign Traders in Emerging Markets. Harvard Business School, Working Paper.
Shan, C. Y., Tang, Y. J., Wang, Q., & Zhang, C. (2018). The Diversification Benefits and Policy Risks of Accessing China’s Stock Market. Working paper.
Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance, 52*(1), 35-55.
Shukla, R. K., & Van Inwegen, G. B. (1995). Do Locals Perform Better than Foreigners? : An Analysis of UK and US Mutual Fund Managers. *Journal of Economics and Business, 47*(3), 241-254.
Sias, R., Turtle, H. J., & Zykaj, B. (2015). Hedge fund crowds and mispricing. *Management Science, 62*(3), 764-784.
Teo, M. (2009). The geography of hedge funds. *The Review of Financial Studies, 22*(9), 3531-3561.
Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies, 21*(4), 1455-1508.
Zhang, R. & Zhang, T. (2019). AH Premium: A Natural Experiment. Working paper.
Zhong, Y. L., & Lu, Z. F. (2018). Can the opening of stock market improve the information content of stock price? Empirical evidence from Shanghai-Hong Kong Stock Connect. *Management World, 34*(1), 169-179 (in Chinese).
Zhong, K., Sun, C. L., Wang, Y. Z., & Wang, H. C. (2018). Stock market liberalization and idiosyncratic return volatility evidence from “Shanghai–Hong Kong Stock Connect” scheme. *Journal of Financial Research, 7*, 174-192 (in Chinese).

Figure 1. Cumulative returns for portfolios sorted by HPC

This table plots the cumulative returns for value-weighted portfolios that we form on the change of northbound capital shareholding percent (HPC) from March 2017 to December 2018. At the beginning of each week, eligible stocks in Stock Connect Scheme are sorted into quintile portfolios on the basis of their HPCs and held for one week. Blue line denotes the cumulative returns for long portfolio where the stocks are in the highest HPC quintile. Red line denotes the cumulative returns for short portfolio where the stocks are in the lowest HPC quintile. Green line denotes the hedged portfolio which is constructed by buying long portfolio and selling short portfolio.



Time Series of Cumulative Long/Short/Hedge Strategy Return: Holding 1 week

Table 1. Summary statistics

This table reports the summary statistics of stock-level and firm-level characteristics from March 2017 to December 2018. Panel A describes the distribution of the main variables in the full sample: the change of shareholding percent (HPC, in percent), denoted as the difference in the ratio of northbound holding to the total shares outstanding between the end of week t and week t-1; level of holding positions among all northbound investors (HP, in percent); the log of market capitalization (Size, in thousand RMB); the book-to-market ratio (BM); the gross profits-to-assets ratio (ROA); stock return volatility (VOL), measured as the standard deviation of daily return over the past 3 months; idiosyncratic volatility (IVOL), computed as the standard deviation of residual from Fama-French three-factor regressions using 3-month daily data; turnover rate (TURN); stock return reversal (REVERSAL), the cumulative return from week t-4 to week t-1; stock return run-up (RUNUP), the cumulative stock return from week t-52 to week t-5. Size, BM, and ROA are quarterly accounting variables; HPC, HP, VOL, IVOL, TURN, RUNUP, and REVERSAL are weekly updated. Summary statistics include mean, median, and the 1th, 25th, 75th, and 99th percentiles. Panel B presents the mean value for main variables in decile portfolios sorted by weekly HPC. Panel C presents the mean value for overseas income in decile portfolios sorted by HP. Detailed definitions about the variables are listed in the Appendix.

Panel A. Summary statistics of full sample

Variables	Mean	Std. Dev	P1	P25	P50	P75	P99
HPC	0.010	0.421	-0.450	-0.010	0.000	0.030	0.540
HP	0.614	1.698	0.000	0.050	0.170	0.460	8.760
Size	16.703	0.933	15.144	16.058	16.524	17.145	19.724
BM	0.544	0.264	0.100	0.332	0.508	0.745	1.104
ROA	0.035	0.043	-0.032	0.009	0.023	0.049	0.192
VOL	0.023	0.008	0.008	0.017	0.022	0.028	0.046
IVOL	0.018	0.009	0.005	0.012	0.016	0.022	0.046
TURN	7.237	8.165	0.465	2.699	4.778	8.618	40.818
RUNUP	-0.048	0.379	-0.588	-0.277	-0.112	0.100	1.135
REVERSAL	-0.016	0.103	-0.259	-0.076	-0.018	0.039	0.271

Panel B. Decile portfolio sorted by HPC

Variables	1	2	3	4	5	6	7	8	9	10
HPC	-0.207	-0.043	-0.023	-0.011	0.000	0.007	0.018	0.029	0.061	0.256
HP	1.202	0.489	0.325	0.300	0.206	0.268	0.316	0.421	0.705	1.987
Size	16.594	16.451	16.360	16.355	16.395	16.379	16.403	16.487	16.568	16.750
BM	0.517	0.549	0.551	0.576	0.534	0.561	0.575	0.557	0.550	0.501
ROA	0.043	0.034	0.032	0.031	0.030	0.031	0.030	0.034	0.037	0.049
VOL	0.025	0.023	0.023	0.023	0.022	0.022	0.023	0.022	0.023	0.024
IVOL	0.020	0.018	0.018	0.018	0.017	0.017	0.017	0.018	0.018	0.019
TURN	9.055	7.216	6.819	6.517	7.081	6.475	6.332	6.784	7.310	8.384
RUNUP	0.050	-0.045	-0.102	-0.126	-0.071	-0.113	-0.119	-0.067	-0.019	0.094
REVERSAL	-0.004	-0.016	-0.019	-0.023	-0.019	-0.025	-0.027	-0.019	-0.012	0.000

Panel C. Overseas income of decile portfolios sorted by HP

HP	0.00	0.02	0.05	0.09	0.14	0.20	0.30	0.46	0.85	3.97
overseas income	806.84	506.90	718.76	1145.73	1701.16	2352.06	2164.28	2456.06	2926.84	3170.37

Table 2. Univariate portfolio sorts

This table presents average weekly excess returns and alphas (in percent) for value-weighted portfolios from March 2017 to December 2018. At the beginning of each week t, eligible stocks in the Stock Connect Scheme are sorted into deciles, quintiles or halves based on HPCs at week t-1 and held for one week. Panel A displays weekly excess returns. Panel B displays the variation of cross-sectional predictability across market and over time. The first four rows examine the decile portfolios and long-short portfolio returns using eligible stocks listed on Shanghai and Shenzhen exchange, respectively. The last four rows in Panel B examine weekly portfolio returns in two sub-periods, March 2017 to January 2018 and February 2018 to December 2018. Panel C displays benchmark adjusted returns, including Carhart four-factor adjusted alphas, DGTW adjusted returns, and excess returns of the strategy that skips one week between formation and holding period. The last column in each panel (High-Low) shows the performance of a long-short portfolio where stocks with HPC in the highest (lowest) quintile are assigned to the long (short) portfolio. Newey and West (1987) three-lag adjusted t-statistics are reported. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A. Excess returns for HPC sorted portfolios											
	Low	2	3	4	5	6	7	8	9	High	High-Low
Deciles	-0.13	-0.17	-0.35	-0.41	-0.29	-0.08	0.00	0.08	0.22	0.48*	0.61***
	(-0.45)	(-0.65)	(-1.31)	(-1.25)	(-1.03)	(-0.31)	(0.02)	(0.32)	(0.94)	(1.83)	(2.99)
	Low	2	3	4	High	High-Low			Low	High	High-Low
Quintiles	-0.15	-0.31	-0.17	0.05	0.35	0.50***		Halves	-0.18	0.17	0.35***
	(-0.54)	(-1.23)	(-0.74)	(0.21)	(1.41)	(3.31)			(-0.75)	(0.75)	(3.10)
Panel B. Excess returns for subsamples across markets and over time											
	Low	2	3	4	5	6	7	8	9	High	High-Low
Shanghai	-0.16	-0.12	-0.41	-0.45	-0.27	0.22	0.01	0.15	0.39	0.41	0.56***
	(-0.55)	(-0.45)	(-1.47)	(-1.49)	(-0.90)	(0.86)	(0.04)	(0.60)	(1.57)	(1.62)	(2.61)
Shenzhen	-0.01	-0.21	-0.26	-0.39	-0.34	-0.50	-0.08	-0.05	-0.04	0.50*	0.51*
	(-0.02)	(-0.70)	(-0.85)	(-0.91)	(-1.13)	(-1.33)	(-0.24)	(-0.18)	(-0.15)	(1.68)	(1.96)
First half	0.34	0.30	0.10	-0.50	0.15	-0.14	0.24	0.14	0.58**	0.84***	0.51*
	(1.31)	(1.17)	(0.36)	(-1.35)	(0.56)	(-0.37)	(0.64)	(0.45)	(2.35)	(2.66)	(1.70)
Second half	-0.59	-0.62	-0.78*	-0.36	-0.62	-0.04	-0.23	0.01	-0.13	0.12	0.71***
	(-1.21)	(-1.47)	(-1.71)	(-0.82)	(-1.50)	(-0.14)	(-0.61)	(0.04)	(-0.34)	(0.31)	(2.61)
Panel C. Benchmark adjusted returns and raw returns of one-week skip strategy											
	Low	2	3	4	5	6	7	8	9	High	High-Low
Carhart 4 factor alphas	0.06	0.02	-0.21*	-0.06	-0.11	0.21*	0.28*	0.32***	0.43***	0.66***	0.59***
	(0.52)	(0.23)	(-1.83)	(-0.30)	(-0.89)	(1.71)	(1.77)	(2.97)	(5.15)	(5.65)	(2.90)
DGTW adjusted return	-0.12	-0.17	-0.31	-0.40	-0.27	0.00	-0.03	0.08	0.24	0.48*	0.60***
	(-0.41)	(-0.68)	(-1.17)	(-1.20)	(-0.89)	(0.01)	(-0.11)	(0.32)	(0.99)	(1.83)	(2.98)
One-week skip	-0.21	-0.10	-0.13	-0.25	-0.41	-0.57*	-0.11	-0.13	0.12	0.15	0.36***
	(-0.77)	(-0.42)	(-0.53)	(-0.86)	(-1.64)	(-1.90)	(-0.39)	(-0.55)	(0.46)	(0.56)	(2.61)

Table 3. Fama-MacBeth regression

This table reports the results of second stage Fama-MacBeth regression. The dependent variable for all the models is the returns of long-short portfolio formed by sorting weekly HPC. In Model (1), only lagged HPC is included as the independent variable. In Model (2), firm-level and stock-level characteristics are added as control variables, including SIZE, BM, ROA, IVOL, TURN, MSCI, past one-week stock return (PastRet (1w)), past one-month stock return (REVERSAL), past one- to three-month stock return (PastRet (1m-3m)), and past twelve-month stock return (RUNUP). In Model (3), contemporaneous HPC is taken into account to control for the effect of capital flow pressure. All regressions are run on a weekly basis from March 2017 to December 2018. The last two rows in the table report the average observations and average R^2 for cross-sectional regressions at the end of each week. Newey-West three-lag adjusted t-statistics are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)
HPC(t-1)	1.12***	0.85***	0.69***
	(6.03)	(6.49)	(5.11)
HPC(t)			0.99
			(3.77)
Size		0.03	0.03
		(0.59)	(0.57)
BM		0.29	0.30
		(1.10)	(1.13)
ROA		1.21	1.17
		(0.70)	(0.69)
PastRet (1w)		-2.38*	-2.38*
		(-1.74)	(-1.74)
REVERSAL		0.73	0.67
		(0.92)	(0.85)
PastRet (1m-3m)		0.60	0.58
		(1.31)	(1.27)
RUNUP		0.10	0.09
		(0.77)	(0.70)
IVOL		-2.56	-2.51
		(-0.56)	(-0.55)
TURN		-0.04***	-0.04***
		(-5.76)	(-5.68)
MSCI		0.22***	0.23***
		(3.67)	(3.72)
Avg. Obs.	1302	1302	1302
Avg.R^2	0.29	12.13	12.49

Table 4. Cross-sectional return predictability in different formation and holding weeks

This table reports the accumulative returns (in percent) for long-short portfolios employing different length of formation weeks (F) and holding weeks (H). In Panel A, portfolios are constructed as follows: at the beginning of each week t, the eligible stocks are sorted into quintiles on the basis of F-week lagged HPC. Then we construct the long-short portfolio by buying stocks with HPC in the highest quintile, and selling stocks with HPC in the lowest quintile, and hold for H weeks. Average cumulative return is calculated by averaging the H-week cumulative returns generated at each week. In Panel B, we employ Jegadeesh and Titman (1993) method to construct portfolios. At the beginning of each week t, the eligible stocks on the connected market are ranked in ascending order on the basis of their HPC in the past F weeks, and sorted into five quintile portfolios. In each week t, we buy the stocks with HPC in the highest quintile (denoted as "Long") and sell the stocks with HPC in the lowest quintile (denoted as "Short"). The position is held for H weeks. In addition, we close out the position initiated in week t-F. Under this trading strategy, we revise the weights on 1/F of the stocks in the entire portfolio in any given week and carry over the rest from the previous week. The values of F and H are set to be 1, 2, and 4, respectively. Sample period is March 2017 to December 2018. Newey-West adjusted t-statistics are given below the portfolio returns in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A. Cross-sectional returns in the long run

	H							
F	1	2	3	4	5	6	7	8
1	0.50***	0.74***	1.11***	1.25***	1.32***	1.38***	1.28***	1.58***
	(3.31)	(4.01)	(4.90)	(4.54)	(4.57)	(4.00)	(3.28)	(3.81)
2	0.44***	0.86***	1.19***	1.34***	1.49***	1.46***	1.62***	1.93***
	(3.20)	(4.82)	(5.47)	(4.86)	(4.71)	(3.93)	(4.01)	(4.36)
3	0.60***	0.96***	1.30***	1.52***	1.53***	1.55***	1.78***	2.05***
	(4.20)	(5.47)	(5.56)	(4.88)	(4.55)	(3.97)	(4.10)	(4.36)
4	0.52***	0.92***	1.30***	1.38***	1.45***	1.65***	1.83***	2.11***
	(3.67)	(5.22)	(5.73)	(4.77)	(4.64)	(4.35)	(4.32)	(4.59)

Panel B. Cross-sectional returns based on Jegadeesh and Titman (1993) method

	F = 1			F = 2			F = 4		
H	1	2	4	1	2	4	1	2	4
Long	0.48	0.28	0.21	0.32	0.24	0.20	0.33	0.23	0.18
	(1.58)	(0.92)	(0.70)	(1.03)	(0.79)	(0.64)	(1.06)	(0.71)	(0.55)
Short	-0.13	-0.19	-0.12	-0.16	-0.23	-0.15	-0.17	-0.19	-0.12
	(-0.39)	(-0.59)	(-0.40)	(-0.51)	(-0.74)	(-0.49)	(-0.55)	(-0.59)	(-0.41)
Long-Short	0.61***	0.47***	0.34***	0.47***	0.48***	0.35***	0.49***	0.42***	0.30**
	(3.32)	(3.98)	(4.07)	(2.91)	(4.06)	(3.51)	(2.99)	(2.77)	(2.15)

Table 5. The determinants of HP and HPC

Panel A and Panel B in this table present the level of northbound capital shareholding (HP), and the potential determinants of the change of northbound capital shareholding percent (HPC), respectively. Sample period is March 2017 to December 2018. Independent variables include Size, the log of market capitalization (in thousand RMB) at the end of last Friday; BM, the latest book-to-market ratio calculated at least 3 month ago; ROA, the latest gross profits-a-assets ratio calculated at least 3 month ago; SOE, a dummy variable with 1 denoting that the firm is state-owned, and 0 denoting non-state-owned; VOL, the standard deviation of daily return over the past 3 months; IVOL, idiosyncratic volatility which is calculated as the standard deviation of return residual from Fama-French three-factor regressions using past 3-month daily data; TURN, the turnover rate in the past week; REVERSAL, the cumulative return from week t-4 to week t-1; RUNUP, return momentum which is the cumulative stock return from week t-52 to week t-5; MSCI, a dummy with 1 indicating the firm is listed in MSCI China index, and 0 indicating the opposite; InstHold, the ratio of shares held by institutional investors to the number of A shares outstanding; AnalyNum, the number of analysts covered the firm during past year. There is no fixed effect in Model (1); Model (2) has industry fixed; Model (3) has year-month fixed effect; Model (4)-Model (7) have both industry and year-month fixed effect. Standard errors are clustered at industry level. *, ** and *** represent statistical significance at the 10%, 5% and 1%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A. The determinants of HP				
Size	0.400***	0.492***	0.431***	0.527***	0.515***	0.430***	0.421***
	(61.27)	(69.49)	(65.98)	(5.36)	(5.31)	(4.10)	(4.04)
BM	-0.711***	-0.611***	-0.977***	-0.969***	-0.982***	-1.008***	-1.018***
	(-31.98)	(-24.29)	(-41.89)	(-3.64)	(-3.70)	(-3.36)	(-3.41)
ROA	3.168***	2.373***	3.248***	2.286***	2.269***	1.485*	1.473*
	(24.61)	(18.72)	(22.65)	(2.96)	(2.98)	(1.77)	(1.77)
SOE	0.093***	0.078***	0.129***	0.114	0.094	0.126	0.109
	(8.47)	(6.77)	(11.93)	(1.56)	(1.23)	(1.51)	(1.25)
VOL	21.322***	22.857***	-3.767***	-3.972	-4.415	-9.499**	-9.858**
	(22.06)	(24.23)	(-3.38)	(-1.00)	(-1.12)	(-2.11)	(-2.19)
IVOL	-0.928	-1.939**	3.218***	2.251	1.763	2.162	1.779
	(-1.07)	(-2.32)	(3.63)	(1.33)	(0.98)	(1.14)	(0.90)
TURN	-2.560***	-2.011***	-1.591***	-1.011***	-0.811***	-0.704**	-0.540
	(-31.03)	(-24.94)	(-18.56)	(-3.93)	(-2.73)	(-2.09)	(-1.49)
REVERSAL	0.459***	0.235***	0.454***	0.257*	0.238	0.283*	0.269*
	(9.10)	(4.85)	(8.58)	(1.88)	(1.66)	(1.94)	(1.77)
RUNUP	0.277***	0.149***	0.435***	0.310***	0.308***	0.349***	0.346***
	(18.26)	(9.85)	(28.27)	(3.24)	(3.17)	(3.45)	(3.38)
MSCI	0.985***	0.944***	0.604***	0.555***	0.554***	0.492***	0.491***
	(42.97)	(42.95)	(24.80)	(3.24)	(3.24)	(2.95)	(2.95)
InstHold					0.156		0.137
					(1.27)		(1.08)
AnalyNum						0.017***	0.017***
						(3.76)	(3.71)

(Continued on the next page)

Table 5 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. The determinants of HP							
Industry FX	No	Yes	No	Yes	Yes	Yes	Yes
Month FX	No	No	Yes	Yes	Yes	Yes	Yes
#observations	112,080	112,080	112,080	112,080	111,710	103,060	102,843
R^2	0.132	0.211	0.155	0.233	0.234	0.245	0.245
Panel B. The determinants of HPC							
HPC(t-1)							-0.068
							(-1.45)
Size	0.002	0.002	0.001	0.001	0.001	0.000	0.000
	(1.10)	(1.11)	(0.74)	(1.54)	(1.38)	(0.29)	(0.26)
BM	0.005	0.005	-0.003	-0.004	-0.004	-0.004	-0.003
	(0.90)	(0.90)	(-0.62)	(-1.66)	(-1.63)	(-1.26)	(-1.20)
ROA	0.177***	0.180***	0.106***	0.108***	0.108***	0.104***	0.113***
	(5.98)	(5.90)	(3.20)	(5.38)	(5.40)	(4.78)	(5.27)
SOE	-0.001	0.000	-0.000	0.000	-0.000	0.000	0.000
	(-0.24)	(0.05)	(-0.16)	(0.34)	(-0.01)	(0.46)	(0.25)
VOL	0.683***	0.675***	-0.141	-0.178	-0.179	-0.188	-0.206
	(3.08)	(2.97)	(-0.55)	(-1.33)	(-1.33)	(-1.19)	(-1.26)
IVOL	-0.580***	-0.591***	0.123	0.121	0.120	0.128	0.123
	(-2.90)	(-2.94)	(0.60)	(0.87)	(0.86)	(0.84)	(0.77)
TURN	-0.020	-0.016	-0.046**	-0.043***	-0.042***	-0.047***	-0.046***
	(-1.08)	(-0.81)	(-2.34)	(-4.08)	(-3.67)	(-3.55)	(-2.87)
REVERSAL	0.026**	0.024**	0.059***	0.058***	0.058***	0.064***	0.066***
	(2.21)	(2.07)	(4.81)	(4.61)	(4.60)	(4.82)	(4.57)
RUNUP	0.005	0.004	0.008**	0.007***	0.007***	0.007***	0.007***
	(1.35)	(1.03)	(2.17)	(4.32)	(4.27)	(4.14)	(3.59)
MSCI	0.011**	0.011**	0.008	0.007**	0.007**	0.007**	0.007**
	(2.02)	(1.99)	(1.35)	(2.42)	(2.42)	(2.35)	(2.22)
InstHold					0.002		0.002
					(1.09)		(1.04)
AnalyNum						0.000	0.000*
						(1.25)	(1.83)
Industry FX	No	Yes	No	Yes	Yes	Yes	Yes
Month FX	No	No	Yes	Yes	Yes	Yes	Yes
# observations	112,080	112,080	112,080	112,080	111,710	103,060	101,367
R^2	0.001	0.001	0.017	0.017	0.017	0.018	0.023

Table 6. Cross-sectional returns around information release

This table reports the cumulative returns (in percent) for portfolios sorted by the change of northbound capital shareholding percent (HPC) around quarterly earnings announcements. We divide previous week's HPC into five quintiles three days before the announcement of quarterly earnings. Holding week varies from one week ([-2,2] around the earnings announcement) to eight weeks to be comparable with the results in Table 4. Cumulative return calculation is the same as that in Table 4. Sample period is March 2017 to December 2018. Newey-West adjusted t-statistics are given below the portfolio returns in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	H							
F	1	2	3	4	5	6	7	8
Low	-0.45*	-0.74**	-1.11***	-1.14***	-1.27***	-1.42**	-1.51**	-1.72**
	(-1.74)	(-2.10)	(-2.70)	(-2.58)	(-2.62)	(-2.50)	(-2.30)	(-2.44)
2	-0.38	-0.71**	-1.19***	-1.47***	-1.57***	-1.97***	-2.37***	-2.75***
	(-1.53)	(-2.09)	(-3.08)	(-3.41)	(-3.17)	(-3.44)	(-3.86)	(-4.42)
3	-0.30	-0.60**	-0.92**	-1.10***	-1.29***	-1.51***	-1.69***	-1.72***
	(-1.36)	(-1.98)	(-2.57)	(-2.73)	(-2.97)	(-3.35)	(-3.45)	(-3.28)
4	-0.25	-0.49	-0.69*	-0.91*	-0.99**	-1.20**	-1.28**	-1.26**
	(-1.03)	(-1.46)	(-1.74)	(-1.94)	(-2.06)	(-2.22)	(-2.26)	(-2.16)
High	0.07	0.09	0.10	0.23	0.26	0.18	0.36	0.59
	(0.28)	(0.25)	(0.24)	(0.46)	(0.51)	(0.32)	(0.59)	(0.86)
High-Low	0.52***	0.83***	1.21***	1.36***	1.53***	1.60***	1.88***	2.31***
	(3.30)	(3.93)	(4.85)	(4.39)	(4.51)	(4.06)	(4.44)	(5.10)

Table 7. HPCs forecast CARs and SUEs around QEAs

This table reports the predictability of HPC on earnings surprises and abnormal returns around earnings announcements. Following Livnat and Mendenhall (2006), standard earnings surprises are defined as: $SUE_{i,t} = \frac{X_{i,t}-X_{i,t-4}}{P_{i,t}}$, where $X_{i,t}$ is earnings per share for firm i in quarter t. Five-day buy-and-hold abnormal returns is measured based on the market model, where the market benchmark is defined as the value-weighted portfolio of all A-share stocks listed on stock exchanges. Estimation window is required to be one quarter (66 trading days) before the announcement of earnings for each stock, with minimum observations of 22 trading days. HPC is the difference of holding ratio between t-1 and t-6, where t is first day of event window. Control variables include Size, BM, ROA, SOE, VOL, IVOL, TURN, REVERSAL, RUNUP, MSCI, InstHold, and AnalyNum. Panel A reports the full sample prediction results. Panel B compares the predictability of HPC between subsamples where firms' overseas income is above and below the sample median. Industry and year-month fixed effect are included. Standard errors are clustered at industry level. Sample period is March 2017 to December 2018.

Panel A.	SUE				CAR[-2,2]			
Full sample	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
HPC	0.0003**	0.0003**	0.0003**	0.0003**	0.0288*	0.0272*	0.0281*	0.0273*
	(2.27)	(2.26)	(2.24)	(2.22)	(1.89)	(1.80)	(1.83)	(1.77)
Size	-0.0004	-0.0004	0.0002	0.0001	-0.0004	-0.0002	-0.0003	-0.0003
	(-1.20)	(-1.35)	(0.44)	(0.21)	(-0.63)	(-0.40)	(-0.44)	(-0.38)
BM	0.0010***	0.0010***	0.0011***	0.0011***	0.0011***	0.0011***	0.0011***	0.0011***
	(3.45)	(3.43)	(3.35)	(3.32)	(2.83)	(2.86)	(3.01)	(3.00)
ROA	0.0213***	0.0213***	0.0248***	0.0247***	0.0292***	0.0279**	0.0296**	0.0289**
	(3.39)	(3.40)	(3.79)	(3.77)	(2.77)	(2.61)	(2.63)	(2.55)
SOE	0.0006	0.0004	0.0006	0.0005	-0.0012	-0.0010	-0.0008	-0.0008
	(1.12)	(0.89)	(1.42)	(1.09)	(-1.28)	(-1.09)	(-0.77)	(-0.76)
VOL	-0.0641	-0.0657	-0.0503	-0.0525	0.1036	0.1124	0.1029	0.1044
	(-1.19)	(-1.22)	(-0.84)	(-0.87)	(1.38)	(1.44)	(1.15)	(1.14)
IVOL	0.0032	0.0036	-0.0028	-0.0021	-0.1560***	-0.1602***	-0.1431**	-0.1429**
	(0.14)	(0.16)	(-0.11)	(-0.08)	(-2.67)	(-2.67)	(-2.10)	(-2.10)
TURN	0.0024	0.0030	0.0014	0.0022	-0.0146***	-0.0152***	-0.0152**	-0.0152**
	(0.73)	(0.92)	(0.39)	(0.59)	(-3.05)	(-3.05)	(-2.47)	(-2.39)
REVERSAL	0.0061***	0.0059***	0.0069***	0.0067***	-0.0365***	-0.0367***	-0.0339***	-0.0342***
	(3.13)	(2.95)	(3.49)	(3.25)	(-6.50)	(-6.59)	(-5.66)	(-5.74)
RUNUP	0.0091***	0.0091***	0.0089***	0.0089***	-0.0013	-0.0015	-0.0025	-0.0025
	(11.47)	(11.30)	(10.62)	(10.58)	(-1.15)	(-1.28)	(-1.57)	(-1.59)
MSCI	-0.0000	-0.0000	-0.0001	-0.0001	-0.0006	-0.0006	-0.0007	-0.0007
	(-0.07)	(-0.06)	(-0.18)	(-0.17)	(-0.38)	(-0.38)	(-0.40)	(-0.42)
InstHold		0.0000		0.0000		-0.0000		-0.0000
		(1.34)		(1.51)		(-1.18)		(-0.26)
AnalyNum			-0.0001***	-0.0001***			-0.0000	-0.0000
			(-3.94)	(-3.95)			(-0.54)	(-0.59)
Industry FX	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FX	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
#Observations	6,954	6,948	6,464	6,459	7,163	7,146	6,604	6,596
R-squared	0.172	0.172	0.186	0.186	0.069	0.069	0.066	0.066

(Continued on the next page)

Table 7 (Continued)

Panel B. Comparison in subsamples divided by the median of firms overseas income

	SUE				CAR[-2,2]			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
HPC (Overseas	0.0004**	0.0004**	0.0004**	0.0004**	0.0639***	0.0632***	0.0582**	0.0572**
Income>Median)	(2.47)	(2.60)	(2.26)	(2.49)	(3.17)	(3.11)	(2.52)	(2.47)
HPC (Overseas	-0.0001	-0.0002	-0.0001	-0.0001	0.0102	0.0072	0.0043	0.0034
Income<Median)	(-0.60)	(-0.68)	(-0.46)	(-0.53)	(0.40)	(0.29)	(0.19)	(0.14)
Diff	0.0005*	0.0006*	0.001	0.0005*	0.0537*	0.056*	0.0539*	0.0538*
P-value	0.08	0.06	0.12	0.08	0.08	0.07	0.07	0.07

Table 8. Double sort portfolio

This table reports average weekly excess returns (in percent) for 2×5 value-weighted portfolios conditional on different firm characteristics over the period March 2017 through December 2018. At the beginning of each week, we form sequential double-sort portfolios based on lagged rank characteristics and hold for one week. The first rank variable in Panel A, Panel B, and Panel C is the level of northbound capital holding (HP), A-H cross list, and overseas income, respectively. The second rank variable in each panel is the change of northbound capital shareholding percent (HPC). The last column in each panel (High-Low) shows return to a long-short portfolio where firms with HPC in the highest (lowest) are assigned to the long (short) portfolio. The last row in each panel shows the average return of quintile portfolios in each column. Newey-West adjusted t-statistics are given below the portfolio returns in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A Conditional on HP						
	HPC					
HP	Low	2	3	4	High	High-Low
High	-0.09	-0.17	-0.03	0.16	0.38	0.56***
	(-0.31)	(-0.58)	(-0.11)	(0.59)	(1.28)	(4.02)
Low	-0.44	-0.50*	-0.32	-0.24	-0.01	0.44***
	(-1.32)	(-1.88)	(-1.29)	(-0.76)	(-0.03)	(2.93)
High-Low	0.35	0.33	0.30*	0.39*	0.46**	0.09
	(1.64)	(1.52)	(1.79)	(1.86)	(2.11)	(0.53)
Panel B. Conditional on A-H cross-listing						
	HPC					
A-H	Low	2	3	4	High	High-Low
Cross-listing	0.69**	0.33	-0.07	-0.28	-0.13	0.81***
	(2.60)	(1.28)	(-0.23)	(-1.03)	(-0.36)	(3.33)
Non-cross-listing	0.28	-0.06	-0.27	-0.29	-0.15	0.43***
	(1.14)	(-0.25)	(-1.04)	(-1.02)	(-0.60)	(2.99)
(Cross-listing)-(Non-cross-listing)	-0.35**	-0.40	-0.20	-0.01	-0.03	0.30
	(-2.03)	(-1.49)	(-0.71)	(-0.02)	(-0.12)	(1.39)
Panel C. Conditional on overseas income						
	HPC					
Overseas income	Low	2	3	4	High	High-Low
High	-0.19	-0.40	0.03	0.27	0.30	0.49**
	(-0.54)	(-1.41)	(0.15)	(0.92)	(0.91)	(2.25)
Low	0.01	-0.37	-0.35	-0.24	-0.03	-0.04
	(0.03)	(-1.04)	(-1.14)	(-0.74)	(-0.10)	(-0.24)
High-Low	-0.20	-0.03	0.38	0.52**	0.33	0.53**
	(-0.91)	(-0.10)	(1.58)	(2.38)	(1.60)	(2.20)

Table 9. Cross-sectional return predictability in Southbound capital

This table reports the weekly returns (in percent) of H-share stock portfolios sorted by their share changes held by southbound investors from China Mainland. Individual stocks on the connected markets are sorted into decile portfolios. In Panel A, both the formation period and holding period are one week. We examine the portfolio performance for the full sample and two sub-samples: firms connected with Mainland, and firms not connected with Mainland. In Panel B, we examine the long-run portfolio performance. Formation period is one week. Holding period varies from one to eight weeks. Sample period is March 2017 to December 2018. Newey-West adjusted t-statistics are given in parentheses. *, ** and *** represent statistical significance at the 10%, 5% and 1%, respectively.

Panel A. Baseline performance

	All firms		Mainland firms		Non-Mainland firms	
Stock portfolios	Excess return (%)	t-stat	Excess return (%)	t-stat	Excess return (%)	t-stat
Low	-0.19	(-0.63)	-0.22	(-0.65)	-0.07	(-0.24)
2	0.11	(0.44)	0.13	(0.44)	0.06	(0.28)
3	0.14	(0.64)	0.15	(0.57)	0.06	(0.30)
4	0.35	(1.19)	0.40	(1.21)	0.36	(1.47)
High	0.45	(1.28)	0.59	(1.67)	0.41	(0.94)
High-Low	0.64***	(3.85)	0.72***	(4.59)	0.48	(1.12)

Panel B. Long-run performance

Holding weeks	1	2	3	4	5	6	7	8
Low	-0.19	-0.16	-0.10	-0.04	-0.14	-0.11	-0.06	-0.04
2	0.11	0.14	0.14	0.17	0.17	0.10	-0.09	-0.07
3	0.14	0.28	0.53	0.63	0.75	0.93	1.22	1.30
4	0.35	0.55	0.83	1.16	1.57	1.67	1.87	2.17
High	0.45	0.65	0.89	1.01	1.20	1.15	1.38	1.69
High-Low	0.64***	0.81**	0.99***	1.05**	1.34***	1.25**	1.44***	1.73**
	(3.85)	(3.69)	(3.73)	(4.08)	(4.45)	(3.69)	(3.74)	(3.88)

Table 10. Cross-sectional returns before and after regulation change

This table compares the cross-sectional return predictability of northbound and southbound HPC before and after the regulation change. On August 24, 2018, HKEX announced to construct Northbound investor identification model for northbound trading in Stock Connect Scheme. Thus, the period before the regulation change is from March 2017 to August 24, 2018, and the period after the regulation change is from September 2018 to the present. Panel A reports the weekly portfolio returns of northbound HPC and Panel B reports the results of southbound HPC. Newey-West adjusted t-statistics are given in parentheses. *, ** and *** represent statistical significance at the 10%, 5% and 1%, respectively.

	Excess return (%)	t-stat	Excess return (%)	t-stat
	Before		After	
Panel A. Sorted by northbound HPC				
Low	-0.15	(-0.47)	0.89*	(1.72)
2	-0.21	(-0.75)	0.70	(1.34)
3	-0.37	(-1.40)	0.82	(1.54)
4	-0.51	(-1.46)	0.65	(1.53)
5	-0.26	(-0.83)	0.57	(1.48)
6	-0.14	(-0.50)	0.76*	(1.82)
7	-0.03	(-0.10)	0.94*	(1.74)
8	0.04	(0.15)	1.20**	(2.14)
9	0.24	(0.96)	1.15**	(2.40)
High	0.53*	(1.98)	1.13**	(2.29)
High-Low	0.68***	(3.14)	0.24	(0.94)
Panel B. Sorted by southbound HPC				
Low	-0.32	(-1.08)	-0.07	(-0.13)
2	0.02	(0.08)	0.09	(0.24)
3	0.03	(0.11)	-0.08	(-0.25)
4	0.15	(0.66)	1.34	(1.51)
5	0.08	(0.33)	-0.39	(-1.60)
6	0.03	(0.12)	1.11	(1.61)
7	0.16	(0.53)	0.27	(0.82)
8	0.10	(0.30)	0.05	(0.13)
9	0.46	(1.61)	0.19	(0.46)
High	0.29	(0.77)	0.60	(1.16)
High-Low	0.60***	(2.76)	0.67***	(2.97)

Table 11. Portfolio performance conditional on ownership before and after regulation change

This table reports average weekly excess returns (in percent) for 2×5 value-weighted portfolios conditional on firm attribute of ownership, namely, SOEs and non-SOEs. Panel A and Panel B compare the weekly excess returns of portfolios before and after the launch of Northbound Investor Identification Model, respectively. The first sub-period spans from March 2017 to August 2018, and the second one spans from September 2018 to June 2019. Newey-West adjusted t-statistics are given below the portfolio returns in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A. Conditional on firm attribute of ownership before regulation change						
	HPC					
	Low	2	3	4	High	High-Low
SOE	-0.16	-0.39	-0.06	-0.01	0.36	0.53***
	(-0.52)	(-1.38)	(-0.21)	(-0.05)	(1.15)	(2.92)
Non-SOE	-0.24	-0.39	-0.35	-0.07	0.32	0.56***
	(-0.76)	(-1.29)	(-1.38)	(-0.22)	(1.00)	(3.32)
(SOE)-(Non-SOE)	0.08	0.00	0.30	0.05	0.05	-0.03
	(0.52)	(-0.01)	(1.63)	(0.34)	(0.31)	(-0.18)

Panel B. Conditional on firm attribute of ownership after regulation change						
	HPC					
	Low	2	3	4	High	High-Low
SOE	0.58	0.63	0.59	1.05**	1.28**	0.70**
	(1.05)	(1.18)	(1.33)	(2.04)	(2.23)	(2.12)
Non-SOE	0.93	0.76	0.67	1.03*	1.00*	0.07
	(1.56)	(1.43)	(1.53)	(1.76)	(1.73)	(0.32)
(SOE)-(Non-SOE)	-0.35	-0.13	-0.08	0.02	0.28	0.63**
	(-1.20)	(-0.60)	(-0.38)	(0.09)	(0.96)	(2.03)

Table 12. Insider trading and northbound capital flows

The table examines whether opportunistic insiders and northbound investors trade similarly using their private information either within or outside of pre-QEA windows. Opportunistic insiders are identified as the top 50% based on the profitability of their past pre-QEA trades. We compare weekly changes in shareholdings of identified opportunistic insiders with contemporaneous HPC of northbound flows across the entire sample. Panel A reports the summary statistics of HPC of northbound investors and changes in shareholdings of opportunistic insiders, respectively. At the bottom of Panel A, we report the Pearson correlation in terms of weekly changes in shareholdings between the two groups of informed traders. Panel B reports the coefficient estimates when regressing HPC of northbound investors on contemporaneous HPC of opportunistic insiders. Firm and month fixed effect are included. Standard errors are clustered at the month level. Newey-West adjusted t-statistics are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Summary statistics		
	HPC	Changes in shareholdings of insiders
Mean	0.006	0.86
Std. Dev	0.101	5.220
Obs.	188 (157 non-SOEs vs 31 SOEs)	188
Unique Firms	154 (127 non-SOEs vs 27 SOEs)	154
Corr. with HPC	1	0.202***

Panel B: Panel regression	
	Northbound HPC
Changes in shareholdings of insiders	0.0021***
	(2.54)
Const.	0.0248
	(1.15)
Firm fixed effect	Yes
Month fixed effect	Yes
Obs.	188
Adj. R-squared	0.145

Table 13. Imitative herding and limited attention

The table presents the results of imitative herding of mainland investors tracking northbound capital. We estimate the daily changes in shareholdings of the mainland investors at the stock level, namely $IMHERD_{i,t}$ in Eq. (2). Panel A reports the summary statistics. In Panel B, we regress $IMHERD_{i,t}$ on the lagged terms of northbound HPCs controlling the contemporaneous HPC as well as stock returns on a daily and weekly basis in Column 1 and 2, respectively. Column 3 reports the effect of investors' limited attention on the imitative herding. Top10, a dummy variable, takes the value of 1 if the stock is on the daily list of Top 10 most actively traded stocks, and 0, otherwise. We extend the setting in Column 1 by including the indicator variable for stocks on the list along with their interaction with HPCs. Stock and day fixed effect are included. Standard errors are clustered at the stock level. The coefficients of lagged return (from the last day (week) to five days (weeks)) before and constant terms are omitted for brevity.

Panel A: Summary statistics

	Mean	Min	P10	Median	P90	Max	Std. Dev	N
Daily *IMHERD*	-0.075	-42.236	-0.417	-0.046	0.220	32.759	0.579	637485
Daily northbound HPC	-0.001	-32.690	-0.020	0.000	0.030	32.620	0.394	671753

Panel B: Imitative herding

	(1) Daily	(2) Weekly	(3) Attention
HPC	-1.4507***	-0.9684***	-1.4029***
	(-53.51)	(-21.20)	(-51.32)
HPC(t-1)	**0.3072*** **	**0.2047*** **	**0.2990*** **
	(22.37)	**(8.06)**	**(21.05)**
HPC(t-1)	**0.2236*** **	0.0196	**0.2167*** **
	(18.66)	(0.87)	**(17.51)**
HPC(t-3)	**0.0612*** **	-0.0217	**0.0589*** **
	(5.58)	(-1.00)	**(5.23)**
HPC(t-4)	0.0083	-0.0351	0.0106
	(0.74)	(-1.61)	(0.91)
HPC(t-5)	-0.0040	0.0018	-0.0057
	(-0.39)	(0.08)	(-0.53)
Ret	8.3645***	6.8924***	8.3693***
	(76.60)	(46.46)	(76.68)
Top10*HPC			-0.7617***
			(-8.12)
Top10(t-1)*HPC(t-1)			**0.3601*** **
			(8.65)
Top10(t-2)*HPC(t-2)			**0.1960*** **
			(5.16)
Top10(t-3)*HPC(t-1)			**0.0663** **
			(1.97)
Top10(t-4)*HPC(t-4)			-0.0354
			(-1.17)
Top10(t-5)*HPC(t-5)			0.0365
			(1.29)
Top10			-0.0248***
			(-4.22)
Lagged stock returns	Control	Control	Control
Stock Fixed Effect	Yes	Yes	Yes
Day Fixed Effect	Yes	Yes	Yes
Observations	606,014	120,653	606,014
Adjusted R2	0.433	0.365	0.433

Table 14. Path Analysis of daily return spreads

The table breaks the correlation between daily northbound HPC and daily individual stock future returns into two parts: the direct effect of HPC and the indirect effect related to imitative herding of the mainland investors. Panel A reports the coefficient estimates for two-step regressions. In Column (1), we regress daily stock returns on HPC(t-1), HPC(t-2) and IMHERD(t-1). In Column (2), we regress daily IMHERD (t-1) on HPC (t-2). HPC and IMHERD are standardized to the mean of zero and the standard deviation of one. Control variables in the two equations are the same, including Size, BM, ROA, SOE, VOL, IVOL, TURN, REVERSAL, RUNUP, MSCI and InstHold. Stock and day fixed effect are included. Standard errors are clustered by firm and trading day. Detailed definitions of control variables are presented in the Online Appendix. The coefficients of constant terms are omitted for brevity. Panel B displays the calculation for proportion of stock returns explained by each channel.

Panel A. Two-step regressions		
	Ret(t)	IMHERD(t-1)
HPC(t-1)	0.0275***	
	(6.77)	
HPC(t-2)	0.0692***	0.0437***
	(15.04)	(13.65)
IMHERD(t-1)	-0.0365***	
	(-7.92)	
Size	-0.7191***	-0.0187
	(-25.82)	(-0.94)
BM	0.0251***	0.0287***
	(3.57)	(6.27)
ROA	0.5469***	-0.3056***
	(3.07)	(-2.78)
SOE	0.0169	0.0156
	(0.31)	(0.62)
VOL	1.3479	-4.1701***
	(1.54)	(-6.90)
IVOL	0.0437	-0.1191
	(0.07)	(-0.36)
TURN	-0.0063***	-0.0138***
	(-6.09)	(-19.09)
REVERSAL	0.0129	-0.1372***
	(0.29)	(-5.00)
RUNUP	0.0343**	0.0098
	(2.28)	(0.77)
MSCI	0.0379**	-0.0319***
	(2.39)	(-4.25)
InstHold	-0.0206	0.0140
	(-0.83)	(1.36)
Stock Fixed Effect	Yes	Yes
Day Fixed Effect	Yes	Yes
Observations	506,483	506,483
Adjusted R2	0.297	0.168

Panel B. Path analysis for the effect of HPC		
(1) Direct effect of HPC on stock return		
HPC → Ret	0.0275***	(6.77)
(2) Indirect effect of HPC on stock return via IMHERD		
HPC → IMHERD	0.0692***	(15.04)
IMHERD → Ret	0.0437***	(13.65)
Herding Path Effect	0.0030	
(3) Total Effect of HPC on Ret		
Total Effect	0.0305	
% Direct Path	90.09%	
% Indirect Path	9.91%	

Table 15. Time-series prediction

This table reports the time-series prediction on a weekly basis from November 2014 to December 2018. We use the lagged net inflows of north bound capital (buy amount minus sell amount) through Shanghai/Shenzhen-Hong Kong Stock Connect Scheme to predict the excess returns of market index labelled in row, respectively. Market index include value weighted portfolio of all A shares, value weighted portfolio of A shares listed on Shanghai stock exchange (Shanghai Composite Index), value weighted portfolio of A shares listed on Shenzhen stock exchange (Shenzhen Composite Index). In-sample prediction utilizes the full sample data from November 2014 to December 2018. For out-of-sample prediction, forecasting model is estimated using a recursive window starting from November 2014. Benchmark is the historical mean model. We examine three sub-periods for out-of-sample test: January 2016 to December 2018, January 2017 to December 2018, January 2018 to December 2018. In-sample R^2, R_{OS}^2, and annualized certainty equivalent return are reported.

Sample	Stat	Value weighted A Shares	Shanghai Composite Index	Shenzhen Composite Index
2014.11.1-2018.12.31	In-sample R^2	2.16%	0.06%	3.39%
2016.1.1-2018.12.31				
	R_{OS}^2	3.48%	1.11%	3.38%
	CER	19.09%	10.58%	13.32%
2017.1.1-2018.12.31				
	R_{OS}^2	4.12%	2.01%	3.75%
	CER	24.31%	15.81%	15.03%
2018.1.1-2018.12.31				
	R_{OS}^2	6.37%	2.30%	6.86%
	CER	42.81%	30.59%	21.44%

Table 16. Time-series trading strategy

This table reports the performance of time-series trading strategy. The design of the strategy is: At the end of each week, the investor will take a long position if the predicted market return is positive. Otherwise the investor will liquidate the market portfolio and use the proceeds to invest in the risk-free asset. Holding period varies from one week to eight weeks. We employ the methodology of Jegadeesh and Titman (1993) with active portfolios to derive the single time series of the weekly returns for all holding periods larger than one week. Excess returns of value-weighted A shares portfolio is predicted by net inflows of northbound capital. Forecasting model is estimated using a recursive window starting from November 2014. Out-of-sample evaluation period extends from January 2016 to December 2018. Benchmark is the buy-and-hold strategy. Panel A reports the annualized excess return (in percent) as well as annualized Sharpe ratio for the benchmark and the market timing strategy. Panel B reports the Carhart (1997) four-factor adjusted alphas of the benchmark and timing strategy where the holding period is one week. Newey-West adjusted t-statistics are given below the portfolio returns in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A. Summary statistics of the timing strategy

	Buy-and-hold	1	2	3	4	5	6	7	8
Annualized excess return	-5.68	12.83	11.17	15.09	7.41	10.95	-0.73	-8.94	-16.39
Sharpe ratio	-0.29	0.82	0.42	0.39	0.15	0.18	-0.01	-0.11	-0.18

Panel B. Carhart (1997) four-factor regression

	Intercept	RMRF	SMB	HML	UMD
Buy-and-hold	0.000	0.960***	0.088***	-0.049***	0.013
t-stat	(0.58)	(94.16)	(4.76)	(-2.56)	(1.09)
Timing strategy	0.002**	0.615***	0.300***	0.073	0.277***
t-stat	(2.32)	(8.49)	(3.40)	(0.95)	(4.15)

Appendix

A.1 Variable Definitions

The variables used for the analysis are described in detail below. They are constructed using data from CSMAR, WIND, Choice, RESSET, HKEX (daily and monthly stock files, quarterly and annual industrial file, northbound capital flows, northbound investors' shareholdings, etc.).

1. Size: Denotes the natural logarithm of the market capitalization at the end of the week for each stock. Size = stock price × the number of shares outstanding, in thousand RMB yuan).
2. BM: Denotes the ratio of book value to market value, where the book value is the common equity plus balance-sheet deferred taxes for the latest fiscal quarter as of the end of the last quarter, and the market value is equity capitalization measured at the end of each week.
3. REVERSAL: Following Da, Liu, and Schaumburg (2014), we define the past short-term reversal as the cumulative return from week $t-4$ to week $t-1$ for each stock-week observation.
4. RUNUP: Following Jegadeesh and Titman (1993), we define the run-up effect as the cumulative return from week $t-52$ to week $t-5$ for each stock-week observation.
5. VOL: Denotes the monthly realized volatility of stock return. We compute this variable as the standard deviation of daily excess market return over the three-month deposit rate.
6. IVOL: Denotes the idiosyncratic volatility of stock return. Following Ang et al. (2006), we measure this variable as the standard deviation of residuals in the Fama-French three-factor model over the past three months.
7. TURN: Denotes the weekly turnover, which is measured using the ratio of weekly trading volume to shares outstanding in the prior one month.
8. ROA: Denotes the return on assets. This variable is measured as the ratio of net income to average total assets for the prior fiscal quarter ending at least three months ago and no longer than one year.
9. SOE: Denotes a dummy variable identifying state-owned enterprise. SOE equals 1 if the stock is classified as state-owned enterprise in CSMAR and 0, otherwise.
10. MSCI: MSCI is a dummy variable to flag firms listed in the MSCI China index. MSCI (Morgan Stanley Capital International) started to include *some* of large-cap A shares in the MSCI Emerging Markets Index on May 31st, 2018. If one stock is included in the MSCI A-share index, the dummy variable MSCI equals one after June 2018.

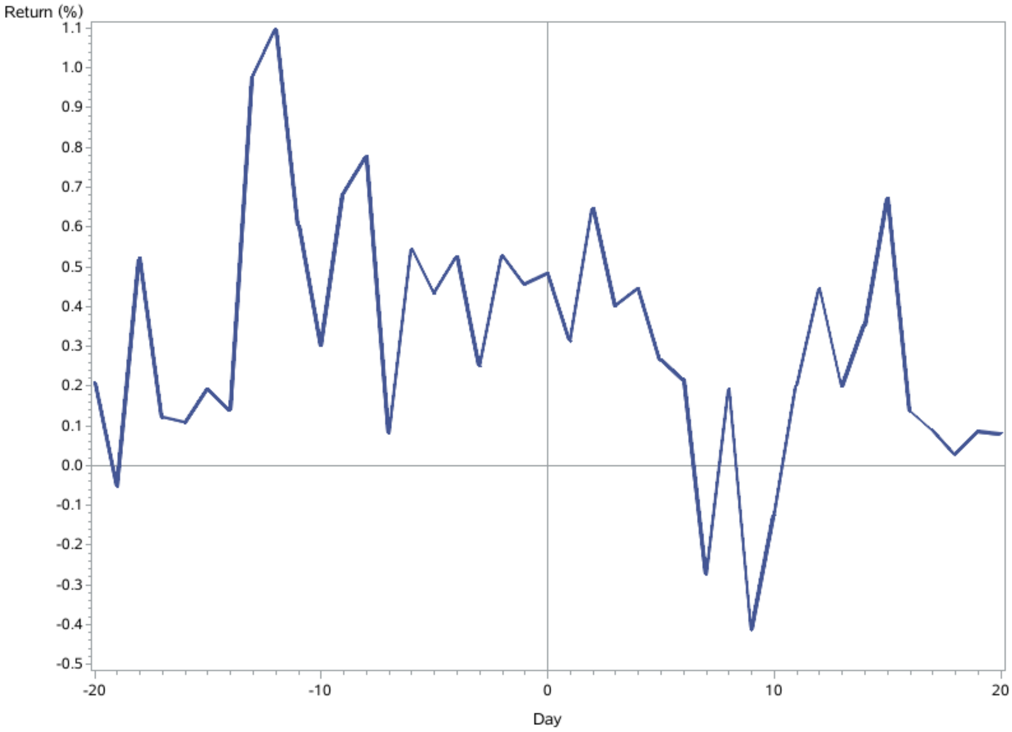
11. InstHold: Denotes the stock share of institutional holding, which is measured as the percentage of stakes in all institutions. The raw measure of institutional holding is disclosed to the public quarterly in China and thus, we use the latest data as of the end of the prior fiscal quarter.
12. AnalyNum: Denotes the number of individual analysts that covering the stock in the year. An analyst is said to be "covering" a stock if he/she has produced a stock recommendation for a given stock in the CSMAR Analyst Prediction database in the past year.

A.2 Robustness checks

We perform a bunch of robustness tests to determine whether the main findings are sensitive to our research design. Some of these tests have been mentioned or footnoted throughout the text, and the rest are discussed here in detail. First, to address the concern on estimation errors in extreme groups, we repeat univariate portfolio sorts using the back-tested procedure suggested by Mamaysky et al. (2008) in Table I.1. More specifically, for a stock to be included in a portfolio in week t+2, its return in week t+1 must be below (above) the cross-sectional median if its HPC estimated at the end of month t is below (above) the cross-sectional median. All qualitative inferences from the paper remain unchanged and the cross-sectional predictability of HPC is even more salient in the back-tested sample. Second, in Table I.2 of the Internet Appendix, we examine daily HPCs. Consistent with main findings, we find that the cross-sectional predictability that we observe is robust to a trading strategy based on daily HPCs. Another concern is that the persistence of return predictability of HPC might be captured by actively seeking out risks characterized by common factors following periods of recent success of northbound investors, we confirm that the benchmark adjusted return spread between top and bottom quintiles sorted by weekly HPC does not reverse in the long run in untabulated results. Third, as we mentioned in the text, to determine whether northbound flows are purely value investors tracking "good" assets without any motivation driven by information edge, our analysis of portfolio sorts is repeated by using HPC residuals after a bunch of firm-level characteristics, as reported in Table I.3. All results are quantitatively similar. In untabulated results, to determine whether our results are robust to the attention-induced demand pressure explanation for HPC's ability to predict future returns, we repeat stock-level analyses in Table 2 and Table 4 after excluding observations related to daily positions on the lists of actively traded stocks. We continue to find no return reversal in the long run.

Figure I.1. Daily cross-sectional returns around earnings announcements.

This figure depicts the daily performance of the cross-sectional trading strategy before and after ten days of firm's earnings announcement dates. At the beginning of each trading day, we sort eligible stocks into quintile portfolios on the basis of their previous day's HPCs and hold for one day. "0" in the x-axis denotes the day of earnings announcement, and the value below (above) 0 denotes the trading day before (after) earnings announcement. Y-axis represents the daily high-minus-low returns (in percent).



Daily cross-sectional returns around earnings announcements

Table I.1 Back-tested portfolio sorts

To address the concern on estimation errors in extreme groups, we repeat univariate portfolio sorts using the back-tested procedure suggested by Mamaysky et al. (2008). Specifically, for a stock to be included in a portfolio in week t+2, its return in week t+1 must be below (above) the cross-sectional median if its HPC estimated at the end of month t is below (above) the cross-sectional median. This table presents average weekly excess returns (in percent) for value-weighted portfolios constructed using back-tested procedure. Newey and West (1987) three-lag adjusted t-statistics are reported. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Excess return (%)	t-stat
Low	-0.30	(-0.86)
2	-0.17	(-0.56)
3	-0.46	(-1.51)
4	-0.63*	(-1.83)
5	-0.07	(-0.22)
6	-0.13	(-0.37)
7	0.11	(0.32)
8	0.11	(0.35)
9	0.02	(0.07)
High	0.32	(1.02)
High-Low	0.62***	(2.83)

Table I.2 Cross-section return predictability using residual HPC

This table presents weekly excess returns (in percent) for portfolios constructed by residual HPC over the period March 2017 through December 2018. Residual HPC is the residual of Model (1) in Table 5 i.e., the residual of regressing HPC on Size, BM, ROA, SOE, VOL, IVOL, TURN, REVERSAL, RUNUP, MSCI, InstHold, and AnalyNum. Formation and holding period for portfolio sorting is one week. *t*-statistics are adjusted for heteroscedasticity and autocorrelation using Newey and West (1987) method. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Excess return (%)	t-stat
Low	-0.15	(-0.51)
2	-0.16	(-0.58)
3	-0.13	(-0.53)
4	-0.07	(-0.26)
5	-0.12	(-0.53)
6	-0.24	(-0.96)
7	-0.16	(-0.61)
8	-0.17	(-0.69)
9	0.15	(0.61)
High	0.41	(1.57)
High-Low	0.56***	(2.61)

Table I.3 Univariate portfolio sorts (daily rebalanced)

This table presents average daily excess returns (in percent) for value-weighted portfolios calculated over the period March 2017 through December 2018. We repeat the univariate portfolio sorts in Table 2 except that portfolios are rebalanced on a daily basis. At the beginning of each day, eligible stocks in Shanghai/Shenzhen-Hong Kong stock connect scheme are sorted into decile portfolios on the basis of their previous day's HPCs and held for one day. Newey and West (1987) three-lag adjusted t-statistics are reported. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Daily excess return (%)	t-stat
Low	-0.08	(-1.24)
2	-0.08	(-1.30)
3	-0.10	(-1.25)
4	-0.10	(-1.24)
5	-0.02	(-0.41)
6	-0.03	(-0.44)
7	-0.14	(-1.52)
8	0.01	(0.23)
9	0.06	(1.12)
High	0.19***	(3.00)
High-Low	0.28***	(7.99)

Table I.4 Portfolio performance sorted by the net buying of the mainland investors

The table presents the results of portfolio sorts based on the changes in shareholdings of the mainland investors *IMHERD*. The procedures of portfolio sorts are the same as the analysis in Table 2. The results for daily rebalance (daily return), weekly rebalance (weekly return), the first half (March 2017 to January 2018, weekly return) and second half period (February 2018 to December 2018, weekly return), Shanghai and Shenzhen stock exchange (weekly return) are reported as follows. Newey-West adjusted t-statistics are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Low	2	3	4	High	High-Low
Daily	-0.04	-0.03	-0.02	-0.01	-0.07	-0.03
	(-0.50)	(-0.55)	(-0.29)	(-0.13)	(-1.18)	(-0.91)
Weekly	-0.20	-0.09	0.14	0.16	-0.12	0.08
	(-0.70)	(-0.36)	(0.64)	(0.74)	(-0.51)	(0.36)
Shanghai	-0.22	-0.05	0.16	0.17	-0.07	0.15
	(-0.79)	(-0.22)	(0.68)	(0.82)	(-0.28)	(0.67)
Shenzhen	-0.22	-0.16	0.06	0.08	-0.15	0.07
	(-0.71)	(-0.58)	(0.23)	(0.31)	(-0.57)	(0.32）
First half	-0.15	0.10	0.51	0.57	0.23	0.38
	（-0.49）	（0. 39）	(1.91)	(2.51)	(1.10)	(1.19)
Second half	-0.28	-0.36	-0.28	-0.35	-0.59	-0.32
	(-0.61)	(-0.92)	(-0.80)	(-1.10)	(-1.55)	(-1.13)

Table I.5 Return reversal in the subsample of the most actively trades stocks

This table reports the portfolio performance based on daily list of Top 10 most actively traded stocks in a two-week rolling window. We use the subsample of A shares which are on the daily list of ten most actively traded stocks to construct the quintile portfolios. The table presents daily returns for quintile portfolios and long-short portfolio within two trading weeks after portfolio formation. Sample period is from March 2017 through December 2018. Newey-West adjusted t-statistics are given below the portfolio returns in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

HPC	1	2	3	4	5	6	7	8	9	10
high	0.30***	0.12	0.11	0.10	0.12	0.09	0.11	0.12	0.09	0.03
	(3.58)	(1.36)	(1.39)	(1.16)	(1.43	(1.03)	(1.38)	(1.46)	(0.97)	(0.31)
3	0.17**	0.17**	0.12	0.08	0.16	0.06	0.10	0.04	0.10	0.09
	(2.09)	(2.07)	(1.41)	(0.96)	(2.11)	(0.70)	(1.17)	(0.51)	(1.22)	(1.16)
2	0.05	0.05	0.09	0.08	0.05	0.09	0.05	0.10	0.05	0.12
	(0.72)	(0.73)	(1.32)	(1.03)	(0.63)	(1.24)	(0.63)	(1.31)	(0.72)	(1.66)
1	0.04	0.05	0.07	-0.01	0.06	0.04	0.04	0.08	0.06	0.06
	(0.54)	(0.64)	(0.85)	(-0.16)	(0.79)	(0.60)	(0.48)	(0.99)	(0.77)	(0.84)
low	-0.07	-0.02	-0.01	0.00	0.05	0.02	0.06	0.01	0.07	0.09
	(-0.87)	(-0.19)	(-0.09)	(-0.04)	(0.55)	(0.26)	(0.73)	(0.16)	(0.78)	(1.05)
high-low	0.37***	0.14**	0.12*	0.09	0.08	0.07	0.05	0.11	0.03	-0.06
	(5.42)	(2.02)	(1.86)	(1.46)	(1.24)	(1.08)	(0.85)	(1.75)	(0.42)	(-0.86)

Table I.6 Portfolio performance conditional on cross listing before and after regulation change

This table reports average weekly excess returns (in percent) for 2×5 value-weighted portfolios in the two subgroups, AH cross-listed firms and non-cross-listed firms, before and after the launch of Northbound Investor Identification Model. The first sub-period is from March 2017 to August 2018 and the results is reported in Panel A. The second one is from September 2018 to June 2019 and the results is reported in Panel B. Newey-West adjusted t-statistics are given below the portfolio returns in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A. Conditional on A-H cross-listing before regulation change						
	HPC					
	Low	2	3	4	High	High-Low
Cross-listing	-0.13	-0.32	0.04	0.35	0.77***	0.88***
	(-0.34)	(-1.10)	(0.12)	(1.24)	(2.87)	(3.31)
Non-cross-listing	-0.18	-0.34	-0.36	-0.13	0.27	0.46***
	(-0.66)	(-1.18)	(-1.41)	(-	(1.05)	(2.97)
(Cross-listing)-(Non-cross-listing)	0.06	0.02	0.4	0.48	0.42**	0.34
	(0.23)	(0.07)	(1.41)	(1.61)	(2.28)	(1.41)

Panel B. Conditional on A-H cross-listing after regulation change						
	HPC					
	Low	2	3	4	High	High-Low
Cross-listing	0.70	0.65	0.19	1.56*	1.32**	0.63*
	(1.15)	(1.46)	(0.59)	(1.79)	(2.51)	(1.71)
Non-cross-listing	0.91	0.81	0.94*	0.97*	1.11**	0.20
	(1.69)	(1.52)	(1.93)	(1.85)	(2.38)	(0.82)
(Cross-listing)-(Non-cross-listing)	-0.21	-0.16	-0.75*	0.59	0.21	0.42
	(-0.63)	(-0.42)	(-1.82)	(1.05)	(0.79)	(1.11)

Table I.7 Path analysis of weekly return spreads

The table breaks the correlation between HPC and weekly individual stock future returns into two parts: the direct effect of northbound HPC and the indirect effect related to imitative herding of the mainland investors. Panel A reports the coefficient estimates for two-step regressions. In Column (1), we regress weekly stock returns on HPC(t-1), HPC(t-2) and IMHERD(t-1). In Column (2), we regress weekly IMHERD (t-1) on HPC (t-2). HPC and IMHERD are standardized to the mean of zero and the standard deviation of one. Control variables in the two equations are the same, including Size, BM, ROA, SOE, VOL, IVOL, TURN, REVERSAL, RUNUP, MSCI and InstHold. Detailed definitions of control variables are presented in the Online Appendix. The coefficients of constant terms are omitted for brevity. Panel B displays the calculation for proportion of stock returns explained by each channel.

Panel A. Two-step regressions		
	Ret(t)	IMHERD(t-1)
HPC(t-1)	0.0010***	
	(5.30)	
HPC(t-2)	0.0002	0.0278***
	(1.18)	(6.99)
IMHERD(t-1)	0.0004**	
	(2.01)	
Size	-0.0376***	-0.1177***
	(-28.08)	(-3.19)
BM	0.0026***	0.0854***
	(5.50)	(6.72)
ROA	0.0390***	-0.5483**
	(4.40)	(-2.52)
SOE	0.0069***	0.0469
	(2.64)	(0.61)
VOL	-0.0629	-11.2426***
	(-1.44)	(-9.72)
IVOL	0.1446***	1.8341***
	(4.91)	(2.88)
TURN	-0.0299***	-3.2425***
	(-5.94)	(-21.60)
REVERSAL	0.0010	-0.5080***
	(0.43)	(-8.95)
RUNUP	0.0024***	0.0750***
	(3.26)	(3.35)
MSCI	0.0022**	-0.0236
	(2.51)	(-1.37)
InstHold	-0.0011	0.0028
	(-0.89)	(0.13)
Stock Fixed Effect	Yes	Yes
Week Fixed Effect	Yes	Yes
Observations	107,083	107,102
Adjusted R2	0.328	0.255

Panel B. Path analysis for the effect of HPC		
(1) Direct effect of HPC on stock return		
HPC → Ret	0.0010***	(5.30)
(2) Indirect effect of HPC on stock return via IMHERD		
HPC → IMHERD	0.0278***	(6.99)
IMHERD → Ret	0.0004**	(2.01)
Herding Path Effect	0.00001	
(3) Total Effect of HPC on Ret		
Total Effect	0.00101	
% Direct Path	98.90%	
% Indirect Path	1.11%	