The China-U.S. Equity Valuation Gap

Geert Bekaert Shuojia Ke Xiaoyan Zhang

May 2021

Abstract

If the capital markets in China and the U.S. are integrated, then the valuation ratios, such as price earnings ratios (PEs), would converge. Before 2009, the market average PE of Chinese firms is significantly higher than that of the U.S. firms, while after 2009, the valuation gap reverses. Using data from 1995 to 2018, we examine the dynamics and sources of valuation differentials between comparable Chinese and U.S. firms. The sectoral composition of the indices, growth expectations, financial openness, financial development, and the investor base, all contribute to the cross-sector and time-series variation of the valuation differentials, but financial openness and changing growth expectations are the most important contributors.

JEL Classification: F36, G15, G18

Keywords: Chinese stock prices, market integration, financial development, stock valuation, earnings yields, price earnings ratios, speculative trading.

* Bekaert is at Columbia Business School and CEPR. Ke and Zhang are at PBC School of Finance of Tsinghua University. We are grateful for the comments of Jennifer Huang, Clark Liu, Christian Lundblad, and participants of the China Finance Research Conference. We thank Pengfei Li and Zhaojing Chen for early research assistance on this project.

The China-U.S. Equity Valuation Gap

Abstract

If the capital markets in China and the U.S. are integrated, then the valuation ratios, such as price earnings ratios (PEs), would converge. Before 2009, the market average PE of Chinese firms is significantly higher than that of the U.S. firms, while after 2009, the valuation gap reverses. Using data from 1995 to 2018, we examine the dynamics and sources of valuation differentials between comparable Chinese and U.S. firms. The sectoral composition of the indices, growth expectations, financial openness, financial development, and the investor base, all contribute to the cross-sector and time-series variation of the valuation differentials, but financial openness and changing growth expectations are the most important contributors.

JEL Classification: F36, G15, G18

Keywords: Chinese stock prices, market integration, financial development, stock valuation, earnings yields, price earnings ratios, speculative trading.

I. Introduction

China is a unique emerging market, featuring the 2nd largest economy and the 2nd largest equity market in the world. What makes its situation even more unique is that, despite being very open to trade since joining the WTO in 2001, the Chinese government has only cautiously and gradually embarked on a path towards global capital market integration, liberalizing inward and outward capital flows, in a controlled manner. The Chinese B-share market intended for foreign investors was tiny, and the government only allowed limited foreign investment in the A-share market through a Qualified Foreign Institutional Investor Program in 2002. In contrast, most emerging markets followed more aggressive paths and liberalized their capital markets in the late 80s or early 90s (Bekaert, Harvey and Lundblad, 2003). This integration process made valuations in emerging markets converge towards developed levels, but, on average they continue to trade at discounts.

Did 20 years of gradual liberalization integrate the Chinese equity market into the global capital market? Figure 1 presents some intuitive valuation evidence showing the price-earnings ratios (PE henceforth) for emerging markets[1], the U.S. and China. As expected, the emerging market valuation discount is clearly visible before 2009-2010, with the PE ratios of emerging markets, shown in dotted line, on average 8.14 lower than the PE ratio of the U.S., shown in dashed line. Intriguingly, before 2009, the China A-share stocks have a high average price-earnings ratio (shown in solid line), 11.49 points higher than that of the U.S., and 19.62 points higher than that of the emerging market index. The valuation ratio patterns change after 2009. Chinese valuations

[1] The data for emerging markets is obtained from Datastream, using the data series of "TOTMKEK", which has 2302 constituents. Of the 2302 constituents in the Emerging Market index, it includes 50 China H shares, but no China A shares.

quickly and significantly decline and become in line with those of emerging markets, but lower than those of U.S. firms. Over the most recent 10 years, the valuation differentials between China and U.S. clearly declines, while significant gaps still remain between the two countries.

Given that the stock market capitalizes growth opportunities and sets the cost of capital for public firms, understanding valuations in the world's second largest economy is of paramount importance. To better understand China's unusual path towards integration, we analyze the time-series and cross-section valuation differentials between China and the U.S. To be specific, we follow Bekaert, Harvey, Lundblad and Siegel (2011, BHLS henceforth) and focus on earnings yield differentials (the reciprocals of the PE ratios), between Chinese and U.S. stock portfolios. We use the U.S. market as a benchmark because the U.S. market is the largest and clearly a highly open and integrated market in the world. BHLS show that, under the assumptions of economic and financial integration, earnings yield differentials across countries for the same sector should be small[2].

We propose three hypotheses to explain the (changes in the) valuation gap between China and the U.S. First, a changing sector decomposition might explain the stylized facts if the relative importance of high multiple (low multiple) industries has decreased (increased) over time. Second, the capitalization of strong growth opportunities in China, not present in a mature economy such as the U.S. may explain a "China Valuation Premium" before 2009. After all, in the decade before 2009, China enjoyed double digit annual GDP growth rates, while the U.S., as well as other developed markets, typically feature annual GDP growth rates of around 2%. Perhaps, later in the

[2] The BHLS's cross-country study includes China over the 1993 to 2005 period, but uses a large set of countries. It does not address whether Chinese firms have comparable valuations to firms priced in global capital markets.

sample, China's growth prospects diminished and market valuations began to be more reflective of the usual emerging market risks. Finally, we hypothesize that a gradual liberalization and financial development process, combined with classes of investors (in particular, retail investors) with unrealistic market valuations (and few alternative investment outlets), may result in the observed valuation gap variation. That is, during the earlier part of the sample, the inability to short sell combined with a significant presence of unsophisticated individual investors in stock trading can potentially lead to speculative excess and generate a Chinese valuation premium (see Mei, Scheinkman, and Xiong, 2009). Meanwhile, the Chinese government's liberalization efforts have narrowed the valuation differential with the U.S., over time.

To test the first hypothesis, we decompose the market level earnings yield differential between China and U.S. into a pure valuation differential part, which measures valuation differences in the same sectors, and an industry structure part, which is driven by sector composition differences between China and the U.S. The valuation gap between China and the U.S. is dominated by valuation differentials rather than industry structure, thereby rejecting the first hypothesis. Yet, we also show that the relative importance and valuation of the Chinese banking sector did contribute in a non-negligible way to the magnitude of the valuation gap and its variation over time.

For the second and third hypothesis, we consider several valuation fundamentals, including measures of growth expectations, financial openness, and financial development, of the importance of different investor bases, and also firm specific measures of liquidity and efficiency. Using a multivariate regression framework, we find that all above channels jointly explain 33% of the total variation in the valuation differentials between China and the U.S. over the past 24 years between 1995 and 2018. The growth expectation channel accounts for around 5% of the total explanatory power, the financial openness channel explains more than 70%, the financial development channel explains almost 10%, and the ownership structure channel explains about 12.5%. If we focus on the shorter 2003-2018 sample, then the growth expectations channel is more important, accounting for 30% of the explained variation with the contribution of the financial openness decreasing commensurately to 43%, but remaining dominant.

The BHLS framework is formulated under the assumption of market integration. We examine the robustness of our main results with alternative modifications to the market integration hypothesis. In particular, we consider the effect of time-varying betas in a (partial) segmentation setting, and the effect of time-varying global discount rates. Our main results stay the same, and the financial openness continues to dominate explained variation.

Our paper is related to the general literature on how market segmentation affects equity valuation. The "China valuation premium" between 1995 and 2009, relative to the U.S., violates the predictions of standard market segmentation/integration theory, as in Errunza and Losq (1985), and Bekaert and Harvey (2003). This unique premium pattern attracts substantial academic interests. For example, Bailey, Chung and Kang (1999) documents that among 11 markets that both issue domestic (like A shares) and foreign securities (like B shares), the Chinese stock market is the only market where domestic securities price richer than foreign securities. Fernald and Rogers (2002) also document the discount received by foreigners relative to the domestic A-share market. Mei, Scheinkman and Xiong (2009) attribute the A-B premium to speculative trading in A-shares, and Chan, Menkveld and Yang (2008) attribute it to information asymmetry. Unlike these previous research on Chinese valuation patterns using same companies in the A- versus Bshare markets, we focus on the general time-series and cross-section valuation gaps, to understand what economic-wide forces are behind the valuation gaps over the past 24 years and across 54 stock portfolios. For these economic-wide forces, our paper connects to the literature examining the factors that affect valuation differentials across countries, focusing mostly on emerging markets. In panel data, the previous literature finds factors such as political risk (Bekaert, 1995; Erb, Harvey and Viskanta, 1996), liquidity (Lesmond, 2005; Bekaert, Harvey and Lundblad, 2007), financial openness restrictions and poor stock market development (Bekaert, Harvey, Lundblad and Siegel, 2011), to effectively segment markets. Finally, there is a rapidly growing literature on the Chinese equity market, studying the cross-section of expected returns (Liu, Stambaugh and Yuan, 2019), institutional features of the A-share market and Chinese equity prices (Allen, Qian, Shan and Zhu, 2019), the return effects of specific episodes of liberalization, such as the Hong Kong Connect program (Chan and Kwok, 2018; Liu, Wang and Wei, 2020) and the efficiency and informativeness of the Chinese stock market (Carpenter, Lu and Whitelaw, 2021).

Compared to the extant literature, we provide two contributions. First, our study is one of the first to detail the times-series and cross-sector dynamics of the valuation gap between China and the U.S., the two largest equity markets in the world, attributing a key role to the financial openness channel. Second, we show how high growth expectations and the presence of speculative retail investors combined to reverse the conventional valuation effects of international accessibility, which decreased rather than increased valuations in China. Our findings are useful for practitioners and regulators alike.

The paper is organized as follows. We describe the data and document the Chinese valuation gap in Section 2. Our empirical framework is introduced in Section 3. In Section 4, we present the empirical results on what drives the time and cross-sectional variation in valuation differences. Section 5 provides robustness checks, and Section 6 concludes.

II. Data

We introduce the data in Section II.A, and construct earnings yields at the portfolio level in Section II.B. Stylized facts about the valuation gap are provided in Section II.C.

II.A Data Sources

Our main sample period is from 1995 to 2018. The Chinese stock exchanges were opened in 1990, but with limited number of listed firms. Here we start our sample in 1995 to have sufficient number of firms in the cross section. Given that the valuation data is available at quarterly frequency, we construct all variables in this study at quarterly frequency.

We obtain Chinese firm level data from DataStream, WIND, CSMAR, and Suntime. Datastream provides Level 4 sector classifications for each firm. From WIND, we collect basic firm level accounting, trading, and institutional ownership data. From CSMAR, we obtain share structure data, such as the number of shares, information on state ownership, and analyst data. Suntime provides additional data on analysts and their expectations on growth. The analysts-based growth expectation and institutional / retail ownership variables are mostly available after 2003, which restricts our sample length when using these variables. Following Liu, Stambaugh, Yuan (2019), we apply the following data filters to the Chinese data: 1) Exclude stocks that have become public within the past 2 quarters; 2) Drop stocks that have less than 45 daily return observations during the most recent quarter; and 3) Drop stocks that have less than 120 daily return observations during the most recent year. We adopt the first filter (the "IPO filter") for Chinese firms because Chinese IPO pricing (and hence valuation) is controlled by the Chinese Security Regulatory

Commission (CSRC), which doesn't necessarily reflect market consensus. After IPO, the valuations gradually incorporate market forces and market expectations. Therefore, we omit the first 2 years right after IPO to only include market related valuation ratios.

Firm level data for U.S. are obtained from DataStream, CRSP, Compustat, and I/B/E/S. Again, we obtain Level 4 sector classifications for each firm from Datastream. From CRSP, and Compustat, we obtain basic firm level information regarding stock trading and accounting variables. I/B/E/S provides analyst data. We adopt filters 2) and 3) described above to the U.S. data. Because the IPO pricing mechanism is already market based in the U.S., we skip the IPO filter for U.S. firms.

II.B Valuation Variables

In this study, our key valuation ratio is the earnings yield (EY), which is the reciprocal of the price earnings ratio. As documented in BHLS, the earnings yield can be directly linked to discount rates and expected cash flows and thus to economic and financial integration. We conduct our empirical investigation at the portfolio level, so here we define the earnings yield ratio for portfolio *j* in quarter *t* as

$$EY_{j,t} = \frac{\sum_{i=1}^{N_j} Total\ annualized\ net\ income_{i,j,t}}{\sum_{i=1}^{N_j} Price_{i,j,t} \times Number\ of\ common\ equity_{i,j,t}}, \tag{1}$$

where N_j is the number of stocks in portfolio j, $Price_{i,j,t}$ is the unadjusted price of stock *i* in portfolio *j* at the end of quarter *t*, $Number\ of\ common\ equity_{i,j,t}$ is the latest reported number of common equity shares in a firm's quarterly or annual report, and $Total\ annualized\ net\ income_{i,j,t}$ is the sum of quarterly net income from quarter *t*-4 to quarter *t-1*. Following the previous literature, if *Total* *annualized net income*$_{i,j,t}$ is negative, we set it to zero.[3] The earnings yield differential between China and the U.S. for portfolio j in quarter t becomes:

$$DIFEY_{j,t} = EY_{j,t}^{CN} - EY_{j,t}^{US} \,. \tag{2}$$

BHLS show that under economic and financial integration, $DIFEY_{j,t}$ defined for portfolios in different countries but same sectors, should be close to zero.

To understand the market level differences in earnings yields, we define the aggregate EY at the market level as,

$$EY_t = \frac{\sum_{i=1}^{N} Total\ annualized\ net\ income_{i,t}}{\sum_{i=1}^{N} Price_{i,t} \times Number\ of\ common\ equity_{i,t}}, \tag{3}$$

where N is the number of stocks in the market. After we calculate the earnings yield measures, we define the portfolio level price earnings ratio (PE) and aggregate market level PE as the reciprocals of portfolio level EY in Equation (1) and market level EY in Equation (3), respectively. More details on the variable constructions are provided in Appendix A.

To capture cross-sectional valuation differences at firm level in a parsimonious way, we group firms into portfolios. Following the literature, we first consider industry sectors by grouping firms into 38 sectors using Level 4 sector classifications from DataStream. In this process, we drop 5 sectors, because China has no listed firms in “Nonlife Insurance”, “Tobacco”, “Real Estate Investment Trust”, “Equity Investment Instrument”, “Non-Equity Investment Instrument” sectors. In addition to the 33 single industry sectors, given that valuation ratios can be significantly affected by firm level characteristics, we also construct 21 characteristics portfolios. To be specific, we

[3] The results are robust when we first aggregate firm level annualized net income into a portfolio level annualized net income and then set negative portfolio level annualized net income to be zero.

organize firms by state ownership, institutional ownership, retail ownership, international accessibility, liquidity and size. Over recent years, the technology related stocks sometimes receive astronomically high valuations, therefore, we construct a "tech portfolio", which includes firms in the TMT sectors ("Fixed and Mobile Telecom," "Media", "Software and Computer Services" and "Technology, Hardware and Equipment"). We also construct a non-tech portfolio which include stocks not in TMT sectors. Our last set of portfolios differentiates stocks according to where they list. There are three listing boards in China with large differences in listing requirements: the Main board, SME board and the ChiNext board. Firms listed on the main board are usually large companies, while the SME board mainly includes small firms and ChiNext is a board which aims to attract innovative firms with lower listing requirements. More details on the construction of these portfolios are provided in Appendix B.

After we construct the China portfolios, we compute the benchmark valuation ratios for matching U.S. firms. We first compute the matched U.S. portfolios sorted on sector classification, institutional and retail ownership, liquidity, size and listing boards. When the sorting variable is not readily available to the U.S. firms, we form the U.S. counterpart to reflect the sectoral composition of the Chinese portfolios. For instance, state ownership is not observable for U.S. firms, so we construct the U.S. counterpart by matching the sector composition of the Chinese state ownership portfolio. That is, with $VW_{j,t}^{CN}$ representing the sector level weight for sector j in the Chinese portfolio, these sector level weights are used to form the corresponding U.S. benchmark for all relevant variables. Therefore, the matching U.S. portfolio's earnings yield is $EY_t^{US} = \sum_j VW_{j,t}^{CN} EY_{j,t}^{US}$.

II.C Stylized Facts on the Valuation Gap

To have a rough idea on the magnitude of the valuation gap between China and U.S., we first turn to Table 1. The top row provides the time-series average of the PE and EY for the markets in China and the U.S. in the top row of Table 1. From 1995 to 2018, the average aggregate PE ratio is 25.9 for China, with a corresponding average earnings yield of 4.94%, while the U.S. aggregate PE ratio is 20.3, with the earnings yield at 5.13%. The average valuation gap over 24 years might not seem substantial, but the time-series plot in Figure 1 tells a different story. Figure 1 plots aggregate price earnings ratios over time, for China, U.S., and an emerging market index. Before 2009, China's PE ratio is mostly above that of the U.S., peaking at a difference of 31.9 points compared to the U.S. in 2001Q1. After 2009, China's PE ratio significantly decreases, and it stays at 11.29 at the end of the sample (2018Q4), which is 6.68 points below that of the U.S. The time-series dynamics of the market aggregate valuations are quite dramatic, with large differences and sudden turns.

In addition to the time-series dynamics, we report the average cross-sectional differences at portfolio level in Table 1. For each sector, we present number of stocks, sector market capitalization, sector market cap as percentage of total market cap, sector PEs and EYs. From the number of stocks and their market capitalization, we can see that China and the U.S. have largely different sector structures. In China, the sector with the largest market cap is "Banks and Life Insurance", representing 14% of the total market capitalization, with as a distinct second the "Real Estate Investment and Services" sector, representing about 8% of the market on average. In the U.S., the sector weights are more balanced. The largest two sectors are "Technology Hardware & Equipment", and "Software and Computing Services," with both representing around 9%. Out of 33 industry sectors, 30 Chinese sectors have higher PE ratios than their U.S. counterparts. The remaining 3 sectors, "Forestry & Paper", "Mining" and "Real Estate Investment & Services", have lower PE ratios than their U.S. counterparts by 21.0, 15.8 and 3.4 respectively. Given the relatively high PE ratios for most of the sectors in China, one would expect earnings yields in China to be generally lower than in the U.S., but Table 1 shows that there are some Chinese industries with higher average PE ratios, yet also higher earnings yields than their U.S. counterparts (e.g. "Banks & Life Insurance" and "Fixed and Mobile Telecom"). One possible reason for this pattern is the usual convexity effect: the average earnings yield is approximately one over the average PE ratio plus a positive function of the variance of the PE ratio. The variability of PE ratios (and earnings yield differentials) is much larger in China than in the U.S. This convexity effect makes the valuation premium smaller in earnings yield terms than in PE ratio terms.

Next we turn to the 21 portfolios formed on firm characteristics. In China, high state ownership (SO) is associated with lower valuations, with a non-negligible difference of 10.5 between the lowest and highest SO groups, possibly because state-owned firms are less profitable. The retail investor dominated stocks are almost twice as valuable; and firms with low institutional ownership are twice as valuable. This could indicate that public companies with higher retail ownership may be irrationally priced higher. Firms with international accessibility have PEs lower than firms not accessible to global investors, which is in sharp contrast with the conventional wisdom that foreign investors increase stock valuations. The PE ratios of liquid firms exceed those of firms with low liquidity by 4.7, and high turnover stocks in China receive 17.9 high PEs than low turnover stocks, both indicating liquidity brings high valuation ratios. Small firms being almost three times as valuable as large firms, perhaps because retail investors prefer small stocks, this may again be due to irrational beliefs of retail investors. The tech stocks average valuation is 41.5, much larger than the average valuation of 25.5 for non-tech stocks. Finally, for stocks on different listing boards, small high-tech firms on ChiNext have average PEs of 52.9, small stocks trading on the SME board have average PEs of 35.8, while firms on Main Board have average Pes of 25.2. These findings are consistent with previous literature.

The stylized facts in Figure 1 and Table 1 show substantial cross-portfolio and time-series variations in valuation ratios, which are the focus of our empirical investigation.

III. Methodology

To explain the valuation gaps between China and the U.S. over time and in the cross-section, Section III.A introduces our empirical specification, building on the valuation framework from BHLS. Given the distinct time-series pattern in the valuation gap, we examine potential structural breaks in Section III.B and make corresponding adjustments in our empirical setting.

III.A Valuation Framework

The classical valuation model, the Gordon model, argues that with constant expected cash flow growth rates and discount rates, and full payout of earnings, the earnings yield reflects the difference between the discount rate and the expected cash flow growth rate. Based on this intuition, we adopt the valuation framework from BHLS (2007, 2011), which contains stochastic growth opportunity and discount rates that links to market integration. That is, under the assumption of financial and economic integration, the portfolio PEs/EYs are identical across countries, and the time-variations and cross-sectional differences are driven by variation in the discount rates and growth opportunities.

We first introduce the dynamics of the portfolio level discount rate and the growth opportunity. The discount rate for industry j in country c, $\delta_{c,j,t}$, under the null of integration, is affected by the world risk-free rate r_f, the world discount rate $\delta_{w,t}$, and the industry's exposure to world discount rate, $\beta_{c,j}$:

$$\delta_{c,j,t} = r_f(1 - \beta_{c,j}) + \beta_{c,j}\delta_{w,t}. \tag{4}$$

This logic of this specification is similar to the widely-adopted CAPM. Given that sector innovations are more likely to be global rather than country specific, we make an assumption that the industry's exposure to the world discount rate risk is not country-specific, we would then have $\beta_{c,j} = \beta_j$. We also define the portfolio level log earnings growth for portfolio j in country c:

$$\Delta \ln(\mathrm{Earn}_{c,j,t}) = GO_{w,j,t} + \epsilon_{c,j,t}\ . \tag{5}$$

Here $GO_{w,j,t}$ is the worldwide stochastic growth opportunity for portfolio j, which is independent of which country the industry belongs to, and $\epsilon_{c,j,t}$ is the country-industry specific disturbance. With simplifying assumptions for world growth opportunity and world discount rate dynamics, BHLS show that the PE ratio for portfolio j in country c, $PE_{c,j,t}$, can be written as:

$$PE_{c,j,t} = \sum_{k=1}^{\infty} exp(a_{c,j,k} + b_{j.k}\delta_{w,t} + g_{j,k}GO_{w,j,t})\,. \tag{6}$$

Equation (6) shows that PE ratio is a function of global discount rate factor and global growth opportunity factor. Given that $b_{j.k}$ and $g_{j,k}$ are portfolio-specific but not country-specific, the PE (and EY) for the same portfolios should be equal under the assumption of economic integration and financial integration, after controlling for the term $a_{c,j,k}$.

From Equation (6), the earnings yield ratio, the reciprocal of PE ratio, is affected by the same set of variables as those affecting the PE ratio. Following BHLS (2011), we take the earnings yield ratio, not the PE ratio, as our main valuation measure for three considerations: PE is highly positively skewed, while EY has better distributional properties; PE is not defined when earnings is zero, while EY is not affected; EY is easier to interpret as a percentage term and can be easily interpreted as the difference between discount rates and growth rates.

With linearization, Equation (6) leads to the following specification for the dynamics of earnings yield:

$$EY_{c,j,t} = \alpha 1 + b1' X_{c,j,t} + c1' Control_{c,j,t} + e1_{c,j,t} \, . \tag{7}$$

The variables $X_{c,j,t}$ contains explanatory variables for earnings yield, which presumably includes measures of growth prospect and discount rate. $Control_{c,j,t}$ represent our control variables to count for the other portfolio level differences, as in term $a_{c,j,k}$ in Equation (6), which will be further explained below. The process of market integration should cause valuation ratios between same portfolios from different countries to converge. Of course, the China-U.S. situation is not quite far from fully integrated mood, confirmed by Figure 1 demonstrating significant variation in the aggregate valuation differentials between China and the U.S. Nevertheless, as a first step we adopt the BHLS framework to understand what drives the time-series and cross-section variation in valuation differentials, using the U.S. as a benchmark. From Equation (7), the valuation gap between China and U.S., measured by earnings yield in Equation (2), can be specified as,

$$DIFEY_{j,t} = a2 + b2' DIFX_{j,t} + c2' Control_{j,t} + e2_{j,t}. \tag{8}$$

where $DIFX_{j,t} = X_{j,t}^{CH} - X_{j,t}^{US}$, captures differences in explanatory variables, such as the growth prospects, discount rates etc.[4] Since our research question focusing on understanding valuation

[4] There are two exceptions. First, for variables that only available for the Chinese market, such as regulation variables, we set their U.S. counterpart values as constants, which will then affect the intercept, but not slope coefficients. Second,

gap between China and U.S., Equation (8) and its modifications are the key specification we estimate and present in this article.

Our parameter estimates are obtained using panel regressions on sector and characteristics portfolios from China and U.S., and the standard errors are computed by double clustering on portfolio and quarter, as in Thompson (2011) and Petersen (2009). Notice that we intentionally exclude the time fixed effects, because we aim to examine whether and how much the various hypotheses can explain the time-series and cross-sector variation in valuation differentials.

Our choices for the X variables mainly reflect two alternative, non-exclusive, hypotheses for the dynamics of the valuation gap, including variations in growth prospects, and variations in discount rates, measured by financial openness, financial development and ownership structure. It is intuitive that higher but slowing growth prospects in China can explain the Chinese valuation premium and its demise. However, the combination of a gradual liberalization and financial development process can only explain the observed valuation changes when combined with certain classes of investors (in particular, retail investors) having unrealistic market valuations (and few alternative investment outlets), which are "corrected" by more rational or less constrained foreign investors.

Following BHLS, we always include three control variables: leverage differential, earnings growth volatility differential, and minimum number of firms (see our Appendix A for exact definitions). We include the leverage differential as our first control variable because higher financial risk should be reflected in higher discount rates even under the null of integration. The

---

for the political risk variables, following previous literature, we calculate differences as the ratios of Chinese over U.S. variables.

second control variable is the earnings growth volatility differential. BHLS use earnings growth volatility to capture variation of $a_{c,j,k}$ in Equation (6). A portfolio with higher (idiosyncratic) earnings volatility may, all else equal, be more valuable than a portfolio with less volatility (see also Pástor and Veronesi, 2006). Finally, we control for the number of firms, which potentially affects the accuracy of the portfolio level measure. We include the minimum number of firms between the two portfolios in the computation as our third control variable. Summary statistics for the control variables are reported in Online Appendix Table OA1.

For the derivation of Equation (8), we assume that $\beta_{c,j}$ is time-invariant and equal (to one) both for China and U.S., so we don't empirically estimate $\beta_{c,j}$ and include them in our regression specifications. This assumption makes our empirical models simple and straightforward, but many previous studies show that time-variation and cross-sectional differences in $\beta_{c,j}$ can be important. For our robustness check in Section V, we make several extensions to incorporate portfolio specific betas and allow they to be time-varying and cross-sectional different. In addition, we also directly estimate Equation (7) for determinants of China's valuation ratios and present these results in Section V.

III.B Structural Break

For the panel regression to work properly, the time-series of the valuation gap needs to be stationary. Given that the valuation gap between China and U.S. switches sign around 2008 and 2009, we examine whether there is a significant structural break in the time-series and modify our specification to accommodate a potential structural break.

We adopt the structural break test from Bai, Lumsdaine and Stock (1998). They assume that the variables before and after a single break follow a stationary process, which is well described by a vector autoregressive (VAR) process. Bai, Lumsdaine and Stock (1998) then find the endogenous break date using a Sup-Wald test. In our case, to find the break in the valuation differential, we estimate the following specification:

$$DIFEY_t = \delta + \theta Break_t + \sum_{j=1}^{n} \rho_j DIFEY_{t-j} + \varepsilon_t, \tag{9}$$

where $DIFEY$ is the earnings yield differential between China and the U.S., $Break$ is a dummy variable equal to one (zero) after (before) the break date detected by the methodology, and ε is the error term. The optimal length n for the AR process is selected by the BIC criterion. A more rigorous and complete discussion is provided in Appendix C.[5] We present the structural break test results in Table 2. In the first row, we examine the structural break in the market level valuation differential. The sup-Wald statistic of 10.03 indicates that the break in the market earnings yield differentials is significant at the 5% significance level. The estimated break date is the third quarter of 2009, with a 90% confidence interval of 2007:02 to 2011:04.

Since we find that there is strong evidence of a structural break around 2009, we therefore modify our empirical specification in equation (8) by adding a break dummy:

$$DIFEY_{j,t} = a3 + \gamma Break_t + b3' DIFX_{j,t} + c3' Control_{j,t} + e3_{j,t}. \tag{10}$$

The break dummy, $Break_t$, is set to be one after 2009:03, and zero otherwise. By adding this break dummy, the time-series of the dependent variables become stationary. Also, the break dummy can also help to separate different hypothesis in later sections, in the sense that if the competing hypothesis can fully explain the time variation in the earnings yield differential, they probably

[5] Following Bekaert, Harvey and Lumsdaine (2002), we also try another specification which in addition allows the lag terms to break. The break dates are robust.

should also account for the break, and render the break dummy coefficient insignificant. In Section IV, we mostly report estimation results based on Equation (10), a modification of Equation (8).

IV. Empirical Results

In this section, we examine three hypotheses for explaining what drives the time-series and cross-sectional variation in valuation differences between China and U.S. We start with industry structure hypothesis in Section IV.A. In Section IV.B., we investigate the role of growth prospects. Section IV.C. focuses on several discount rate factors, including financial openness, financial development and the investor base. In Section IV.D., we combine all these channels together and use a PcGets model selection method to pick up the most relevant variables.

IV.A Hypothesis I: Changes in Industry Structure

It is conceivable that the variation and sign switch of the China-U.S. valuation differential is driven by an increase of the market shares of low PE firms in the Chinese stock market. In this section, we only consider the 33 industry sectors, not the 21 characteristics portfolios. To understand whether particular sectors play an outsized role in valuation differential changes, Figure 2, Panel A shows earnings yield differentials between Chinese and U.S. industries both in the first 5 years (horizontal axis) and in the last 5 years (vertical axis) of the sample. If the valuations of Chinese industries decrease compare with their U.S. counterparts, they should be above the diagonal line, and otherwise, they should be below the diagonal. 26 industries show up on the left of the vertical line at zero, indicating their earnings yields are lower than that of their U.S. counterparts between 1995 and 1999. In the last 5 years of the sample, 25 industries are below the horizontal zero line, featuring lower earnings yields than their U.S. counterparts. The distance from the diagonal line reveals how substantial the valuation change over the sample period is. Now the “banks & life insurance” portfolio separates from the other portfolios, and is a clear outlier. Its earnings yield differential moves from a negative -2.5% in 1995-1999 to a large positive 7.2% in 2014-2018. That is, the banking portfolio shows a valuation premium at the early sample and a deficit at the end of our sample, consistent with the pattern in the overall market.

For banking industry to be the main driver of the valuation gap, this industry must comprise a large portion of the market. Figure 2 Panel B presents the market share of the banking sector. Before 2007, the banks and life insurance sector constituted a small fraction (lower than 10%) of the market capitalization. But, its market share increase substantially after several important IPOs of state-owned banks in the mid-2000s, increasing its total market share to around 30%. This relative increase in the importance of the banking sector is even more dramatic when market shares are computed in terms of earnings (which drive the sector weights in PE ratios). Given the high average earnings yield of the banking sector, it is unlikely to play a role in the unusual negative earnings yield differences observed in the early part of the sample, but it can definitely have contributed to the rise of the earnings yield in the later part of the sample. To examine this conjecture, we present the earnings yield differentials with and without the banking portfolio in Figure 2 Panel C. Up until 2007, the two lines greatly overlap, reflecting the banking portfolio constituting a negligible part of the market. But after 2007, especially after 2009, the two lines diverge with the increase in the earnings yield more pronounced for the overall statistic than for the one without the banking portfolio.

Is it the banking sector behind the structural break around 2009? We present the structural break test results in the second row of Table 2. When we exclude the banking portfolio from the market yields, the sup-Wald test is 3.64, which is not significant at the 10% level. Clearly, this indicates the banking portfolio valuation is an important contributor to the structural break.

To formally examine the role of industry structure for the valuation gap, we consider the following decomposition of the earnings yield differential between China and the U.S.:

$$DIFEY_t = EY_t^{CN} - EY_t^{US} = \sum_{j=1}^{33} w_{j,t}^{CN}\left(EY_{j,t}^{CN} - EY_{j,t}^{US}\right) + \sum_{j=1}^{33}\left(w_{j,t}^{CN} - w_{j,t}^{US}\right)EY_{j,t}^{US}$$
$$= DIF\_VAL_t + DIF\_STRUC_t. \quad (11)$$

Here $w_{j,t}^{CN}$ and $w_{j,t}^{US}$ are the weights of industry j in terms of market capitalization in China and the U.S. respectively, and 33 is the total number of sectors. The first component, DIF_VAL, represents the earnings yield differential within the same sector between China and the U.S., and thus it constitutes a pure valuation differential. The second component, DIF_STRUC, captures sectoral weight differences between China and the U.S. and represents the valuation effect of a different industry structure. This decomposition exercise is conducted each quarter.

We present the decomposition results in Table 1, Panel B. The earnings yield differential at the market level has a time-series average of -0.19%, with the first component has a time-series average of -0.69% and the second 0.50%. That is, the portfolio composition component, DIF_STRUC, partially mitigates the negative pure valuation differential of DIF_VAL. We then compute how much each component contributes to the total variance of the overall differential. The valuation component, DIF_VAL, accounts for 99% of the variation of total earnings yield differentials, while the structure component, DIF_STRUC, contributes only 1%. This result suggests that variation in the valuation gap between China and U.S. is dominated by valuation changes within sector rather than changes in sector structure.

Moreover, relying on the decomposition, we further investigate whether the previously recognized break is driven by the valuation component or the portfolio composition component, as defined in equation (11). In the last three rows of Table 2, we present break tests for the valuation component, the valuation component without the banking sector, and the compositional component. When focusing on the valuation part, the break is significant at the 10% level, regardless whether we compute the differential including or excluding the banking portfolio, with an identical break date of 2009Q2. When focusing on the compositional component, we fail to detect a structural break.

Given the above results, we conclude that time-variation in the China-U.S. valuation gap is mostly driven by valuation changes in the same sectors rather than by changes in industry structure. Yet, the banking industry played an important role in the sign switch in earnings yield differentials post 2009. Note that the banking sector features the second highest international accessibility out of all 33 sector portfolios, and its international accessibility increases dramatically due to the dual listings of big state-owned banks in the A-share market and Hongkong Stock Exchange. Therefore, the valuation change in the banking industry may be largely explained by the financial openness channel, which we examine below.

IV.B Hypothesis II: Changes in Growth Prospects

As shown in equation (6), a reduction in growth opportunities decreases equity valuations. China's growth differential relative to the U.S. has certainly decreased over time, so that growth prospect differentials may explain the valuation dynamics we document. We consider one market level and two portfolio level growth prospect measures: GDP growth rate[6], expected sales growth,

[6] As expected, on average, China's GDP growth is 8.8%, much higher than average growth in the U.S., which is 2.5% over the sample period.

and expected earnings (net income) growth. We collect GDP data and firm level analyst data from the China National Bureau of Statistics, Bureau of Economic Analysis, CSMAR, Suntime, and I/B/E/S. For the two portfolio level measures, we first aggregate analysts' firm level median sales and earnings forecast estimates into portfolio-level measures, and then calculate the portfolio-level expected sales growth and earnings growth over a three-year horizon. Notice that the analyst data is only available after 2003, the statistics and estimation results involving these measures all start from 2003. Figure 3 Panel A shows time-series of GDP growth rate and market level earnings/sales growth expectation. All these three measures show that the China's growth prospect are slowing down, especially in the short sample of 2003-2018.

Table 3 reports the estimation results for Equation (10). The first two regressions include only GDP growth, and pertain to the sample of 1995 to 2018. A 1% increase in GDP growth differential, decreases the average earnings yield differential by 15 basis points, and the effect is significant at the 5% level, and the adjusted R^2 is at 6.8%. In column (2), we add the break dummy, which doubles the R^2, and break dummy coefficient is statistically significant, indicating that GDP growth differential itself cannot fully explain the valuation gap between China and U.S. In columns (3) to (5), we introduce the analyst expectations for sales and earnings growth, with sample starting in 2003. In column (3), both earnings and sales growth expectations show negative coefficients, but only the sales growth coefficient is significant (at the 1% level). Intuitively, the negative coefficient means that sectors/portfolios that have higher expected future growth usually have higher valuations. The R^2 in the regression is 10%. Column (4) again shows that the break dummy remains significant. In column (5), we add GDP growth differential to this regression, and its coefficient remains negative and highly statistically significant. The break dummy now reduces in value and is no longer statistically significant, suggesting that, at least for the 2003-2018 sample, growth prospect differentials can help explain the time-series valuation break observed in China. The adjusted R^2 increases to 17.4%.

For the analyst variables, the explanatory power of the forecasts may depend on their quality. Forecast quality should increase with the number of analysts and decrease with forecast dispersion, and thus the dependence of earnings yields on earnings growth expectations should increase in absolute magnitude with the number of analysts and decrease with forecast dispersion. Forecast dispersion is calculated as the standard deviation of analyst earnings forecasts, standardized by the absolute value of the average forecast. Column (6) shows the direct effect of earnings growth expectations to be insignificant, but the interaction effect with the number of analysts is indeed significantly negative and the interaction effect with forecast dispersion is significantly positive. These interaction effects increase the R^2 to 28%. Note that the direct effects of the number of analysts and forecast dispersion are also significant. The positive coefficient on analyst variable is consistent with the fact that large firms with better analyst coverage are less dominated by retail investors, which tends to lead to lower prices (see below). Higher forecast dispersion may effectively function as an indicator of optimism and cause current over-pricing (See Diether, Malloy and Scherbina, 2002) explaining its negative sign.

To summarize, we find that growth prospects are clearly part of the explanation for the time variation in the China-U.S. earnings yield gap, especially when we focus on the 2003 to 2018 sample.

IV.C Hypothesis III: Changes in Discount Rates Factors

Other than growth prospects, the discount rates can also significantly affect dynamics of valuation ratios, as in equation (6). We distinguish three sets of factors that cause deviations from global discount rates in China. First, China's financial market development (e.g. short selling regulations) and variation in liquidity across time and portfolios should affect discount rates. Second, there is substantial cross-firm and time-series variation in financial openness, which is intended to attract foreign investors. Third, with local investors likely the marginal investors for most stocks, the unusual and time-varying mix of state, institutional and retail investors is an important factor in local valuations. We now investigate the valuation effects of financial development/liquidity, financial openness, and the investor base.

Stock Market Development, Liquidity

It is widely recognized that a stock market's development and efficiency should affect its valuation multiples through improved allocative efficiency (see e.g. Wurgler, 2000), relaxing financial constraints (see e.g. Love, 2003) or improving market liquidity (see BHLS, 2011). Moreover, illiquidity itself is also an important priced factor in the cross-section (see Amihud, 2002; Amihud, Mendelson and Pedersen, 2006).

To measure the market development from regulators' perspective, we create a discrete variable, REGDEV, which captures the stage by stage market modernization process. It starts from zero and is set to 1 after the Split-share reform in 2005Q1. It then increases by 0.5 with the following 3 events, the announcement of the Margin Trading and Short-selling Program, its official start, and the start of a registration-based IPO system. We choose to set the split-share reform to have value 1, and other three regulation changes to 0.5, because the former is widely considered more impactful, see Liao, Liu and Wang (2014).

As direct development indicators, we use the log of the number of public firms and a modified market capitalization to GDP indicator, as in BHLS. While market capitalization to GDP is often used as a development measure, its numerator is affected by stock market fluctuations which obviously also affect our dependent variable. We therefore create a relative development indicator that controls for recent stock market returns. Specifically, we first calculate the ratio of Chinese market capitalization to GDP over U.S. market capitation to GDP. We also calculate the one year past cumulative market return in China and divide by one-year cumulative market return in the U.S. Our "Adjusted market development" measure is the difference of the full sample Z-score of these two variables. We also include two indirect development indicators. The first is Morck, Yeung and Yu (2000)'s synchronicity measure. This is the value weighted average R^2 for a local market model applied to individual stocks in both countries. The higher the R^2, the less stock specific information is embedded in stock prices and the less efficient the market is. Analogously, the second indirect measure is based on the average idiosyncratic volatility in the two markets. A higher level of idiosyncratic volatility may indicate a more efficient stock market (see also Bartram, Brown and Stulz, 2012). Finally, we measure sector concentration using the four firm concentration ratio (see Appendix A).[7] While high concentration may reflect a poorly developed market, where only very large firms list publicly, it may also be an indication of market power, and more highly concentrated industries may fetch higher valuations (and thus lower earnings yields).

As direct (il)liquidity indicators, we use two variables, zeros (the proportion of daily zero returns per quarter, see Lesmond, Ogden and Trzcinka (1999)), and turnover (value traded over market capitalization). The latter indicator has also been used in the development literature as a stock market development indictor (see e.g. Levine and Zervos, 1998). We plot the market level

[7] For non-sector portfolios, this measure is computed as the weighted sum of corresponding sector level variables, using the market share of the sector in the portfolio as weights.

turnover rate in Figure 3 Panel B, and we observe that the Chinese stock market turnover rate is quite high, especially during stock market booms. The average quarterly market turnover rate is 1.00. During market booms, the quarterly market level turnover rates can be easily higher than 2.00.

While the various firm-specific variables may reflect general stock market and liquidity developments, they also reflect different liquidity conditions across stocks and portfolios, which may be priced (see Amihud, 2002). Therefore, our explanatory power may be primarily due to cross-sectional, rather than time-series variation.[8]

Table 4 reports the estimation results of equation (10), using development and liquidity variables. In the univariate regressions in specification (1) to (8), four variables are statistically significant: REGDEV, zeros, the turnover rate and the number of listed companies. The development measures (REGDEV and number of public firms) have positive coefficients, consistent with more development leading to lower overall prices. While this may, at first glance, be counterintuitive, a key reform in this regard was the 2010 permission for short-selling, which presumably suppress price bubbles and lower prices. Illiquidity (zeros) differentials has the expected positive coefficient, which could arise from cross-sectional liquidity pricing. Turnover differential has a negative coefficient, which may suggest that it is more likely to measure excess speculative trading in the Chinese stock market in the sense that higher speculative trading can

[8] We report summary statistics of these development and liquidity variables in Online Appendix Table OA4, and they are overall consistent with China being an emerging and the U.S. a developed market. In terms of Zeros, the Chinese stock market is significantly less liquid than the U.S., but this is partly due to some Chinese A-share listed firms frequently suspending their trading because of a merger or acquisitions, seasoned equity offerings etc. The Chinese stock market also has a lower number of public firms and a smaller stock market size in relation to the size of its economy. Portfolios in China are also typically much more concentrated than their U.S. counterparts.

lead to irrationally higher equity valuation (see Mei, Scheinkman, and Xiong, 2009).

In the multivariate specification in column (9) and (10), the coefficient on the number of public firms differential becomes negative, and the alternative market development and R^2 differential measures become significant. The adjusted market development measure is associated with lower earnings yield differentials, which is in line with expectations if it measures improved liquidity and efficiency. More synchronicity differential is associated with higher earnings yield differentials as one would expect, though the flipped coefficients may suggest our have multicollinearity problem to some extent.

The multivariate regression in column (9) features an R^2 of 24.4%. When we add the break dummy, it essentially retains its value and remains highly statistically significant, suggesting that changing liquidity and development conditions did not play a meaningful role in explaining time-series variation in the China-U.S. valuation gap, in terms of the structural break. All in all, our results suggest that these variables, with R^2 of 24.4%, mainly explain cross-sectional differences in valuation.

Financial Openness

Our first financial openness variable is REGOPEN, a discrete variable measuring China's regulation process towards more financial openness. We list the major events in China regulation changes in Appendix D. The REGOPEN variable is set to zero at the beginning of the sample. After B-shares become investable for Chinese investors in 2000Q1, REGOPEN variable is set to one. For the next 6 events, which include allowing "Qualified Foreign Institutional Investors" to invest in A-shares, allowing "Qualified Domestic Institutional Investors" to invest in foreign markets, allowing "Renminbi Qualified Institutional Investors" to invest in A-shares, setting up the Shanghai-Hongkong Connect, setting up the Shenzhen-Hongkong Connect and incorporating A-shares into the MSCI Index, we add one to the REGOPEN variable after each event. For those events that are announced but implemented in different quarters, we separately incorporate announcement and actual implementation effects when we define the REGPEN variable. The REGOPEN variable increases by 0.5 when the QFII (QDII) is announced and another 0.5 when the QFII (QDII) is implemented. We give slightly higher weight (0.67) to the announcement effects of Shanghai-Hong Kong Connect, Shenzhen-Hong Kong Connect and MSCI incorporation because it is likely that these events have more impactful announcement effects (see Liu, Wang and Wei, 2020), and we assign weights of 0.33 to the implementation date of above three events. While the REGOPEN variable captures all major market-opening regulatory events, it varies over time but is the same for all firms, and does not differentiate across Chinese firms with different degrees of accessibility to foreign investors. Therefore, we expect REGOPEN to capture time-series variations in valuation gap.

To measure international accessibility in the cross section, we construct three international accessibility (IA) variables using firm level information. The first variable, IA1, is a discrete variable, adding 4 dummy variables, indicating the presence of B shares, H shares, an ADR and membership of the Mainland - Hong Kong Connects. The second variable, IA2, is the ratio of the market capitalization of B shares, H shares and ADRs to the firm's total market capitalization. To construct portfolio level IA1/IA2, we value weight firm level IA1s/IA2s within the portfolio, using the firm's last quarter market capitalization as weight. The third variable, IA3, measures the market share of firms with positive firm level IA1 within the portfolio, which is particularly relevant if there are strong sectoral spillover effects in terms of international pricing. These 3 variables, not surprisingly, are highly correlated, showing an average correlation coefficient of 0.78. We plot the time-series of market level international accessibility measure in Figure 3 Panel C. The three international accessibility measures in general increases over time, except for the early sample where many domestic firms went public.

Apart from these direct financial openness variables, we consider several indirect financial openness variables. Following Frankel (1992) and BHLS, we employ the "real interest differential", calculated as the difference between the real interest rates in China and the U.S.[9] The real interest rate is one component of the discount rate, thus a higher real interest rate should be associated with higher earnings yields. BHLS also suggest that high levels of political risk can effectively segment an emerging market from international investment. To measure political stability, we obtain data from ICRG, including the "overall rating", and two measures reflecting relevant sub-categories, namely "Quality of Institutions" and "Investment Profile". We compute these political ratings as the ratio of the Chinese values over the U.S. values, so that higher numbers represent less political risk (higher stability). Less risk should be associated with higher valuations and lower earnings yields. Finally, we collect direct information on differential pricing between local and foreign investors, by using the A-B and A-H premiums. Following the methodology in Mei, Scheinkman, and Xiong (2009), we calculate the A-B premium (A-H premium) as the price of the A share divided by the price of B share (H share) minus 1, which is then market cap weighted across all international shares within each portfolio. Mei, Scheinkman and Xiong (2009) interpret these premiums as measures of excess speculation in Chinese A shares.

In Table 5, we report the estimation results for equation (10) using all the direct and indirect financial openness measures. Columns (1) through (10) report univariate results. The Regulatory

---

[9] We initially also used trade openness as an indicator, potentially associated with increased international spillovers (see e.g. Baele, 2005), but found that the variable is dominated by low frequency movements in US trade-openness.

Openness measure receives a positive and statistically significant coefficient, consistent with financial openness being associated with lower prices and higher earnings yields. The portfolio specific international accessibility variables all receive statistically significant positive coefficients. In other words, internationally accessible firms feature higher earnings yield differentials than do domestic firms. In other emerging markets, this would constitute a surprising result as domestic firms trade at lower multiples and being priced internationally results in lower discount rates and higher valuations. The effect in China is the opposite, and international accessibility reduces the valuation premium! Higher real interest rate differentials are associated with higher earnings yield differentials, but the effect is not statistically significant. The political risk measures all have statistically significant negative coefficients, suggesting that higher political risk (a lower rating) is associated with lower prices and higher yields. The coefficients on the A-B and A-H premiums are both significantly negative, showing that the difference in domestic and foreign investor pricing is an important factor which affects the China-U.S. valuation gap.

When we add all variables together in a multivariate regression in column (11), four variables remain statistically significant, with the right signs: REGOPEN, IA2, the overall political rating, and the A-H premium. The adjusted R^2 of the regression now reaches 27.9%. When we add the break dummy in column (12), it is still positive but no longer statistically significant. Note that all regressions here use the 1995-2018 sample. Thus, variation in international accessibility across time may have helped remove the Chinese-U.S. valuation gap, by driving down, not up as in other emerging markets, Chinese equity market valuations.

Domestic Investor Base

One important difference between the Chinese and U.S. stock markets is the important role of the government as a shareholder. Moreover, as an emerging market, institutional ownership show substantial time-series and cross-portfolio variation. Finally, with few outside investment options, the speculative behavior of retail investors may potentially lead to unrealistically high market valuations.

We use the available data to construct proximate measures of state, institutional and retail ownership. State ownership is measured as the fraction of total shares that are owned by the state, while institutional and retail ownership are measured as the fraction of tradable shares that are owned by institutions, and retail investors. We measure state ownership as faction of total shares because previous literature find that the state ownership can affect firms through affecting the corporate government. In this sense, it is more proper to measure the state ownership in terms ratio of total shares. In China, very few institutional investors and retail investors play roles in corporation decisions (see Jiang and Kim, 2015). They are more likely to affect firm valuations through their trading, which makes ownership measured by tradable shares a better choice. For Chinese firms, we estimate state ownership from information on the ten largest shareholders from CSMAR and more precise information regarding state-owned shares among the non-tradable shares. Then we use institutional holdings data from WIND regarding the firms' top ten holders of tradable shares and more precise information for some categories of institutional investors (including mutual funds). Finally, retail investor ownership is defined as (1 - institutional ownership - tradable state ownership - insider ownership). Because institutional and retail ownership are measured in terms of percentage of tradable shares, the subtracted state ownership here is the fraction of *tradable* shares that are owned by the state (i.e. tradable state-owned shares over total tradable shares), which is different from the previous state ownership measure that we used as a explanatory variable. Insiders are defined as directors, supervisors or managers in a company, or large individual shareholders who show up in the firms' ten largest shareholders profile, with data from CSMAR. Notice that only mutual funds and wealth products of security companies have an obligation to report their holdings in China, but for other types of institutions, WIND can only collect the holdings of the ten largest tradable shareholders disclosed in a firm's quarterly financial statements. Therefore, our institutional ownership measure for China is a lower bound to the true estimate, making our retail ownership an upper bound to true retail ownership.

For the U.S., the state ownership data is not available, and government typically doesn't hold large quantity of firm stocks. For institutional ownership, we follow Ferreira and Matos (2008), and use Factset Lion shares data to calculate institutional holdings. Insider information for U.S. firms are from Thomson Reuters. The U.S. retail investor ownership is defined as (1 - institutional ownership - insider ownership).

For China, we also compute the "Standardized Number of Shareholders (SNS)" to proxy for retail ownership which is the number of shareholders divided by the number of total tradable shares and multiplied by 1000. The data are obtained from Wind. The rationale here is that retail ownership should be positively correlated with the number of shareholders, especially in the A share market where individual investors own 99.78% of stock trading accounts according to the 2019 Shanghai Stock Exchange Statistical Year book.

Other than state ownership, which is available from 1995, most of the data are only available from 2003. We re-use the turnover variable over the long 1995-2018 sample, but interpret it as an indirect indicator of retail ownership. According to the 2018 Shanghai Stock Exchange Statistical Year book, retail trading accounted for 82.01% of total trading in 2017, which indicates retail investors trade aggressively and have high turnover rates.

We present the market average investor share over time in Figure 3 Panel D. Average state ownership in China is 45.6%, and this share has not changed much over time. Institutional ownership in the U.S. is on average 80.3%, while in China it is on average around 15.7%.[10] Over time, Chinese institutional ownership increased sharply up to about 2008, but decrease afterwards. The decrease can be caused by higher state ownership in tradable shares after the Split-share reform.[11] As mentioned earlier, due to data coverage issues, our institutional ownership measure for China is a lower bound to the true estimate, making our retail ownership an upper bound to true retail ownership. To partially verify the impact of data issue, we compare our institutional and retail ownership data with those from Shanghai stock exchange. Indeed, our estimates under-estimate institutional ownership on average by 4.3% and over-estimate retail ownership on average by 8.3%. We also find that the correlation between the exchange measures and our proxy measures is quite high, being 97.7% for the institutional ownership series, and 93.9% for the retail ownership series. As long as the bias does not show strong cross-sectional or temporal variation, our panel regressions should still provide useful information.

Table 6 presents the panel regression results for investor bases variables to explain valuation gap. The first 5 columns report results for the individual ownership variables, while the last 4 columns report results from multivariate regressions. All variables are individually statistically significant, indicating the strong explanatory power of the investor base variables. The coefficient on state ownership is significantly positive, indicating high state ownership leads to relatively lower valuations. This could be due to private-owned companies being run more

---

[10] Summary statistics on these variables can be found in the Online Appendix Table OA6. Not surprisingly, compared with the U.S., China has significant higher retail ownership and lower institutional ownership. However, the Chinese stock market turnover rate is more than twice as higher as that of the U.S.

efficiently than state-owned companies (see Boardman and Vining, 1989; Megginson et. al, 1994), or because state-owned companies have weaker corporate governance (Shleifer and Vishny, 1997; Bai, Liu, Lu, Song and Zhang, 2004) or face heavier policy burdens (Lin and Tan, 1999). Sun and Tong (2003) and Liao, Liu and Wang (2014) show that the SOEs' performance improves after reforms that reduce state ownership. Institutional ownership in contrast is associated with higher relative valuations and thus lower earnings yield differentials. The same is true for all three proxies to retail ownership, suggesting retail investors drive up valuations relative to international valuations.

In column (6), we include the state ownership and the turnover rate differential, available over the longer sample, finding both to retain their explanatory power of the univariate regressions. When we add the break dummy in column (7), it remains significant, however, suggesting these variables cannot account for the change in the valuation gap. In column (8), we use all the ownership variables jointly for the short sample. The coefficients are consistent with the univariate results and all are significantly different from zero. Jointly, these variables explain close to 29% of the variation in China-U.S. earnings yield differentials. When we add the break dummy, it is no longer statistically significant and much smaller in magnitude, suggesting that time variation in ownership may help explain the valuation gap change.

IV.D Model Selection Using PcGets

Our analysis of various determinants of the Chinese-U.S. earnings yield differentials suggests that growth expectations, financial openness and other ownership variation not only explain a non-trivial fraction of their overall variation, but may also have contributed to the change in valuation differentials from negative to positive over time. Stock market development and liquidity variables explain variation in earning yield differentials as well, but do not help explain the change in the valuation gap. We now set out to run a horse race between the various potential explanations. In order to obtain a parsimonious set of factors, we employ the general-to-specific search algorithm of Hendry (1995) and Hendry and Krolzig (2001), implemented in PcGets. This algorithm eliminates insignificant variables through an intricate "testing-down" process and generates a final set of variables with significant coefficients. A detailed discussion is provided in Appendix E.[12]

In Table 7, we report the results of applying the PcGets procedure to the 1995-2018 and the 2003-2018 samples. There are a total of 22 (31) variables for the long (short) sample, respectively. For each specification, we report the selected variables, the final coefficients and t-statistics in the first column, and a variance decomposition in the second column for both samples. The variance decomposition reports the covariance between the product of each coefficient and the corresponding independent variable with the fitted value in the regression, divided by the variance of the fitted value. The numbers therefore add up to 100%.

For the 1995-2018 sample, 9 out of 23 variables are selected, and the adjusted R^2 is 33%. Given that it was 39% for the regression with all variables, the selection procedure preserves much of the variables' explanatory power. Interestingly, all 4 groups of variables are represented. Least important are growth expectations, only accounting for about 5% of the explained variation. Of course, one reason is that for this long sample, we only have the GDP growth available, and the explanatory power of GDP growth actually increased over time. The "Financial Development" category accounts for about 9% of the explained variation, with the bulk of it accounted for by the

[12] We also examine robustness using an alternative, simpler model selection procedure which selects variables using a two-step procedure based only on univariate t-stats. The results are largely robust (see Online Appendix Table OA7).

"Zeros" illiquidity variable. This, of course, could reflect mostly a cross-sectional effect, with less liquid portfolios being priced lower. The R^2 and adjusted market development measures account for a negative or negligible amount of the total variation. The selected "Financial Openness" variables are unmistakably the main driver of the explained variation. Least important is the "A-H premium" which accounts for 5.3% of the variation. "REGOPEN" accounts for 13.2% of the explained variation. Most important by far is the IA2 variable, measuring the market capitalization represented by B, H and ADR shares. Finally, for domestic investor base variables, state ownership does not survive the model selection procedure but the turnover rate, a proxy for retail ownership, does, and accounts for 12.4% of the explained variation.[13]

We conclude that the most important driver of the cross-sectional and time-series variation in the China-U.S. earnings yield differentials is the cross-sectional and time-series variation in foreign ownership. Note that the break dummy is not selected, but the intercept term is still -1.3%, and significant, indicating that there is a stable component in the China valuation premium that is not explained by our variables.

For the more recent 2003-2018 sample, the PcGets procedure also selects 9 variables out of a total of 31 variables, but there are some differences in the variables selected. There is now a more even contribution of the various groups of variables to the total explained variation, but the main category remains "Financial Openness", with the IA2 variable now accounting for 42.5% of the total explained variation. REGOPEN and the "A-H" premium do not survive the model selection procedure. The second most important group is now "growth expectations", accounting

[13] Although the break dummy is not selected in our PcGets procedure, nevertheless, in unreported results, we additionally add the break dummy to our selected variables and re-run the regressions. In the long sample, the break dummy is insignificant at 5% confidence level but still significant at 10% level. In the short sample, it is not significant even in the 10% confidence level.

for more than 30% of the explained variation. However, this conclusion must be qualified as one of the variables selected is analyst forecast dispersion, which may be correlated with optimism biases in Chinese valuations. Still, GDP and sales growth expectations together account for about 22% of the explained variation. The turnover rate is still the only ownership variable selected, and accounts for about 16% of the explained earnings yield variation. For the liquidity variables, zeros remain the most important variable, with the R^2 variable contributing a small positive fraction and idiosyncratic volatility contributing negatively to the explained variation. The total explained variation by financial development is about 8.5%. The intercept term is no longer statistically significant. The overall adjusted R^2 is now 42%. Note that the explanatory power decreases only slightly (to 40.2%) if we replace the turnover rate by retail ownership, suggesting these variables measure the same economic phenomenon.[14]

Our results show that growth expectations, financial openness, financial development, and the investor base all contribute to the cross-sector and time-series variation of the valuation differentials, but financial openness and growth expectations are the most important contributors. Our selected variables account for a significant part of the observed earnings yield differential variations (33.0% for the long sample and 41.8% for the short sample). Figure 5 further plots the data earnings yield differential at market level and the earning yield differential predicted from our PcGets model. In both the long sample and short sample, the time series of the predicted values closely match the data time series. If we regress the fitted market level valuation gap on the predicted values, our model predicted values are close to the actual values, generating adjusted R-

[14] The turnover rate and retail ownership are significantly correlated with the correlation being 24.92%. Note that this represents the panel correlation of the China-U.S. differenced turnover and retail ownership variables. The correlation between the Chinese turnover rate and Chinese retail ownership is 34.67% while it is -24.41% for the U.S. counterparts.

squares of 0.64 for the long sample and 0.71 for the short sample.

V. Robustness Checks and Extensions

As we discussed in section III, The BHLS framework makes a number of implicit assumptions, such as portfolio specific betas being equal between China and the U.S., and betas being time-invariant. Most importantly, the whole system is estimated under the null of market integration. This full integration hypothesis may undermine the ability to explain the Chinese earnings yield, as China, given its extensive capital controls, is likely much closer to being segmented from global capital markets than being fully integrated.

To relax the assumption, rather than using the U.S. market as a benchmark, we directly estimate our equation (7). In this case, we regard the Chinese market as a segmented standard-alone market. To estimate equation (7), we still keep the assumption that betas are time-invariant and equal to one, and will relax this assumption later.

Table 8 contains the results for the PcGets exercise for fitting the China valuation ratios using the four groups of variables. We observe the following three interesting patterns. First, the adjusted R^2 is 55.6% for the 1995-2018 sample, and 60.8% for the shorter 2003-2018 sample. These R^2's are considerably higher than the regressions for the differenced earnings yield, where it is 33.0% for the longer, 41.8% for the shorter sample. This indicates that fitting one country's valuation ratios can be easier than fitting two. Second, for the long sample, the selected variables are largely the same as before with some small differences and one more substantive difference. REGDEV now survives and accounts for 25.5% of the explained variation. This increases the relative contribution of the financial development variables to almost 40%, which comes mostly at the cost of the relative contribution of the financial openness variables which decreases from over 70% to a little less than 50%. Third, for the shorter sample, starting in 2003, the SNS variable (the number of shareholders) now also gets selected in terms of the investor base variables. The growth expectations variables are somewhat different (earnings growth rate expectations and number of analysts being selected instead of sales growth expectations). The break dummy survives model selection; however, its break coefficient is now negative suggesting the selected variables "over-controlled" for the overall empirical increase in the Chinese earnings yield. It also delivers a negative contribution to the fitted variance. The contribution of the growth expectations variables remains around 30%. The contribution of financial openness decreases from 42.5% to 33.3% with the slack picked up by financial development and ownership structure (investor base). Financial openness and growth expectations remain the dominant variables explaining the temporal and cross-portfolio variation in Chinese earnings yields.

Now to remove the assumption of unit beta and verify whether accommodating "free" beta variation improves model fit, we consider several extensions of the basic regression, relying on the beta estimates of the CAPM regression models. To allow free beta to affect valuation ratios, we multiple our discount rate variables with the estimated free beta (except for the real interest rate which is multiplied by one minus beta) before estimating equation (7).

To estimate the free betas, we adopt two alternative models. In the first "full segmentation model," we assume a local CAPM model, so that only the beta with respect to the Chinese stock market matters. In the second "partial segmentation model", the discount rate now also depends on the U.S. discount rate, requiring the addition of a term capturing the U.S. discount rate times the portfolio specific beta. To implement this, we need a proxy for the U.S. discount rate. We consider three proxies. First, if we assume the U.S. discount rate is constant over time, we can simply add the U.S. beta to the regression. Second, we extract an estimate for the U.S. discount rate from the U.S. earnings yield. In particular, we add the analysts expected earnings growth to the earnings yield, both at the portfolio level, and use their sum as a measure of U.S. discount rate. This is inspired by the approach in Ferreira and Santa Clara (2011). Finally, we use the discount rate provided by Martin (2017), who shows that a measure closely related to the VIX is a lower bound to the equity risk premium. We use his measure with a horizon of 12-months plus the U.S. risk-free rate as a measure of the U.S. discount rate. Since Martin's measure of the equity risk premium lower bound is only available from January 1996 to January 2016, our sample period here is different from our main results. In the latter two cases, we simply add a series multiplying the portfolio specific U.S. betas with the U.S. discount rate to the regression.

For all above variations, we consider both an unconditional version and a version with time-varying betas. For the unconditional version, we estimate the beta from a full sample regression, running Chinese portfolio level excess returns on Chinese market level excess returns. For the conditional version, we obtain betas using the following procedure: in each quarter, we run the excess portfolio level excess return on the Chinese market excess return (for totally segmented models, also U.S. market excess return for partially segmented models) using data of the past 52 weeks.

For growth expectations measures, we directly use the Chinese counterparts to the growth expectation variables in Table 3 as our measures. To further accommodate different portfolios to have different sensitivities to GDP growth, we also measure how sensitive earnings yields and earnings growth for a particular portfolio are to changes in GDP. Specifically, we regress portfolio earnings yields (earnings growth rates) on GDP growth over the full sample, and then define the coefficient as the EY beta (earnings beta). We include both the betas and their interaction with the GDP growth rate into our model[15].

In Table 9, we present how different versions of betas affecting China's valuation ratios, focusing on the variance contributions of the various variable groups. Similar to our Table 7, the variance contribution is calculated as the covariance between the product of each coefficient and the corresponding independent variable with the fitted value in the regression, divided by the variance of the fitted value. Panel A reports the results for the long sample. Column 1 reports the benchmark results, showing financial openness to dominate, followed by financial development. Column 2 adjusts for GDP betas (of earnings yields and earnings growth) which yields identical results. The next two columns examine the segmented model, with unconditional and time-varying betas. Importantly, the adjusted R^2 of these models is lower than for the benchmark model. Thus, adding information about beta noises up the regression and does not improve its fit. The same is true for the 6 cases (representing three different models for U.S. discount rates and constant or time-varying betas) accommodating a partially segmented model. Thus, the model maximizing explanatory power for the Chinese earnings yield is the benchmark model.

Nevertheless, the robustness of the main results across specifications is striking. The contribution of financial openness varies between 38.4% and 48.4% and is the dominant group of variables in half the cases, but is outperformed by the financial development variables in the other 5 cases. Recall that the financial development variables mostly explain cross-sectional variation but do not help explain the temporal variation in Chinese earnings yields. The contribution of the

[15] There is enormous variation in both the magnitude and precision of the direct portfolio specific EY betas and the earnings betas. To use these betas in the panel regression, we apply a Bayesian model which shrinks less precisely estimated betas towards zero. Appendix F describes the exact procedure.

other variables is below 10% with the exception of the break variable which survives in some partial segmentation models and accounts for a substantial part of the explained variation.

In Panel B, we focus on the short sample. The first column repeats our main result that financial openness and growth expectations are the most important variable groups, with financial development and the investor base still accounting for about 20% of the explained variation each. The results from the other models largely confirm what we find for the longer sample. First, adding the GDP betas gives identical PcGets results. Second, all the (partial) segmented models generate lower adjusted R^2s than the benchmark model. Yet, the main results remain robust. There is a bit more variation on the contribution of financial openness with its contribution varying between 20.8% and 37.2%, whereas the contribution of growth expectations varies between 22.4% and 42.8%. In 6 out of the 10 cases, they remain the two most important variable groups, and in only two cases do they represent jointly less than 60% of the explained variation. The investor base and financial development variables sometimes account for around 30% of explained variation.

The PcGets model never selects U.S. discount rate variation as an important independent variable in either panels. This confirms that valuations in China are still mostly driven by domestic factors. To better understand the fitting of the PcGets model using only Chinese variables, we also plot the fitted value of valuation gap for the market level data in Figure 5, presented in dashed lines. That is, we first estimate the Chinese EY from the PcGets model using only Chinese variables, and then subtract the data U.S. earnings yield. The presented time-series are aggregated from portfolio level earnings yields. Compared to the dotted line using both China and U.S. data, estimated from models in Section V, the China models relax several model constraints, such as market integration, and unit betas, and should fit the data better. That is exactly what we observe in Figure 5. The earnings yield differentials generated from the Chinese earnings yield model are closer to the actual values than those generated from the China-U.S. differential models. If we regress the fitted market level valuation gap on the predicted values using the China model, the adjusted R-squares are 0.74 for the long sample and 0.78 for the short sample, slightly better than using the China-U.S. differential model. This moderate better fit implies that the relaxation of constraints is useful for data fitting, and also implies the China-U.S. differential models are good benchmarks, which capture majority of the variations in valuation gaps between China and U.S. already.

VI. Conclusion

We study valuation differentials in China and the U.S. over the past 24 years at the portfolio level. We first document a curious valuation gap, with Chinese price earnings ratios being substantially higher than those of the U.S. in the first half of our sample (which starts in 1995), in contrast to the usual discount observed for emerging markets. The valuation gap disappears in the second half of the sample. There is also a cross-portfolio dimension to these valuation gaps, both in terms of magnitude and sign. With valuations linked to costs of capital and signals about future growth opportunities, it is important to understand what drives these valuation differentials and their evolution over time.

Focusing on earnings yield differentials, we examine a number of potential explanations. First, we examine differences in industry structure across China and the U.S. We find that sector differences generally play a minor role in driving valuation differentials across the two countries. However, the banking sector's increased prominence going hand in hand with its increased earnings yields did play a non-trivial role in engineering higher earnings yields for the Chinese market as a whole. Second, we then focus on valuation differentials across industries and consider 4 groups of potential explanations: differential growth expectations; the investor base, which we split into two, one focusing on foreign ownership, the other on state, institutional and retail ownership, and finally, financial development and liquidity. For the longer sample, we have insufficient information on growth prospects, and growth expectations account for less than 5% of the explained variation. The most important variable group by far is foreign ownership, accounting for more than 73% of the explained variation, followed by the investor base (12.4%) and financial development (9.3%). For the shorter sample, with more accurate measurement of growth expectations, its explanatory power increases to 33.0%, whereas financial openness still accounts for 42.5% of the explained variation. The investor base still accounts for nearly 16% and financial development for 8.6% of the explained variation. This result is rather robust across different specifications, with some resulting in a larger relative role for financial development. Note that the role of the banking sector is partially driven by its increased foreign ownership over time. It is the portfolio with the second highest international accessibility out of all 33 portfolios.

China witnessed a gradual opening of its shares to foreign investors, and foreign investors price Chinese stocks at lower valuations than do domestic investors, especially retail investors. This gradual integration of Chinese into global capital markets helped eliminate the Chinese valuation gap. However, the valuation gap has not disappeared completely for all industries.

If foreigners' value Chinese stocks at more realistic valuations than do domestic investors, the increased foreign ownership may in fact make stock market valuations in China more informative to economic policy makers, e.g. in predicting economic activity. We defer testing this conjecture to future research, but it is noteworthy that the explanatory power of growth related variables has increased over time.

References

Allen, F., Qian, J., Shan, C., & Zhu, J. (2019). Dissecting the long-term performance of the Chinese stock market. *SSRN Working Paper*.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31-56.

Amihud, Y., Mendelson, H., & Pedersen, L. H. (2006). *Liquidity and asset prices*. Now Publishers Inc.

Baele, L. (2005). Volatility spillover effects in European equity markets. *Journal of Financial and Quantitative Analysis*, 373-401.

Bai, J., Lumsdaine, R. L., & Stock, J. H. (1998). Testing for and dating common breaks in multivariate time series. *The Review of Economic Studies*, 65(3), 395-432.

Bai, C. E., Liu, Q., Lu, J., Song, F. M., & Zhang, J. (2004). Corporate governance and market valuation in China. *Journal of Comparative Economics*, *32*(4), 599-616.

Bailey, W., Chung, Y. P., & Kang, J. K. (1999). Foreign ownership restrictions and equity price premiums: what drives the demand for cross-border investments?. *Journal of Financial and Quantitative Analysis*, 489-511.

Bartram, S. M., Brown, G., & Stulz, R. M. (2012). Why are US stocks more volatile?. *The Journal of Finance*, 67(4), 1329-1370.

Bekaert, G. (1995). Market integration and investment barriers in emerging equity markets. *The World Bank Economic Review*, 9(1), 75-107.

Bekaert, G., & Harvey, C. R. (2003). Emerging markets finance. *Journal of Empirical Finance*, 10(1-2), 3-55.

Bekaert, G., Harvey, C. R., & Lumsdaine, R. L. (2002). Dating the integration of world equity markets. *Journal of Financial Economics*, *65*(2), 203-247.

Bekaert, G., Harvey, C. R., & Lundblad, C. T. (2003). Equity market liberalization in emerging markets. *Journal of Financial Research*, 26(3), 275-299.

Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and expected returns: Lessons from emerging markets. *The Review of Financial Studies*, *20*(6), 1783-1831.

Bekaert, G., Harvey, C. R., Lundblad, C. T., & Siegel, S. (2011). What segments equity markets?. *The Review of Financial Studies*, 24(12), 3841-3890.

Boardman, A. E., & Vining, A. R. (1989). Ownership and performance in competitive environments: A comparison of the performance of private, mixed, and state-owned enterprises. *The Journal of Law and Economics*, 32(1), 1-33.

Carpenter, J. N., Lu, F., & Whitelaw, R. F. (2021). The real value of China's stock market. *Journal of Financial Economics*, *139*(3), 679-696.

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.

Chan, M. K., & Kwok, S. (2018). Connecting the markets? Recent evidence on China's capital account liberalization. *Economic Modelling*, 70, 417-428.

Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, *57*(5), 2113-2141.

Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1996). Political risk, economic risk, and financial risk. *Financial Analysts Journal*, *52*(6), 29-46.

Errunza, V., & Losq, E. (1985). International asset pricing under mild segmentation: Theory and test. *The Journal of Finance*, 40(1), 105-124.

Fernald, J., & Rogers, J. H. (2002). Puzzles in the Chinese stock market. *Review of Economics and Statistics*, *84*(3), 416-432.

Ferreira, M. A., & Matos, P. (2008). The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics*, 88(3), 499-533.

Ferreira, M. A., & Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*, *100*(3), 514-537.

Frankel, J. A. (1992). Measuring international capital mobility: a review. *The American Economic Review*, *82*(2), 197-202.

Hendry, D. F. (1995). Dynamic Econometrics. *Oxford University Press.*

Hendry, D. F. & Krolzig, H. M. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, 25(6-7), 831-866.

Jiang, F., & Kim, K. A. (2015). Corporate governance in China: A modern perspective. *Journal of Corporate Finance*, *32*, 190-216.

Lesmond, D. A. (2005). Liquidity of emerging markets. *Journal of Financial Economics*, *77*(2), 411-452.

Lesmond, D. A., Ogden, J. P., & Trzcinka, C. A. (1999). A new estimate of transaction costs. *The Review of Financial Studies*, 12(5), 1113-1141.

Levine, R., & Zervos, S. (1998). Stock markets, banks, and economic growth. *American Economic Review*, 537-558.

Liao, L., Liu, B., & Wang, H. (2014). China′ s secondary privatization: Perspectives from the split-share structure reform. *Journal of Financial Economics*, 113(3), 500-518.

Lin, J. Y., & Tan, G. (1999). Policy burdens, accountability, and the soft budget constraint.

American Economic Review, 89(2), 426-431.

Liu, J., Stambaugh, R. F., & Yuan, Y. (2019). Size and value in China. *Journal of Financial Economics*, 134(1), 48-69.

Liu, C., Wang, S., & Wei, K. C. (2020). Demand shock, speculative beta, and asset prices: Evidence from the Shanghai-Hong Kong stock connect program. Speculative Beta, and Asset Prices: Evidence from the Shanghai-Hong Kong Stock Connect Program. *SSRN Working Paper*.

Love, I. (2003). Financial development and financing constraints: International evidence from the structural investment model. *The Review of Financial Studies*, 16(3), 765-791.

Martin, I. (2017). What is the Expected Return on the Market?. *The Quarterly Journal of Economics*, 132(1), 367-433.

Megginson, W. L., Nash, R. C., & Van Randenborgh, M. (1994). The financial and operating performance of newly privatized firms: An international empirical analysis. *The Journal of Finance*, 49(2), 403-452.

Mei, J., Scheinkman, J. A., & Xiong, W. (2009). Speculative Trading and Stock Prices: Evidence from Chinese AB Share Premia. *Annals of Economics & Finance*, *10*(2), 225-255.

Morck, R., Yeung, B., & Yu, W. (2000). The information content of stock markets: why do emerging markets have synchronous stock price movements?. *Journal of Financial Economics*, 58(1-2), 215-260.

Pástor, Ľ., & Veronesi, P. (2006). Was there a Nasdaq bubble in the late 1990s?. *Journal of Financial Economics*, 81(1), 61-100.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22(1), 435-480.

Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. *The Journal of Finance*, 52(2), 737-783.

Sun, Q., & Tong, W. H. (2003). China share issue privatization: the extent of its success. *Journal of Financial Economics*, 70(2), 183-222.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, *99*(1), 1-10.

Wurgler, J. (2000). Financial markets and the allocation of capital. *Journal of Financial Economics*, *58*(1-2), 187-214.

Figure 1. Time-series of price earnings ratios

This figure plots the time-series of price earnings (PE) ratios for the Datastream Emerging Market Index, China and the U.S., during the period of 1995Q1 - 2018Q4. The data for the Emerging Market Index is obtained from Datastream using the data series of "TOTMKEK", which has 2302 constituents. Of the 2302 constituents in Emerging Market index, it includes 50 China H shares, but no China A shares. For China and the U.S., we generate the market-level PE ratio from individual firm data following the Datastream method which is also applied in BLHS (2011). We first calculated firm-level earnings at quarter t as the trailing annualized net income by summing up net income from quarter t-4 to quarter t-1. Negative values of firm earnings are set as zero before being aggregated into market level. Total market value is calculated as the summation of all the stocks' price multiplied by common shares outstanding at the end of each quarter. Market level PE ratios are calculated as total market value divided by total earnings.

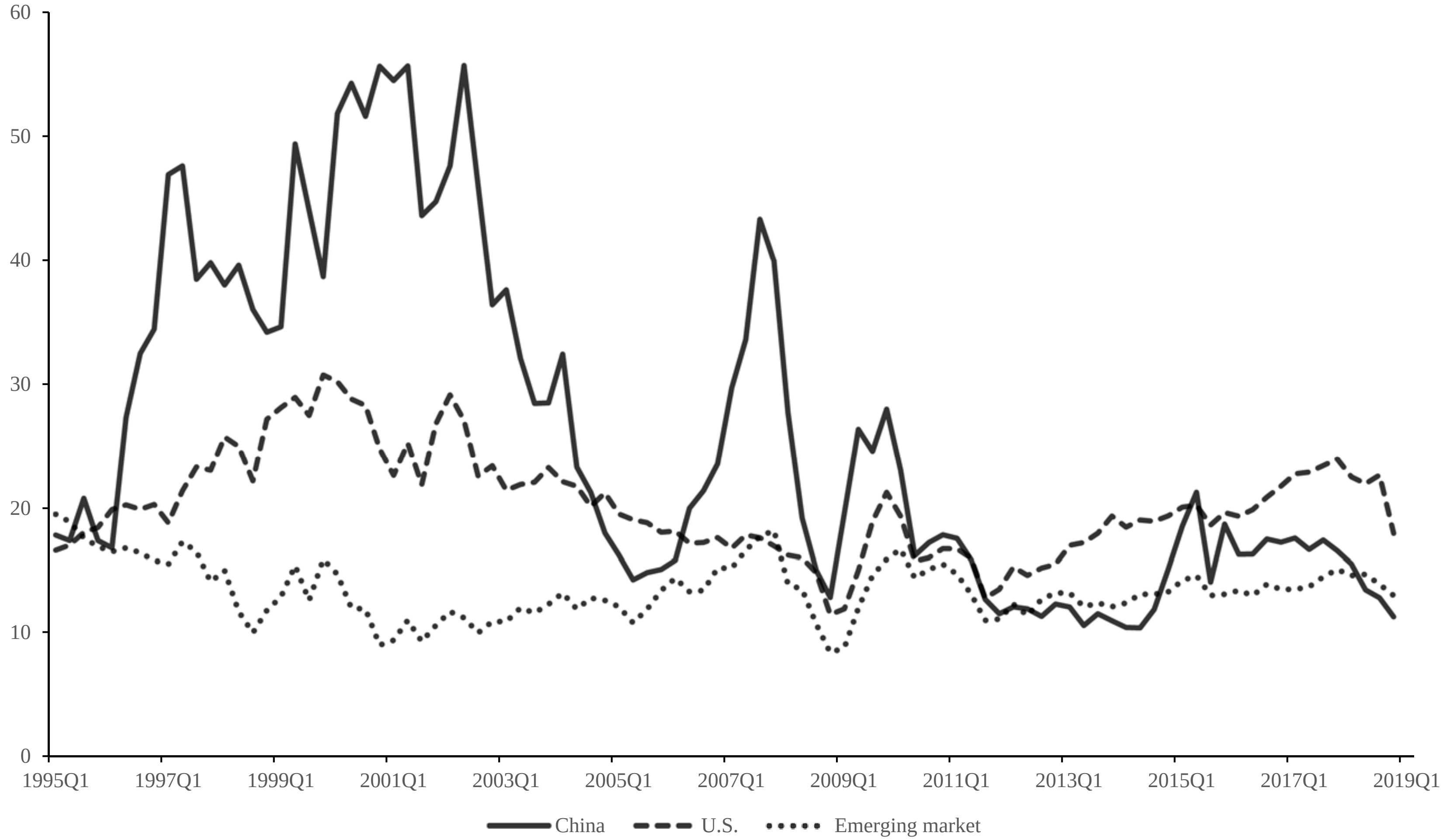
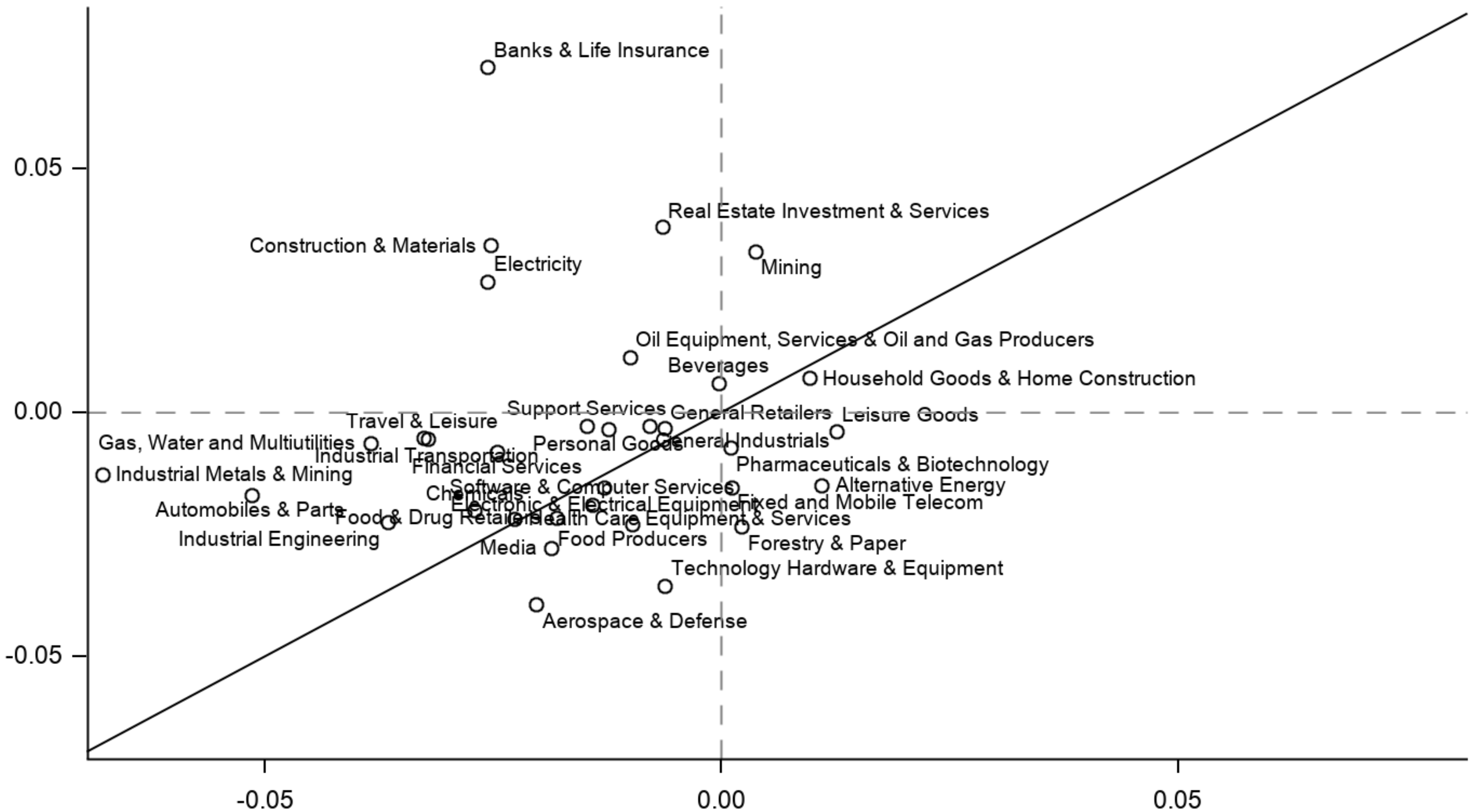


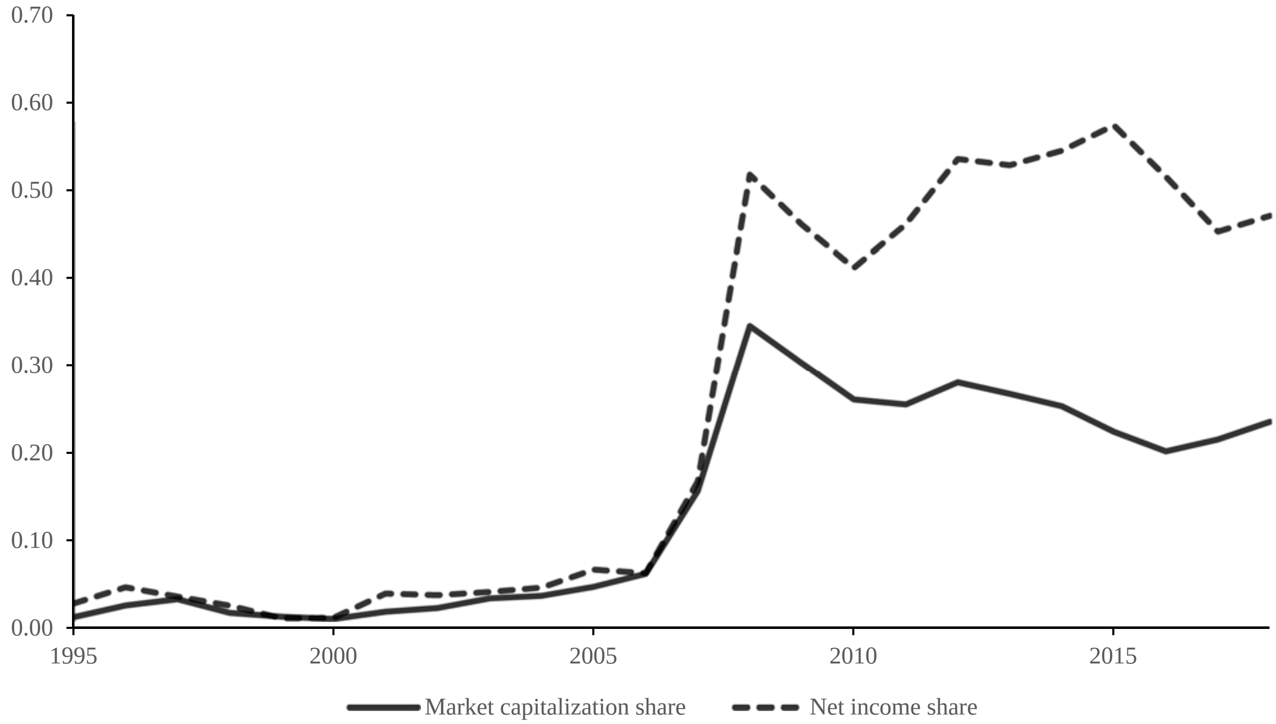
Figure 2. Earnings yield differentials in the cross-section

This figure shows earnings yield differentials in the cross-section. Panel A shows the evolution of earnings yield differentials for different sectors. The X-axis shows the average earnings yield differentials (Chinese sector level earnings yields - U.S. sector level earnings yields) during the first five years of our sample (1995-1999). The Y-axis shows the earnings yield differentials during the last five years of our sample (2014-2018). Panel B shows the time-series of the market share for the "Banks & Life Insurance" sector, both in terms of market capitalization and net income, during our sample period of 1995-2018. The solid line shows the market share of the "Banks & Life Insurance" sector in terms of market capitalization, calculated as the sum of market capitalization of firms which belong to the sector divided by the total market capitalization of the entire market. The dash line shows the market share in terms of net income, calculated as the sum of net income of firms which belong to the sector divided by the total net income of the entire market. In panel C, the solid line shows earnings yield differentials for the whole market while the dash line shows earnings yield differentials constructed using all firms except for firms in the "Banks & Life Insurance" sector.

Panel A. Changes of earnings yield differential by industries



Panel B. Market share of Banks & Life Insurance sector



Panel C. Earnings yield differential with/without Banks & Life Insurance sector

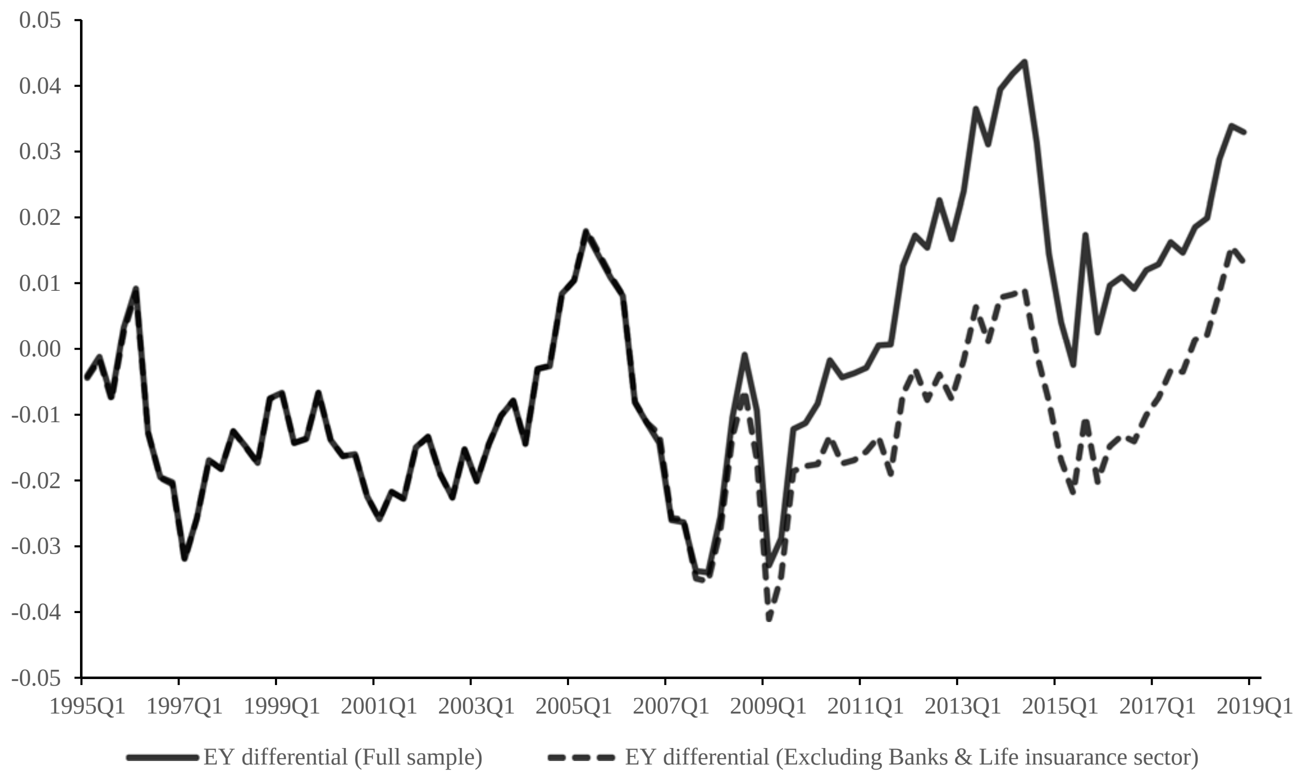
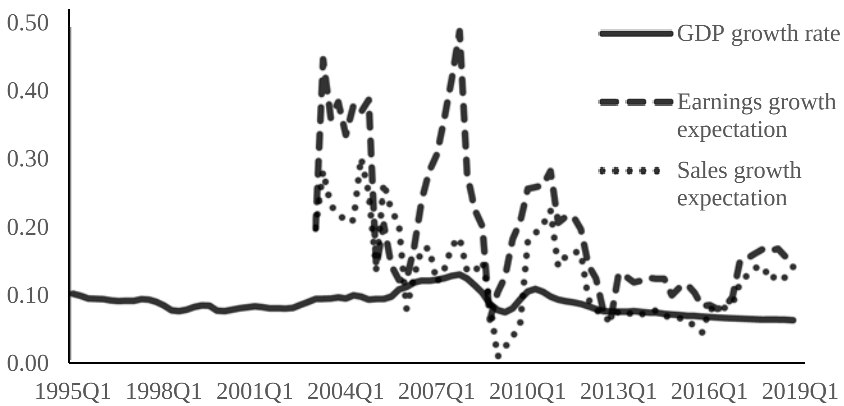


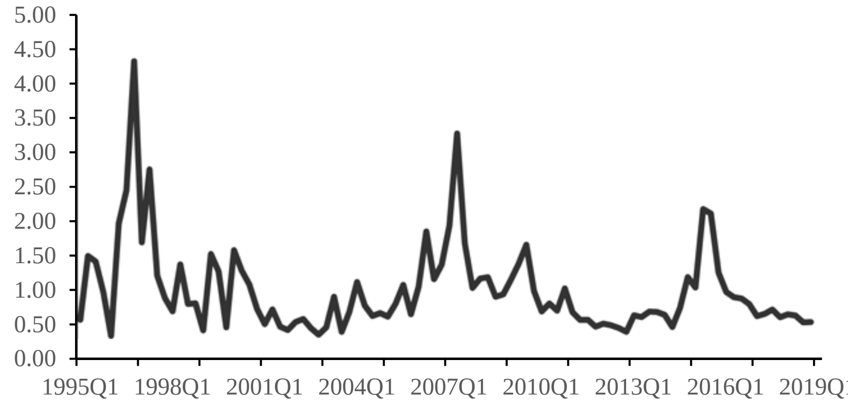
Figure 3. Time series of China Aggregate Variables

This figure shows the time-series of Chinese growth prospect, turnover, international accessibility and ownership measures. Firm level variable construction details are shown in the Appendix A. To obtain market level values, we value weight firm level variables.

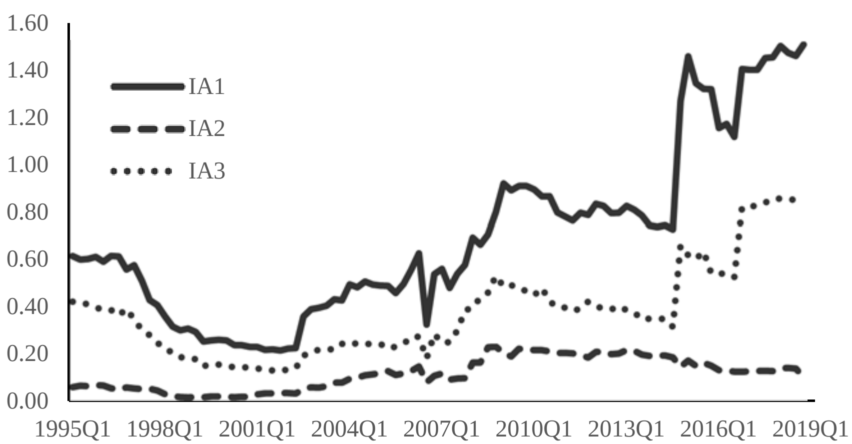
Panel A. Growth Prospect



Panel B. Turnover rate



Panel C. International Accessibility



Panel D. Ownership

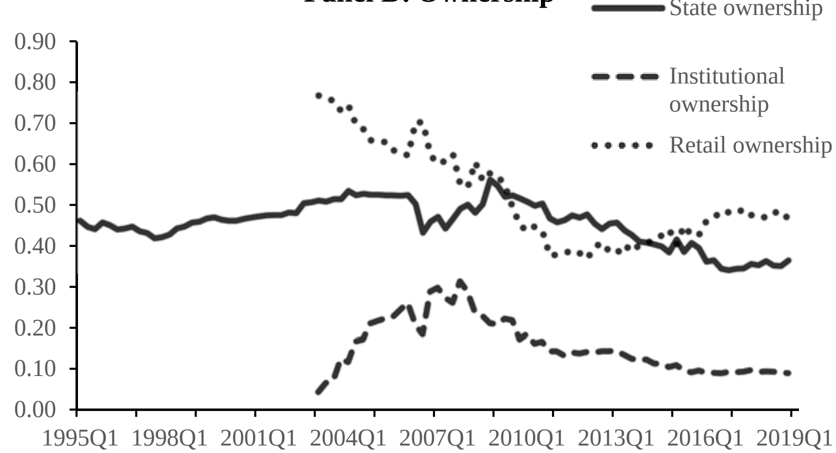
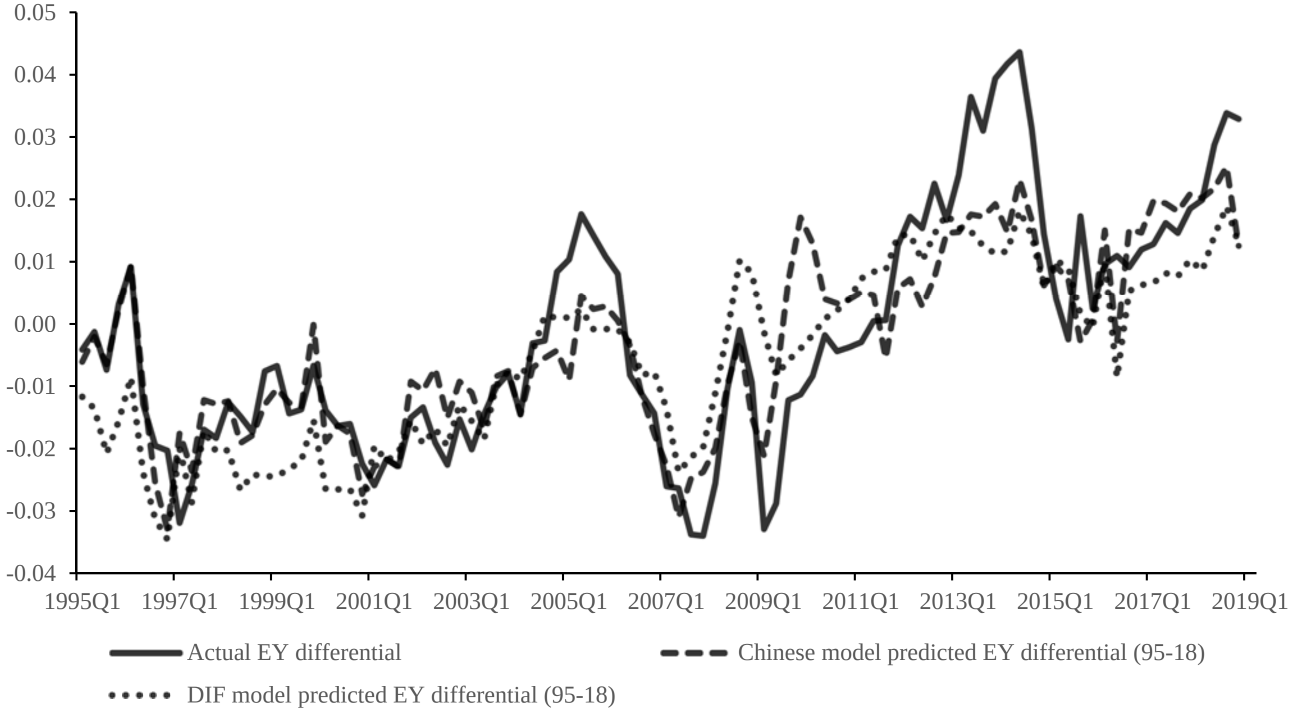


Figure 4. Model fitness comparison

This figure plots the observed earnings yield differentials, with the fitted value of the differenced model and the fitted value of the Chinese earnings yield model. For the latter model we simply subtract the observed U.S. earnings yield from the predicted value for the Chinese earnings yield. For the differenced model, we use the model selected by PcGets in Table 7. For Chinese earnings yield models, we use the model that features the highest adjusted R-square (benchmark model in Table 9). Panel A shows the results for long sample while panel B shows results for the short sample.

Panel A. Long sample (1995-2018)



Panel B. Short sample (2003-2018)

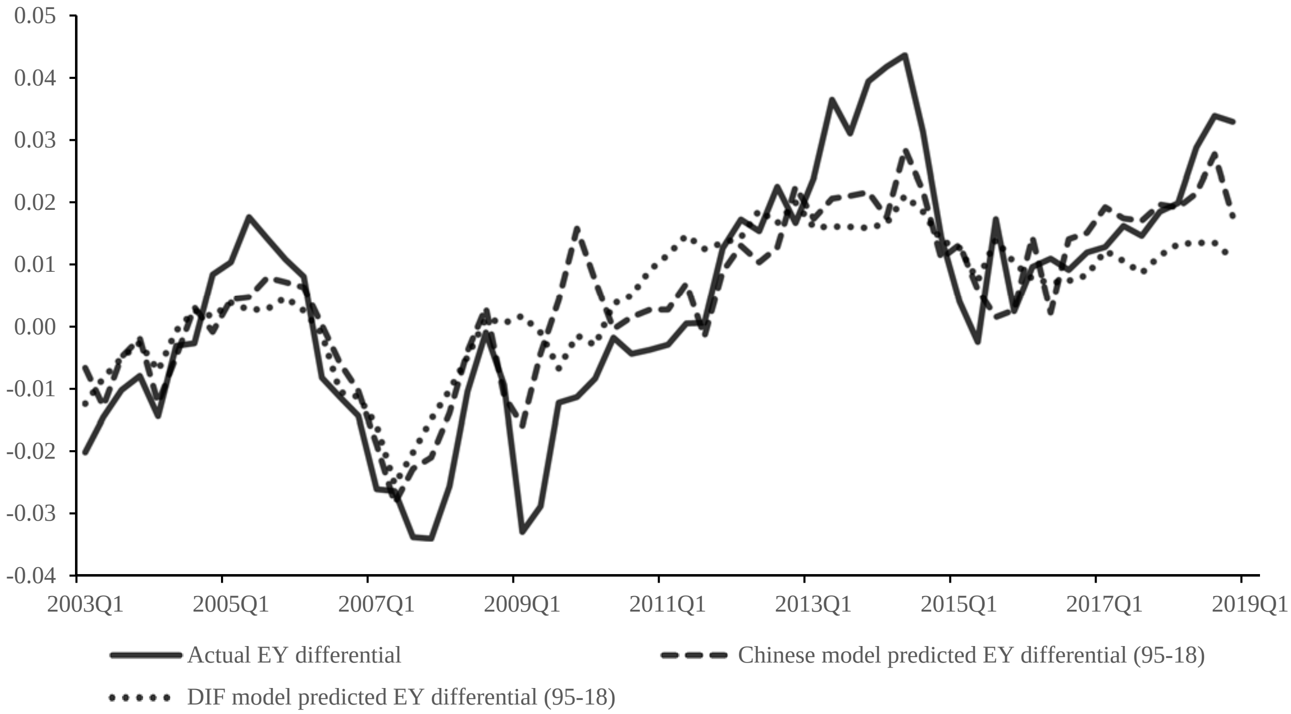


Table 1. Summary statistics by sectors for China and the U.S.

This table reports the time-series average of number of stocks, market values in billion U.S. dollars, sector market shares in the whole market (%), PE ratios and earnings yields in each sector/portfolio and for the market of China and the U.S., from 1995Q1 to 2018Q4. The Chinese sample covers all firms that listed in the A share market. The U.S. sample includes all common stocks listed in New York Stock Exchange, NASDAQ and American Stock Exchanges. Following Liu, Stambaugh, Yuan (2018), we apply following filters to Chinese stocks: (1) Exclude stocks that have become public within the past 2 quarters; (2) Drop stocks that have less than 45 daily return observations during the most recent quarter; (3) Drop stocks that have less than 120 daily return observations during the most recent one year. For U.S. stocks, we apply the following filters: (1) Drop stocks that have less than 45 daily return observations during the most recent quarter; (2) Drop stocks that have less than 120 daily return observations during the most recent one year. All variables in this table are constructed on a quarterly basis for each sector and the whole market. We calculated earnings at quarter t as the trailing annualized net income by summing up net income from quarter t-4 to quarter t-1. In each quarter, for both sector and market level calculations, MV, market value for common equities in billion U.S. dollars, is calculated as the sum of all stocks' price multiplied by common shares outstanding, converted to U.S. dollars using the quarter-end exchange rate. For both China and the U.S., PE ratio is market value for common equity divided by total net income. EY, earnings yield, is total earnings divided by market value for common equity. Negative values of firm earnings are set to zero before being aggregated into sector level. In addition to the industrial sectors, we construct 21 additional portfolios based on state ownership, international accessibility, illiquidity, market size, technology sector and listing boards. The detailed description of the portfolio formations is in Appendix C. The top row of this table reports the time-series summary statistics of the market in China and the U.S. Panel B shows the decomposition of the market level earnings yield differential. The decomposition is performed using the following formula: $Dif\_EY_t = EY_t^{CN} - EY_t^{US} = \sum_{j=1}^{N} w_{j,t}^{CN}\left(EY_{j,t}^{CN} - EY_{j,t}^{US}\right) + \sum_{j=1}^{N}\left(w_{j,t}^{CN} - w_{j,t}^{US}\right)EY_{j,t}^{US}$. Dif_VAL and Dif_STRUC are defined as the first and second component of the decomposition.

Panel A. Summary statistics by sectors/portfolios for China and U.S.

	China					U.S.				
	n(stocks)	MV ($ billion)	Sector MV (%)	PE	EY(%)	n(stocks)	MV ($ billion)	Sector MV (%)	PE	EY(%)
Market	1,400	2,578	100	25.9	4.94	3,976	13,833	100.00	20.3	5.13
Industrial sectors										
Aerospace & Defense	9	14	0	64.7	1.83	62	302	2.10	19.4	5.54
Alternative Energy	6	10	0	68.5	2.23	12	8	0.06	54.9	3.10
Automobiles & Parts	59	86	4	26.9	4.90	41	145	1.16	20.7	7.45
Banks & Life Insurance	10	589	14	22.2	7.82	536	1,223	9.13	14.6	7.15
Beverages	24	65	3	33.6	3.50	23	329	2.47	25.0	4.29
Chemicals	127	110	6	36.8	3.46	83	265	2.01	18.6	5.85
Construction & Materials	79	110	4	32.0	4.58	79	108	0.75	20.5	5.21
Electricity	44	79	5	22.6	5.44	57	351	2.64	16.3	6.42
Electronic & Electrical Equipment	115	119	4	45.4	2.52	190	181	1.38	26.8	4.30
Financial Services	14	83	2	44.9	3.39	142	718	5.01	16.2	6.37
Fixed and Mobile Telecom	3	12	1	59.9	5.23	58	516	4.23	23.3	5.20
Food & Drug Retailers	7	5	0	46.7	2.85	37	201	1.46	21.8	4.92
Food Producers	57	61	2	39.8	2.84	84	284	2.10	18.7	5.59
Forestry & Paper	17	10	1	35.4	3.82	16	31	0.27	56.4	5.21
Gas, Water and Multiutilities	14	14	1	39.3	3.07	51	137	1.02	18.2	5.72
General Industrials	19	17	1	28.5	4.04	48	489	3.71	22.5	4.91
General Retailers	65	51	3	36.8	3.18	205	881	6.15	23.4	4.52
Health Care Equipment & Services	10	11	0	82.3	2.16	261	523	3.65	23.4	4.53
Household Goods & Home Construction	26	40	1	22.6	5.16	101	287	2.10	18.7	5.51

	China					U.S.				
	n(stocks)	MV ($ billion)	Sector MV (%)	PE	EY(%)	n(stocks)	MV ($ billion)	Sector MV (%)	PE	EY(%)
Industrial Engineering	114	130	5	37.0	3.37	129	218	1.55	18.4	5.98
Industrial Metals & Mining	68	114	6	52.8	4.21	35	77	0.57	19.6	6.99
Industrial Transportation	43	78	4	30.7	4.35	67	212	1.49	17.4	5.98
Leisure Goods	19	19	2	45.6	3.30	49	72	0.51	30.8	4.05
Media	18	26	1	59.4	2.15	125	600	4.32	29.3	4.36
Mining	28	93	3	28.2	4.84	32	40	0.34	44.0	3.21
Oil Equipment, Services & Oil and Gas Producers	12	200	7	27.0	5.44	178	1,076	7.84	20.7	6.21
Personal Goods	55	42	2	32.6	3.45	80	194	1.44	21.2	4.94
Pharmaceuticals & Biotechnology	92	109	4	37.7	3.00	232	1,202	8.78	27.0	4.08
Real Estate Investment & Services	101	109	8	33.4	4.51	30	21	0.13	36.8	3.96
Software & Computer Services	39	44	1	65.8	1.88	279	1,281	8.46	29.0	3.92
Support Services	31	25	2	41.0	3.06	212	293	2.11	25.7	4.07
Technology Hardware & Equipment	47	60	2	43.4	2.57	280	1,208	8.64	30.6	4.39
Travel & Leisure	31	49	2	41.9	3.38	162	360	2.43	20.4	5.10
Other portfolios										
State-own Portfolio 1 (SO=0)	344	357	10	33.7	3.50	3,008			20.0	5.15
State-own Portfolio 2 (0<SO<=10%)	215	318	8	33.7	3.61	2,987			20.3	5.14
State-own Portfolio 3 (10%<SO<=50%)	480	646	27	29.7	4.36	3,757			19.0	5.42
State-own Portfolio 4 (SO>50%)	360	1,257	54	23.2	5.96	3,583			17.6	5.92
Retail Portfolio 1(Retail ownership <= country level lower 30%)	535	2,140	56	16.6	7.04	1,038			19.9	5.16
Retail Portfolio 2(Retail ownership >= country level upper 30%)	535	535	14	32.1	3.53	798			22.8	4.60
IO Portfolio 1(IO <= country level lower 30%)	559	685	16	35.2	4.46	844			22.3	4.69
IO Portfolio 2(IO > country level upper 30%)	535	1,665	52	17.8	6.22	1,032			19.9	5.14
International Accessibility Portfolio 1 (IA1=0)	1,135	1,013	56	32.4	3.65	3,920			19.6	5.29
International Accessibility Portfolio 2 (IA1>0)	264	1,565	44	26.1	5.91	2,905			17.3	6.02
Illiquidity Portfolio 1(Zeros <= country level lower 30%)	622	1,044	42	29.1	4.06	1,418			20.4	5.06
Illiquidity Portfolio 2(Zeros > country level upper 30%)	549	1,172	44	24.4	5.74	1,242			20.3	5.17
Turnover Portfolio 1(Turnover <= country level lower 30%)	420	1,571	49	24.1	5.69	1,088			20.6	4.99
Turnover Portfolio 2(Turnover >= country level upper 30%)	420	357	20	42.0	3.02	1,080			20.2	5.15
Size Portfolio 1(Market value <= country level lower 30%)	420	141	7	61.5	1.87	904			23.0	4.70
Size Portfolio 2(Market value >= country level upper 30%)	420	2,053	74	22.9	5.60	1,088			19.9	5.19
Tech portfolio 1(Not TMT industry)	1,292	2,436	95	25.5	5.06	3,193			19.7	5.23
Tech portfolio 2(TMT industry)	107	142	5	41.5	2.88	742			25.2	4.47
Listing board portfolio 1(Main board)	1,048	2,169	93	25.2	5.30	3,934			18.2	5.69
Listing board portfolio 2(SME board)	422	516	9	35.8	3.06	2,707			18.9	5.44
Listing board portfolio 3(GEM board)	321	330	5	52.9	2.10	2,405			20.1	5.11

Panel B. Decomposition of earning yield differentials

					Variance Decomposition	
	DIFEY (%) Mean	DIF_VAL (%) Mean	DIF_STRUC (%) Mean	DIFEY (%) variance	COV(DIF_VAL,DIFEY) /VAR(DIFEY)	COV(DIF_STRUC, DIFEY) /VAR(DIFEY)
Market	-0.190	-0.690	0.500	3.549	0.99	0.01

Table 2. Bai, Lumsdaine and Stock (1998) break point test on earnings yield differentials

This table shows the Bai, Lumsdaine and Stock (1998) break tests for different portfolio settings. We include both intercept and lag terms in our models and test whether there is a break in the intercept. “Market EY” is the market earnings yield differential. We also show break point tests for the market level earnings yield constructed using all firms except for firms in the “Banks & Life Insurance” sector. We do break point tests for the valuation part and the structural part of the market level earnings yield differential. The decomposition is performed using the following formula: $DIFEY_t = EY_t^{CN} - EY_t^{US} = \sum_{j=1}^{N} w_{j,t}^{CN}\left(EY_{j,t}^{CN} - EY_{j,t}^{US}\right) + \sum_{j=1}^{N}\left(w_{j,t}^{CN} - w_{j,t}^{US}\right)EY_{j,t}^{US}$. “DIF_VAL” and “DIF_STRUC” are defined as the first component and the second component of the decomposition. “DIF_VAL_NB” is a version of “DIF_VAL” excluding the “Banks & Life Insurance” sector during the process of calculation. The sample period ranges from 1995 to 2018. The maximum allowed lag is 6. The optimal lag is chosen by the BIC criterion. We report the Sup-Wald statistic from Bai, Lumsdaine and Stock (1998). Critical values are from Bekaert, Harvey and Lumsdaine (2002), Table 10 in their appendix. ***, ** and * indicate significances at the 1%, 5% and 10% levels using two-tailed tests.

Variable	Sup-Wald statistic	Estimated break point	90% confidence interval
Market EY	10.03**	2009:03	2007:02-2011:04
Market EY (Excluding banking sector)	3.64	2011:04	2003:02-2018:04
DIF_VAL	11.68**	2009:02	2007:03-2011:01
DIF_VAL (Excluding banking sector)	8.24*	2009:02	2005:01-2013:03
DIF_STRUC	4.27	2000:04	1996:03-2006:01

Table 3. Valuation differentials and growth prospect

This table reports results for pooled OLS sector/portfolio level regressions of earnings yield differentials on growth prospect variables from 1995Q1 to 2018Q4. The growth prospect variables include GDP growth rate, earnings growth expectation and sales growth expectation. GDP growth rate starts from 1995Q1. Due to availability of analyst forecast data, analyst-related variables including earnings growth expectation, sales growth expectation are only available after 2003Q1. "Dummy: 2009Q3" is a dummy variable which equals to one for dates after 2009, quarter 3, otherwise zero. Definitions of all other variables are described in detail in Appendix A. The dependent variable is the sector/portfolio level earning yield differentials between China and the U.S., DIFEY. All independent variables are differences between China and the U.S. except for "Dummy: 2009Q3" which is a dummy variable equal to 1 (0) during period after (before) 2009Q3 and political risk variables (including overall political rating, quality of institutions and investment profile) which are constructed by taking the ratio of Chinese over U.S. variables. Control variables include leverage differentials, earnings growth differential and minimum number of stocks. Standard errors are double clustered by sector and time. We report t statistics under the coefficient estimates. ***, ** and * indicate significances at the 1%, 5% and 10% levels using two-tailed tests.

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.008	-0.011**	-0.011	-0.016**	0.007	0.032***
	(-1.482)	(-2.389)	(-1.386)	(-2.105)	(0.877)	(2.78)
Dummy: after 2009Q3		0.015***		0.012***	0.007	-0.001
		(3.721)		(2.767)	(1.598)	(-0.148)
GDP growth rate	-0.158**	-0.118**			-0.264***	-0.396***
	(-2.458)	(-2.285)			(-4.482)	(-5.312)
Earnings growth expectation			-0.002	-0.002	-0.002	0.002
			(-1.234)	(-1.134)	(-1.296)	(0.510)
Sales growth expectation			-0.055***	-0.048***	-0.046***	-0.034**
			(-3.269)	(-3.061)	(-2.888)	(-2.568)
Earnings growth expectation * number of analyst						-0.012***
						(-4.839)
Earnings growth expectation * forecast dispersion						0.043***
						(4.734)
Number of analysts						0.010***
						(4.268)
Forecast dispersion						-0.030***
						(-3.010)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	4,873	4,873	3,317	3,317	3,317	3,204
Adjusted R-square	0.068	0.133	0.102	0.140	0.174	0.277

Table 4. Valuation differentials and market development

This table reports regression results related to sector/portfolio level market development variables in 1995Q1 to 2018Q4. Market development variables include regulatory financial development (REGDEV), zeros, turnover rate, number of public firms, adjusted market development, MYY R^2 synchronicity, idiosyncratic volatility, industry concentration ratio. The dependent variable is the sector/portfolio level earning yield differential between China and the U.S., DIFEY. All independent variables are differences between China and the U.S. except for "Dummy: 2009Q3" which is a dummy variable equal to 1 (0) during period after (before) 2009Q3 and political risk variables (including overall political rating, quality of institutions and investment profile) which are constructed by taking the ratio of Chinese over U.S. variables. Control variables include leverage differentials, earnings growth differential and minimum number of stocks. Definitions of all the variables are described in detail in Appendix A. The standard errors are double clustered by sector and time. We report t statistics under the coefficient estimates. ***, ** and * indicate significances at the 1%, 5% and 10% levels using two-tailed tests.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	-0.020***	-0.019***	-0.012***	-0.005	-0.019***	-0.018***	-0.018***	-0.019***	-0.045***	-0.047***
	(-4.757)	(-4.473)	(-3.055)	(-0.772)	(-4.511)	(-3.900)	(-4.524)	(-4.600)	(-5.732)	(-6.221)
Dummy: after 2009Q3										0.018***
										(4.084)
REGDEV	0.007***								0.014***	0.007**
	(3.532)								(4.414)	(2.231)
Zeros		0.135***							0.130***	0.157***
		(4.236)							(4.886)	(5.616)
Turnover rate			-0.009***						-0.009***	-0.008***
			(-4.494)						(-4.380)	(-4.234)
Number of public firms				0.007***					-0.012***	-0.014***
				(2.654)					(-3.133)	(-3.700)
Adjusted market development					-0.001				-0.003***	-0.003***
					(-0.824)				(-2.754)	(-2.752)
MYY R^2 synchronicity						-0.003			0.019***	0.018***
						(-0.364)			(2.900)	(2.808)
Idiosyncratic volatility							0.001		0.011	0.010
							(0.074)		(1.194)	(1.247)
Industry concentration ratio								0.004	0.007	0.008*
								(0.976)	(1.614)	(1.822)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	4,873	4,873	4,873	4,873	4,873	4,873	4,873	4,873	4,873	4,873
Adjusted R-square	0.107	0.107	0.124	0.084	0.052	0.052	0.051	0.053	0.244	0.264

Table 5. Valuation differentials and financial openness

This table reports regression results related to sector/portfolio level financial openness variables from 1995Q1 to 2018Q4. Financial openness variables include real interest rate, three international accessibility measures (IA1, IA2 and IA3), regulatory financial openness (REGOPEN), overall political rating, quality of institutions, investment profile, AB premium and AH premium. The dependent variable is the sector/portfolio level earning yield differential between China and the U.S., DIFEY. All independent variables are differences between China and the U.S. except for "Dummy: 2009Q3" which is a dummy variable equal to 1 (0) during period after (before) 2009Q3 and political risk variables (including overall political rating, quality of institutions and investment profile) which are constructed by taking the ratio of Chinese over U.S. variables. Control variables include leverage differentials, earnings growth differential and minimum number of stocks. Definitions of all these variables are described in detail in Appendix A. Standard errors are double clustered by sector and time. We report t statistics under the coefficient estimates. ***, ** and * indicate significances at the 1%, 5% and 10% levels using two-tailed tests.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	-0.022***	-0.022***	-0.022***	-0.022***	-0.018***	0.023	0.014	-0.003	-0.016***	-0.017***	-0.015	-0.027
	(-5.076)	(-6.021)	(-5.706)	(-5.761)	(-4.486)	(1.165)	(0.781)	(-0.291)	(-3.839)	(-4.259)	(-0.717)	(-1.377)
Dummy: after 2009Q3												0.007
												(1.234)
REGOPEN	0.004***										0.002**	0.001
	(4.293)										(2.254)	(1.053)
IA1		0.022***									0.012	0.013
		(5.900)									(1.488)	(1.594)
IA2			0.133***								0.111***	0.102***
			(6.049)								(4.687)	(4.085)
IA3				0.030***							-0.020	-0.019
				(4.891)							(-1.566)	(-1.509)
Real interest rate					0.016						0.000	-0.004
					(0.288)						(0.004)	(-0.066)
Overall political rating						-0.050**					-0.057**	-0.037
						(-2.179)					(-2.219)	(-1.439)
Quality of institutions							-0.049*				0.072	0.065
							(-1.815)				(1.276)	(1.211)
Investment profile								-0.022*			-0.011	-0.009
								(-1.800)			(-0.526)	(-0.448)
A-B premium									-0.003***		-0.001	-0.001
									(-3.147)		(-0.871)	(-0.903)
A-H premium										-0.002***	-0.001*	-0.001*
										(-3.439)	(-1.760)	(-1.646)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	4,873	4,873	4,873	4,873	4,873	4,873	4,873	4,873	4,873	4,873	4,873	4,873
Adjusted R-square	0.111	0.197	0.228	0.148	0.051	0.067	0.062	0.060	0.078	0.085	0.279	0.282

Table 6. Valuation differentials and investor base

This table reports time-series sector/portfolio level panel regression results on investor base variables. Ownership variables include Chinese state ownership, institutional ownership, retail ownership, Chinese standardized number of shareholders (Chinese SNS) and turnover rate. Institutional ownership, retail ownership and Chinese SNS are available from 2003Q1 to 2018Q4 while Chinese state ownership and turnover rate start from 1995Q1 to 2018Q4. The dependent variable is the sector/portfolio level earning yield differential between China and the U.S., DIFEY. All independent variables are differences between China and the U.S. except for "Dummy: 2009Q3" which is a dummy variable equal to 1 (0) during period after (before) 2009Q3 and political risk variables (including overall political rating, quality of institutions and investment profile) which are constructed by taking the ratio of Chinese over U.S. variables. Control variables include leverage differentials, earnings growth differential and minimum number of stocks. Definitions of all the variables are described in detail in Appendix A. Standard errors are double clustered by sector and time. We report t statistics under the coefficient estimates. ***, ** and * indicate significances at the 1%, 5% and 10% levels using two-tailed tests.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	-0.028***	-0.037***	-0.003	-0.001	-0.012***	-0.019***	-0.025***	-0.027***	-0.027***
	(-5.789)	(-3.572)	(-0.364)	(-0.111)	(-3.055)	(-4.480)	(-5.008)	(-2.814)	(-2.629)
Dummy: after 2009Q3							0.017***		0.0005
							(3.991)		(0.098)
Chinese state ownership	0.024***					0.016**	0.028***	0.022**	0.023**
	(2.711)					(2.050)	(2.897)	(2.221)	(2.089)
Institutional ownership		-0.025**						-0.038***	-0.037***
		(-2.007)						(-2.963)	(-2.867)
Retail ownership			-0.036***					-0.028**	-0.028**
			(-3.546)					(-2.435)	(-2.233)
Chinese SNS				-0.086***				-0.047**	-0.046**
				(-3.478)				(-2.375)	(-2.099)
Turnover rate					-0.009***	-0.008***	-0.006***	-0.008***	-0.008***
					(-4.494)	(-4.551)	(-4.171)	(-4.810)	(-4.757)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	4,873	3,408	3,408	3,408	4,873	4,873	4,873	3,408	3,408
Adjusted R-square	0.075	0.069	0.114	0.131	0.124	0.134	0.207	0.287	0.287

Table 7. Valuation differentials under the PcGets model selection method

This table reports the PcGets model selection results on all variables that we discussed in Table 3 - Table 6. The left panel shows results for variables available from 1995 to 2018 while the right panel shows results for variables from 2003 to 2018. The dependent variable is the sector/portfolio level earning yield differential between China and the U.S., DIFEY. All independent variables are differences between China and the U.S. except for “Dummy: 2009Q3” which is a dummy variable equal to 1 (0) during period after (before) 2009Q3 and political risk variables (including overall political rating, quality of institutions and investment profile) which are constructed by taking the ratio of Chinese over U.S. variables. Standard errors are double clustered by sector and time. We apply the PcGets procedure to pick up the most important independent variables. The overall variance contribution of each selected variable is reported. A detailed description of the PcGets procedure is provided in Appendix E. We report t statistics under the coefficient estimates. ***, ** and * indicate significances at the 1%, 5% and 10% levels.

	1995-2018		2003-2018	
	(1)	(2)	(3)	(4)
Dependent Variable: DIFEY				
Intercept	-0.013***		0.001	
	(-3.267)		(0.173)	
Growth Expectations				
GDP growth rate	-0.176***	4.7%	-0.284***	15.8%
	(-3.327)		(-4.713)	
Sale growth expectation			-0.027***	6.4%
			(-2.930)	
Forecast dispersion			-0.034***	10.8%
			(-5.973)	
Financial Development				
Zeros	0.073***	10.0%	0.205***	12.9%
	(3.884)		(5.366)	
MYY R^2 synchronicity	0.018***	-1.3%	0.017***	1.4%
	(3.934)		(2.797)	
Idiosyncratic volatility			0.023***	-5.7%
			(2.670)	
Adjusted market development	-0.002***	0.6%		
	(-2.967)			
Financial Openness				
IA2: MV(B,H and ADR)/total MV	0.118***	55.1%	0.101***	42.5%
	(5.570)		(4.743)	
Regulatory financial openness	0.002***	13.2%		
	(3.702)			
A-H premium	-0.001**	5.3%		
	(-2.496)			
Investor base				
Turnover rate	-0.005***	12.4%	-0.007***	15.9%
	(-4.202)		(-3.859)	
Total Variance Contribution		100%		100%
Number of observations	4,873		3,204	
Adjusted R-square	0.330		0.418	

Table 8. Chinese earnings yields under the PcGets method

This table shows results for the benchmark regression of Equation (11), in which we impose unit betas. The dependent variable is the Chinese earnings yield (EY_China). We include all variables we discussed in table 3 - table 6. All variables are Chinese based. The left panel shows the result for variables available from 1995 to 2018 while the right panel shows the result for variable from 2003 to 2018. Standard errors are double clustered by sector and time. We apply the PcGets procedure to pick out the most important independent variables. The overall variance contribution of each selected variable is reported. Detailed description of the PcGets procedure is provided in Appendix E. We report t statistics under the coefficient estimates. ***, ** and * indicate significances at the 1%, 5% and 10% levels.

	1995-2018		2003-2018	
	(1)	(2)	(3)	(4)
Dependent Variable: EY_China				
Intercept	-0.026		0.058***	
	(-1.633)		(6.805)	
Dummy: 2009Q3			-0.009**	-4.0%
			(-2.487)	
Growth Expectations				
GDP growth rate	-0.142**	2.1%	-0.267***	6.3%
	(-2.433)		(-3.970)	
Earnings growth expectation			-0.014***	9.2%
			(-4.772)	
Forecast dispersion			-0.049***	6.3%
			(-4.879)	
Number of analysts			0.005***	7.6%
			(2.697)	
Financial Development				
Regulatory financial development	0.010***	25.5%		
	(7.943)			
Zeros	0.188***	12.3%	0.260***	14.7%
	(5.683)		(6.545)	
MYY R^2 synchronicity	0.020***	2.0%	0.033***	3.2%
	(3.666)		(4.652)	
Adjusted market development	-0.003***	-0.2%	-0.002***	1.8%
	(-4.557)		(-3.892)	
Financial Openness				
IA2: MV(B,H and ADR)/total MV	0.128***	48.8%	0.090***	33.3%
	(7.575)		(5.052)	
Quality of institution	0.005***	-5.3%		
	(4.425)			
Investment profile	-0.003***	2.5%		
	(-3.340)			
A-H premium	-0.001***	2.0%		
	(-2.856)			
Investor base				
Chinese SNS			-0.048***	8.3%
			(-2.899)	
Turnover rate	-0.006***	10.3%	-0.007***	13.6%
	(-4.637)		(-3.848)	
Total Variance Contribution		100%		100%
Number of observations	4,873		3,212	
Adjusted R-square	0.556		0.608	

Table 9. Chinese earnings yield models

This table shows PcGets results for different Chinese earnings yield model specifications. In model 1, we assume that the Chinese stock market is totally segmented. Chinese betas are estimated by running Chinese portfolio level excess returns on Chinese market level excess returns. In model 2, model 3 and model 4, we assume that the Chinese market is partially segmented, and we obtain Chinese betas and U.S. betas by running regressions of Chinese portfolio level excess returns on Chinese market level excess returns and U.S. market level excess returns. To measure portfolio-specific discount rates, variables in the financial openness, financial development and investor base groups are multiplied by Chinese betas, except for real interest rate which is multiplied by one minus Chinese betas. To incorporate U.S. expected return into the PcGet procedure. In model 2, we assume the U.S. expected return is constant and add the estimated U.S. beta into the PcGets procedure. In model 3, we add "U.S. beta * Earnings yield based U.S. expected return" into the PcGets. In model 4, we add "U.S. beta * Martin based U.S. expected return" into the procedure. For each model, we both estimate an unconditional version and a conditional version. In an unconditional version, we estimate the beta from a full sample regression. In the conditional version, we obtain betas using the following procedure: in each quarter, we run the excess portfolio level excess return on the Chinese market excess return (for model 1, also U.S. market excess return for model 2-4) using data of the past 52 weeks. "Earnings yield based U.S. expected return" is calculated as U.S. market level earnings yield plus analyst earnings expectation. "Martin based U.S. expected return" is calculated as Ian Martin's lower bound of U.S. expected return measure plus U.S. risk-free rate. The "Martin based U.S. expected return" is only available between 1996Q1 and 2012Q1.

Panel A. 1995-2018

	Benchmark		Totally segmented		Partially segmented					
	No GDP beta, Unit betas	Include GDP beta, Unit betas	Model 1. Un-cond.	Model 1. Cond.	Model 2. Un-cond.	Model 2. Cond.	Model 3. Un-cond.	Model 3. Cond.	Model 4. Un-cond.	Model 4. Cond.
Break (2009:Q3)	0.0%	0.0%	0.0%	14.4%	0.0%	14.9%	0.0%	14.9%	-14.5%	0.0%
Growth expectation - measures and accuracy	2.1%	2.1%	0.0%	4.4%	0.0%	4.5%	0.0%	4.5%	-2.8%	0.0%
Growth expectation - betas	N/A	0.0%	0.0%	6.7%	0.0%	6.6%	0.0%	6.6%	0.0%	0.0%
Financial development	39.6%	39.6%	51.4%	26.5%	53.0%	29.6%	51.6%	29.6%	69.4%	51.3%
Financial openness	48.0%	48.0%	48.6%	42.0%	47.0%	38.4%	48.4%	38.4%	47.9%	43.0%
Investor base	10.3%	10.3%	0.0%	5.9%	0.0%	6.1%	0.0%	6.1%	0.0%	5.7%
U.S. expected returns	N/A	N/A	N/A	N/A	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Number of observations	4,873	4,873	4,873	4,857	4,873	4,857	4,873	4,857	3,261	3,245
Adjusted R-square	0.556	0.556	0.529	0.523	0.528	0.524	0.530	0.524	0.472	0.429

Panel B. 2003-2018

	Benchmark		Totally segmented		Partially segmented					
	No GDP beta, Unit betas	Include GDP beta, Unit betas	Model 1. Un-cond.	Model 1. Cond.	Model 2. Un-cond.	Model 2. Cond.	Model 3. Un-cond.	Model 3. Cond.	Model 4. Un-cond.	Model 4. Cond.
Break (2009:Q3)	-4.0%	-4.0%	-5.5%	-4.9%	-4.3%	-5.8%	-5.6%	0.0%	0.0%	0.0%
Growth expectation - measures and accuracy	29.3%	29.3%	32.7%	38.6%	32.9%	42.8%	32.9%	33.9%	22.4%	39.9%
Growth expectation - betas	N/A	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Financial development	19.6%	19.6%	26.9%	10.5%	20.6%	9.7%	26.8%	9.8%	29.2%	28.6%
Financial openness	33.3%	33.3%	35.5%	26.2%	37.2%	26.0%	35.5%	24.6%	23.6%	20.8%
Investor base	21.8%	21.8%	10.4%	29.6%	13.6%	27.2%	10.5%	31.7%	24.8%	13.7%
U.S. expected returns	N/A	N/A	N/A	N/A	0.0%	0.0%	0.0%	0.0%	0.0%	-2.9%
Number of observations	3,212	3,212	3,212	3,199	3,212	3,199	3212	3,199	1,754	1,741
Adjusted R-square	0.608	0.608	0.594	0.580	0.592	0.574	0.593	0.574	0.514	0.516

Appendix

Appendix A. Variable descriptions

Variable	Description
Earnings yield differential (DIFEY) (Sector/portfolio level, 95-18)	$DIFEY_{j,t} = EY_{j,t}^{CN} - EY_{j,t}^{US}$ This variable measures the sector level earnings yield differentials between China and the U.S. In each country, sector valuation EY is the sum of earnings across all firms in the sector over sector market capitalization. Earnings at quarter t is calculated as the trailing annualized net income by summing up net income from quarter t-4 to quarter t-1. Negative earnings are set to be 0 before aggregating into the sector level. Because Chinese firms only reported semi-annual reports before 2002 and they reported accumulated net income in their semi-annual reports, for missing quarterly earnings data before 2002, we assume that earnings in the first and second quarter of the year are one half of the earnings reported in the semi-annual reports and the earnings for the third and fourth quarter are one half of the total earnings generated in the second half of the year, which is the difference between the earnings reported in the firm's annual reports and that in the semi-annual reports. Frequency: Quarterly.
Control Variables	
Leverage (Sector/portfolio level, 95-18)	For non-financial firms, sector level leverage is calculated as the value-weighted (using last-quarter market cap as weight) ratios of long-term debt plus short-term debt over total assets. For China, the direct items measuring short-term debt and long-term debt are not available, so we add up four items: short-term borrowing, long term borrowing, debt due in future one-year and bond payable to measure total debt. For financial firms, sector level leverage is calculated as the value-weighted (using last-quarter market cap as weight) ratios of total liability over total assets. We winsorize leverage at the 1 and 99 percentiles. Frequency: Quarterly. Sources: WIND and COMPUSTAT.
Earnings growth volatility (Sector/portfolio level, 95-18)	To compute earnings growth volatility, we first calculate annualized firm level net income. Annualized firm level net income at quarter t is calculated by summing up firm level net income from quarter t-4 to quarter t-1. In each quarter, we then compute sector level annualized net income (NI) by adding up firm level annualized net income within each sector. The sector earnings growth at quarter t is calculated as $\log\left(\frac{NI_t * CPI_{t-4}}{NI_{t-4} * CPI_t}\right)$. We calculate the volatility of sector NI growth each quarter by calculating the standard deviation of the log growth rate over the past twenty quarters. For the 10^{th} -19^{th} observation of each sector, we use all available observations to calculate the standard deviation. For the first 10 observations, we use the standard deviation computed for the 10^{th} observation. Frequency: Quarterly. Source: WIND and COMPUSTAT.
Minimum number of stocks (Sector/portfolio level, 95-18)	Natural logarithm of the minimum of the number of stocks in each sector of China and the U.S. Frequency: Quarterly.
Growth Expectations	
GDP growth rate (Market level, 95-18)	GDP growth rate in quarter t is calculated as [GDP(t)+GDP(t-1)+GDP(t-2)+GDP(t-3)]/[GDP(t-4)+GDP(t-4)+GDP(t-6)+GDP(t-7)]-1, using quarterly real GDP. Frequency: Quarterly. Source: China National Bureau of Statistics and Bureau of Economic Analysis.

Sales growth expectation (Sector/portfolio level, 03-18)	We use analysts' sales forecasts to calculate the sales growth expectation. Sales growth expectation is calculated as the weighted average of the annualized sales growth rate expectation in the next 3 years using the most recent comparable CPI growth rates as deflators. Specifically, in each quarter, we first simply sum up the median firm-level sales forecasts by sectors. Then, for the most nearby fiscal year t, the real sales growth rate expectation SG_t is calculated as (median analyst sales forecast for fiscal year t * CPI_{t-2}) / (actual sales in fiscal year t-1* CPI_{t-1}) - 1. For fiscal year t+1, the sales growth rate expectation SG_{t+1} is defined as [(median analyst sales forecast for fiscal year t+1$^*CPI_{t-3}$/actual sales in fiscal year t-1$^*CPI_{t-1}$) - 1]/2. For fiscal year t+2, the sales growth rate expectation SG_{t+1} is defined as [(median analyst sales forecast for fiscal year t+2 $^*CPI_{t-4}$ / actual sales in fiscal year t-1$^*CPI_{t-1}$) - 1]/3. We use the number of quarters that actually have to be predicted as weight. In the first quarter of every year, the weighted earnings growth expectation is defined as $4/12^*SG_t + 4/12^*SG_{t+1} + 4/12^*SG_{t+2}$. For the second quarter, it is $3/11^*SG_t + 4/11^*SG_{t+1} + 4/11^*SG_t$, $2/10^*SG_t + 4/10^*SG_{t+1} + 4/10^*SG_t$ for the third quarter and $1/9^*SG_t + 4/9^*SG_{t+1} + 4/9^*SG_t$ for the fourth quarter. For China, we obtain the analyst forecast data from CSMAR and supplement it using analyst data from Suntime. Suntime have more forecast records for a given firm than CSMAR, but it has a shorter sample started from 2006. For sample 2003-2005, we use CSMAR data. For sample 2006-2018, we use Suntime data. U.S. analyst data is collected from I/B/E/S. Frequency: Quarterly. Source: CSMAR, Suntime and I/B/E/S.
Earnings growth expectation (Sector/portfolio level, 03-18)	We use analysts' earnings forecasts to calculate earnings growth expectation. Earnings growth expectation is calculated as the weighted average of annualized earnings growth rate expectation in the next 3 years using the most recent comparable CPI growth rates as deflators. Specifically, in each quarter, we first sum up the median firm-level earnings forecasts with setting those negative values to zeros by sectors. Then, for most nearby fiscal year t, the real earnings growth rate expectation EG_t is calculated as (median analyst earnings forecast for fiscal year t * CPI_{t-2}) / (actual earnings in fiscal year t-1* CPI_{t-1}) - 1. For fiscal year t+1, the earnings growth rate expectation EG_{t+1} is defined as [(median analyst earnings forecast for fiscal year t+1$^*CPI_{t-3}$/actual earnings in fiscal year t-1$^*CPI_{t-1}$) - 1]/2. For fiscal year t+2, the earnings growth rate expectation EG_{t+1} is defined as [(median analyst sales forecast for fiscal year t+2 $^*CPI_{t-4}$ / actual earnings in fiscal year t-1$^*CPI_{t-1}$) - 1]/3. We use the number of quarters that actually have to be predicted as weight. In the first quarter of every year, the weighted earnings growth expectation is defined as $4/12^*EG_t + 4/12^*EG_{t+1} + 4/12^*EG_{t+2}$. For the second quarter, it is $3/11^*EG_t + 4/11^*EG_{t+1} + 4/11^*EG_t$, $2/10^*EG_t + 4/10^*EG_{t+1} + 4/10^*EG_t$ for the third quarter and $1/9^*EG_t + 4/9^*EG_{t+1} + 4/9^*EG_t$ for the fourth quarter. For China, we obtain the analyst forecast data from CSMAR and supplement it using analyst data from Suntime. Suntime have more forecast records for a given firm than CSMAR, but it has a shorter sample started from 2006. For sample 2003-2005, we use CSMAR data. For sample 2006-2018, we use Suntime data. U.S. analyst data is collected from I/B/E/S. Frequency: Quarterly. Source: CSMAR, Suntime and I/B/E/S.
Number of analysts (Sector/portfolio level, 03-18)	Number of analysts that reported forecasts for a given firm in each quarter. We take the value-weighted average number of analysts across all firms in each sector to obtain the sector-level measure. For China, we obtain the analyst forecast data from CSMAR and supplement it using analyst data from Suntime. Suntime have more forecast records for a given firm than CSMAR, but it has a shorter sample started from 2006. For sample 2003-2005, we use CSMAR data. For sample 2006-2018, we use Suntime data. U.S. analyst data is collected from I/B/E/S. Frequency: Quarterly. Source: CSMAR, Suntime and I/B/E/S.

Forecast dispersion (Sector/portfolio level, 03-18)	We calculate this measure as the standard deviation of reported EPS forecast for Fiscal year 1 (forecast period indictor, FPI=1) in each quarter, standardized by the absolute value of the average forecast across analysts for a given firm in each quarter. We take the value-weighted forecast dispersion across all firms in each sector to obtain the sector-level measure. We winsorize this variable at the 1 and 99 percentiles. For China, we obtain the analyst forecast data from CSMAR and supplement it using analyst data from Suntime. Suntime have more forecast records for a given firm than CSMAR, but it has a shorter sample started from 2006. For sample 2003-2005, we use CSMAR data. For sample 2006-2018, we use Suntime data. U.S. analyst data is collected from I/B/E/S. Frequency: Quarterly. Source: CSMAR, Suntime and I/B/E/S.

Financial Development

Number of public firms (Market level, 95-18)	The log of the number of publicly traded firms at the end of each quarter in a given country. Frequency: Quarterly. Source: World Bank
Adjusted market development (Market level, 95-18)	Let MC_t^{CN} be the stock market capitalization of China relative to GDP and MC_t^{US} be the stock market capitalization of the US relative to GDP. Let $MCRatio_t = MC_t^{CN}/MC_t^{US}$.Then, standardize $MCRatio_t$by subtracting its mean and divided by standard deviation over 1995Q1 to 2018Q4. For the next step, take one year past cumulative market return in China and divide by one-year cumulative market return in the US. This ratio is denoted as $RetRatio_t$. This variable measures recent trends in returns. Then, standardize $RetRatio_t$ by subtracting its mean and divided by standard deviation over 1995Q1 to 2018Q4. The Adjusted market development is the difference between standardized $MCRatio_t$ and standardized $RetRatio_t$. Frequency: Quarterly. Sources: WIND and CRSP.
REGDEV (Market level, 95-18)	Based on the major events listed in Table A1, this cumulative regulation dummy variable is constructed as follows: take the value of 0 from 1995Q1 to 2005Q1, the value of 1 from 2005Q2 to 2008Q3 (the Split-share Reform), the value of 1.5 from 2008Q4 to 2009Q4 (the announcement of the Margin Trading and Short-selling Program), the value of 2 from 2010Q4 to 2015Q3 (the official start of the Margin Trading and Short-selling Program), the value of 2.5 from 2015Q4 to 2018Q4 (The Standing Committee of the People's Congress authorize the central government to apply a registration-based initial public offering (IPO) system Frequency: Quarterly. Source: constructed by authors.
Industry concentration ratio (Sector/portfolio level, 95-18)	Industry concentration ratio is calculated by adding up the market share of the top four largest firms (in terms of market capitalization) within the sector. $\text{Industry concentration ratio} = \sum_{i=1}^{4} s_i$, where s_i is the market share of the ith largest firm in a given sector. Frequency: Quarterly.
Zeros (Sector/portfolio level, 95-18)	Following Bekaert, Harvey, and Lundblad (2007), we calculate zeros as the proportion of zero daily returns observed over the relevant quarter for each security. We obtain security-level daily return data from WIND (China) and CRSP (US). For each sector/portfolio in each quarter, we calculate the market capitalization-weighted (using the market cap from last quarter) proportion of zero daily returns across all firms. Frequency: Quarterly.
Turnover rate (Sector/portfolio level, 95-18)	We first calculate firm-level turnover rate as the ratio of market value traded to total tradable shares market capitalization in each quarter. For sector-level and market-level, we take the value-weighted average of all the firms in the sector and in the market. Frequency: Quarterly. Source: WIND, CRSP
MYY R^2 synchronicity	Following Morck, Yeung, and Yu (2000), the synchronicity measure is a quarterly value-weighted R^2 obtained from regressing each firm's daily returns

(Sector/portfolio level, 95-18)	on the local market portfolio return over each quarter. For China, the local market portfolio return is the value-weighted return for all A shares. The U.S. market portfolio is the value-weighted return of all CRSP stocks. Frequency: Quarterly. Sources: WIND and CRSP.
Idiosyncratic volatility (Sector/portfolio level, 95-18)	We obtain security level idiosyncratic volatility by calculating the standard deviation of the residuals after regressing daily stock returns on local market portfolio returns in each quarter, annualized by multiplying by $\sqrt{250}$. For sector-level and market-level, we take the value-weighted average of all the firms in the sector and in the market. Frequency: Quarterly.

Financial Openness

Real Interest rate (Market level, 95-18)	The difference between the real interest rate between China and the U.S. For the nominal interest rate in China, we use the 1-year institution and individual deposit rate, obtained from People's Bank of China. For US, we use the 1-year Treasury constant maturity Rate from FRED Economic Data. The real interest rate is calculated by subtracting inflation from the nominal interest rate. The inflation rate is calculated as the percentage change of quarterly CPI over the same quarter in the previous year. Inflation rate(t) = CPI(t)/CPI(t-4)-1. We obtained the quarterly CPI data from China National Bureau of Statistics and US Bureau of Labor Statistics. Frequency: Quarterly. Source: People's Bank of China, China National Bureau of Statistics, FRED and US Bureau of Labor Statistics.
International accessibility (Sector level, 95-18)	There are two IA variables at the firm level. Firm IA1 is calculated by adding up four dummy variables, Bshare, Hshare, ADR and CHconnect. Variable Bshare (Hshare, ADR) is equal to 1 if the stock has B shares (H shares, ADR) issued. Variable CHconnect is equal to 1 if the stock is included in the China-HK connect program. Firm IA1 takes the minimum value of 0, meaning the stocks have no international accessibility, and takes the maximum of 4, which indicates that the stocks have B shares, H shares, ADRs and are incorporate into the China-HK connect program. Firm IA2 is the ratio of market capitalization of B shares, H shares and ADRs to firm total market capitalization. There are three sector-level IA variables. Sector IA1 (IA2) is the weighted average of the firm-level IA1 (IA2) within the sector, using the firm market capitalization of last quarter as weight. Sector IA3 is the market share of firms with positive firm-level IA1 within the sector. Frequency: Quarterly. Source: calculated by author using data from WIND.
REGOPEN (Market level, 95-18)	Based on the major events listed in Table A1, this cumulative regulation dummy variable is constructed as follows: take the value of 0 from 1995Q1 to 2000Q4, the value of 1 from 2001Q1 to 2002Q3 (Bshares), the value of 1.5 from 2002Q4 to 2003Q2 (the announcement of QFII), the value of 2 from 2003Q3 to 2006Q1 (the first transaction by QFII), the value of 2.5 in 2006Q2 (the announcement of QDII), the value of 3 from 2006Q3 to 2011Q3 (market execution of QDII), the value of 4 from 2011Q4 to 2014Q1 (the announcement and market execution of RQFII), the value of 4.67 from 2014Q2 to 2014Q3 (the announcement and regulation execution of Shanghai-Hong Kong Connect), the value of 5 from 2014Q4 to 2016Q2 (the official start of Shanghai-Hong Kong Connect), the value of 5.67 in 2016Q3 (the announcement and regulation execution of Shenzhen-Hong Kong Connect), the value of 6 from 2016Q4 to 2017Q1 (the official start of Shenzhen-Hong Kong Connect), the value of 6.67 from 2017Q2 to 2018Q1 (the announcement of incorporating A share into MSCI index), and the value of 7 from 2018Q2-2018Q4 (233 stocks listed in A-share market was

	officially incorporated MSCI emerging markets index and MSCI All Country World Index). Frequency: Quarterly. Source: constructed by authors.
Overall political rating (Market level, 95-18)	The sum of all 12 ICRG subcomponents, with a total score of 100 and the maximum score for each subcomponent displayed in parenthesis: Government Stability (12), Socioeconomic Conditions (12), Investment Profile (12), Internal Conflict (12), External Conflict (12), Corruption (6), Military in Politics (6), Religious Tensions (6), Law and Order (6), Ethnic Tensions (6), Democratic Accountability (6), and Bureaucracy Quality (4). ICRG currently only provides data till 2018Q3. We fill in the 2018Q4 numbers making an assumption that they are equal to that in 2018Q3. Frequency: Annual. Source: ICRG.
Quality of institutions (Market level, 95-18)	The sum of ICRG subcomponents, with a maximum score of 28: Corruption, Law and Order, Bureaucratic Quality, and Investment Profile, following Bekaert, Harvey and Lundblad, 2005. ICRG currently only provides data till 2018Q3. We fill in the 2018Q4 numbers making an assumption that they are equal to that in 2018Q3 Frequency: Annual. Source: ICRG.
Investment profile (Market level, 95-18)	An ICRG measure of the factors affecting the risk to investment that are not covered by other political, economic and financial risk components. ICRG currently only provides data till 2018Q3. We fill in the 2018Q4 numbers making an assumption that they are equal to that in 2018Q3. Frequency: Annual. Source: ICRG.
A-B premium (Sector/portfolio level, 95-18)	For China stocks that have both A and B shares, A-B premium is calculated as the price of A share divided by the price of B share minus 1. B shares listed in Shanghai Stock Exchange are priced by US dollars, and B shares listed in Shenzhen exchange are priced by Hong Kong dollars. B share prices are converted into RMB prices using quarter-end exchange rates. Frequency: Quarterly. Source: WIND.
A-H premium (Sector/portfolio level, 95-18)	For China stocks that have both A and H shares, A-H premium is calculated as the price of A share divided by the price of H share minus 1. H share prices are converted into RMB prices using quarter-end exchange rates. Frequency: Quarterly. Source: WIND.

Ownership

Chinese state ownership (Sector/portfolio level, 95-18)	State ownership is measured as fraction of total shares that are owned by the state. It is calculated by three steps. First, from CSMAR, we collect the ten largest shareholder information (including their numbers of holding shares, the nature of the shares) and use this information to calculate a measure of state ownership of a given company. Second, since the financial statement will disclose how many shares are state-owned shares among the non-tradable shares, we use this information to calculate another version of state ownership by only taking the non-tradable shares into account. Then, we take the larger value of the first and second measure to proxy for the state ownership of a given firm. This variable is only available for China. Frequency: Quarterly. Source: CSMAR.
Institutional ownership (Sector/portfolio level, 03-18)	Institutional ownership is measured as fraction of tradable shares that are owned by institutions. For China's case, we use institutional holding data from WIND to calculate the institutional ownership. Institutions are defined as professional money managers like mutual fund holdings, insurance company holdings, banks, hedge fund, investment trust company, pension fund holdings and security fund holdings. For U.S.'s case, following Ferreira, Miguel and Matos (2008), we use Factset data to calculate institutional holdings. Specifically, Institutional holding is calculated as the market value of aggregate of 13f

	holdings and non-13f fund holding, divided by total market value. Frequency: Quarterly. Source: WIND, FactSet Lion Shares
Retail ownership (Sector/portfolio level, 03-18)	Retail ownership is measured as fraction of tradable shares that are owned by the retail investors. For China, retail investor ownership is defined as follows: 1 - institutional ownership - state ownership - insider ownership. State ownership here is the fraction of tradable shares that are owned by the state which should be separated from our Chinese state ownership measure mentioned above; For U.S., retail investor ownership is defined as follows:1 - institutional ownership - insider ownership. Insiders are defined as directors, supervisors or managers in a company, or large individual shareholders shown up in the firms' ten largest shareholders profile. Insider information for China and U.S. are extracted from CSMAR and Thomson Reuters respectively. Frequency: Quarterly. Source: WIND, FactSet Lion Shares, CSMAR, Thomson Reuters
Turnover rate (Sector/portfolio level, 95-18)	We first calculate firm-level turnover rate as the ratio of market value traded to total tradable shares market capitalization in each quarter. For sector-level and market-level, we take the value-weighted average of all the firms in the sector and in the market. Frequency: Quarterly. Source: WIND, CRSP
Chinese standardized number of shareholders (Sector/portfolio level, 03-18)	Standardized number of shareholders is only available for China. It is calculated as number of shareholders divided by total tradable shares and multiplied by 1000. Frequency: Quarterly. Source: WIND

Appendix B. Portfolio formations

We split the whole sample of firms with available quarterly earnings into portfolios based on their state ownership (SO), international accessibility (IA, whether a firm has B share, H share or ADR), zeros (Illiquidity), market capitalization (Size), attribute of industries (Tech, whether a firm belongs to TMT industry) and listed boards (Board)

1 State ownership, International accessibility and Listing board portfolios

We formed 4 state ownership portfolios based on firm-level state ownership (0, 0-10%, 10%-50% and >50%), 2 portfolios based on firm-level IA1 (0, and >0) and 3 listing board portfolio based on the board in which they are listed. Since these three variables are only available for Chinese firms, we only form the portfolios for China. After the SO portfolios, IA portfolios and listing board portfolios are formed, within each portfolio, portfolio-level variables for China are generated. The calculation procedure for China is as follows: among all the variables, portfolio-level leverage, IA1, IA2, R-squared, idiosyncratic volatility, zeros, state ownership, A-B premium, A-H premium, number of analyst, SUE, forecast dispersion, retail ownership, turnover rate, standardized number of shareholders and institutional ownership are calculated as the weighted average of the corresponding firm-level variables, using the lagged firm market value as weights. Portfolio-level industry concentration ratio is calculated as the weighted sum of the corresponding sector-level industry concentration ratio, using the market share of the sector in the portfolio as weights. For other variables (earnings yield, earnings growth volatility, minimum number of stocks, IA3, earnings growth expectation and sales growth expectation), we calculate their portfolio-level measures directly in the same way that we applied for sectors using individual firm data.

Next, we match each China portfolio with a U.S. portfolio benchmark, which has the same sector composition. Specifically, within each portfolio in China, for each sector j, we sum up the market capitalization of firm i and get the sector-level weight $VW_{j,t}^{CN} = \sum_i VW_{i,j,t}^{CN}$. For the next step, we use the sector level weight to form the U.S. benchmark as $EY_t^{US} = \sum_j VW_{j,t}^{CN} EY_{j,t}^{US}$. We carry out this procedure for all the variables for the benchmark U.S. portfolios.

2 Illiquidity portfolios, Size, Turnover, Retail, IO and Tech portfolios

We formed 2 illiquidity portfolios based on quarterly firm-level zeros: ≤ Lower country 30% percentile of zeros and >= Higher country 30% percentile of zeros, 2 size portfolio based on quarterly firm-level market value: ≤ Lower country 30% percentile of market value and >= Higher country 30% percentile of market value, 2 turnover portfolio based on quarterly firm-level turnover: ≤ Lower country 30% percentile of turnover rate and >= Higher country 30% percentile of turnover rate, 2 retail portfolio based on quarterly firm-level retail ownership: ≤ Lower country 30% percentile of retail ownership and >= Higher country 30% percentile of retail ownership, 2 IO portfolio based on quarterly institutional ownership: ≤ Lower country 30% percentile of institutional ownership and >= Higher country 30% percentile of institutional ownership, 2 tech portfolio based on whether firms are in high-tech industry. High-tech industry is defined as "TMT" industry, including "Fixed and Mobile Telecom", "Media", "Software & Computer Services", "Technology Hardware & Equipment" 4 sectors. Since both China and the U.S. have these classifications, we can form these portfolios for both China and US. After these portfolios are formed, within each portfolio, we calculate portfolio-level variables for China and the U.S.

Portfolio-level leverage, IA1, IA2, R-squared, idiosyncratic volatility, zeros, state ownership, A-B premium, A-H premium, number of analyst, SUE, forecast dispersion, retail ownership, turnover rate, standardized number of shareholders and institutional ownership are calculated as the weighted average of the corresponding firm-level variables, using the lagged firm market value as weights. Portfolio-level industry concentration ratio is calculated as the weighted sum of the corresponding sector-level industry concentration ratio, using the market share of the sector in the portfolio as weights. For other variables (earnings yield, earnings growth volatility, minimum number of stocks, IA3, earnings growth expectation and sales growth expectation), we calculate their portfolio-level measures directly in the same way that we applied for sectors using individual firm data.

Next, we match each China portfolio with a U.S. portfolio benchmark, which has the same "within-portfolio" sector composition. Specifically, within each portfolio in China, for each sector j, we sum up the market capitalization of firm i and get the sector-level weight $VW_{j,t}^{CN} = \sum_i VW_{i,j,t}^{CN}$. For the next step, we use the sector level weight to form the U.S. benchmark as $EY_t^{US} = \sum_j VW_{j,t}^{CN} EY_{j,t}^{US}$, where $EY_{j,t}^{US}$ is the "within-portfolio" sector-level earnings yield. We carry out this procedure for all the variables for the benchmark U.S. portfolios.

Appendix C. Bai, Lumsdaine and Stock (1998) structural break test

Considering the following specification[1]:

$$y_t = (G_t' \otimes I_n)\theta + d_t(k)(G_t' \otimes I_n)S'S\delta + \varepsilon_t \tag{1}$$

Where y_t is n×1, G_t' is a row vector containing a constant, lags of y_t, and row t of the matrix of exogenous regressors X, I_n is a n×n identity matrix. $d_t(k) = 0$ for $t < k$ and $d_t(k) = 1$ for $t \geq k$, and $\sum$ is the covariance matrix of error term ε_t. θ and δ are parameter vectors with dimension r. S is a selection matrix containing zero and ones. It is used to identify (via the placement of the ones) which of the r parameters are allowed to change in the regression. For our case, we consider two specifications $S = s \otimes I_n, s = (1,0, \ldots 0)$ when only the intercept is allowed to break and $S = I_r$ (all parameters break).

We can write the system more compactly as

$$y_t = Z_t'(k)\beta + \varepsilon_t \tag{2}$$

Where $Z_t'(k) = (\left((G_t' \otimes I_n), d_t(k)(G_t' \otimes I_n)S'\right)$ and $\beta = (\theta', (S\delta)')'$. If we let $R = (0, I)$ be the selection matrix associated with β, then $R\beta = S\delta$ and the F-statistic testing $S\delta = 0$ is

$$\hat{F}_T(k) = T\{R\hat{\beta}(k)\}' \left\{R\left(T^{-1}\sum_{t=1}^{T} Z_t \hat{\Sigma}_k^{-1} Z_t'\right)^{-1} R'\right\}^{-1} \{R\hat{\beta}(k)\} \tag{3}$$

Where $\hat{\beta}(k)$ and $\hat{\sum}_k$ denote the estimators of β and $\sum$, respectively, evaluated at $\hat{k}$. We here focus on max $\hat{F}_T(k)$ which is our sup-Wald statistics.

Bai, Lumsdaine and Stock (1998) show that the confidence interval is as follows:

$$\hat{k} \pm \alpha_{\frac{\pi}{2}}[(S\hat{\delta}_T)'S(\hat{Q} \otimes \hat{\Sigma}_k^{-1})S'(S\hat{\delta}_T)^{-1}]$$

Where $\hat{Q} = (1/T)\sum_{t=1}^{T} G_t G_t'$

[1] These notations are largely based on Bai, Lumsdaine and Stock (1998), Bekaert, Harvey and Lumsdaine (2001). See their papers for more details.

Appendix D. Major events related to stock market development in China

Category	Date	Description	Key words
Policy	2001.02.21	Citizens in mainland China were permitted to invest in B shares.	B shares
Policy	2002.11.05	People's Bank of China (PBOC) and China Securities Regulatory Commission (CSRC) jointly published "the Administration of Domestic Securities Investments of Qualified Foreign Institutional Investors (QFII) (Trial)", indicating the official start of QFII.	QFII
Market	2003.07.09	The first investment of QFII was generated by UBS.	QFII
Policy	2005.04.29	CSRC announced the official start of the trial run of Split-Share Structure Reform.	Split-Share
Policy	2006.04.13	People’s Bank of China announced for the first time that qualified funds and other fund-raising institutions can trade in stocks, bonds and funds and other securities outside of China, indicating the official start of QDII.	QDII
Market	2006.08	The first trial QDII fund, Hua’an International Fund, was established by Hua’an Fund.	QDII
Policy	2008.10.05	CSRC announced that the program of dual margin trading and short selling in stock market would start at some point in the future.	Margin Trading and Short Selling
Market	2010.03.31	It was the first day of margin trading and short selling. Hundreds of transactions went through and the total trading value of margin trading and short selling was about 6.59 million RMB.	Margin Trading and Short Selling
Policy	2011.12.16	CSRC announced "Administration of Domestic Securities Investment by Fund Management Companies and Securities Companies as RMB Qualified Foreign Institutional Investors (RQFII) (Trial)"	RQFII
Market	2014.04.10	Shanghai-Hong Kong Stock Connect was announced to be started in the future.	Shanghai-Hong Kong Connect
Policy	2014.06.13	"Regulations of Shanghai-Hong Kong Stock Connect" was published and executed.	Shanghai -Hong Kong Connect
Policy	2014.11.17	Shanghai-Hong Kong Stock Connect officially started.	Shanghai -Hong Kong Connect
Market	2015.12.27	The Standing Committee of the People's Congress authorize the central government to apply a registration-based initial public offering (IPO) system.	Registration-based IPO
Market	2016.08.16	Shenzhen-Hong Kong Stock Connect was announced to be started in the future	Shenzhen -Hong Kong Connect

Category	Date	Description	Key words
Policy	2016.08.26	"Regulations of China mainland-Hong Kong Stock Connect" was published and executed.	Shenzhen -Hong Kong Connect
Policy	2016.12.05	Shenzhen-Hong Kong Stock Connect officially started.	Shenzhen -Hong Kong Connect
Market	2017.06.21	A-share was announced to be incorporated into MSCI index in June, 2018	MSCI Index
Market	2018.06.01	233 stocks listed in A-share market was officially incorporated MSCI emerging markets index and MSCI All Country World Index	MSCI Index

Appendix E. Model selection: general-to-specific search algorithm (PcGets procedure)

Steps	Significance Levels
1 Eliminating collinear variables	0.800
Going variable by variable, if multiple variables are correlated more highly than the cut-off value, we select the one variable that features the highest absolute t statistics in the univariate regression and drop other variables.	
2 Estimate a general model with all variables (M1)	
a. Test significance of individual coefficient estimates: t-test. If all coefficients are individually significant, M1 is the final model.	0.025
b. Test M1 against the null of "all coefficients are zero" and the null of "all coefficients but intercept are zero": F-test. If the null is not rejected, M1 is the final model.	0.500
3 Pre-search tests	
a. Top-down tests.	
Rank the p-values of all coefficients in M1 from largest to smallest. Test joint significance of expanding list of coefficient estimates from largest p-value (least significant) to smallest p-value (most significant): F-test. If F-test is not rejected, remove variables on the current list. (M2)	0.500
b. Repeat top-down tests.	
Estimate M2 and rank the p-values of all coefficients from largest to smallest. Test joint significance of expanding list of coefficient estimates from largest p-value (least significant) to smallest p-value (most significant): F-test. If F-test is not rejected, remove variables on the current list. (M3)	0.250
c. Bottom-up tests	
Rank the p-values of all coefficients in M3 from smallest to largest. Test joint significance of decreasing list of coefficient estimates from smallest p-value (most significant) to largest p-value (least significant): F-test. If F-test is not rejected, remove variables on the current list. (M4)	0.025
4 Multiple-path tests	
a. Estimate M4. If all estimates are individually significant, M4 is the final model.	0.025
b. Otherwise, initiate search paths.	
Remove blocks of variables with increasing p-values of t-statistics and reestimate the model: remove one insignificant variables each time until all insignificant variables are removed and commenced a path.	
c. Repeat step 3b as long as insignificant variables survives.	0.025
d. The algorithm arrives to a terminal model if all coefficients are individually significant: t-test (M5)	0.025

Appendix F. Bayesian adjustment

We estimate EY beta and Earnings beta as follows:

$$EY_{it} = \alpha_0 + \hat{\beta}_i^{ey_{ols}} g_t + \varepsilon_{it} \quad (1)$$

$$G_{earnings_{it}} = \alpha_0 + \hat{\beta}_i^{ea_{ols}} g_t + \varepsilon_{it} \quad (2)$$

Where EY_{it} is the earnings yield of industry i at quarter t, $G\_earnings_{it}$ is the earnings growth rate of industry i at quarter t, g_t is the GDP growth rate at quarter t which is calculated as [GDP(t) + GDP(t-1) + GDP(t-2) + GDP(t-3)] / [GDP(t-4) + GDP(t-5) + GDP(t-6) + GDP(t-7)] - 1, $\hat{\beta}_i^{ey}$ is the EY beta, $\hat{\beta}_i^{ey}$ is the earnings beta.

We apply the following Bayesian method to adjust the EY beta and earnings beta.

For the EY beta, we define

$$\hat{\beta}_i^{ey} = w_i * \hat{\beta}_i^{ey_{ols}} + (1 + w_i) * \hat{\beta}_i^{ey_{prior}} \quad (3)$$

Where $\hat{\beta}_i^{ey}$ is the Bayesian adjusted EY beta, $\hat{\beta}_i^{ey_{ols}}$ is the EY beta estimated directly using OLS, $\hat{\beta}_i^{ey_{prior}}$ is our prior on EY beta (we set it to be zero),

$$w_i = \left[{v_i^{ols}}^{-1} + \Omega^{-1}\right]^{-1} * \frac{1}{v_i^{ols}} \quad (4)$$

Where v_i^{ols} is the variance of $\hat{\beta}_i^{ey_{ols}}$, Ω is the variance around the $\hat{\beta}_i^{ey_{prior}}$.

1) Rank the industry/portfolio in terms of R-square
2) Take the median R-square sector/portfolio, verify that $\hat{\beta}_i^{ey_{ols}} < 0$ and its associated absolute t-statistic larger than 1.65. If not, move upward in terms of R-square until the criterion is satisfied.
3) Set $\Omega = 9v_i^{ols}$ for the industry/portfolio selected in 2) so that we can calculated the weight for every industry/portfolio. This implies that the weight for the portfolio in 2) is 90%, with noisier $\hat{\beta}_i^{ey}$ getting shrunk more.

For the Earnings beta, we define

$$\hat{\beta}_i^{ea} = w_i * \hat{\beta}_i^{ea_{ols}} + (1 + w_i) * \hat{\beta}_i^{ea_{prior}} \quad (5)$$

Where $\hat{\beta}_i^{ea}$ is the Bayesian adjusted earnings beta, $\hat{\beta}_i^{ea_{ols}}$ is the earnings beta estimated directly using OLS, $\hat{\beta}_i^{ea_{prior}}$ is our prior on earnings beta (we set it to be zero),

$$w_i = \left[{v_i^{ols}}^{-1} + \Omega^{-1}\right]^{-1} * \frac{1}{v_i^{ols}} \quad (6)$$

Where v_i^{ols} is the variance of $\hat{\beta}_i^{ea_{ols}}$, Ω is the variance around the $\hat{\beta}_i^{ea_{prior}}$.

1) Rank the industry/portfolio in terms of R-square
2) Take the median R-square sector/portfolio, verify that $\hat{\beta}_i^{ea_{ols}} > 0$ and its associated absolute t-statistics larger than 1.65. If not, move upward in terms of R-square until the criteria is satisfied.
3) Set $\Omega = 9v_i^{ols}$ for the industries/portfolios selected in 2) so that we can calculate the weight for every industry/portfolio.

Online Appendix

Table OA1. Summary statistics of control variables

This table reports the time-series summary statistics of the control variables of China and the U.S. from 1995Q1 to 2018Q4. The control variables include leverage differentials, earnings growth volatility differentials, and the minimum of the number of firms in China and the U.S. Definitions of all the variables are described in detail in Appendix A. The panel includes 33 sectors and 21 additional portfolios formed on state ownership, international accessibility, illiquidity, market size, retail ownership, institutional ownership, turnover rate, technology sectors and listing boards.

	China			U.S.		
	Leverage	Earnings growth volatility	Number of Firms	Leverage	Earnings growth volatility	Number of Firms
Industrial sectors						
Aerospace & Defense	0.170	0.213	9	0.221	0.244	62
Alternative Energy	0.214	0.523	6	0.137	0.903	12
Automobiles & Parts	0.187	0.485	59	0.325	0.735	41
Banks & Life Insurance	0.922	0.166	10	0.906	0.191	536
Beverages	0.093	0.197	24	0.307	0.135	23
Chemicals	0.255	0.366	127	0.286	0.356	83
Construction & Materials	0.256	0.230	79	0.249	0.258	79
Electricity	0.353	0.317	44	0.383	0.127	57
Electronic & Electrical Equipment	0.224	0.285	115	0.196	0.349	190
Financial Services	0.625	0.533	14	0.788	0.201	142
Fixed and Mobile Telecom	0.098	0.585	3	0.335	0.549	58
Food & Drug Retailers	0.224	0.311	7	0.216	0.106	37
Food Producers	0.240	0.150	57	0.325	0.127	84
Forestry & Paper	0.327	0.369	17	0.359	1.177	16
Gas, Water and Multiutilities	0.264	0.262	14	0.344	0.151	51
General Industrials	0.272	0.285	19	0.407	0.149	48
General Retailers	0.205	0.195	65	0.227	0.110	205
Health Care Equipment & Services	0.130	0.158	10	0.233	0.156	261
Household Goods & Home Construction	0.109	0.134	26	0.300	0.186	101
Industrial Engineering	0.195	0.228	114	0.316	0.297	129
Industrial Metals & Mining	0.264	0.683	68	0.262	0.621	35
Industrial Transportation	0.199	0.243	43	0.248	0.160	67
Leisure Goods	0.196	0.372	19	0.165	0.250	49
Media	0.190	0.479	18	0.293	0.289	125
Mining	0.180	0.380	28	0.259	0.554	32
Oil Equipment, Services & Oil and Gas Producers	0.201	0.423	12	0.184	0.435	178
Personal Goods	0.234	0.258	55	0.258	0.140	80
Pharmaceuticals & Biotechnology	0.191	0.144	92	0.207	0.137	232
Real Estate Investment & Services	0.274	0.146	101	0.275	0.589	30

	China			U.S.		
	Leverage	Earnings growth volatility	Number of Firms	Leverage	Earnings growth volatility	Number of Firms
Software & Computer Services	0.152	0.496	39	0.121	0.148	279
Support Services	0.258	0.254	31	0.208	0.127	212
Technology Hardware & Equipment	0.190	0.225	47	0.115	0.341	280
Travel & Leisure	0.260	0.424	31	0.354	0.207	162
Other portfolios						
State-ownership Portfolio 1 (SO=0)	0.230	0.168	344	0.289	0.272	3008
State-ownership Portfolio 2 (0<SO<=10%)	0.308	0.254	215	0.337	0.315	2987
State-ownership Portfolio 3 (10%<SO<=50%)	0.359	0.163	480	0.377	0.308	3757
State-ownership Portfolio 4 (SO>50%)	0.349	0.140	360	0.375	0.339	3583
Retail Portfolio 1(Retail ownership <= country level lower 30%)	0.414	0.119	535	0.435	0.421	1038
Retail Portfolio 2(Retail ownership >= country level upper 30%)	0.269	0.178	535	0.268	0.634	798
IO Portfolio 1(IO <= country level lower 30%)	0.334	0.187	559	0.322	0.595	844
IO Portfolio 2(IO >=country level upper 30%)	0.381	0.114	535	0.420	0.392	1032
International Accessibility Portfolio 1 (IA1=0)	0.250	0.159	1135	0.303	0.316	3920
International Accessibility Portfolio 2 (IA1>0)	0.387	0.168	264	0.393	0.338	2905
Illiquidity Portfolio 1(Zeros <= country level lower 30%)	0.278	0.163	622	0.317	0.378	1418
Illiquidity Portfolio 2(Zeros >= country level upper 30%)	0.377	0.150	549	0.408	0.658	1242
Turnover Portfolio 1(Turnover <= country level lower 30%)	0.362	0.119	420	0.375	0.400	1088
Turnover Portfolio 2(Turnover >= country level upper 30%)	0.246	0.247	420	0.283	0.566	1080
Size Portfolio 1(Market value <= country level lower 30%)	0.237	0.216	420	0.223	0.641	904
Size Portfolio 2(Market value >= country level upper 30%)	0.348	0.137	420	0.379	0.352	1088
Tech portfolio 1(Not TMT industry)	0.337	0.140	1292	0.368	0.323	3193
Tech portfolio 2(TMT industry)	0.166	0.283	107	0.170	0.314	742
Listing board portfolio 1(Main board)	0.348	0.136	1048	0.371	0.326	3934
Listing board portfolio 2(SME board)	0.168	0.151	422	0.265	0.315	2707
Listing board portfolio 3(GEM board)	0.103	0.087	321	0.264	0.244	2405
Market	0.329	0.135	1400	0.324	0.235	3976

Table OA2. Summary statistics of macroeconomic variables

This table reports the time-series summary statistics of macroeconomic variables of China and the U.S. from 1995Q1 to 2018Q4. Definitions of all the variables are described in detail in Appendix A. The mean differences between China and the U.S. are reported, with significances marked by stars. ***, ** and * indicate significances at the 1%, 5% and 10% levels using two-tailed tests.

	China		U.S.		Difference
	Mean	Std.Dev	Mean	Std.Dev	(China-U.S.)
Growth Expectations					
GDP growth rate (%)	8.83	1.65	2.51	1.58	6.32***
Financial Development					
Number of public firms	1759	896	5438	1325	-3679***
Adjusted market development	0.00	1.34			
REGDEV	1.04	1.00			
Financial Openness					
Real interest rate(%)	0.51	2.66	0.36	1.95	0.14
REGOPEN	2.59	2.09			
Overall political rating	64.21	4.79	82.81	3.08	-18.6***
Quality of institutions	15.33	1.08	24.67	1.03	-9.34***
Investment profile	7.03	0.88	11.15	1.45	-4.12***

Table OA3. Summary statistics of growth expectations measures

This table reports summary statistics related to sector/portfolio level growth expectations. Due to availability of analyst forecast data, earnings growth expectation, sales growth expectation, number of analyst and forecast dispersion are only available after 2003Q1. Definitions of all these variables are described in detail in Appendix A. The panel includes 33 sectors and 21 additional portfolios formed on state ownership, international accessibility, illiquidity, market size, retail ownership, institutional ownership, turnover rate, technology sectors and listing boards.

	China				U.S.			
	Earning growth expectation	Sales growth expectation	Numbers of analysts	Forecast dispersion	Earning growth expectation	Sales growth expectation	Numbers of analysts	Forecast dispersion
Industrial sectors								
Aerospace & Defense	0.408	0.237	6	0.164	0.075	0.029	15	0.036
Alternative Energy	0.867	0.329	6	0.167	0.122	0.318	14	0.394
Automobiles & Parts	0.303	0.160	10	0.119	0.177	0.001	13	0.172
Banks & Life Insurance	0.230	0.206	15	0.061	0.158	0.038	19	0.105
Beverages	0.275	0.151	16	0.080	0.042	0.022	13	0.015
Chemicals	0.420	0.159	4	0.177	0.127	0.033	12	0.057
Construction & Materials	0.333	0.178	7	0.099	0.124	0.045	10	0.108
Electricity	0.235	0.111	7	0.142	0.038	0.014	9	0.050
Electronic & Electrical Equipment	0.411	0.267	6	0.140	0.126	0.038	10	0.070
Financial Services	0.336	0.178	11	0.162	0.100	0.057	17	0.078
Fixed and Mobile Telecom	0.652	0.092	13	0.162	0.048	0.019	25	0.118
Food & Drug Retailers	0.266	0.178	6	0.097	0.081	0.050	15	0.031
Food Producers	0.360	0.193	8	0.150	0.056	0.016	11	0.041
Forestry & Paper	0.332	0.164	4	0.160	0.161	-0.002	12	0.215
Gas, Water and Multiutilities	0.159	0.150	3	0.090	0.027	0.019	8	0.031
General Industrials	0.288	0.185	5	0.131	0.053	0.018	12	0.026
General Retailers	0.289	0.169	7	0.147	0.112	0.066	22	0.095
Health Care Equipment & Services	0.311	0.255	7	0.071	0.098	0.072	14	0.052
Household Goods & Home Construction	0.237	0.168	11	0.074	0.054	0.031	13	0.051
Industrial Engineering	0.349	0.200	7	0.144	0.118	0.034	13	0.059
Industrial Metals & Mining	0.389	0.121	7	0.235	0.368	0.052	13	0.259
Industrial Transportation	0.158	0.111	7	0.122	0.102	0.050	18	0.034
Leisure Goods	0.346	0.151	5	0.201	0.122	0.045	14	0.139
Media	0.260	0.168	8	0.136	0.104	0.037	16	0.142
Mining	0.198	0.093	10	0.147	0.288	0.064	14	0.332
Oil Equipment, Services & Oil and Gas Producers	0.168	0.059	11	0.134	0.122	0.007	21	0.177
Personal Goods	0.221	0.163	5	0.102	0.056	0.036	14	0.030
Pharmaceuticals & Biotechnology	0.257	0.170	6	0.090	0.052	0.040	17	0.063

	China				U.S.			
	Earning growth expectation	Sales growth expectation	Numbers of analysts	Forecast dispersion	Earning growth expectation	Sales growth expectation	Numbers of analysts	Forecast dispersion
Real Estate Investment & Services	0.341	0.275	7	0.118	0.231	0.085	5	0.337
Software & Computer Services	0.384	0.239	5	0.118	0.102	0.065	25	0.055
Support Services	0.340	0.139	4	0.113	0.095	0.047	11	0.052
Technology Hardware & Equipment	0.439	0.231	7	0.128	0.162	0.054	29	0.099
Travel & Leisure	0.475	0.121	9	0.164	0.112	0.045	17	0.099
Other portfolios								
State-own Portfolio 1 (SO=0)	0.336	0.238	7	0.117	0.115	0.042	14	0.098
State-own Portfolio 2 (0<SO<=10%)	0.364	0.216	8	0.132	0.141	0.045	15	0.116
State-own Portfolio 3 (10%<SO<=50%)	0.241	0.167	9	0.120	0.141	0.043	15	0.117
State-own Portfolio 4 (SO>50%)	0.158	0.105	11	0.102	0.158	0.034	16	0.130
Retail Portfolio 1(Retail ownership <= country level lower 30%)	0.187	0.125	12	0.103	0.169	0.045	14	0.141
Retail Portfolio 2(Retail ownership >= country level upper 30%)	0.267	0.155	5	0.142	0.466	0.158	4	0.262
IO Portfolio 1(IO <= country level lower 30%)	0.359	0.119	5	0.151	1.509	0.145	5	0.284
IO Portfolio 2(IO >= country level upper 30%)	0.219	0.161	12	0.096	0.163	0.050	13	0.123
International Accessibility Portfolio 1 (IA1=0)	0.320	0.178	7	0.130	0.143	0.043	14	0.114
International Accessibility Portfolio 2 (IA1>0)	0.169	0.109	12	0.096	0.147	0.032	17	0.128
Illiquidity Portfolio 1(Zeros <= country level lower 30%)	0.248	0.153	10	0.114	0.146	0.048	15	0.111
Illiquidity Portfolio 2(Zeros >= country level upper 30%)	0.182	0.131	10	0.102	0.467	0.047	9	0.255
Turnover Portfolio 1(Turnover <= country level lower 30%)	0.172	0.122	12	0.101	0.207	0.052	6	0.178
Turnover Portfolio 2(Turnover >= country level upper 30%)	0.373	0.191	5	0.155	0.246	0.067	14	0.184
Size Portfolio 1(Market value <= country level lower 30%)	0.487	0.198	1	0.168	0.649	0.120	1	0.387
Size Portfolio 2(Market value >= country level upper 30%)	0.185	0.127	12	0.103	0.153	0.037	17	0.118
Tech portfolio 1(Not TMT industry)	0.195	0.132	10	0.107	0.151	0.038	15	0.123
Tech portfolio 2(TMT industry)	0.338	0.174	8	0.126	0.123	0.051	26	0.099
Listing board portfolio 1(Main board)	0.188	0.125	10	0.107	0.152	0.038	16	0.126
Listing board portfolio 2(SME board)	0.350	0.265	8	0.117	0.094	0.039	15	0.090
Listing board portfolio 3(GEM board)	0.464	0.408	8	0.121	0.080	0.043	18	0.070
Market	0.197	0.132	10	0.108	0.101	0.042	18	0.083

Table OA4. Summary statistics of financial development measures

This table reports summary statistics related to sector/portfolio level financial development from 1995Q1 to 2018Q4. Definitions of all these variables are described in detail in Appendix A. The panel includes 33 sectors and 21 additional portfolios formed on state ownership, international accessibility, illiquidity, market size, retail ownership, institutional ownership, turnover rate, technology sectors and listing boards.

	China				U.S.			
	Zeros	MYY R^2	Idiosyncratic volatility	Industry concentration ratio	Zeros	MYY R^2	Idiosyncratic volatility	Industry concentration ratio
Industrial sectors								
Aerospace & Defense	0.033	0.370	0.350	0.741	0.018	0.313	0.224	0.598
Alternative Energy	0.036	0.373	0.362	0.859	0.042	0.167	0.588	0.826
Automobiles & Parts	0.040	0.397	0.326	0.413	0.028	0.298	0.291	0.718
Banks & Life Insurance	0.063	0.436	0.230	0.810	0.026	0.417	0.231	0.437
Beverages	0.029	0.352	0.303	0.645	0.019	0.238	0.186	0.867
Chemicals	0.045	0.410	0.328	0.309	0.022	0.351	0.235	0.477
Construction & Materials	0.045	0.418	0.320	0.339	0.033	0.291	0.283	0.341
Electricity	0.055	0.419	0.273	0.508	0.041	0.197	0.195	0.292
Electronic & Electrical Equipment	0.039	0.388	0.358	0.257	0.025	0.349	0.314	0.360
Financial Services	0.040	0.473	0.321	0.701	0.018	0.416	0.257	0.371
Fixed and Mobile Telecom	0.055	0.449	0.285	0.989	0.025	0.262	0.246	0.734
Food & Drug Retailers	0.038	0.387	0.323	0.808	0.026	0.217	0.252	0.610
Food Producers	0.040	0.371	0.332	0.369	0.029	0.204	0.225	0.405
Forestry & Paper	0.046	0.426	0.310	0.538	0.022	0.326	0.269	0.843
Gas, Water and Multiutilities	0.047	0.440	0.305	0.641	0.038	0.218	0.194	0.474
General Industrials	0.038	0.417	0.321	0.531	0.020	0.453	0.186	0.832
General Retailers	0.042	0.400	0.324	0.260	0.023	0.267	0.278	0.500
Health Care Equipment & Services	0.036	0.338	0.356	0.799	0.024	0.215	0.298	0.284
Household Goods & Home Construction	0.043	0.402	0.305	0.598	0.023	0.255	0.223	0.650
Industrial Engineering	0.040	0.417	0.328	0.263	0.025	0.361	0.257	0.455
Industrial Metals & Mining	0.057	0.453	0.305	0.402	0.030	0.294	0.342	0.607
Industrial Transportation	0.052	0.435	0.278	0.469	0.024	0.332	0.239	0.590
Leisure Goods	0.044	0.399	0.332	0.578	0.032	0.233	0.336	0.618
Media	0.039	0.366	0.368	0.571	0.024	0.305	0.265	0.481
Mining	0.035	0.444	0.314	0.629	0.033	0.151	0.416	0.665
Oil Equipment, Services & Oil and Gas Producers	0.055	0.454	0.251	0.935	0.019	0.309	0.241	0.539
Personal Goods	0.042	0.401	0.325	0.300	0.021	0.226	0.250	0.653
Pharmaceuticals & Biotechnology	0.036	0.361	0.325	0.214	0.017	0.263	0.254	0.492
Real Estate Investment & Services	0.043	0.410	0.334	0.275	0.057	0.239	0.338	0.677

	China				U.S.			
	Zeros	MYY R^2	Idiosyncratic volatility	Industry concentration ratio	Zeros	MYY R^2	Idiosyncratic volatility	Industry concentration ratio
Software & Computer Services	0.034	0.364	0.383	0.421	0.015	0.350	0.290	0.655
Support Services	0.038	0.412	0.336	0.459	0.031	0.285	0.290	0.305
Technology Hardware & Equipment	0.034	0.399	0.354	0.424	0.015	0.339	0.332	0.509
Travel & Leisure	0.046	0.413	0.317	0.563	0.027	0.246	0.298	0.408
Other portfolios								
State-own Portfolio 1 (SO=0)	0.041	0.394	0.337	0.712	0.029	0.295	0.272	0.481
State-own Portfolio 2 (0<SO<=10%)	0.042	0.397	0.332	0.772	0.033	0.303	0.284	0.474
State-own Portfolio 3 (10%<SO<=50%)	0.040	0.420	0.312	0.591	0.030	0.310	0.275	0.459
State-own Portfolio 4 (SO>50%)	0.056	0.431	0.271	0.718	0.031	0.308	0.275	0.461
Retail Portfolio 1(Retail ownership <= country level lower 30%)	0.056	0.395	0.276	0.692	0.009	0.346	0.269	0.275
Retail Portfolio 2(Retail ownership > country level upper 30%)	0.047	0.407	0.326	0.541	0.033	0.190	0.426	0.475
IO Portfolio 1(IO <= country level lower 30%)	0.065	0.367	0.311	0.581	0.033	0.184	0.415	0.459
IO Portfolio 2(IO > country level upper 30%)	0.044	0.398	0.297	0.628	0.009	0.351	0.267	0.286
International Accessibility Portfolio 1 (IA1=0)	0.042	0.411	0.326	0.410	0.030	0.295	0.276	0.373
International Accessibility Portfolio 2 (IA1>0)	0.055	0.438	0.270	0.820	0.031	0.315	0.275	0.373
Illiquidity Portfolio 1(Zeros <= country level lower 30%)	0.014	0.428	0.319	0.606	0.015	0.309	0.275	0.324
Illiquidity Portfolio 2(Zeros > country level upper 30%)	0.085	0.411	0.276	0.671	0.123	0.191	0.384	0.376
Turnover Portfolio 1(Turnover <= country level lower 30%)	0.055	0.434	0.252	0.673	0.058	0.211	0.313	0.459
Turnover Portfolio 2(Turnover > country level upper 30%)	0.035	0.373	0.405	0.579	0.029	0.253	0.384	0.406
Size Portfolio 1(Market value <= country level lower 30%)	0.043	0.383	0.362	0.374	0.119	0.057	0.637	0.160
Size Portfolio 2(Market value > country level upper 30%)	0.050	0.428	0.281	0.632	0.026	0.320	0.263	0.266
Tech portfolio 1(Not TMT industry)	0.049	0.424	0.292	0.483	0.031	0.305	0.274	0.245
Tech portfolio 2(TMT industry)	0.038	0.401	0.346	0.511	0.018	0.332	0.302	0.244
Listing board portfolio 1(Main board)	0.050	0.426	0.288	0.513	0.031	0.308	0.276	0.234
Listing board portfolio 2(SME board)	0.039	0.342	0.372	0.596	0.009	0.359	0.238	0.235
Listing board portfolio 3(GEM board)	0.032	0.324	0.391	0.513	0.008	0.363	0.207	0.194
Market	0.048	0.423	0.295	0.484	0.022	0.317	0.259	0.408

Table OA5. Summary statistics of financial openness measures

This table reports summary statistics and regression results related to sector/portfolio level financial openness in 1995Q1 to 2018Q4. Definitions of all these variables are described in detail in Appendix A. The panel includes 33 sectors and 21 additional portfolios formed on state ownership, international accessibility, illiquidity, market size, retail ownership, institutional ownership, turnover rate, technology sectors and listing boards.

	IA1: Weighted average of sum of dummies for B, H,ADR or China-HK Connect	IA2: MV(B,H and ADR)/total MV	IA3: Market share of firms with B share, H share, ADR or China-HK Connect	A-B Premium	A-H Premium
Industrial sectors					
Aerospace & Defense	0.142	0.000	0.141	0.000	0.000
Alternative Energy	0.206	0.014	0.148	0.000	0.287
Automobiles & Parts	0.456	0.049	0.272	1.365	0.168
Banks & Life Insurance	0.938	0.201	0.446	0.000	0.050
Beverages	0.402	0.051	0.268	1.251	0.896
Chemicals	0.517	0.066	0.285	2.463	1.812
Construction & Materials	0.573	0.084	0.349	1.481	2.812
Electricity	0.587	0.082	0.348	0.986	0.431
Electronic & Electrical Equipment	0.246	0.017	0.207	1.709	3.697
Financial Services	0.292	0.019	0.182	0.000	0.071
Fixed and Mobile Telecom	0.144	0.000	0.144	0.000	0.000
Food & Drug Retailers	0.089	0.000	0.089	0.000	0.000
Food Producers	0.197	0.007	0.185	2.330	0.012
Forestry & Paper	0.257	0.040	0.134	0.485	0.335
Gas, Water and Multiutilities	0.350	0.014	0.253	0.000	3.084
General Industrials	0.374	0.087	0.271	0.740	0.123
General Retailers	0.156	0.008	0.136	1.501	0.085
Health Care Equipment & Services	0.081	0.000	0.081	0.000	0.000
Household Goods & Home Construction	0.277	0.032	0.225	0.902	1.553
Industrial Engineering	0.526	0.065	0.357	1.774	2.582
Industrial Metals & Mining	0.638	0.069	0.330	2.139	1.559
Industrial Transportation	0.598	0.070	0.386	1.642	0.606
Leisure Goods	0.323	0.034	0.267	2.152	0.000
Media	0.103	0.000	0.103	0.000	0.000
Mining	0.826	0.137	0.356	0.000	0.708
Oil Equipment, Services & Oil and Gas Producers	1.141	0.232	0.626	0.000	0.344
Personal Goods	0.362	0.042	0.300	1.595	0.013
Pharmaceuticals & Biotechnology	0.230	0.017	0.180	1.903	2.637

	IA1: Weighted average of sum of dummies for B, H,ADR or China-HK Connect	IA2: MV(B,H and ADR)/total MV	IA3: Market share of firms with B share, H share, ADR or China-HK Connect	A-B Premium	A-H Premium
Real Estate Investment & Services	0.434	0.041	0.300	1.317	0.476
Software & Computer Services	0.199	0.011	0.169	2.668	0.000
Support Services	0.369	0.016	0.235	2.095	0.142
Technology Hardware & Equipment	0.303	0.026	0.219	1.006	2.564
Travel & Leisure	0.829	0.158	0.477	1.107	1.304
Other portfolios					
State-own Portfolio 1 (SO=0)	0.212	0.025	0.155	1.439	1.631
State-own Portfolio 2 (0<SO<=10%)	0.346	0.040	0.216	1.268	0.147
State-own Portfolio 3 (10%<SO<=50%)	0.556	0.080	0.325	1.335	1.377
State-own Portfolio 4 (SO>50%)	0.943	0.169	0.482	1.800	1.466
Retail Portfolio 1(Retail ownership <= country level lower 30%)	1.113	0.221	0.543	0.837	0.351
Retail Portfolio 2(Retail ownership > country level upper 30%)	0.330	0.039	0.252	0.910	1.214
IO Portfolio 1(IO <= country level lower 30%)	0.721	0.124	0.396	1.295	1.333
IO Portfolio 2(IO > country level upper 30%)	0.828	0.169	0.431	0.617	0.279
International Accessibility Portfolio 1 (IA1=0)	0.000	0.000	0.000	0.000	0.000
International Accessibility Portfolio 2 (IA1>0)	1.519	0.263	1.000	1.551	1.438
Illiquidity Portfolio 1(Zeros <= country level lower 30%)	0.557	0.087	0.329	1.556	1.958
Illiquidity Portfolio 2(Zeros > country level upper 30%)	0.817	0.142	0.417	1.492	1.315
Turnover Portfolio 1(Turnover <= country level lower 30%)	0.797	0.135	0.427	1.290	1.274
Turnover Portfolio 2(Turnover > country level upper 30%)	0.375	0.048	0.233	2.002	2.503
Size Portfolio 1(Market value <= country level lower 30%)	0.027	0.002	0.022	1.931	0.842
Size Portfolio 2(Market value > country level upper 30%)	0.866	0.146	0.472	1.481	1.403
Tech portfolio 1(Not TMT industry)	0.722	0.122	0.387	1.548	1.396
Tech portfolio 2(TMT industry)	0.223	0.014	0.173	1.974	2.564
Listing board portfolio 1(Main board)	0.769	0.130	0.409	1.551	1.437
Listing board portfolio 2(SME board)	0.158	0.004	0.139	0.000	0.476
Listing board portfolio 3(GEM board)	0.174	0.000	0.175	0.000	0.000
Market	0.697	0.117	0.376	1.551	1.438

Table OA6. Summary statistics of investor base measures

This table reports time-series summary statistics on investor base variables. Definitions of all these variables are described in detail in Appendix A. Institutional ownership, retail ownership and standardized numbers of shareholders are available from 2003Q1 to 2018Q4 while Chinese state ownership and turnover rate start from 1995Q1 to 2018Q4. The panel includes 33 sectors and 21 additional portfolios formed on state ownership, international accessibility, illiquidity, market size, retail ownership, institutional ownership, turnover rate, technology sectors and listing boards.

	China					U.S.		
	Chinese state ownership	Institutional ownership	Retail ownership	Standardized numbers of shareholders	Turnover	Institutional ownership	Retail ownership	Turnover
Industrial sectors								
Aerospace & Defense	0.519	0.152	0.546	0.147	1.095	0.892	0.107	0.364
Alternative Energy	0.132	0.100	0.736	0.244	1.387	0.646	0.347	1.285
Automobiles & Parts	0.441	0.142	0.574	0.128	1.096	0.805	0.193	0.639
Banks & Life Insurance	0.411	0.205	0.354	0.083	0.548	0.750	0.247	0.394
Beverages	0.518	0.213	0.430	0.131	0.805	0.703	0.294	0.203
Chemicals	0.405	0.123	0.634	0.162	1.319	0.829	0.170	0.391
Construction & Materials	0.407	0.126	0.586	0.159	1.148	0.874	0.119	0.477
Electricity	0.597	0.115	0.463	0.104	0.851	0.772	0.226	0.338
Electronic & Electrical Equipment	0.302	0.127	0.657	0.170	1.384	0.881	0.108	0.477
Financial Services	0.364	0.122	0.586	0.127	1.088	0.863	0.128	0.486
Fixed and Mobile Telecom	0.576	0.120	0.485	0.110	0.981	0.697	0.302	0.321
Food & Drug Retailers	0.293	0.186	0.582	0.129	1.028	0.888	0.109	0.394
Food Producers	0.303	0.143	0.654	0.162	1.261	0.771	0.225	0.356
Forestry & Paper	0.305	0.097	0.747	0.164	1.414	0.957	0.042	0.472
Gas, Water and Multiutilities	0.475	0.059	0.632	0.170	1.199	0.737	0.259	0.303
General Industrials	0.182	0.105	0.722	0.153	1.208	0.736	0.263	0.273
General Retailers	0.269	0.160	0.641	0.143	1.073	0.753	0.222	0.517
Health Care Equipment & Services	0.319	0.181	0.654	0.149	1.301	0.920	0.077	0.472
Household Goods & Home Construction	0.144	0.196	0.643	0.160	0.945	0.771	0.225	0.379
Industrial Engineering	0.393	0.137	0.611	0.168	1.216	0.862	0.129	0.461
Industrial Metals & Mining	0.540	0.127	0.556	0.148	1.174	0.745	0.252	0.878
Industrial Transportation	0.502	0.132	0.505	0.114	0.863	0.852	0.144	0.407
Leisure Goods	0.331	0.075	0.714	0.202	1.279	0.898	0.096	0.695
Media	0.363	0.164	0.575	0.150	1.460	0.772	0.225	0.383
Mining	0.500	0.155	0.507	0.156	1.126	0.912	0.084	0.967
Oil Equipment, Services & Oil and Gas Producers	0.654	0.139	0.340	0.087	0.807	0.777	0.221	0.411
Personal Goods	0.264	0.099	0.708	0.170	1.259	0.884	0.110	0.440
Pharmaceuticals & Biotechnology	0.296	0.181	0.636	0.149	1.101	0.823	0.176	0.358

	China					U.S.		
	Chinese state ownership	Institutional ownership	Retail ownership	Standardized numbers of shareholders	Turnover	Institutional ownership	Retail ownership	Turnover
Real Estate Investment & Services	0.344	0.136	0.637	0.167	1.084	0.848	0.146	0.407
Software & Computer Services	0.179	0.139	0.645	0.179	1.506	0.795	0.144	0.583
Support Services	0.381	0.087	0.662	0.205	1.317	0.898	0.091	0.453
Technology Hardware & Equipment	0.328	0.148	0.625	0.184	1.278	0.824	0.173	0.852
Travel & Leisure	0.458	0.139	0.543	0.166	1.045	0.831	0.158	0.613
Other portfolios								
State-own Portfolio 1 (SO=0)	0.000	0.148	0.745	0.156	1.269	0.816	0.176	0.469
State-own Portfolio 2 (0<SO<=10%)	0.040	0.135	0.744	0.151	1.174	0.812	0.181	0.479
State-own Portfolio 3 (10%<SO<=50%)	0.334	0.144	0.565	0.143	1.078	0.815	0.179	0.484
State-own Portfolio 4 (SO>50%)	0.663	0.162	0.386	0.118	0.849	0.802	0.194	0.504
Retail Portfolio 1(Retail ownership <= country level lower 30%)	0.560	0.205	0.294	0.095	0.670	0.973	0.022	0.757
Retail Portfolio 2(Retail ownership > country level upper 30%)	0.252	0.037	0.931	0.208	1.443	0.256	0.737	0.331
IO Portfolio 1(IO <= country level lower 30%)	0.441	0.004	0.547	0.170	1.163	0.246	0.736	0.301
IO Portfolio 2(IO > country level upper 30%)	0.416	0.269	0.511	0.114	0.926	0.972	0.026	0.743
International Accessibility Portfolio 1 (IA1=0)	0.374	0.144	0.633	0.151	1.221	0.819	0.175	0.503
International Accessibility Portfolio 2 (IA1>0)	0.527	0.179	0.425	0.115	0.868	0.799	0.196	0.488
Turnover Portfolio 1(Turnover <= country level lower 30%)	0.485	0.163	0.443	0.115	0.498	0.491	0.493	0.106
Turnover Portfolio 2(Turnover > country level upper 30%)	0.356	0.106	0.719	0.209	2.387	0.903	0.092	0.902
Illiquidity Portfolio 1(Zeros <= country level lower 30%)	0.414	0.183	0.557	0.149	1.153	0.827	0.167	0.528
Illiquidity Portfolio 2(Zeros > country level upper 30%)	0.491	0.131	0.488	0.120	0.868	0.681	0.310	0.373
Size Portfolio 1(Market value <= country level lower 30%)	0.269	0.030	0.757	0.213	1.628	0.416	0.562	0.274
Size Portfolio 2(Market value > country level upper 30%)	0.500	0.181	0.473	0.115	0.882	0.815	0.181	0.508
Tech portfolio 1(Not TMT industry)	0.462	0.158	0.514	0.131	0.982	0.810	0.185	0.487
Tech portfolio 2(TMT industry)	0.345	0.151	0.615	0.161	1.307	0.795	0.189	0.648
Listing board portfolio 1(Main board)	0.482	0.155	0.493	0.128	0.944	0.807	0.188	0.496
Listing board portfolio 2(SME board)	0.112	0.179	0.572	0.137	1.494	0.838	0.155	0.607
Listing board portfolio 3(GEM board)	0.042	0.156	0.672	0.118	1.830	0.865	0.124	0.515
Market	0.456	0.157	0.520	0.133	1.000	0.803	0.188	0.467

Table OA7. Valuation differentials under simplified model selection method

This table reports sector regression results on all related variables discussed in Table 3 - Table 6. The left panel shows results for variables available from 1995 to 2018 while the right panel shows results for variables from 2003 to 2018. The dependent variable is the sector/portfolio earning yield difference between China and US, DIFEY. All independent variables are differences between China and the U.S. except for "Dummy: 2009Q3" which is a dummy variable equal to 1 (0) during period after (before) 2009Q3 and political risk variables (including overall political rating, quality of institutions and investment profile) which are constructed by taking the ratio of Chinese over U.S. variables. Standard errors are double clustered by sector and time. We apply simplified model selection procedure to select the most important independent variables included in the regression. The overall variance contribution of each selected variable is reported. We perform the simplified model selection method by 3 steps. First, if multiple variables are correlated more highly than 0.8, we select the one variable that features the highest absolute t statistic in univariate regression and drop the other variables. Second, we perform a multivariate regression with the remaining variables from step 1 and drop all variables with absolute t-stats less than 1.0. Third, we repeat the regression with the remaining variables from step 2 and drop all variables with absolute t-stats less than 1.5. A regression with the remaining variables constitutes the final version of the model. We report t statistics under the coefficient estimates. ***, ** and * indicate significance at the 1%, 5% and 10% levels.

	1995-2018		2003-2018	
	(1)	(2)	(3)	(4)
Dependent Variable: DIF_EY				
Intercept	-0.043***		-0.005	
	(-3.195)		(-0.688)	
Growth Expectations				
GDP growth rate	-0.238***	6.2%	-0.260***	14.3%
	(-3.980)		(-4.398)	
Sale growth expectation			-0.027***	6.4%
			(-3.023)	
Forecast dispersion			-0.034***	10.7%
			(-6.067)	
Financial Development				
Zeros	0.081***	10.9%	0.193***	12.1%
	(4.458)		(4.794)	
R-square	0.018***	-1.3%	0.018***	1.5%
	(4.095)		(2.907)	
Idiosyncratic volatility			0.019**	-4.6%
			(2.013)	
Adjusted market development	-0.002**	0.5%	-0.001*	1.2%
	(-2.313)		(-1.667)	
Financial Openness				
IA1: Weighted average of sum of dummies for B, H,ADR or China-HK Connect	-0.008*	-17.2%		
	(-1.664)			
IA2: MV(B,H and ADR)/total MV	0.149***	68.3%	0.102***	42.8%
	(5.941)		(4.725)	
REGOPEN	0.003***	20.8%		
	(4.308)			
Quality of institutions	0.050**	-4.7%		
	(2.264)			
A-H premium	-0.001**	4.3%		
	(-2.090)			

Investor base				
Institutional ownership			-0.008	0.9%
			(-0.931)	
Turnover rate	-0.005***	12.3%	-0.007***	14.8%
	(-4.418)		(-3.554)	
Total Variance Contribution		100%		100%
Number of observations	4,873		3,204	
Adjusted R-square	0.337		0.421	