How do Funds Deviate from Benchmarks? Evidence from MSCI's Inclusion of Chinese A-shares

Lennart Dekker   Jasmin Gider   Frank de Jong*

October 29, 2021

Abstract

An increasing amount of assets is managed by benchmark-tracking investment funds. This study investigates how benchmarking changes affect portfolio compositions in the cross-section of different investor types and stock characteristics. To that end, we exploit the phased introduction of Chinese A-shares to the MSCI Emerging Markets index, which was announced in June 2017 and implemented over the period from May 2018 to November 2019. This change presents a rare opportunity to estimate the impact of index changes and to shed light on cross-sectional implications. We document that particularly passive funds systematically deviate from the benchmark. Market capitalization, stock liquidity and stock volatility affect how benchmark changes translate to portfolio adjustments of mutual funds and ETFs. We then study how the changes in benchmark weights affect financial market outcomes, more specifically the comovement of returns. We find that these characteristics moderate the impact of benchmarking changes on financial market outcomes, suggesting that deviations from benchmarks matter.

*The authors gratefully acknowledge financial support by INQUIRE Europe. *All authors are at the Department of Finance, Tilburg University, Warandelaan 2, 5037 AB Tilburg, Netherlands. Phone: +31 13 466 3238. Email addresses: L.P.Dekker@tilburguniversity.edu, J.Gider@tilburguniversity.edu, F.deJong@tilburguniversity.edu.*

1 Introduction

The overwhelming majority of investments in financial assets nowadays are delegated to institutional asset managers, such as mutual funds, pension funds and life insurers. The investment performance of funds and their managers are typically evaluated relative to benchmarks, or indices, which are created and maintained by a small number of professional index providers. The importance of indices has risen substantially in recent years because many investors now follow passive strategies and track these benchmarks. According to an estimate by the Financial Times as of January 2020, there are approximately USD 20 trillion assets under management in index-tracking strategies. The asset demand created for benchmark securities has substantial pricing implications for underlying constituents as documented by, e.g., Barberis and Shleifer (2003), Barberis et al. (2005), and Greenwood and Sosner (2007), or recently Pavlova and Sikorskaya (2021). Furthermore, recent work by Kashyap et al. (2021) shows theoretically and empirically that pricing effects have real implications for corporate investment and policies. The existing literature commonly assumes that passive portfolios closely track the benchmark compositions. Predictions by theory and also the interpretations of empirical evidence depend on this assumption. However, passive funds and ETFs are likely to engage in sampling when replicating a given benchmark (Brogaard et al., 2021), which might result in systematic deviations from benchmarks.

In this paper, we study the heterogeneity in the extent to which funds select stocks from the universe of benchmark securities. We investigate how passive and active investment funds respond to changes in the underlying benchmark and how stock characteristics related to sampling explain these changes. If sampling is a persistent phenomenon, changes in the benchmark will not affect all securities homogenously, but the impact of benchmarking on underlying securities will be moderated by the stock characteristics associated with sampling. After documenting systematic deviations, we investigate the implications of these deviations for market outcomes such as comovement.

We use the recent inclusion of Chinese A-shares in the MSCI Emerging Markets index (MSCI EM) to study how changes in index rules affect portfolio compositions. The MSCI Emerging Markets Index is one of the most intensely tracked emerging market benchmarks (see Figure 1). As of June 2021, it comprises more than 1,400 constituents and approximately 85% of free-float adjusted market capitalization across 27 Emerging Markets countries. On June 20^{th}, 2017 MSCI announced to partially include large-cap Chinese A-shares in the MSCI EM during the index rebalancing on May 31st, 2018. On February 28th, 2019, MSCI announced to increase the weights of the included large-cap A-shares during the quarterly rebalancings in May, August, and November of 2019, and to add mid-cap A-shares in November 2019. These changes increased the weight of Chinese equity in the MSCI Emerging Markets Index by approximately 5 percentage points, namely from 28% to 33%.[1] Table 1 illustrates the phased introduction and the respective inclusion factors over time. The inclusion of Chinese A-shares also affects the compositions of the MSCI All Country World Index (MSCI ACWI), since it nests the MSCI EM and the MSCI China Index, the Chinese subset of the MSCI EM. MSCI's decision was motivated by reflecting China's rising weight in the global economy. In the past decade the Chinese financial market has grown tremendously, now presenting the second largest capital market after the U.S. (Hu et al., 2021).[2]

[INSERT FIGURE 1 HERE]

[INSERT TABLE 1 HERE]

The phased introduction of the Chinese A-shares in the MSCI EM index exhibits several features that make it an attractive laboratory to study the impact of changes in benchmark weights. First, the introduction in several stages allows for a difference-in-differences analysis with multiple events and treatment intensities, which help to disentangle pure inclusion effects from index weight changes. Second, the setting allows for a rich pool of potential control securities, providing close proxies for the counterfactual outcome in the absence of index rebalancings: There is a large number of ineligible A-shares which is not included in the MSCI, and a number of Chinese H-shares that have already been included in the MSCI EM. Third, the setting helps with progress on studying heterogenous benchmark effects. Addressing heterogeneity is challenging as many settings where we observe benchmark changes are not well-suited to study the cross-section of stocks, either because of the small quantity of affected stocks or limited variation across stocks.[3] A large number of securities are simultaneously treated by the intervention, creating a rare opportunity to exploit cross-sectional heterogeneity. More specifically, 225 large cap A-shares wered added in May 2018 and 189 mid-cap A-shares were added in November 2019.

[1] H-shares, i.e., stocks listed in Hong Kong, as well as American Depository Receipts (ADRs) had been already included in the MSCI EM.

[2] Recent work by Liu et al. (2019); Carpenter et al. (2021) study size and value effects and price informativeness in the Chinese equity market.

[3] For example in the case of the Russell 1000 and Russell 2000 rebalancings are based on market capitalization, leading to little variation in the stocks that are added to or removed from the benchmark. S&P changes mostly involve the addition or deletion of individual stocks, where the majority of changes are triggered by corporate events.

Our main findings are as follows. First, we observe that fund portfolios respond closely to the timing of the staggered index inclusion. Our paper documents substantial heterogeneity in benchmark effects and that passive funds deviate from their benchmarks. Even the median ETF marketed as "fully replicating" in its fund prospectus holds less than 400 out of the 481 included A-shares. Funds with smaller tracking errors hold more eligible securities and funds for which the China A-shares have a greater benchmark weight (such as funds tracking the MSCI China) hold a larger share of eligible stock.

Next, we document a strong relation between aggregate stock ownership by passive funds and stock characteristics that are associated with sampling. More specifically, we find that (free-float adjusted) market capitalization, illiquidity and stock return volatility help predict which benchmark constituents will be added to the funds' portfolios. Collectively, these stock characteristics help explain more than 40% of the variation in portfolio holdings relative to the entire benchmarking universe, while the benchmark change alone explains only 24%. Active funds are more likely to deviate from the weights mandated by the benchmark, which is plausible as active funds are not bound to track the benchmark performance closely. Given the benchmark, the portfolio holdings of active funds are much less predictable, with an R-squared of just 6%.

We then study the implications of the deviations created by sampling. In this paper, we are mainly interested in the effect on comovement with respect to the MSCI Emerging Markets Index. We measure comovement using betas and R-squared from the one-factor market model in which the return on the MSCI EM serves as the market factor, which is standard in the comovement literature. Our main tests consist of panel regressions in which we use the set of non-included A-shares as the counterfactual. Importantly, we control for common time trends in the outcome variables by including time fixed effects. Our results are strongest for R-squared as a measure of comovement: we find that the predictors of the security selection within the benchmark also predict changes in R-squared after the benchmark inclusion. In a final specification which closely resembles the instrumental variable (IV) setup in Pavlova and Sikorskaya (2021), we confirm that cross-sectional differences in the response by passive funds moderate the impact of benchmark inclusion on comovement when measured by R-squared. We find weaker results when comovement is measured by beta.

Lastly, we zoom in on a subset of dual-listed stocks, which have both A- and H-shares. These double-listings in principle allow to control for any fundamental characteristics of the underlying securities, helping us to isolate the effect of the MSCI inclusion from any other unobserved time-varying factors that might affect eligible stocks differently from ineligible ones. We find however that ownership by passive funds benchmarked against the MSCI EM index also increases for H-shares after the inclusion of their dual-listed A-shares. This suggests that at least to some extent, passive funds use a firm's H-shares as a substitute for its A-shares. Nevertheless, we find that after index inclusion of the A-share, the gap in comovement with respect to the MSCI EM index compared with its dual-listed H-share narrows substantially.

The remainder of the paper is organized as follows: Section 2 summarizes related literature, while Section 3 presents our data. In Section 4 we present the findings on the impact of benchmarking changes for fund portfolios. Section 5 studies the market outcomes subsequent to benchmarking changes and explores the implications of systematic deviations from the benchmarks. In Section 6, we focus on the subset of A-shares with a dual-listing traded in Hong Kong. Section 7 concludes.

2 Literature review

2.1 Benchmarking

First, our work adds to a vast literature analyzing the consequences of the addition and deletion of individual securities to indices, such as the S&P 500. Harris and Gurel (1986) are among the first to examine price and volume effects of index additions. They find that after a stock has been added to the S&P500, its price and trading volume increase. Furthermore, Shleifer (1986), Kaul et al. (2000) and Greenwood (2005) exploit index changes to analyze demand curves. The results of these studies are consistent with a downward-sloping demand curve, as they find significant price effects as a result of the index changes. However, more recent evidence by Patel and Welch (2016) suggests that stocks no longer experience permanent shifts in investor demand upon addition or deletion from the S&P 500.

Pavlova and Sikorskaya (2021) document that index membership lowers the stock's risk premium, consistent with inelastic demand of the funds tracking those benchmarks. The strength of the index effect increases with the aggregate benchmark weight of the stock based on the aggregation of 34 different U.S. equity indices. The authors construct a stock-specific benchmarking intensity measure, capturing the number of different benchmarks that include a given stock and the amount of assets invested in these benchmarks, thereby obtaining a more precise measure of stock-specific benchmark exposure. They study Russell rebalancings between the Russell 1000 and Russell 2000 to lend support to a causal relationship.

Kashyap et al. (2021) provide theoretical evidence that the demand effects of benchmarking have real effects for investment and economic growth. In their model, index membership creates inelastic demand, which leads to a lower discount rate for risky investment projects of benchmark firms relative to outside firms, effectively presenting a subsidy.

In addition to a shift in the demand for benchmark securities, the extant literature has also considered the implications of benchmarking on comovement. This strand of the literature goes back to a theoretical framework set out in Barberis et al. (2005), Barberis and Shleifer (2003), or Greenwood and Sosner (2007). These models consider a market with a finite number N of risky assets, which present claims to a distant cash flow in the future. The cash flows are uncertain and exposed to a joint market factor and to an idiosyncratic component of risk. The risk-free asset is in elastic supply. There are M index stocks with $M < N$. Two types of investors interact in this market: index traders and fundamental traders, who act as arbitrageurs. Driven by exogenous demand for index products, index traders direct funds in and out of the M index stocks. Arbitrageurs trade against them, while they are limited in their arbitrage possibilities, e.g., because of limited capital or short investment horizons. The main prediction of this theoretical framework is that index stocks are subject to correlated demand shocks, which results in greater comovement of their stock returns. A number of papers (e.g., Vijh, 1994; Greenwood and Sosner, 2007; Barberis et al., 2005; Harford and Kaul, 2005; Chan et al., 2013) provide empirical support for this prediction. Chordia et al. (2000) document commonality in liquidity, which appears to be driven by asymmetric information and inventory risk. In a related study, Harford and Kaul (2005) investigate commonality in order flow, returns, and trading costs. They argue that indexing is the main driver of commonality, as common effects are higher for stocks included in the S&P500 than for non-index stocks.

We add to the existing index literature by investigating the nuanced portfolio responses by funds in the cross-section of stocks resulting from benchmark changes. These insights matter, because the causal impact of index rules crucially depends on how investors respond.

Our work is also related to Koijen and Yogo (2019), who develop a demand system approach to asset pricing that is based on asset-specific characteristics and stresses heterogeneity in asset demand. Our results indicate that systematic sampling by passive investors explains part of the heterogeneity in demand across securities within a given benchmark.

2.2 Passive investment

Second, our work is related to the understanding of the impact of ETF ownership and ownership of passive mutual funds on underlying securities, such as, e.g., Saglam et al. (2020); Ben-David et al. (2018); Brogaard et al. (2021). Most theoretical work modelling the impact of ETFs on information in securities, comovement and liquidity such as Malamud (2016); Bhattacharya and O'Hara (2018); Cong and Xu (2019); Breugem and Buss (2019); Bhattacharya and O'Hara (2020) rely on the assumption that the funds replicate the entire index. However, recent evidence challenges this assumption: Easley et al. (2021)'s findings suggest that ETFs are quite active in that they deviate from the market portfolio. Brogaard et al. (2021) are the first to document sampling behavior of ETFs in their work. They find that sampling leads to a heterogenous effect of ETF trading activity on stocks with ETF ownership: ETF trading reduces liquidity of ex-ante liquid securities only, while illiquid securities are unaffected. They provide a theoretical model illustrating the main tradeoffs that lead to sampling and document this behavior empirically using the Vanguard Total Stock Market ETF as an example (see Table 2 in their paper). Building on their work, we complement this line of research by investigating sampling behavior systematically using a larger sample of funds. Further, we explicity analyze how different types of funds and different types of stocks respond to changes in the underlying benchmark.

Lastly, our findings help discriminate between heterogeneity in the response to treatment by benchmarking versus heterogeneity in treatment. Our results have implications for any studies in asset pricing or corporate finance that rely on index rebelancings for causal inference by suggesting to pay attention to sampling when interpreting cross-sectional differences and try to isolate them from the determinants of sampling.

2.3 Financial integration

Third, our work contributes to work on the determinants of financial integration of emerging markets. The addition of Chinese A-shares was preceded by several initiatives by the Chinese governments to open up their financial markets to foreign investors. In 2014, the Chinese government launched Shanghai-Hong Kong Connect, and in 2016 Shenzhen-Hong Kong Stock Connect. These links enable foreign investors to buy Chinese A-shares, thereby relaxing restrictions imposed by having a market that is split between shares available to local and foreign investors respectively. Existing literature analyzes the segmentation of emerging markets from world stock markets (De Jong and de Roon, 2005) and the role of market liberalization (Bekaert et al., 2005). These papers show that opening emerging markets to foreign investors implies that expected returns are determined more by the covariance of local returns with global market returns, rather than with domestic pricing factors.

Existing work investigate the role of foreign institutional investors' portfolio choices in emerging markets. Boyer et al. (2006) show that stock market crises spread globally through the asset holdings of international investors. Jotikasthira et al. (2012) study the impact of investor flows to investment funds, which are domiciled in developed markets and trade in emerging markets. They find that fund-flow induced trades imply changes in emerging market portfolio allocations, which in turn affect emerging market equity prices, correlations, and betas. Hau and Lai (2017) show that during the 2007/08 financial crisis, re-allocations of equity fund portfolios due to losses in their bank-stock holdings affected the prices of non-bank stocks owned by these funds. The fund flows therefore contributed to contagion from developed to emerging markets. We contribute to this strand of literature by investigating how global benchmarks affect passive fund's equity allocations, and how changes in these allocations affects financial market outcomes such as comovement.

3 Data

3.1 Fund data

The data in this paper is obtained from multiple sources. We obtain fund holdings from FactSet and complement them with additional fund characteristics from Morningstar Direct. For a detailed description of the data selection and filters applied, we refer to Appendix A. To the best of our knowledge, FactSet presents the most comprehensive source for international fund holdings.[4] We use Morningstar to complement the fund characteristics available in FactSet. Most importantly, Morningstar Direct provides prospectus information on underlying benchmarks, which allows us to identify the funds that are directly affected by the index changes. We merge FactSet with Morningstar using fund tickers. Unfortunately, this leaves a number of unmatched funds, so we

[4]Most existing research on fund holdings relies on the CRSP Mutual Funds database and Thomson Reuters Mutual Fund Holdings. Unfortunately, these databases are limited to U.S. funds and holdings in U.S. stock, while our study requires holdings data on international stocks by international funds.

expand the intersection between FactSet and Morningstar by manually matching the funds by name. As mentioned before, several indices maintained by MSCI are affected by the inclusion of Chinese A-shares: the MSCI Emerging Markets Index, the MSCI All Country World Index, and the MSCI China Index. We only keep funds in our sample that are benchmarked to one of the affected indices. This results in a sample of 965 funds.[5]

Table 2 shows the composition of our fund sample. Open-end funds make up the vast majority of our sample, and more than half of the funds in our sample follow a growth-related investment style. Note that there are also some open-end funds that have the objective to track a certain index (style "Index"), which we denote as "passive" funds. Nearly 60 percent of the funds are domiciled in the U.S. Finally, 55 percent of the funds are benchmarked against the MSCI ACWI, 42 percent against the MSCI Emerging Markets Index, and 3 percent against the MSCI China Index.

[INSERT TABLE 2 HERE]

FactSet collects holdings data from a variety of different sources: regulatory filings, self-reported information by funds, and company disclosures. Differences in national reporting standards and the frequency at which funds choose to self-report result in different frequencies at which we observe the portfolio constituents. Table 3 shows these frequencies of funds in our sample for the years 2018 and 2019. For approximately 90 percent of assets under management, we observe holdings at least at a quarterly frequency. For ETFs, we observe holdings at a monthly frequency for 90 to 95 percent of assets under management.

[INSERT TABLE 3 HERE]

3.2 Stock data

We obtain stock data from Thomson Reuters Eikon, including information on the share type (A-share or H-share), exchanges, number of free float shares, prices, and volumes. We also use information by Eikon to identify dual-listed firms that have both A-shares and H-shares.

[5]There exist other affected indices, such as the MSCI Emerging Markets Asia Index, which is a subindex of the MSCI Emerging Markets Index. We keep funds tracking subindices in our sample when the subindex is affected in a similar way as the MSCI Emerging Markets Index. In contrast, we delete funds from our sample that track a selective subindex, such as the MSCI Emerging Markets ESG Leaders Index. To be included in this index, a company needs to meet certain Environmental, Social, and Governance (ESG) criteria. As we do not observe constituent lists for subindices, we ignore funds tracking subindices in which an unkown subset of A-shares is included.

We hand-collect the identity of Chinese shares added to the MSCI EM Index from MSCI's Quarterly Index Reviews, which list additions to and deletions from their indices for each quarterly rebalancing date. 225 A-shares were included in the MSCI EM Index in May 2018, while 204 stocks were included in November 2019. This combined set presents our sample of treated stocks. From the set of ineligible A-shares, we select the stocks that are accessible to foreign investors through the Stock Connect programs. We delete stocks that have been issued during our sample period to make sure that the composition of the control group is the same before and after the index events. The resulting control group consists of 629 A shares that are not included in the MSCI EM.

Table 4 shows summary statistics. We find that the large-cap A-shares that are included in the MSCI EM Index in May 2018 have the lowest volatility, the lowest free float percentage, the largest average daily trading volume, and show the lowest number of days at which the stock is suspended from trading. Over our 5-year sample period, the average stock in the control group was suspended from trading on 42 days, versus 14 and 39 days for the large-cap inclusions and mid-cap inclusions respectively. The vast majority of trading suspensions took place in the early half of our sample period, as hardly any stock has been suspended during the second half of our sample period. We also find that the large-cap A-shares added in May 2018 are most liquid as proxied by Amihud's illiquidity measure.

[INSERT TABLE 4 HERE]

4 Impact of index changes on fund portfolios

In this section we study how fund holdings respond to changes in the MSCI EM around the Chinese A-shares inclusion dates. As argued by Brogaard et al. (2021), passive funds choose portfolios that minimize tracking error, while also minimizing trading costs. Trading costs decrease with stock liquidity, while tracking errors are smaller for stocks that exhibit a high correlation with the tracked benchmark and little volatility. As a result, passive funds are expected to engage in sampling and concentrate their holdings in a subset of the included stocks. The aim of this section is to document the responses and to determine the stock characteristics that determine the sampling behavior of the funds.

4.1 The effect of index changes on portfolio weights

In examining how changes in index rules affect portfolio compositions, we distinguish between active and passive funds. The objective of passive funds is to closely track a given benchmark at the lowest possible costs, whereas active funds have more freedom to deviate from their benchmarks. We therefore expect passive funds to respond differently to the index inclusions than active funds. We define passive funds as the set of ETFs, and open-end funds for which the style is equal to “Index”. The remainder of the funds is classified as active.

Figure 2 shows that the timing of the response by passive funds moves closely with the rebalancing dates, rejecting the notion that these funds largely anticipate index changes. Our analysis focuses on the main inclusion rounds in May 2018 (Panels (a) and (b)) and November 2019 (Panels (c) and (d)), since these rebalancing dates involve the largest number of simultaneous stock inclusions. Since the included stocks have different weights in the three affected indices, we plot median portfolio weights after grouping funds by benchmark. In Panel (a), we do not find any evidence of anticipation as the increases in portfolio weights are perfectly aligned with the initial inclusion in May 2018 and the weight increases in August 2018, May 2019, August 2019, and November 2019. Similarly, in Panel (c), we find that portfolio weights exhibit a pronounced jump when the inclusion of mid-cap A-shares became effective.

The evolution of the portfolio weights of active funds shows an entirely different pattern. First of all, we now do observe anticipation, mainly by funds benchmarked against the MSCI China Index. On top of that, the variation across funds is wider, indicated by the wider interquartile range. Finally, even though portfolio weights have been increasing over time, there is no clear relation between any portfolio weight changes and the exact timing of the index events.

[INSERT FIGURE 2 HERE]

To assess the economic magnitude of these passive funds’ allocation towards the included A-shares, we aggregate ownership by the passive funds in our sample at the stock level. We scale this by free-float adjusted market capitalization such that we observe which fraction of the included A-shares’ free-float adjusted market capitalization is held by the passive funds in our sample after the index events. Panel (a) of Figure 3 shows that for the median A-share included in May 2018, ownership by the passive funds in our sample rises to about 30 basis points. For the 95^{th} percentile, ownership increases to almost 60 basis points. In Panel (c), we find that passive ownership in the median A-share that was included in November 2019 increases by slightly more than 10 basis points, and by 40 basis points for the 95^{th} percentile. For comparison, Panels (b) and (d) show the same plots for ownership by the active funds in our sample. In both panels, active ownership in the median A-share rose by no more than 10 basis points. However, for some stocks, active ownership rose to 2-4 percent of free-float adjusted market capitalization, as indicated by the 95^{th} percentiles. At first sight, the economic magnitude of the change in passive ownership seems limited. However, Pavlova and Sikorskaya (2021) report that ownership by passive funds benchmarked against the Russell 2000 increases by 77 basis points for stocks added to the Russell 2000, which is in a similar order of magnitude. Further, the relative change is substantial, as the level of passive ownership in A-shares was close to zero prior to the inclusion events. Moreover, recent evidence by Gabaix and Koijen (2021) shows that even small flows can have a meaningful impact on asset prices.

[INSERT FIGURE 3 HERE]

To investigate the commonly made assumption that index investors perfectly replicate their benchmark, we take a closer look at the heterogeneity in the way passive funds respond to the index inclusions. Under this assumption, the models by Barberis and Shleifer (2003), Barberis et al. (2005), and Greenwood and Sosner (2007) predict that the effect of index inclusion is homogeneous across stocks. To test this assumption, we focus on the number of shares held by passive funds in December 2019, right after the second main inclusion was completed. At that point in time, 481 Chinese A-shares were eligible for the MSCI Emerging Markets Index (and, hence, for the MSCI ACWI and MSCI China as well).

Figure 4 shows that passive funds deviate from their benchmark index. ETFs that claim to fully replicate their benchmark actually deviate from their benchmarks, since they do not hold all 481 eligible stocks. In fact, the median ETF labelled as "fully replicating" holds less than 400 Chinese A-shares. We find that ETFs that engage in sampling hold even less Chinese A-shares than traditional index funds and fully-replicating ETFs, in line with their objectives. We also split funds by benchmark, and find that funds tracking the MSCI China hold a larger number of A-shares than funds tracking the MSCI ACWI or MSCI EM. Hence, the larger the weight of A-shares in the benchmark, the more A-shares these funds tend to hold. Next, we find that funds with lower historical tracking errors hold a larger number of A-shares than funds with an above-median historical tracking error. Larger funds tend to hold a greater number of A-shares than smaller funds. Finally, funds with a low expense ratio hold a similar number of A-shares as funds with a high expense ratio.

The results in Figure 4 are consistent with the tradeoff that passive funds face: minimizing tracking errors versus minimizing costs. Fully replicating the benchmark would lead to the lowest possible tracking error, but this involves higher transaction costs. Alternatively, funds can buy a subset of the index to lower transaction costs, but this comes at the expense of a greater tracking error.

[INSERT FIGURE 4 HERE]

In sum, the findings document that the timing of the response to index inclusions by passive funds aligns with the rebalancing dates. However, there is heterogeneity in the number of included A-shares these passive funds buy.

4.2 Stock characteristics and sampling

Given the tradeoff between tracking errors and transaction costs, we hypothesize that passive funds concentrate their holdings in the largest, most liquid and least volatile A-shares. It follows from the previous section that passive funds benchmarked against one of the affected MSCI indices did not buy all eligible A-shares. Rather, these funds select a subset of eligible A-shares. Buying the largest stocks will be important to reduce tracking errors, whereas trading the most liquid stocks will limit transaction costs.

We construct a stock-time panel in which the dependent variable represents the aggregate holdings across the funds in our sample in a given stock. Specifically, the aggregate holdings are given by:

$$y_{it} = \frac{\sum_{j=1}^{J} h_{ijt}}{s_{it}}. \tag{1}$$

Here, h_{ijt} denotes the number of shares of stock i held by fund j at time t. After summing h_{ijt} across funds, we divide by the number of shares outstanding of firm i (denoted by s_{it}), which is adjusted for its free float and the Foreign Ownership Limit. The Foreign Ownership Limit is 30% for Chinese stocks, meaning that foreign ownership can not exceed 30%. The fraction of shares outstanding that MSCI considers as eligible is then given by the minimum of:

- 100% minus non-free float shareholdings, including domestic and foreign shareholdings.

- Foreign Ownership Limit (30%) minus foreign non-free float shareholdings.[6]

Then, s_{it} is given by the total number of shares outstanding of firm i multiplied with the adjustment fraction as defined above. We observe the number of shares outstanding and the free float percentages from Datastream. We do not observe foreign non-free float holdings, but we assume this to be zero. Given the difficulties to trade Chinese A-shares as a foreign investor, we believe this is a reasonable assumption. Importantly, scaling by s_{it} ensures that we already control for cross-sectional differences in index weights. This yields the natural interpretation that the dependent variable would be constant across stocks under the condition that all funds $j \in \{1, ..., J\}$ would perfectly replicate the benchmark. We separately consider aggregate holdings by passive and active funds, and use a quarterly frequency as motivated by the reporting frequency in Table 3. Table 5 shows the number of stock observations per quarter in our resulting panel dataset. We distinguish between the set of included A-shares, the set of A-shares that will be included at some future date, and the control group. The number of stock-quarter observations corresponding to the treated group equals $3430 + 4669 = 8099$ observations, and adding the control group yields a total of $17,864$ observations.

[INSERT TABLE 5 HERE]

We estimate a panel regression where we regress the aggregate holdings from (1) on a set of stock characteristics:

$$y_{it} = \alpha_t + \beta IF_{it} + \gamma' \left(IF_{it} \cdot X_{i,t-1}\right) + \delta' X_{i,t-1} + \eta Ineligible_i + \varepsilon_{it} \tag{2}$$

Specifically, we consider a stock's MSCI inclusion factor (IF_{it}), from Table 1. Secondly, we consider several stock characteristics, summarized by the vector $X_{i,t-1}$. First, we consider the logarithm of a stock's free-float adjusted market capitalization ($log(MC_{i,t-1})$). We also consider the lagged Amihud's measure to capture stock illiquidity ($Illiq_{i,t-1}$) and the lagged volatility of daily returns ($Vol_{i,t-1}$). Amihud's measure and volatility are estimated on a quarterly basis using daily returns, and implemented using a 1-quarter lag. All stock characteristics are standardized to have zero mean and unit standard deviation. Finally, $Ineligible_i$ is a dummy variable that equals 1 if stock i is never included in the MSCI Emerging Markets Index during our sample, and 0 otherwise.

[6] To obtain information on MSCI's benchmark rules, we rely on the MSCI Global Investable Markets Indexes Methodology dated January 2020.

This dummy captures time-invariant differences in the level of passive ownership in non-eligible A-shares.

Table 6 reports the results. In Panel A, we estimate Equation (2) using only those A-shares that were at some point included in the MSCI Emerging Market index.[7] Column 1 shows that aggregate passive ownership is strongly related to index eligibility, confirming the results from Figure 2. We do not have any theoretical prior on the coefficient magnitude, as this depends on the assets under management of the affected funds. The adjusted R^2 of the panel regression in which we solely include the index inclusion factor equals 24%, indicating that index eligibility explains a substantial fraction of the variation in passive ownership. Column 2 shows evidence in favor of concentration in the largest stocks. The largest stocks have a higher weight in the index. Hence, omitting these stocks will lead to a larger tracking error than omitting stocks with a smaller weight. This is confirmed by the significantly positive coefficient on the interaction between the index inclusion factor and the logarithm of market capitalization. Including the interaction between the logarithm of market capitalization and the inclusion factor leads to an increase in the adjusted R-squared from 24% to 41%. In column 3, we also find evidence consistent with the tradeoff between tracking errors and transaction costs. The interaction between the index inclusion factor and Amihud's illiquidity measure is negative and highly significant. This implies that a lower price impact per dollar traded makes an included A-share more attractive to passive investors. In the fourth column, we find a negative relationship between passive ownership and realized volatility, suggesting that passive funds avoid the most volatile stocks. Finally, in the fifth column, the coefficients on the interactions between the index inclusion factor and market cap, Amihud's illiquidity measure, and realized volatility have remained highly significant.

In Panel B of Table 6, we estimate Equation (2) using both the set of treated A-shares as well as the set of non-included A-shares that serves as our control group. Since ownership in non-eligible A-shares by the passive funds in our sample is virtually zero, the conclusions remain the same as in Panel A.

[INSERT TABLE 6 HERE]

The results in Table 6 are consistent with the theoretical predictions of Brogaard et al. (2021), namely that passive investors omit illiquid stocks in order to limit transaction costs. Brogaard

[7] This specification obviously excludes the $Ineligible_i$ dummy.

et al. (2021) confirm this prediction empirically by examining the holdings of a single popular ETF, namely the Vanguard Total Stock Market ETF (VTI) that seeks to track the performance of the CRSP US Total Market Index. We generalize this result by showing a systematic deviation from the index portfolio across a wide set of passive funds that are benchmarked against the same index.

We repeat these panel regressions for aggregate holdings across active funds in Table 7. First of all, note that the adjusted R^2 is well below 9% in every specification, whereas this ranges between 24% and 71% for passive ownership. This already indicates that the inclusion factor and the same set of stock characteristics are only able to explain a small fraction of the variation in active ownership, consistent with the notion that benchmarking effects are substantially diluted for active funds. Columns 2 and 3 suggest that active funds also seem to favor larger, more liquid stocks. However, the coefficients on the standalone variables $log(MC)$ and $Illiq$ are statistically significant as well. This suggests a general preference for these shares, while the timing of the transactions by active funds is decoupled from the index events, consistent with Figure 2. Column 5 contains the specification including all explanatory variables. Comparing this with the results for passive ownership, we also find a positive association between active ownership and the logarithm of market cap, but there is no significant relation between active ownership and the interaction with volatility or liquidity.

The results in Panel B, where we estimate Equation (2) using both treated as well as control stocks, are largely similar to the results in Panel A. The coefficient on the non-eligibility dummy is negative in all specifications, and statistically significant in three out of five cases. This indicates that even before the index inclusion, active ownership in A-shares that at some point were included in the MSCI Emerging Markets Index was on average higher than for non-included A-shares.

[INSERT TABLE 7 HERE]

Finally, we examine the overlap between passive and active holdings. We regress stock-level active ownership on (scaled) passive ownership to test whether active and passive investors concentrate their trades in the same subset of stocks. In column 1 of Table 8, we find a coefficient of 0.32 with a t-statistic of 11.96, which implies that stocks with a high level of passive ownership also tend to have higher levels of active ownership. When controlling for stock characteristics in column 2, we continue to find a positive relation between passive and active ownership, while the economic magnitude drops to 0.11. This suggests that for a one-standard deviation change in passive ownership, the amount of active ownership increases by 11 basis points. The conclusions are the same in columns 3 and 4, where we add the control stocks to the panel of observations as well. However, the adjusted R-squared of these regressions are low, so the level of passive ownership does not explain a large part of the variation in active ownership. Further, these results reject the notion that active funds exploit the predictable demand effects created by the benchmark by trading against passive funds.

[INSERT TABLE 8 HERE]

Summarizing, we document that passive fund portfolios deviate from benchmarks, while these deviations are systematic and, hence, highly predictable. Aggregate passive ownership is strongly related to market capitalization, liquidity, and volatility. In contrast, the same set of stock characteristics barely explains the response of active ownership to benchmark changes. We document a positive association between the level of passive and active ownership, but even the level of passive ownership explains a minor fraction of the variation in active ownership. These results have several implications. The role of index investors is crucial in the transmission of index inclusion to financial market outcomes (Barberis and Shleifer, 2003; Barberis et al., 2005; Greenwood and Sosner, 2007). Further, our results cast doubt on the commonly made assumption that index investors perfectly replicate the index.

5 Impact of index changes on market outcomes

In this section, we investigate the effect of benchmark changes on comovement. A large literature has documented that after a stock joins an index, it starts to comove more with that index (e.g., Barberis et al., 2005). We test this prediction on the full set of included A-shares.

5.1 Changes in beta and correlation

We start with an event-type approach that is, e.g., also employed by Barberis et al. (2005) and Boyer et al. (2006). We consider 10 quarterly index rebalancing dates, where the first one took place on May 31^{st}, 2018, and the final one on August 31^{st}, 2020. For each rebalancing date, we consider a 3-month pre-event window that ends the day before the rebalancing date. Similarly, we consider a 3-month post-event window that starts the day after the rebalancing date. We estimate the following regression on a quarterly frequency over the period between December 2016 and November 2020, which nests the pre- and post-event windows:

$$r_{it} - r_t^f = \alpha_i + \beta_i(r_t^{\text{MSCI EM}} - r_t^f) + \epsilon_{it}, \tag{3}$$

where r_{it} is the return on stock i at time t, r_t^f denotes the risk-free rate, and $r_t^{\text{MSCI EM}}$ is the return on the MSCI Emerging Markets Index. We consider weekly overlapping returns to avoid non-synchronous trading issues arising from differences in time zones. We are interested in the changes in betas and R-squared, which we use to measure comovement. We also consider idiosyncratic volatility as an outcome variable, which we define as the standard deviation of the residuals from the regression in Equation (3). We winsorize r_{it} using the methodology suggested by Welch (2019) to obtain more robust estimates of betas.

We compare the results for the included A-shares with the set of non-included A-shares that are eligible for the Stock Connect programs. This set of 629 stocks serves as a proxy for the counterfactual outcome in the absence of index inclusion. Figure 5 shows the evolution of the outcome variables over time, where we distinguish between the A-shares included in May 2018, the A-shares included in November 2019, and the control group. Panels A and B show that there is substantial time variation in betas and R-squared with respect to the MSCI Emerging Markets Index, of which a large part is likely to be unrelated to the index events. Reassuringly, we find that the set of treated A-shares and the set of non-included A-shares appear to exhibit a common time trend. Hence, by using the non-included A-shares as control group, we aim to isolate the impact of the index events.

[INSERT FIGURE 5 HERE]

First of all, we compute the average change in betas for included A-shares pooled across all 10 event dates as follows:

$$\overline{\Delta\beta}_{Included} = \sum_t w_t \left(\frac{\sum_{i=1}^{I_t} \Delta\beta_{it}}{I_t} \right), \tag{4}$$

where t runs over the 10 rebalancing dates, I_t equals the number of A-shares included to the MSCI Emerging Markets Index at rebalancing date t, and w_t is a weight proportional to the number of

A-shares included during round t. Specifically, w_t is given by:

$$w_t = \frac{I_t}{\sum_{s=1}^{10} I_s}. \tag{5}$$

We compute average changes for the control group in a similar way. Finally, the difference-in-differences estimate is given by:

$$\overline{\Delta\beta}_{Included} - \overline{\Delta\beta}_{Control} = \sum_t w_t \left(\frac{\sum_{i=1}^{I_t} \Delta\beta_{it}}{I_t} - \frac{\sum_{j=1}^{J_t} \Delta\beta_{jt}}{J_t} \right), \tag{6}$$

where J_t denotes the number of A-shares in the control group surrounding rebalancing date t. The variance of the difference in betas used to calculate the t-statistic is derived in Appendix B. We repeat the same procedure for changes in the R-squared of Equation (3) and changes in idiosyncratic volatility, defined as the variance of the residual in Equation (3).[8]

Table 9 first examines the two main inclusion rounds separately. In total, 534 A-shares have been included during these 10 rebalancing dates, of which 225 got included in May 2018, and 204 in November 2019. Hence, these two rebalancing dates account for roughly 80 percent of all inclusions. For the inclusion round in May 2018 (Panel A), we find a positive difference-in-differences estimate for changes in betas, which is statistically significant and economically meaningful with a point estimate of 0.1141. Regarding R-squared, we also find a significantly positive change relative to the control group, equal to more than 6 percentage points. We reach similar conclusions for the inclusion in November 2019 (Panel B), as both betas as well as R-squared again show a significantly positive increase relative to the control group. Panel C of Table 9 shows the average changes pooled across all 10 event dates. The resulting difference-in-differences estimates are economically meaningful and statistically significant, for both changes in beta as well as R-squared. Hence, the results in Table 9 are in line with existing results on index inclusion in the S&P 500 (e.g., Barberis et al., 2005). We finally find a significantly negative difference-in-differences estimate for changes in idiosyncratic volatility.

[INSERT TABLE 9 HERE]

Another prediction is that comovement with respect to a given index decreases after a stock leaves that index (Boyer et al., 2006). Hence, we consider a two-factor market model that includes

[8] To save on notation, we shall in the text only specify the regression equations for β, but the regressions for R-squared and idiosyncratic volatility are always identical.

the return on the MSCI Emerging Markets Index as well as the value-weighted average return on the set of non-included Chinese A-shares in our control group. We then test whether betas with respect to the set of non-included A-shares decreases after inclusion.[9] Note that this is a somewhat different setting as in Boyer et al. (2006), because in their setting stocks explicitly leave one index when joining the other. In contrast, investors that focus on the market for Chinese A-shares are likely to keep investing in the A-shares that are included in the MSCI Emerging Markets Index. We therefore do not have a strong prior on the expected change in betas with respect to the set of non-included A-shares.

[INSERT TABLE 10 HERE]

Table 10 shows that after the index review in May 2018, betas with respect to the set of non-included A-shares decreased significantly. However, we do not find such an effect after the index review in November 2019. Yet, pooled across all the event dates, the dif-in-dif estimate remains significantly negative.

5.2 Cross-sectional heterogeneity in the treatment effect

We have shown in Section 4 that passive investors concentrate their holdings in larger, more liquid and less volatile stocks. Since passive investors are the mechanism through which index inclusion affects comovement, we test whether the same variables that explain the response by passive investors are related to the treatment effect.

Next, we consider panel regressions in which we relate betas and R-squared estimated from Equation (3) to a set of lagged stock characteristics. For a number of reasons, we believe that this is a more appropriate setup to test our hypotheses than analyzing the changes around the event in the cross-section. First, by expanding the time series, our results are less dependent on quarter-specific circumstances unrelated to the index events. For example, the quarter following the index review of November 2019 contains the corona crisis, which severely affected the market outcomes that we study (see Figure 5). Second, this setup allows us to incorporate characteristics of the control group as well. In the event-type tests, we only considered the average effect across control stocks, not exploiting what we can learn about the relation between stock characteristics and the market outcomes for the control stocks. Third, the panel regressions are able to capture

[9] Given that beta on the non-included A-shares index does not change if we orthogonalize the non-included A-shares index with respect to MSCI EM, there is no need for orthogonalization.

the effect of the staggered weight increases, which we ignored in the event-type tests. For example, during the index review of May 2018, MSCI included a selection of large-cap A-shares with an inclusion factor of 2.5%, which gradually rose to 20% in November 2019. Hence, the size of the impact of the index inclusion for those large-cap A-shares is expected to vary with the growing inclusion factor.

The panel estimates are based on the combined set of eligible and non-eligible stocks. First of all, suppose that we consider a set of N stocks, from which the first M stocks ($i \in [1, M]$) are at some point included in the MSCI Emerging Markets Index. The remaining $N - M$ stocks ($i \in [M + 1, N]$) are never included in the MSCI Emerging Markets Index and hence form the control group. We then consider the following regression equations:

$$Y_{it} = \alpha_t + \delta' X_{i,t-1} + \beta IF_{it} + \gamma' \left(IF_{it} \cdot X_{i,t-1} \right) + \varepsilon_{it} \qquad \text{for } i \in [1, M] \qquad (7)$$

$$Y_{it} = \alpha_t + \delta' X_{i,t-1} + \eta + \varepsilon_{it} \qquad \text{for } i \in [M + 1, N]. \qquad (8)$$

Here, Y_{it} denotes either the beta or the R-squared of stock i in quarter t, α_t is a quarterly dummy (fixed time effect) and $X_{i,t-1}$ is a vector of lagged firm characteristics. This set of characteristics includes the logarithm of free-float adjusted market capitalization ($log(MC_i)$), Amihud's illiquidity measure ($Illiq_i$), and realized volatility (Vol_i). The variable IF_{it} denotes the inclusion factor of stock i at time t, which is only non-zero for the eligible stock. Finally, the coefficient η in equation (8) captures any permanent difference between the betas of the eligible and ineligible shares which are not explained by the firm characteristics. Importantly, we impose some cross-equation restrictions on Equations (7) and (8). First of all, we impose the time fixed effects to be identical in both equations, as motivated by the common time trend that is visible in Figure 5. Secondly, the coefficients corresponding to the lagged characteristics are also identical across both equations. By defining a dummy variable $Ineligible_i$ for stocks that are never included in the MSCI Emerging Markets Index, we can pool Equations (7) and (8) into the following panel regression for all stocks in the sample:

$$Y_{it} = \alpha_t + \beta IF_{it} + \gamma' \left(IF_{it} \cdot X_{i,t-1} \right) + \delta' X_{i,t-1} + \eta Ineligible_i + \varepsilon_{it} \qquad (9)$$

The details of the estimation are as follows. The dependent variable Y_{it} denotes the beta or the R-squared from the one-factor market model in Equation (3), where the market factor is given by the return on the MSCI Emerging Markets Index. Both β_{it} and R^2_{it} are estimated on a quarterly frequency over the period between December 2016 and November 2020. α_t is a quarterly dummy (fixed time effect) and $X_{i,t-1}$ is a vector of lagged firm characteristics, and consists of the logarithm of free-float adjusted market cap, Amihud's illiquidity measure, and the realized volatility of daily returns. The variable IF_{it} denotes the inclusion factor of stock i at time t, which varies between 0 and 0.20. Finally, $Ineligible_i$ is a time-invariant dummy variable that equals 1 if stock i is never included in the MSCI Emerging Markets Index during our sample.

Table 11 reports the results.[10] In column 1, we find that the level of beta is on average lower for the stocks in our control group, as indicated by a significantly negative coefficient of -0.064. We do not find support for our hypothesis that betas increase after index inclusion or after increases in index weights, as the coefficient on the inclusion factor is indistinguishable from zero. In column 2, we find a negative coefficient on the interaction between the inclusion factor and the logarithm of free-float adjusted market capitalization. This result suggests that there is no average increase in betas stemming from the benchmark changes. We documented that passive funds concentrate their holdings in larger, more liquid, and less volatile stocks. Hence, we would expect the treatment effect to be stronger for larger stocks, which would lead to the prediction that the coefficient on the interaction term should be positive. However, the coefficient is negative in column 2. We do find support for the notion that sampling moderates the benchmarking effect in columns 3 and 4: the interactions with illiquidity and volatility obtain negative coefficients, implying that the impact of benchmarking changes is indeed larger for more liquid and less volatile stocks. Both coefficients remain statistically significant in the multivariate specification in column 5.

Column 6 shows that the average level of R-squared is lower for ineligible A-shares. Moreover, we now find a significantly positive treatment effect: after index inclusion or after index weight increases, R-squared with respect to the MSCI EM increases markedly. Since the inclusion factor varies between 0 and 0.2, R-squared on average increases by $0.2 \times 0.280 = 0.056$ over the full implementation phase, so close to 6 percentage points. In columns 7-10, we find a positive coefficient on the interaction between the inclusion factor and firm size, and negative coefficients on the interactions between the inclusion factor and illiquidity or volatility. The coefficients on the interaction terms are statistically significant for illiquidity and volatility. These results are in line

[10] The number of observations slightly differs from the number of observations in Table 5. The reason is that in Table 11, we shifted the quarters by 1 month to make sure that each quarter lies exactly between two subsequent index reviews. Therefore, the number of stocks for which we do not have sufficient history to obtain lagged characteristics differs slightly from Table 5.

with our hypotheses: larger stocks, more liquid stocks, and less volatile stocks exhibit a higher increase in R-squared after treatment.

[INSERT TABLE 11 HERE]

Summarizing, we find that the treatment effect on R-squared is related to the same characteristics that also explained a substantial fraction of the variation in passive ownership (Table 6). Regarding betas as a measure of comovement, we find similar results except for the counterintuitive relation between market capitalization and the treatment effect.

5.3 Understanding the mechanism

Finally, we want to further understand the mechanism through which the stock characteristics lead to a heterogeneous response to the index inclusion. According to the theoretical framework of Barberis et al. (2005), index inclusion affects return comovement through passive ownership. In Section 4, we have shown that passive ownership did not change homogeneously for stocks included in the MSCI Emerging Markets Index, suggesting that different stocks have been treated with different intensities. We therefore hypothesize that the heterogeneous effect on market outcomes is a result of the heterogeneity in the treatment intensity. To model this mechanism, we use the benchmarking changes and the interaction with stock characteristics associated with sampling in an instrumental variable setting.

We consider the following regression equation:

$$Y_{it} = \alpha_t + \gamma Passive_{it} + \delta' X_{i,t-1} + \eta Ineligible_i + \varepsilon_{it} \tag{10}$$

where Y_{it} again is either the beta with respect to the MSCI Emerging Markets Index or the R-squared, and $Passive_{it}$ denotes the ownership by passive funds benchmarked against the MSCI Emerging Markets index. This equation models the beta (or R-squared) as a function of a quarterly fixed effect, the holdings by passive investors and a set of firm characteristics. As before, the firm characteristics are log market cap, illiquidity, volatility and a dummy for the ineligible firms. In contrast to Equation (9), this model does not contain the inclusion factor IF nor its interactions with the characteristics. Instead, the index inclusion is assumed to impact the comovement via the change in fund holdings, as measured by the aggregate passive holdings. This interpretation rests upon the assumption that the benchmarking changes did not affect the market outcomes through alternative channels beyond their impact on passive ownership.

Because of the sampling practices by the passive funds, the variable $Passive_{it}$ variable may be endogenous. We therefore use an IV estimation which is similar to the approach of Pavlova and Sikorskaya (2021). Specifically, we instrument passive ownership by the index inclusion factor of the stock and the the set of stock characteristics interacted with the inclusion factor. Hence, the reduced-form first-stage regression is given by[11]

$$Passive_{it} = \tilde{\alpha}_t + \tilde{\beta} IF_{it} + \tilde{\gamma}' \left(IF_{it} \cdot X_{i,t-1} \right) + \tilde{\delta}' X_{i,t-1} + \tilde{\eta} Ineligible_i + \tilde{\varepsilon}_{it} \tag{11}$$

This regression is identical to Equation (2) in Section 4. Since we shifted quarters by one month to align the timing with the index rebalancing dates, the results differ slightly from the corresponding results reported in Panel B of Table 6. Namely, the number of observations for which we lack sufficient history to obtain lagged stock characteristics is slightly different from the reported number of observations in Table 5. Nevertheless, the economic magnitude and the statistical significance of the estimated coefficients is virtually the same as in Table 6. Using an F-test, we strongly reject the hypothesis that $\tilde{\beta}$ and all elements of $\tilde{\gamma}$ are jointly equal to zero as the resulting p-value is smaller than 0.001. On top of that, we also reject the hypothesis that only the elements of $\tilde{\gamma}$ are jointly equal to zero ($p < 0.001$), stressing the importance of taking into account sampling behavior.

Table 12 reports the results of estimating Equation (10). In Panel A, we estimate Equation (10) using OLS. We find in all 4 columns that the coefficient on passive ownership is positive, but only statistically significant when R-squared is the outcome variable. In Panel B, we consider the IV regression where $Passive_{it}$ is replaced by the fitted values $\widehat{Passive}_{it}$ from Equation (11). In columns 5 and 6, we find a negative coefficient on instrumented passive ownership. Hence, we cannot confirm the hypothesis that higher passive ownership would translate in higher comovement as measured by beta. Regarding R-squared as an outcome variable (columns 7 and 8), we do find a significantly positive coefficient on (instrumented) passive ownership, which is robust to the inclusion of stock characteristics. The differences in the coefficients between Panels A and B highlight the relevance of instrumenting passive ownership, as the coefficients on passive ownership

[11] We also consider a variant of Equation (10) without control variables, i.e., without $X_{i,t-1}$. In that case, we also remove the standalone characteristics from the first-stage regression in Equation (11).

are quite different. The results in this subsection summarize the results above, by directly linking the stock characteristics to changes in passive ownership that form the mechanism through which benchmarking affects financial market outcomes.

[INSERT TABLE 12 HERE]

6 Dual-listed A/H-shares

We acknowledge that MSCI's selection of which A-shares to include in its indices is not exogenous. It could be the case that the included A-shares and non-included A-shares differ in terms of unobservable characteristics, exposing them to differential time trends. In order to alleviate these concerns, we zoom in on a particular subset of included A-shares that have a dual-listed share traded in Hong Kong. Focusing on this set of stocks provides us with a suitable proxy for the counterfactual outcome, as the underlying firm fundamentals of the treated stocks and the control group are identical. We view this as a reverse difference-in-differences analysis, because both types of stocks are different in the pre-treatment period but we expect them to be more similar in the post-treatment period. H-shares have been already integrated well in international financial markets, indicated also by the fact that H-shares had already been included in the MSCI EM before. Consequently, after the A-share has been included in the MSCI EM index, we expect both the A and the corresponding H-share to be subject to demand shocks by investors tracking the MSCI EM index. Hence, we examine whether index inclusion reduces any pre-existing wedge in market outcomes between A-shares and their dual-listed H-shares. In total, we identify 69 firms with a listing in Hong Kong that also have an A-share which is at some point included in the MSCI Emerging Markets Index. Table 13 shows summary statistics for this subset of stocks. The included A-shares with a dual-listing in Hong Kong tend to be larger than the average included A-share, as the average market cap equals 8,606 million USD versus the full-sample average of 6,363 million USD (Table 4). We find that the distributions of returns for A- and their dual-listed H-shares are very similar. Even though the percentage of free float is lower for A shares, they still have a higher free-float adjusted market capitalization on average. Trading volume is way higher for A-shares than for H-shares, which also results in lower values for Amihud's illiquidity measure.

[INSERT TABLE 13 HERE]

We next examine changes in holdings in these dual-listed A-shares. Importantly, we address the question whether investors use a firm's H-share as a substitute for its A-share. Despite the imperfect correlation between the returns on a firm's A and H-shares due to market segmentation (see, e.g., Carpenter and Whitelaw, 2017), both share classes have identical firm fundamentals. This feature makes a firm's H-shares a potentially attractive substitute, for instance when the firm's A-shares are relatively illiquid. Of the 69 included A-shares, 58 had an H-share that was already part of the MSCI EM at inclusion. For these stocks in particular, the H-share might be a feasible alternative to the A-share, as mandates might prevent funds from buying stocks that are not part of their benchmark index.

We first consider the following panel regression:

$$y_{it} = \alpha + \beta \cdot Elig_{it}^{H} + \varepsilon_{it}, \tag{12}$$

where the outcome variable (y_{it}) denotes aggregate passive or active ownership in the H-share of firm i. We scale the aggregate dollar holdings of all funds in stock i by the free-float adjusted market capitalization of stock i. The explanatory variable, denoted by $Elig_{it}^{H}$, is a dummy equal to 1 if the H-share of firm i is eligible for the MSCI Emerging Markets Index at time t. Table 14 contains the results. We find in column 1 that this eligibility dummy explains a major part of the variation in passive ownership, as indicated by the high adjusted R-squared of 73%. The difference in passive ownership between eligible and noneligible H-shares on average amounts to 2.8 percentage points of the free-float adjusted market capitalization. The economic magnitude is even larger in column 4, which shows that active ownership is on average 3.9 percentage points higher for eligible H-shares than for noneligible H-shares. Yet, the eligibility dummy only explains about 10% of the variation in active ownership.

In columns 2 and 5 of Table 14, we extend the regression from Equation (12) as follows:

$$y_{it} = \alpha + \beta \cdot Elig_{it}^{H} + \gamma \cdot Elig_{it}^{A} + \delta \cdot \left(Elig_{it}^{H} \cdot Elig_{it}^{A}\right) + \varepsilon_{it}. \tag{13}$$

Compared with Equation (12), we added an interaction with $Elig_{it}^{A}$, which is a dummy equal to 1 if the A-share of firm i is eligible for the MSCI Emerging Markets Index at time t. We are mainly interested in the coefficient δ, as we want to test whether investors buy a firm's H-shares as a substitute for the firm's A-shares in response to the index inclusion of A-shares. We do not expect passive funds to buy stocks that are not within their benchmarks, so the dual-listed H-share should be index-eligible as well in order to be a good substitute. In column 2, we find a significantly positive coefficient on the interaction term, which indicates that passive ownership in a firm's H-shares increases after the A-shares of the same firm have been included in the MSCI Emerging Markets Index. This implies that the set of dual-listed H-shares is at least to some extent affected by the index event as well. We do not find significant evidence in favor of a substitution effect by active funds, as the coefficient on the interaction term is insignificant in column 5.

Finally, in columns 3 and 6 of Table 14, we consider a variant of Equation (13):

$$y_{it} = \alpha + \beta \cdot Elig_{it}^{H} + \gamma \cdot IF_{it}^{A} + \delta \cdot \left(Elig_{it}^{H} \cdot IF_{it}^{A}\right) + \varepsilon_{it}. \tag{14}$$

Here, we replaced the eligibility dummy for the A-share of firm i ($Elig_{it}^{A}$) by the inclusion factor that takes into account MSCI's weight increases (IF_{it}^{A}). The results in columns 3 and 6 are consistent with our findings in columns 2 and 5. In fact, when the inclusion factor of the A-share equals 0.2, passive ownership in the corresponding (index-eligible) H-share is on average $0.2 \cdot -0.410 + 0.2 \cdot 1.577 = 0.2334$ percentage points higher compared to the case when the A-share is not included in the MSCI EM index.

[INSERT TABLE 14 HERE]

We repeat the regressions from Equations (12), (13), and (14) for ownership in the corresponding A-shares. Again, we scale ownership in stock i by its free-float adjusted market capitalization. In contrast to the earlier results from Equation (2) in Tables 6 and 7, we do not take into account the Foreign Ownership Limit of 30%. The reason is that we are interested in the relative size of changes in ownership in A-shares versus H-shares, hence we focus on ownership as a percentage of the regular free-float adjusted market capitalization.

Table 15 contains the results. We are mainly interested in column 3: when the H-share is already index-eligible, inclusion of the A-share at a 20% inclusion factor leads to an average increase in passive ownership of $0.2 \cdot 1.582 - 0.2 \cdot -0.184 = 0.2796$ percentage points. Hence, we conclude that the economic magnitude of the changes in holdings for A-shares and their dual-listed H-shares is similar in magnitude, at least in absolute terms. Therefore, even though the set of dual-listed H-shares has the appealing feature that the underlying firm fundamentals are identical to the treated A-shares, these H-shares have been affected by the inclusion of their A-share counterparts as well.

[INSERT TABLE 15 HERE]

In unreported tests, we find that adding characteristics to Equation (14) does not yield a better fit. Including both the A-share and H-share characteristics, as well as the interaction between the Inclusion Factor and these characteristics, led to an increase in the adjusted R-squared of about 1 percentage point only compared to column 3 in Table 15. Concluding, we do not establish the same cross-sectional patterns for changes in ownership for dual-listed A-shares. This could be a result of the limited cross-sectional variation in characteristics in this sample, the fact that passive funds to some extent substitute these A-shares with H-shares, or both.

Next, we report the average treatment effect for the dual-listed A-shares. Since the vast majority of them has been included in the MSCI EM index during the inclusion round of May 2018 (54 of the 69 included dual-listed A-shares), we focus on these. For these 54 treated A-shares, we compute changes in market outcomes relative to their twin stocks traded in Hong Kong. Panels (c) and (d) of Figure 5 show the outcome variables over time for both the A-shares as well as their dual-listed H-shares.

It follows from Table 16 that we fail to find a significant increase in comovement relative to the control group, either measured by beta or by R-squared. We therefore fail to reject the null hypothesis that the covariance between returns of newly included stocks and the index return is unaffected by the index inclusion. However, as argued above, the set of H-shares has experienced inflows by the same set of passive investors as well, so it is well possible that these H-shares started to comove more with the MSCI EM index as well. We indeed find that betas and R-squared rose significantly for both the treated shares as well as the control group.

[INSERT TABLE 16 HERE]

We finally consider a panel regression such that we can examine changes in market outcomes over longer time periods. Specifically, we consider the following panel regression:

$$\beta_{A,it} - \beta_{H,it} = \alpha + \delta IF_{it} + \varepsilon_{it}, \tag{15}$$

where $\beta_{A,it}$ and $\beta_{H,it}$ denote the beta for the A- and H-shares of firm i in quarter t, respectively (we do the same regression for the R-squared). As before, IF_{it} denotes MSCI's inclusion factor for the A-shares, which ranges between 0 and 0.2. We abstain from characteristics, as we did not establish any cross-sectional pattern in changes in passive ownership. The results are shown in Table 17. Since the majority of the dual-listed A-shares has been included in the MSCI EM index during the index review in May 2018, time fixed effects would absorp almost all variation in the Inclusion Factor (IF_{it}), which is why we omit time fixed effects. Table 17 shows that both beta as well as R-squared was significantly lower for A-shares than for their dual listings in Hong Kong in the pre-treatment period, as indicated by the significantly negative intercepts. The majority of the H-shares was already included in the MSCI Emerging Markets Index, which could explain the higher level of comovement. In both columns, we find a positive coefficient on the Inclusion Factor, which suggests that the gap in comovement has narrowed after the inclusion of the A-shares. Regarding betas, this coefficient is insignificant, but for R-squared we find a statistically significant coefficient with a t-statistic of 3.69. After full inclusion of the A-share, the gap in R-squared with respect to the MSCI EM index on average decreases by $0.2 \cdot 0.260 = 0.052$, so roughly 5 percentage points.

[INSERT TABLE 17 HERE]

7 Conclusion

This paper examines how a variety of investors respond to changes in the benchmark index. We find that the responses by passive funds closely follow the rebalancing dates at which the index provider implements the changes. Contrary to the common assumption that index-tracking investors closely replicate their benchmark index, we find passive funds engage in sampling, resulting in substantial, yet predictable deviations from the benchmark. Deviations are more pronounced for funds that track indices that span a larger amount of stocks and for funds with a lower amount of AUM. Passive funds concentrate their holdings in larger, more liquid, and less volatile stocks, which is consistent with the tradeoff they face between minimizing tracking error and replication costs. Active funds show a distinct response: some funds anticipate the index changes, the share of benchmark stocks bought is lower than for passive funds, and the heterogeneity in holdings across active funds is larger than across passive funds.

Our main contribution is that we provide empirical insights about heterogeneity in the treatment effect of benchmarking, lending support to the notion that understanding deviations from benchmarking matters: especially when comovement is measured by R^2, we find that larger, more liquid, less volatile stocks experience a higher increase in comovement with respect to the MSCI

Emerging Markets Index. This suggests that the effect of benchmarking critically depends on the way investors respond.

Our results have important implications for research that uses changes in benchmarks to study the causal impact of institutional ownership or demand-driven effects of asset pricing. First, the systematic evidence about sampling helps to get a better, more precise instrument for passive ownership based on benchmark changes. This is because the interactions with stock characteristics associated with sampling add substantial explanatory power to variation in ownership. Second, our results guide the interpretation of cross-sectional results in benchmarking changes. Cross-sectional heterogeneity in stock- or firm-level outcomes to benchmarking could stem from either heterogenity in the treatment or heterogeneity in the response to the treatment. Our work provides insights that help to disentangle these two interpretations.

Further, our results offer an explanation for the existing empirical evidence (e.g., Easley et al., 2021) about the activeness of passive funds: sampling can give rise to systematic deviations of portfolio weights from the benchmarking weights, while this explanation also suggests that these deviations are predictable and thereby to a certain extent “mechanical”.

References

Barberis, N. and A. Shleifer (2003). Style Investing. *Journal of Financial Economics 68*, 166–191.

Barberis, N., A. Shleifer, and J. Wurgler (2005). Comovement. *Journal of Financial Economics 75*(2), 283–317.

Bekaert, G., C. Harvey, and C. Lundblad (2005). Does Financial Liberalization Spur Growth? *Journal of Financial Economics 77*, 3–55.

Ben-David, I., F. Franzoni, and R. Moussawi (2018). Do ETFs Increase Volatility? *Journal of Finance 73*, 2471–2535.

Bhattacharya, A. and M. O'Hara (2018). Can ETFs Increase Market Fragility? Effect of Information Linkages in ETF Markets. Working Paper.

Bhattacharya, A. and M. O'Hara (2020). ETFs and Market Runs. Working Paper.

Boyer, B., T. Kumagain, and K. Yuan (2006). How Do Crises Spread? Evidence From Accessible and Inaccessible Stock Indexes. *Journal of Finance 61*, 957–1003.

Breugem, M. and A. Buss (2019). Institutional Investors and Information Acquisition: Implications for Asset Prices and Informational Efficiency. *Review of Financial Studies 32*, 2260–2301.

Brogaard, J., D. Heath, and D. Huang (2021). The Heterogeneous Effects of Passive Investing on Asset Markets. Working Paper.

Carpenter, J., F. Lu, and R. Whitelaw (2021). The Real Value of China's Stock Market. *Journal of Financial Economics 139*, 679–696.

Carpenter, J. N. and R. F. Whitelaw (2017). The development of china's stock market and stakes for the global economy. *Annual Review of Financial Economics 9*, 233–257.

Chan, K., A. Hameed, and W. Kang (2013). Stock price synchronicity and liquidity. *Journal of Financial Markets 16*(3), 414–438.

Chordia, T., R. Roll, and A. Subrahmanyam (2000). Commonality in liquidity. *Journal of Financial Economics 56*(1), 3–28.

Cong, L. and D. Xu (2019). Rise of Factor Investing: Asset Prices, Informational Efficiency, and Security Design. Working Paper.

De Jong, F. and F. de Roon (2005). Time-Varying Market Integration and Expected Returns in Emerging Markets. *Journal of Financial Economics 78*, 583–613.

Easley, D., D. Michayluk, M. O'Hara, and T. Putnins (2021). The Active World of Passive Investing. *Review of Finance 25*(5), 1433—-1471.

Gabaix, X. and R. S. Koijen (2021). In search of the origins of financial fluctuations: The inelastic markets hypothesis. Technical report, National Bureau of Economic Research.

Greenwood, R. (2005). Short-and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics 75*(3), 607–649.

Greenwood, R. and N. Sosner (2007). Trading Patterns and Excess Comovement of Stock Returns. *Financial Analyst Journal 63*, 69–81.

Harford, J. and A. Kaul (2005). Correlated order flow: Pervasiveness, sources, and pricing effects. *Journal of Financial and Quantitative Analysis 40*(1), 29–55.

Harris, L. and E. Gurel (1986). Price and volume effects associated with changes in the s&p 500 list: New evidence for the existence of price pressures. *The Journal of Finance 41*(4), 815–829.

Hau, H. and S. Lai (2017). The Role of Equity Funds in the Financial Crisis Propagation. *Review of Finance 21*, 77–103.

Hu, G., J. Pan, and J. Wang (2021). Chinese Capital Market: An Empirical Overview. *Critical Finance Review 10*(2), 125–206.

Jotikasthira, C., C. Lundblad, and T. Ramadorai (2012). Asset Fire Sales and Purchases and the International Transmission of Funding Shocks. *Journal of Finance 67*, 2015–2050.

Kashyap, A., Kovrijnykh, J. Li, and A. Pavlova (2021). The Benchmark Inclusion Subsidy. *Journal of Financial Economics, forthcoming*.

Kaul, A., V. Mehrotra, and R. Morck (2000). Demand curves for stocks do slope down: New evidence from an index weights adjustment. *The Journal of Finance 55*(2), 893–912.

Koijen, R. and M. Yogo (2019). A Demand System Approach to Asset Pricing. *Journal of Political Economy 127*, 1475–1515.

Liu, J., R. Stambaugh, and Y. Yuan (2019). Size and Value in China. *Journal of Financial Economics 134*, 48–69.

Malamud, S. (2016). A Dynamic Equilibrium Model of ETFs. Working Paper.

Patel, N. and I. Welch (2016). Extended Stock Returns in Response to S&P 500 Changes. Working Paper.

Pavlova, A. and T. Sikorskaya (2021). Benchmarking Intensity. Working Paper.

Saglam, M., T. Tuzun, and R. Wermers (2020). Do ETFs Increase Liquidity? Working Paper.

Shleifer, A. (1986). Do demand curves for stocks slope down? *The Journal of Finance 41*(3), 579–590.

Vijh, A. M. (1994). S&P 500 trading strategies and stock betas. *The Review of Financial Studies 7*(1), 215–251.

Welch, I. (2019). Simpler better market betas. Technical report, National Bureau of Economic Research.

Tables

Table 1: Staggered inclusion of Chinese A-shares

Year	2018										2019											
Month	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Large-cap A-shares	0	0	2.5	2.5	2.5	5	5	5	5	5	5	5	5	5	10	10	10	15	15	15	20	20
Mid-cap A-shares	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	20
Small-cap A-shares	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ineligible A-shares	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Included H-shares	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Notes: This graph shows the MSCI inclusion factor of different types of Chinese stocks in percent of free-float adjusted market capitalization over the period from March 2018 to December 2019.

Table 2: Composition of the fund universe

							OEF	**ETF**
Type							92.54	7.46
	Aggr. Growth	**Growth**	**Index**	**Value**	**GARP**	**Yield**	**Deep Value**	**Other**
Style	33.26	22.07	8.81	6.32	5.28	2.59	0.73	20.93
	US	**LU**	**IE**	**DK**	**CA**	**GB**	**AU**	**Other**
Country	58.45	13.68	9.53	4.35	3.63	3.52	3.11	3.73
				MSCI ACWI		**MSCI EM**		**MSCI China**
Benchmark				55.13		42.07		2.80

Notes: This table shows the composition of our fund universe, in percentage of the total number of funds in our sample.

Table 3: Reporting frequency in 2018 and 2019.

	2018			2019		
	Total	OEF	ETF	Total	OEF	ETF
A. % of number of funds						
Only once	5.48	5.82	1.49	5.23	5.43	2.90
Semiannually or higher	89.86	89.63	92.54	91.58	91.23	95.65
Quarterly or higher	68.53	66.62	91.04	73.38	71.98	89.86
Monthly or higher	22.03	18.08	68.66	31.74	28.02	75.36
B. % of AuM						
Only once	1.17	1.31	0.01	1.38	1.50	0.41
Semiannually or higher	97.88	97.66	99.69	98.01	97.82	99.54
Quarterly or higher	87.05	85.54	99.31	92.62	91.81	99.08
Monthly or higher	19.35	10.46	91.79	25.32	16.55	94.91

Notes: This table shows the frequency at which the funds in our sample reported in 2018 and 2019. We distinguish between open-end funds (OEF) and exchange-traded funds (ETF). In Panel A, we present percentages of the total number of funds in our sample. In Panel B, we present percentages of the total assets under management (AuM) of the funds in our sample.

Table 4: Summary statistics stock universe.

	Mean	St. Dev.	5%	50%	95%
Daily return (%)					
May 18	0.0491	1.7698	-2.4049	-0.0894	3.0371
Nov 19	0.0848	2.1515	-2.8857	-0.1246	3.8400
Ineligible	0.0377	2.2084	-2.9043	-0.1943	3.8772
Free float (%)					
May 18	48.50	17.74	22.25	46.38	78.75
Nov 19	55.61	20.61	26.91	52.47	100
Ineligible	59.92	25.12	21.06	57.10	100
Free-float market cap (mln USD)					
May 18	6,363	9,346	1,207	3,431	21,312
Nov 19	2,008	1,347	699	1,650	4,429
Ineligible	908	570	278	786	2,007
Daily volume (mln USD)					
May 18	81.38	110.89	9.13	44.89	273.32
Nov 19	45.90	54.23	6.68	28.32	149.46
Ineligible	24.06	31.96	3.50	14.07	77.83
Number of days suspended					
May 18	13.87	32.88	0.00	1.00	79.25
Nov 19	39.44	78.39	0.00	5.50	180.30
Ineligible	42.04	75.58	0.00	9.00	175.15
Volatility of daily return (%)					
May 18	1.7698	0.4254	1.2648	1.7740	2.7566
Nov 19	2.1515	0.4541	1.6159	2.1669	3.0544
Ineligible	2.2084	0.4808	1.9988	2.6196	3.4826
Amihud's illiquidity					
(% per mln USD traded)					
May 18	0.4997	0.4196	0.0759	0.3843	1.2580
Nov 19	2.7251	8.2741	0.3228	0.7248	15.7556
Ineligible	2.7131	8.5842	0.6775	1.5356	4.1230

Notes: This table presents summary statistics of the Chinese A-shares in our sample over the period between January 2016 and December 2020. We distinguish between the set of A-shares included in the MSCI EM Index in May 2018, the set of A-shares included in November 2019, and the set of ineligible A-shares that forms our main control group. Regarding returns, free float, market capitalization, and trading volume, we first computed cross-sectional summary statistics for each time period separately, after which we calculated the time averages. The number of trading suspension days, volatility of daily returns, and Amihud's illiquidity measure are computed on the stock level, for which we show cross-sectional statistics. Everything is expressed in USD.

Table 5: Number of observations per quarter.

	Included A-shares	To be included A-shares	Control group	Total
2016 Q4	0	465	596	1061
2017 Q1	0	475	597	1072
2017 Q2	0	479	590	1069
2017 Q3	0	480	592	1072
2017 Q4	0	489	596	1085
2018 Q1	0	493	597	1090
2018 Q2	224	269	596	1089
2018 Q3	233	269	599	1101
2018 Q4	235	276	611	1122
2019 Q1	242	283	624	1149
2019 Q2	269	261	628	1158
2019 Q3	277	254	628	1159
2019 Q4	480	51	628	1159
2020 Q1	482	49	628	1159
2020 Q2	493	40	628	1161
2020 Q3	495	36	627	1168
Total	3430	4669	9765	17864

Notes: This table shows the number of observations in our panel over time. We distinguish between the set of stocks that is included in the MSCI EM index at time t, the set of stocks that will be included in the MSCI EM index at some date in the future, and the set of stocks from our control group. We have deleted stock-quarter observations for which we miss information on market capitalization, Amihud's illiquidity measure, or volatility. This happens for stocks suspended from trading, and also because for some treated A-shares, the IPO took place during our sample period.

Table 6: The determinants of passive holdings.

	Dependent variable: stock-level ownership by all passive funds in our sample.									
	A. Treated stocks only					**B. Treated + control stocks**				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inclusion Factor (IF)	1.834**	1.653**	1.683**	1.839**	1.637**	1.944**	1.359**	1.130**	1.936**	1.095**
	(50.390)	(49.569)	(49.584)	(50.109)	(49.740)	(140.257)	(91.436)	(58.455)	(140.082)	(59.256)
Ineligible						0.002	0.002	0.001	0.002	0.001
						(1.090)	(1.126)	(0.616)	(1.121)	(0.872)
log(MC)		0.001			0.002		−0.001			−0.00003
		(0.621)			(1.225)		(−0.737)			(−0.028)
IF:log(MC)		0.554**			0.380**		0.641**			0.472**
		(37.317)			(20.251)		(62.776)			(37.048)
Illiq			0.001		0.002			0.0003		0.001
			(0.663)		(1.407)			(0.366)		(1.030)
IF:Illiq			−0.690**		−0.380**			−1.404**		−0.702**
			(−35.572)		(−15.550)			(−55.840)		(−22.294)
Vol				−0.002	−0.001				−0.001	−0.0002
				(−0.861)	(−0.417)				(−0.856)	(−0.298)
IF:Vol				−0.128**	−0.191**				−0.130**	−0.181**
				(−7.219)	(−12.192)				(−13.373)	(−20.941)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,099	8,099	8,099	8,099	8,099	17,864	17,864	17,864	17,864	17,864
Adjusted R^2	0.238	0.407	0.361	0.246	0.435	0.612	0.699	0.671	0.617	0.712

Notes: This table reports the results of panel regressions in which we regress aggregate passive holdings in a given stock on a set of stock characteristics. In Panel A, the set of stocks only includes Chinese A-shares that at some point got included in the MSCI Emerging Markest Index. In Panel B, we also include the set of control stocks, i.e., Chinese A-shares that did not get included in the MSCI Emerging Markets Index during our sample period. The set of stock characteristics includes the stock's Inclusion Factor in the MSCI Emerging Markets Index (IF) which ranges between 0 and 20%, the logarithm of the stocks free-float adjusted market cap ($log(MC)$), Amihud's illiquidity measure ($Illiq$), and realized volatility of daily returns (Vol). In Panel B, we also add a time-invariant non-eligibility dummy ($Non-eligible$) that equals 1 for stocks that did not get included in the MSCI Emerging Markets Index during our sample period. T-statistics are shown between brackets. * p<0.05; ** p<0.01.

Table 7: The determinants of active holdings.

	Dependent variable: stock-level ownership by all active funds in our sample.									
	A. Treated stocks only					**B. Treated + control stocks**				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IF	3.473**	0.803	2.362**	3.328**	0.814	2.360**	−0.121	−0.548	2.319**	−0.053
	(7.139)	(1.636)	(4.810)	(6.749)	(1.641)	(12.296)	(−0.530)	(−1.900)	(12.052)	(−0.181)
Ineligible						−0.290**	−0.041	−0.241**	−0.286**	−0.039
						(−13.721)	(−1.592)	(−10.662)	(−13.533)	(−1.502)
log(MC)		0.343**			0.359**		0.211**			0.220**
		(14.291)			(12.307)		(15.761)			(14.598)
IF:log(MC)		0.992**			1.020**		2.077**			2.106**
		(4.539)			(3.607)		(13.258)			(10.544)
Illiq			−0.155**		0.026			−0.064**		0.015
			(−7.045)		(1.012)			(−6.104)		(1.275)
IF:Illiq			−1.647**		0.094			−4.911**		0.253
			(−5.874)		(0.256)			(−13.083)		(0.513)
Vol				−0.054*	−0.013				−0.024*	−0.023*
				(−2.053)	(−0.529)				(−2.308)	(−2.230)
IF:Vol				0.015	−0.190				−0.088	−0.107
				(0.064)	(−0.805)				(−0.648)	(−0.784)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,099	8,099	8,099	8,099	8,099	17,864	17,864	17,864	17,864	17,864
Adjusted R^2	0.004	0.062	0.023	0.005	0.062	0.040	0.082	0.053	0.040	0.083

Notes: This table reports the results of panel regressions in which we regress aggregate active holdings in a given stock on a set of stock characteristics. In Panel A, the set of stocks only includes Chinese A-shares that at some point got included in the MSCI Emerging Markest Index. In Panel B, we also include the set of control stocks, i.e., Chinese A-shares that did not get included in the MSCI Emerging Markets Index during our sample period. The set of stock characteristics includes the stock's Inclusion Factor in the MSCI Emerging Markets Index (IF) which ranges between 0 and 20%, the logarithm of the stocks free-float adjusted market cap ($log(MC)$), Amihud's illiquidity measure ($Illiq$), and realized volatility of daily returns (Vol). In Panel B, we also add a time-invariant non-eligibility dummy ($Non-eligible$) that equals 1 for stocks that did not get included in the MSCI Emerging Markets Index during our sample period. T-statistics are shown between brackets. * p<0.05; ** p<0.01.

Table 8: The relation between active and passive ownership.

	Dep. var.: stock-level ownership by all active funds in our sample.			
	A. Treated stocks only		**B. Treated + control stocks**	
	(1)	(2)	(3)	(4)
Passive	0.321^{**}	0.107^{**}	0.265^{**}	0.121^{**}
	(11.959)	(3.749)	(27.045)	(11.073)
log(MC)		0.412^{**}		0.285^{**}
		(17.313)		(24.045)
Illiq		0.042		0.034^{**}
		(1.787)		(2.961)
Vol		−0.018		-0.022^{*}
		(−0.869)		(−2.260)
Time FE	Yes	Yes	Yes	Yes
Observations	8,099	8,099	17,864	17,864
Adjusted R^2	0.015	0.061	0.039	0.077

Notes: This table reports the results of panel regressions in which we regress aggregate active holdings in a given stock on a set of stock characteristics. The set of stock characteristics includes the level of passive ownership ($Passive$), the logarithm of the stocks free-float adjusted market cap ($log(MC)$), Amihud's illiquidity measure ($Illiq$), and realized volatility of daily returns (Vol). T-statistics are shown between brackets. * p<0.05; ** p<0.01.

Table 9: Changes in market outcomes for A-shares included in the MSCI EM Index.

A. May 2018 rebalancing.

	$\Delta\beta$			ΔR^2			$\Delta IVOL$		
	Included	Control	DiD	Included	Control	DiD	Included	Control	DiD
Mean	0.2701**	0.1560**	0.1141**	0.1242**	0.0628**	0.0614**	0.0062**	0.0074**	−0.0012*
s.e.	0.0309	0.0246	0.0395	0.0122	0.0090	0.0151	0.0004	0.0003	0.0005

B. November 2019 rebalancing.

	$\Delta\beta$			ΔR^2			$\Delta IVOL$		
	Included	Control	DiD	Included	Control	DiD	Included	Control	DiD
Mean	−0.4406**	−0.6549**	0.2143**	−0.0681**	−0.1411**	0.0730**	0.0053**	0.0061**	−0.0008
s.e.	0.0479	0.0297	0.0564	0.0179	0.0103	0.0207	0.0004	0.0003	0.0005

C. Changes pooled across event dates.

	$\Delta\beta$			ΔR^2			$\Delta IVOL$		
	Included	Control	DiD	Included	Control	DiD	Included	Control	DiD
Mean	−0.0012	−0.1251**	0.1240**	0.0358**	−0.0215**	0.0574**	0.0035**	0.0048**	−0.0013**
s.e.	0.0258	0.0155	0.0301	0.0099	0.0055	0.0113	0.0003	0.0002	0.0004

Notes: We consider the following one-factor market model: $r_{it} = \alpha_i + \beta_i r_t^{\text{MSCI EM}} + \varepsilon_{it}$, where r_{it} is the return on stock i at time t, r_t^f denotes the risk-free rate, and $r_t^{\text{MSCI EM}}$ is the return on the MSCI Emerging Markets Index. We estimate these regressions using weekly overlapping returns. We consider 3-month pre- and post-event windows, and estimate the regression for both windows separately. This table shows the average changes in betas ($\Delta\beta$), the average changes in the regression's R-squared (ΔR^2), and the average changes in idiosyncratic volatility ($\Delta IVOL$), pooled across all 9 event dates in our sample. Here, idiosyncratic volatility is computed as the standard deviation of the residuals from the regression. We distinguish between the set of A-shares included in the MSCI Emerging Markets Index ($Included$) and the set of non-included A-shares ($Control$), and provide the corresponding difference-in-difference estimate (DiD). Panels A and B show the results for the main inclusion rounds separately (May 31^{st}, 2018 and November 30^{th}, 2019, respectively), while Panel C pools the results across all dates.

Table 10: Changes in betas with respect to the return index of non-included Chinese A-shares in a two-factor model.

	Included	Control	DiD
A. May 2018			
Mean	0.2196^{**}	0.3248^{**}	-0.1053^{**}
s.e.	0.0201	0.0135	0.0242
B. November 2019			
Mean	-0.0371	-0.0830^{**}	0.0459
s.e.	0.0294	0.0151	0.0330
C. Full Sample			
Mean	0.0488^{**}	0.0863^{**}	-0.0375^{*}
s.e.	0.0159	0.0083	0.0179

Notes: We consider the following two-factor market model: $r_{it} = \alpha_i + \beta_i^{EM} r_t^{\text{MSCI EM}} + \beta_i^{CN} r_t^{\text{Non-included A}} + \varepsilon_{it}$, where r_{it} is the return on stock i at time t, r_t^f denotes the risk-free rate, $r_t^{\text{MSCI EM}}$ is the return on the MSCI Emerging Markets Index, and $r_t^{\text{Non-included A}}$ is the return on the set of Chinese A-shares that are not included in the MSCI Emerging Markets Index. We estimate these regressions using weekly overlapping returns. We consider 3-month pre- and post-event windows, and estimate the regression for both windows separately. Panel A shows the average changes in betas with respect to the set of non-included Chinese A-shares ($\Delta\beta^{CN}$) pooled across all 9 event dates in our sample. We distinguish between the set of A-shares included in the MSCI Emerging Markets Index (*Included*) and the set of non-included A-shares (*Control*), and provide the corresponding difference-in-difference estimate (*DiD*). Panels B and C show the results for the main inclusion rounds separately (May 31^{st}, 2018 and November 30^{th}, 2019, respectively).

Table 11: Panel data analysis: cross-sectional heterogeneity in changes in market outcomes.

	Dep. var.: β_i					Dep. var.: R_i^2				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ineligible	−0.064**	−0.012	−0.033**	−0.070**	−0.015	−0.011**	0.007	−0.0001	−0.009**	0.010**
	(−7.719)	(−1.149)	(−3.551)	(−8.307)	(−1.447)	(−3.462)	(1.726)	(−0.016)	(−2.940)	(2.578)
Inclusion Factor (IF)	−0.112	−0.107	−0.434**	−0.043	−0.345**	0.280**	0.222**	0.223**	0.254**	0.075
	(−1.466)	(−1.156)	(−3.513)	(−0.563)	(−2.714)	(9.989)	(6.521)	(4.884)	(9.046)	(1.609)
log(MC)		0.045**			0.031**		0.015**			0.007**
		(8.338)			(5.045)		(7.481)			(3.219)
IF:log(MC)		−0.124*			−0.213**		0.022			0.021
		(−2.025)			(−2.755)		(0.988)			(0.748)
Illiq			−0.038**		−0.021**			−0.013**		−0.012**
			(−8.879)		(−4.300)			(−8.111)		(−6.661)
IF:Illiq			−0.511**		−0.684**			−0.088		−0.225**
			(−3.228)		(−3.220)			(−1.502)		(−2.891)
Vol				0.038**	0.035**				−0.017**	−0.019**
				(8.924)	(7.998)				(−10.634)	(−11.683)
IF:Vol				−0.132*	−0.160**				−0.213**	−0.226**
				(−2.426)	(−2.837)				(−10.697)	(−10.910)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,974	17,938	17,694	17,694	17,681	17,974	17,938	17,694	17,694	17,681
Adjusted R^2	0.003	0.007	0.008	0.007	0.013	0.011	0.016	0.015	0.031	0.039

Notes: This table shows the results of panel regressions in which we relate the treatment effects of inclusion to the MSCI Emerging Markets Index to a set of stock characteristics. In columns 1-5, the dependent variable is given by the beta of the treated share i. In columns 6-10, the dependent variable is given by the R-squared of the treated share i. Betas and R-squared in the dependent variable are computed with respect to the MSCI Emerging Markets Index. The set of stock characteristics includes the stock's Inclusion Factor in the MSCI Emerging Markets Index (IF) which ranges between 0 and 20%, the logarithm of free-float adjusted market cap ($log(MC)$), Amihud's illiquidity measure ($Illiq$), and realized volatility of daily returns (Vol). T-statistics are shown within brackets. * p<0.05; ** p<0.01.

Table 12: Results IV regressions.

	A. OLS results				B. IV results			
	β_i		R_i^2		β_i		R_i^2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ineligible	-0.053^{**}	-0.004	-0.012^{**}	0.015^{**}	-0.060^{**}	-0.009	-0.008^{**}	0.016^{**}
	(-6.642)	(-0.359)	(-4.268)	(4.262)	(-7.113)	(-0.880)	(-2.647)	(4.475)
Passive	0.042	0.004	0.137^{**}	0.098^{**}	-0.014	-0.089^{*}	0.173^{**}	0.114^{**}
	(1.422)	(0.125)	(12.877)	(8.856)	(-0.396)	(-2.273)	(13.092)	(8.078)
log(MC)		0.034^{**}		0.015^{**}		0.039^{**}		0.015^{**}
		(6.056)		(7.502)		(6.722)		(6.942)
Illiq		-0.023^{**}		-0.010^{**}		-0.023^{**}		-0.011^{**}
		(-4.823)		(-5.948)		(-4.666)		(-6.016)
Vol		0.048^{**}		-0.013^{**}		0.047^{**}		-0.013^{**}
		(11.921)		(-9.105)		(11.639)		(-8.956)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,538	17,538	17,538	17,538	17,538	17,538	17,538	17,538
Adjusted R^2	0.003	0.018	0.015	0.028	0.003	0.018	0.015	0.027

Note: This table shows the results of panel regressions in which we relate the treatment effects of inclusion to the MSCI EM index to (instrumented) ownership by passive funds that are benchmarked against the MSCI EM index. In columns 1-2 and 5-6, the dependent variable is given by the beta of the treated share i. In columns 3-4 and 7-8, the dependent variable is given by the R-squared of the treated share i. Betas and R-squared in the dependent variable are computed with respect to the MSCI Emerging Markets Index. The main independent variable is given by passive ownership ($Passive$). The set of additional stock characteristics includes the logarithm of free-float adjusted market cap ($log(MC)$), Amihud's illiquidity measure ($Illiq$), and realized volatility of daily returns (Vol). Panel A (columns 1-4) shows the results of OLS regressions. In Panel B (columns 5-8), we run IV regressions in which passive ownership is instrumented. Specifically, the first stage regression for columns 5 and 7 is given by:

$$Passive_{it} = \tilde{\alpha}_t + \tilde{\beta} IF_{it} + \tilde{\gamma}' \left(IF_{it} \cdot X_{i,t-1}\right) + \tilde{\eta} Ineligible_i + \tilde{\varepsilon}_{it}.$$

For columns 6 and 8, the first stage regression is given by:

$$Passive_{it} = \tilde{\alpha}_t + \tilde{\beta} IF_{it} + \tilde{\gamma}' \left(IF_{it} \cdot X_{i,t-1}\right) + \tilde{\delta}' X_{i,t-1} + \tilde{\eta} Ineligible_i + \tilde{\varepsilon}_{it}.$$

T-statistics are shown within brackets. * p<0.05; ** p<0.01.

Table 13: Summary statistics dual listings.

	Mean	St. Dev.	5%	50%	95%
Daily return (%)					
A	0.0630	1.6353	-2.2388	-0.0585	2.8762
H	0.0654	1.6958	-2.3994	-0.0416	2.9645
Free float (%)					
A	44.39	23.42	8.67	43.22	89.80
H	77.78	22.23	29.77	85.08	100
Free-float market cap (mln USD)					
A	8,606	12,609	1,102	5,016	24,548
H	6,376	13,009	467	2,062	34,025
Daily volume (mln USD)					
A	98.55	125.64	10.51	60.25	315.27
H	35.29	64.10	1.51	13.70	160.22
Volatility of daily return (%)					
A	1.64	0.50	1.07	1.70	2.78
H	1.70	0.51	1.44	1.76	2.22
Amihud's illiquidity					
(% per mln USD traded)					
A	0.4706	0.7777	0.0720	0.2651	1.0582
H	4.7332	13.6581	0.0725	1.3568	11.5521

Notes: This table presents summary statistics of the Chinese A-shares included in the MSCI EM index that have a dual listing in Hong Kong. Regarding returns, free float, market capitalization, and daily trading volume, we first computed cross-sectional summary statistics for each time period separately, after which we calculated the time averages. The volatility of daily returns and Amihud's illiquidity measure are computed on the stock level, for which we show cross-sectional statistics. Everything is expressed in USD.

Table 14: Ownership in dual-listed H-shares

	Dependent variable:					
	Passive Ownership			Active Ownership		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.073	0.106	0.095	1.498**	1.430**	1.365**
	(1.658)	(1.916)	(1.801)	(4.848)	(3.636)	(3.680)
$Elig^H$	2.780**	2.552**	2.669**	3.879**	3.482**	3.695**
	(54.482)	(37.356)	(42.295)	(10.872)	(7.175)	(8.334)
$Elig^A$		−0.085			0.177	
		(−0.956)			(0.280)	
$Elig^A : Elig^H$		0.406**			0.588	
		(3.942)			(0.804)	
Inclusion Factor (IF)			−0.410			2.490
			(−0.746)			(0.645)
$IF : Elig^H$			1.577*			1.654
			(2.508)			(0.374)
Observations	1,104	1,104	1,104	1,104	1,104	1,104
Adjusted R^2	0.729	0.738	0.732	0.096	0.098	0.098

Notes: This table reports the results of panel regressions in which we regress aggregate passive holdings in the set of dual-listed H-shares on a set of stock characteristics. The characteristics include a dummy that equals 1 when the H-share of firm i is included in the MSCI EM index at time t ($Elig^H_{it}$), a dummy that equals 1 when the A-share of firm i is included in the MSCI EM index at time t ($Elig^A_{it}$), and the inclusion factor for the MSCI EM index that applies to the A-share of firm i at time t (IF_{it}). T-statistics are shown between brackets. * p<0.05; ** p<0.01.

Table 15: Ownership in dual-listed A-shares

	Dependent variable:					
	Passive Ownership			Active Ownership		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.0004	0.0001	−0.003	0.296**	0.134**	0.142**
	(0.085)	(0.009)	(−0.531)	(7.200)	(2.592)	(2.917)
$Elig^H$		0.001	0.006	−0.134**	−0.025	−0.022
		(0.051)	(0.865)	(−2.828)	(−0.389)	(−0.378)
$Elig^A$	0.186**	0.211**			0.421**	
	(26.903)	(14.983)			(5.046)	
$Elig^A : Elig^H$		−0.031			−0.335**	
		(−1.899)			(−3.478)	
Inclusion Factor (IF)			1.582**			2.883**
			(24.560)			(5.706)
$IF : Elig^H$			−0.184*			−2.338**
			(−2.489)			(−4.039)
Observations	1,104	1,104	1,104	1,104	1,104	1,104
Adjusted R^2	0.396	0.398	0.659	0.006	0.030	0.036

Notes: This table reports the results of panel regressions in which we regress aggregate passive holdings in the set of dual-listed A-shares on a set of stock characteristics. The characteristics include a dummy that equals 1 when the H-share of firm i is included in the MSCI EM index at time t ($Elig_{it}^H$), a dummy that equals 1 when the A-share of firm i is included in the MSCI EM index at time t ($Elig_{it}^A$), and the inclusion factor for the MSCI EM index that applies to the A-share of firm i at time t (IF_{it}). T-statistics are shown between brackets. * p<0.05; ** p<0.01.

Table 16: Changes in market outcomes for dual-listed A-shares included in the MSCI EM Index.

	May 2018 rebalancing.								
	$\Delta\beta$			ΔR^2			$\Delta IVOL$		
	Included	Control	DiD	Included	Control	DiD	Included	Control	DiD
Mean	0.3456**	0.3087**	0.0369	0.1637**	0.1914**	−0.0276	0.0061**	0.0020*	0.0041**
s.e.	0.0571	0.0596	0.0652	0.0235	0.0263	0.0290	0.0007	0.0008	0.0007

Notes: We consider the following one-factor market model: $r_{it} = \alpha_i + \beta_i r_t^{\text{MSCI EM}} + \varepsilon_{it}$, where r_{it} is the return on stock i at time t, r_t^f denotes the risk-free rate, and $r_t^{\text{MSCI EM}}$ is the return on the MSCI Emerging Markets Index. We estimate these regressions using weekly overlapping returns. We consider 3-month pre- and post-event windows, and estimate the regression for both windows separately. Panel A shows the average changes in betas ($\Delta\beta$), the average changes in the regression's R-squared (ΔR^2), and the average changes in idiosyncratic volatility ($\Delta IVOL$) corresponding to the index rebalancing on May 31st, 2018. Here, idiosyncratic volatility is computed as the standard deviation of the residuals from the regression. We distinguish between the set of dual-listed A-shares that are included in the MSCI Emerging Markets Index (*Included*) on May 31st, 2018 and the corresponding dual-listed H-shares (*Control*), and provide the corresponding difference-in-difference estimate (*DiD*).

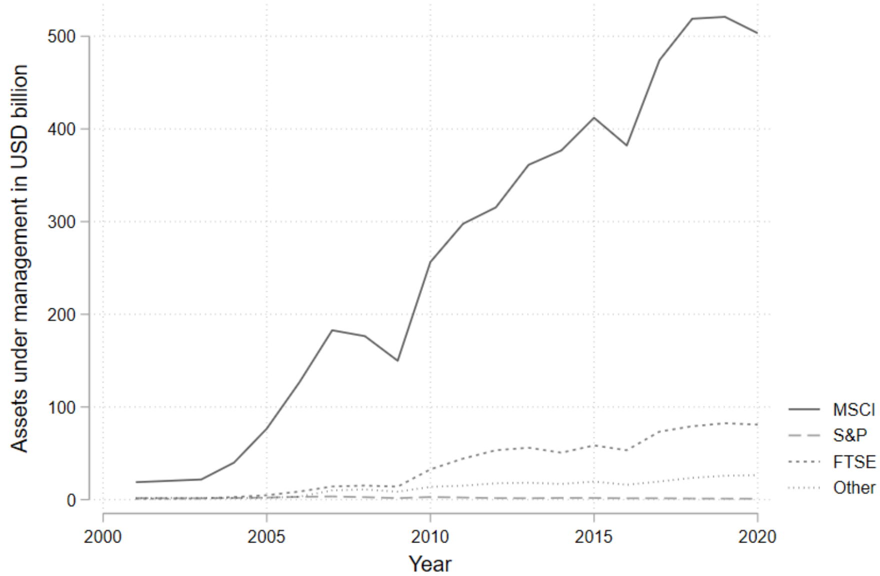
Table 17: Panel analysis on changes in market outcomes for dual-listed A-shares included in the MSCI EM Index.

	Dependent variable:	
	$\beta_{A,i} - \beta_{H,i}$	$R^2_{A,i} - R^2_{H,i}$
	(1)	(2)
Constant	-0.279^{**}	-0.139^{**}
	(-16.718)	(-17.480)
Inclusion Factor (IF)	0.130	0.260^{**}
	(0.883)	(3.690)
Time FE	No	No
Observations	989	989
Adjusted R^2	-0.0002	0.013

Notes: This table shows the results of panel regressions in which we relate financial market outcomes to the A-shares inclusion to the MSCI Emerging Markets Index. We focus on the set of A-shares being included in the MSCI Emerging Markets Index that have a dual-listed H-share traded in Hong Kong. In column 1, the dependent variable is given by the beta of the treated share i minus the beta of the dual-listed H-share ($\beta_{A,i} - \beta_{H,i}$). In column 2, the dependent variable is given by the R-squared of the treated share i minus the R-squared of the dual-listed H-share ($R^2_{A,i} - R^2_{H,i}$). Betas and R-squared in the dependent variable are computed with respect to the MSCI Emerging Markets Index. The independent variable is the stock's Inclusion Factor in the MSCI Emerging Markets Index (IF), which ranges between 0 and 20%. T-statistics are shown within brackets. * p<0.05; ** p<0.01.

Figures

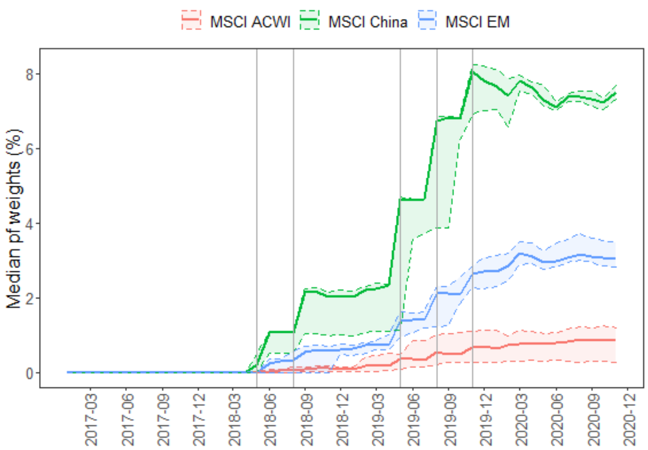
Figure 1: Assets under management in passive Emerging Markets funds by benchmark provider.



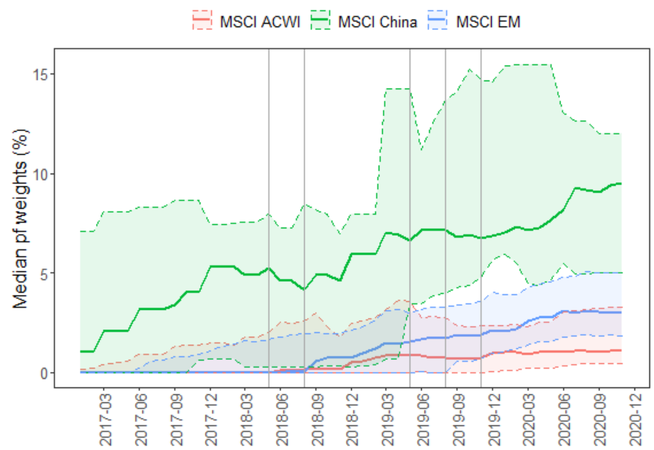
Notes: This graph shows the assets under mananagement in passive funds in Emerging Markets as of December of each year of global exchange traded funds broken down by the provider of the benchmark that these funds track.

Figure 2: Portfolio weights over time.

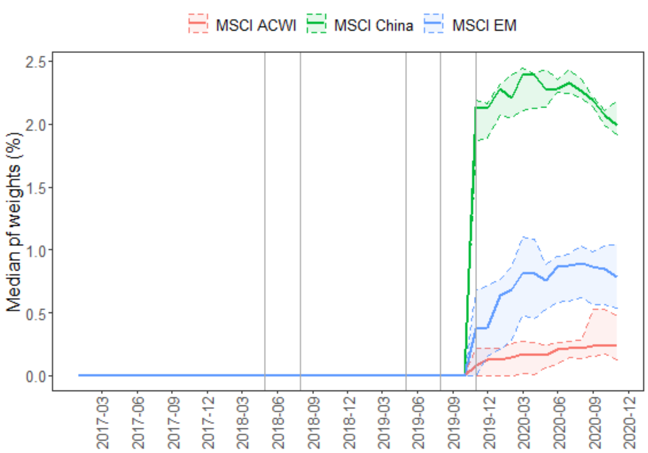
(a) Passive funds, A-shares included in May '18



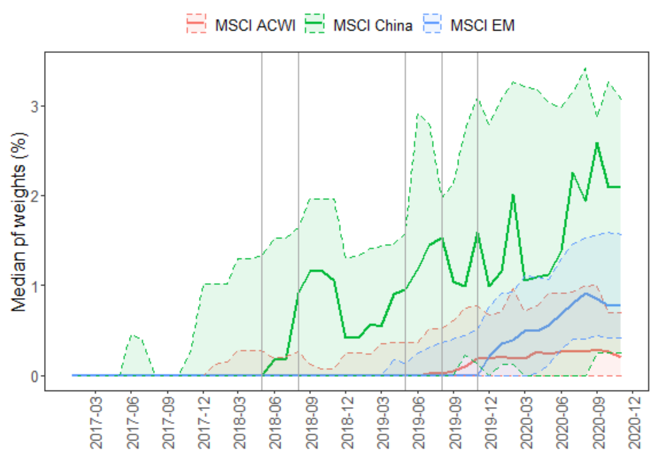
(b) Active funds, A-shares included in May '18



(c) Passive funds, A-shares included in Nov '19



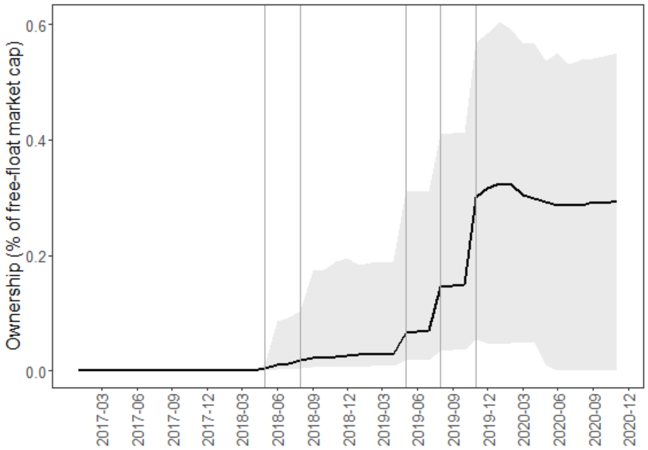
(d) Active funds, A-shares included in Nov '19



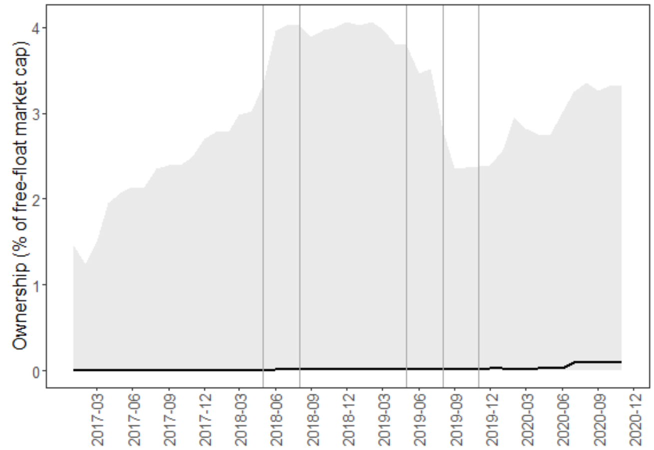
Notes: This figure shows the median and 25^{th} and 75^{th} percentiles of the portfolio weights in Chinese A-shares included in the MSCI Emerging Markets Index. Panels A and C show portfolio weights of passive funds, whereas Panels B and D show portfolio weights of the active funds in our sample. Panels A and B consider portfolio weights in the set of Chinese A-shares that are included in the MSCI Emerging Markets Index on May 31^{st}, 2018, whereas Panels C and D consider portfolio weights in the set of Chinese A-shares that are included in the MSCI Emerging Markets Index on November 30^{th}, 2019.

Figure 3: Ownership over time.

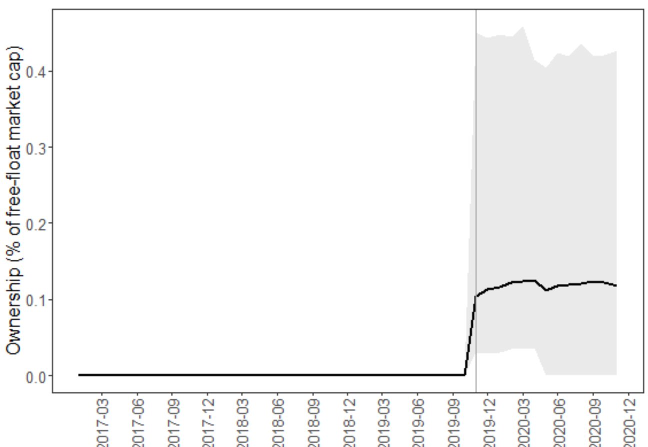
(a) Passive funds, A-shares included in May ’18



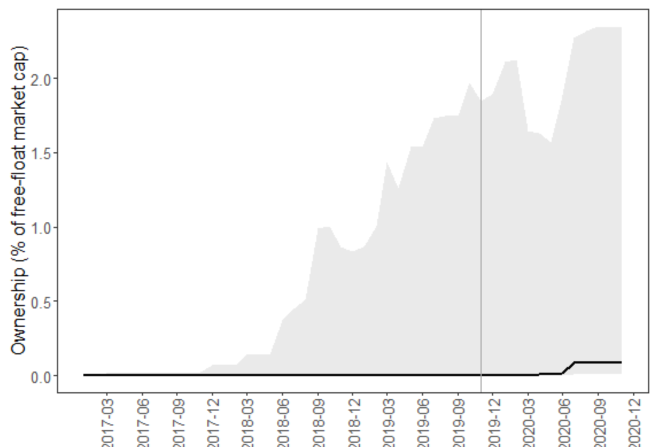
(b) Active funds, A-shares included in May ’18



(c) Passive funds, A-shares included in Nov ’19

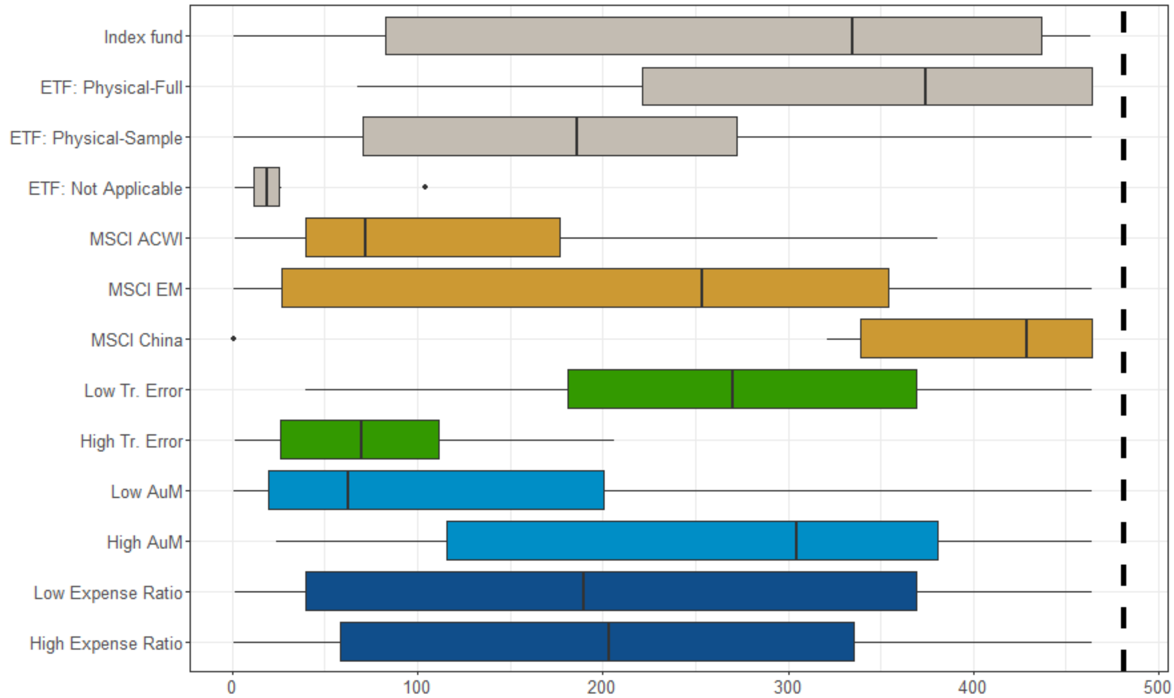


(d) Active funds, A-shares included in Nov ’19



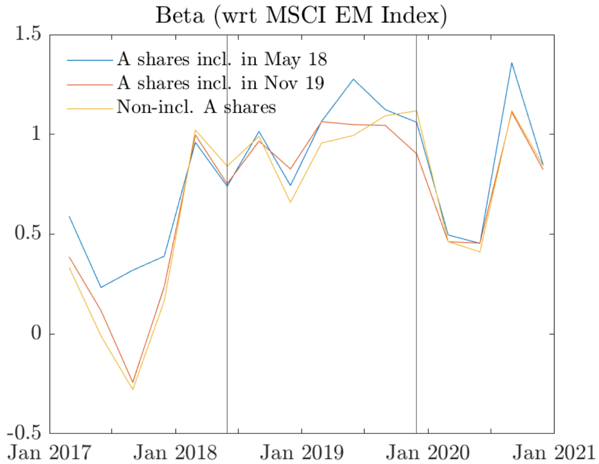
Notes: This figure shows the median and 5^{th} and 95^{th} percentiles of aggregate ownership in Chinese A-shares included in the MSCI Emerging Markets Index, expressed as percentage of free-float adjusted market capitalization. Panels A and C show stock-level ownership by passive funds, whereas Panels B and D show stock-level ownership by the active funds in our sample. Panels A and B consider ownership in the set of Chinese A-shares that are included in the MSCI Emerging Markets Index on May 31^{st}, 2018, whereas Panels C and D consider ownership in the set of Chinese A-shares that are included in the MSCI Emerging Markets Index on November 30^{th}, 2019.

Figure 4: Number of shares held by passive funds in December 2019.

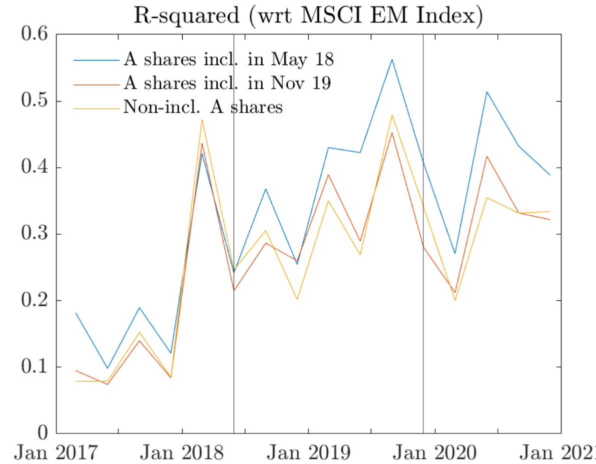


Notes: This graph shows the number of stocks held by passive funds in December 2019. At this point in time, 481 Chinese A-shares were included in the MSCI Emerging Markets Index (indicated by the vertical dashed line). We first split the passive funds in our sample by fund type, where we distinguish between traditional index funds, ETFs that physically replicate the full benchmark index, ETFs that involve in sampling, and ETFs for which the replication strategy is unknown. Secondly, we split funds by benchmark, where we distinguish between funds tracking the MSCI ACWI Index, the MSCI EM Index, and the MSCI China Index. Third, we split funds by their historical tracking error. Fourth, we split funds by their assets under management (AuM). Finally, we split funds by their reported expense ratios.

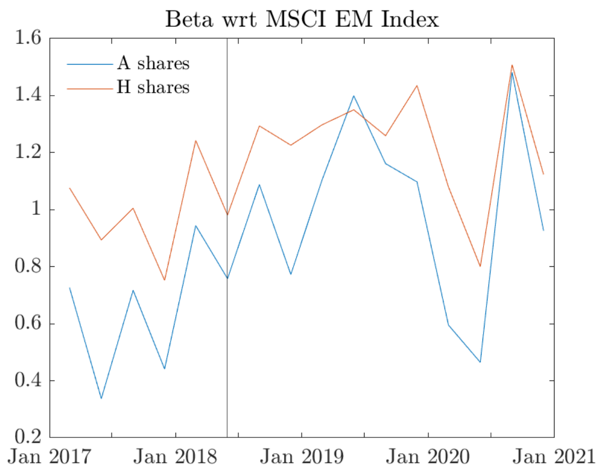
Figure 5: Trends in outcome variables.



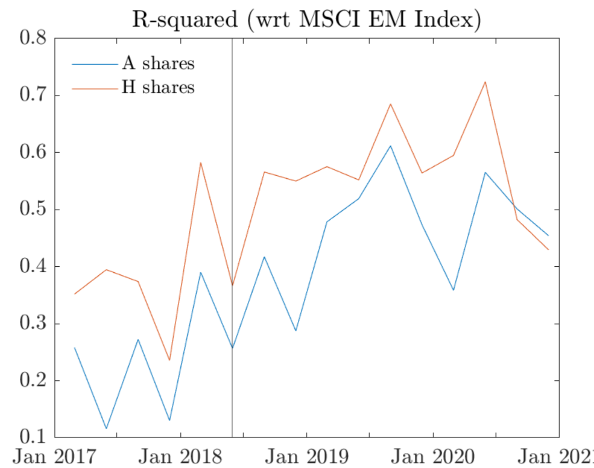
(a) Beta of A-shares



(b) R^2 of A-shares



(c) Beta of dual-listed A-shares



(d) R^2 of dual-listed A-shares

Notes: This graph shows the evolution of betas and R-squared, corresponding to the one-factor market model with the return on the MSCI EM index as the market factor. Betas and R-squared are obtained on a quarterly frequency using weekly overlapping returns. In Panels (a) and (c), we focus on betas, whereas Panels (b) and (c) show the evolution of R-squared over time. In Panels (a) and (b), we distinguish between the set of A-shares included in the MSCI EM index in May 2018, the set of A-shares included in November 2019, and the control group consisting of A-shares that have not been included in the MSCI EM index during our sample period. In Panels (c) and (d), we focus on pairs of dual-listed A- and H-shares, for which the A-share got included in the MSCI EM index during the index review of May 2018.

A Appendix: data

A.1 The universe of stocks

We first used Datastream to obtain the universe of Chinese stocks. Afterwards, we complemented the stock universe with stocks from Compustat.

A-shares

The Shenzhen Stock Exchange and the Shanghai Stock Exchange provide constituent lists for Chinese A-shares. The Datastream codes are as follows:

- **LCHSASHR** (A-shares listed on the Shanghai Stock Exchange)
- **LCHZASHR** (A-shares listed on the Shenzhen Stock Exchange)

We downloaded these constituent lists for each month starting in January 2016 to December 2020. The resulting number of unique A-shares with available ISIN equals **3928**.

H-shares

For Chinese stocks listed in Hong Kong (also called H-shares), we downloaded constituents using the Datastream code **HSHARES**. Unfortunately, it is impossible to download historical consituent lists using this code. We restricted the resulting set of stocks to stocks with exchange name equal to *"Hong Kong"*, resulting in **252** unique ISINs. On top of this, we extracted a list of H-shares from Datastream by selecting *Equities* for which *Exchange* equals *Hong Kong*, *Market* equals *China*, and *Currency* equals *Hong Kong Dollar*. This results in a selection of 406 stocks. After removing records with unavailable ISIN or Sedol, we end with **300** unique ISINs. After merging the two search results, we have **310** unique ISINs. Finally, we restricted the country of incorporation to be equal to 'CHINA' to remove Red Chips and P Chips, resulting in 283 unique ISINs.

Complementing the data with CompuStat

We complemented the stock universe from Datastream with stocks present in CompuStat but missing from Datastream. First, we downloaded all securities from CompuStat Global Daily. Using the query variable for the headquarter location ("LOC"), we restricted the search to only those securities for which the headquarter location is in China (LOC=CHN). In addition, we selected the month-end indicator and set it equal to 1, such that we extract information on a monthly frequency rather than daily to speed up the process. After filtering out securities with SIC code equal to 6722 (ETFs and other investment funds rather than equities), this yields 5005 stocks. Out of these, we extract 86 additional A-shares and 40 additional H-shares. Note that for the H-shares, I restricted the country of incorporation (FIC) to be equal to China to exclude any P Chips or Red Chips (which are stocks from firms incorporated outside China, with their main activities in China, and listed in Hong Kong).

Then, we also downloaded all securities from CompuStat Global Daily for which the country of incorporation ("FIC") is equal to "CHN", again with the additional filter that the month-end indicator is equal to 1 to speed up the process. This did not yield any additional A- and H-shares that were not in our sample yet.

In total, we thus end up with 4014 unique A-shares, and 323 unique H-shares.

Dual listings

We detect dual-listed A- and H-shares using the constituent list for the Hang Seng China AH Index (Datastream code **LHKHCHAH**). This results in 89 pairs of dual-listed A- and H-shares.

A.2 Stock characteristics

Static Variables

Next, we downloaded the following static variables for each ISIN (and for each Sedol in case of H-shares).[12] We downloaded the company name, the equity status, the currency in which the price is given, the name of the exchange on which the stock is listed, a China Connect Eligibility variable, the delisting date, the industry code, and the industry name.

Time Series Variables

Next, we used the ISINs (or Sedols for H-shares) to download historical data from January 1, 2016 to December 31, 2020 on a daily frequency for the following variables:

- Total Return Index (Datastream code **RI**).

[12]For listings abroad (in Hong Kong or in a different place), it happens that multiple listings share the same ISIN. Hence, when downloading additional characteristics based on ISINs, you do not always get data for the right listing in Datastream, which is why we use Sedols instead for downloading H-share characteristics.

- Market Value (Datastream code **MV**).
- Percentage of free float shares (Datastream code **NOSHFF**).
- Trading volume (Datastream code **VA**).

A.3 Matching FactSet with Morningstar

Matching open-end funds

1. We restricted the set of open-end funds in Morningstar to only those funds whose benchmark is one of the affected MSCI indices. This yields 985 unique funds, as measured by Morningstar's *FundId*.

2. We then matched these Morningstar funds from Step 1 with funds in FactSet using fund tickers. From the 985 unique funds in Step 1, 666 have been successfully matched (68%).

3. For each entity in FactSet that has one or more matched funds in Step 2, we checked whether it has additional funds that should be matched but are not. In the example below, I manually added a match for the Luxembourg variant of Morgan Stanley's Asia Opportunity fund.

Fund ID	Entity ID	Fund Name	Domicile	Matched
04JQYV-E	0CK66Z-E	Morgan Stanley Instl. Fund - Asia Opportunity Portfolio	US	Yes
04JN5J-E	0CK66Z-E	Morgan Stanley Invt. Fds. - Asia Opportunity Fund	LU	No

4. Not all Morningstar funds from Step 1 have been matched to FactSet. For these remaining funds in Morningstar, we manually matched them with FactSet based on fund names.
    - This initially yields a match for 164 additional funds from Step 1.
    - We then conducted two checks:
        - There could be an overlap in the manual matching in Steps 3 and 4.
        - Some manually matched FactSet funds could find a match to a Morningstar fund that tracks a different benchmark than the set of affected MSCI indices, thereby not belonging to the set of funds in Step 1.
    - After a small number of corrections resulting from these two checks, the total number of Morningstar funds matched to FactSet equals 825 (84%).

The example above already shows that multiple funds in FactSet can be matched with one unique Morningstar fund. The resulting dataset contains 1272 unique FactSet funds from 364 unique entities.

Matching exchange-traded funds

We proceed anologously to the matching of open-end funds as described above. We restrict attention to funds affected by the benchmark event, and try to match the funds by ticker and the exact name first, before engaging in manual matching. There are 144 unique ETFs in the Morningstar Direct database. Taking into account several fund classes, there are 464 ETFs with mostly unique tickers and names. Of these funds, 151 can be merged by ticker, whereas 14 can be merged by the exact fund names, whereas the remaining 279 funds in the Morningstar Direct database require manual matching.

In the FactSet database, we can identify the mapping of 120 of these 144 ETFs from Morningstar Direct. In some cases multiple FactSet funds match with the same MorningStar fund, which is why we can identify 127 FactSet funds with a mapping to Morningstar Direct. We drop some of these funds because they do not report any Chinese equity holdings or because these funds stopped reporting any holdings data before 2016. Overall, the set of ETFs included in our sample for which we observe holdings in FactSet over the sample period and for which we can identify a mapping to Morningstar Direct is 72.

A.4 MSCI benchmarks

In the following, we list the MSCI benchmarks tracked by open-end funds and ETFs, which were affected by the A-shares inclusion. We exclude all composite benchmarks that present a combination of stock and bond indices, restricting attention to stock indices.

MSCI All Country World

MSCI AC Asia Ex Japan NR AUD; MSCI AC Asia Ex Japan NR USD; MSCI AC Asia Pac Ex JPN NR USD; MSCI AC Asia Pacific USD; MSCI AC Far East ex Japan NR EUR; MSCI AC Far East Ex Japan NR USD; MSCI ACWI; MSCI ACWI 100% Hdg NR USD; MSCI ACWI Ex Australia GR AUD; MSCI ACWI Ex Australia Hdg NR AUD; MSCI ACWI ex Canada IMI NR CAD; MSCI ACWI Ex Japan GR USD; MSCI ACWI ex US 100% Hedged NR USD; MSCI ACWI

Ex USA GR USD; MSCI ACWI Ex USA IMI NR USD; MSCI ACWI Ex USA NR USD; MSCI ACWI IMI NR USD; MSCI ACWI NR AUD; MSCI ACWI NR CAD; MSCI ACWI NR EUR; MSCI ACWI NR USD; MSCI ACWI PR USD; MSCI ACWI with Dev Mkts 100% Hdg NR USD; MSCI All Country World; MSCI ACWI All Cap NR USD; MSCI ACWI Ex Australia NR AUD; MSCI ACWI Ex USA PR USD; MSCI ACWI GR USD

MSCI Emerging Market

MSCI EM NR USD; MSCI EM Asia NR USD; MSCI EM Asia NR LCL; MSCI EM NR CAD; MSCI EM Hdg NR USD; MSCI EM Far East NR USD; MSCI EM IMI NR USD; MSCI EM Leveraged 2X Daily NR USD; MSCI EM GR CAD; MSCI EM Equal Country Weighted GR USD; MSCI EM NR GBP; MSCI EM GR USD; MSCI EM PR JPY; MSCI EM NR EUR; MSCI EM IMI NR CAD; MSCI EM 100% Hdg NR USD; MSCI EM IMI GR USD; MSCI EM PR USD; MSCI EM Asia NR EUR; MSCI EM IMI PR USD; MSCI Emerging Markets; MSCI Emerging Markets (KRW); iShares MSCI EM UCITS ETF (Acc)

MSCI China

MSCI China NR HKD; MSCI China NR USD; MSCI China PR HKD; MSCI China PR USD; MSCI China TR ZAR

B Appendix: derivation of the variance to compute t-statistics

The variance of $\overline{\Delta\hat{\beta}}_{Included}$ is given by:

$$V(\overline{\Delta\hat{\beta}}_{Included}) = V\left(\sum_t w_t\left(\frac{\sum_{i=1}^{I_t}\Delta\hat{\beta}_{it}}{I_t}\right)\right) \tag{16}$$
$$= \sum_t w_t V\left(\frac{\sum_{i=1}^{I_t}\Delta\hat{\beta}_{it}}{I_t}\right) \tag{17}$$
$$= \sum_t w_t \frac{1}{(I_t)^2} V\left(\sum_{i=1}^{I_t}\Delta\hat{\beta}_{it}\right) \tag{18}$$
$$= \sum_t w_t \left(\frac{V\left(\Delta\hat{\beta}_{it}\right)}{I_t}\right). \tag{19}$$

We allow $V\left(\Delta\hat{\beta}_{it}\right)$ to differ across inclusion rounds. The variance of $\overline{\Delta\hat{\beta}}_{Included} - \overline{\Delta\hat{\beta}}_{Control}$ is given by:

$$V(\overline{\Delta\hat{\beta}}_{Included} - \overline{\Delta\hat{\beta}}_{Control}) = V\left(\sum_t w_t\left(\frac{\sum_{i=1}^{I_t}\Delta\beta_{it}}{I_t} - \frac{\sum_{j=1}^{J_t}\Delta\beta_{jt}}{J_t}\right)\right) \tag{20}$$
$$= V\left(\sum_t w_t\left(\frac{\sum_{i=1}^{I_t}\Delta\beta_{it}}{I_t}\right)\right) + V\left(\sum_t w_t\left(\frac{\sum_{j=1}^{J_t}\Delta\beta_{jt}}{J_t}\right)\right) \tag{21}$$
$$= \sum_t w_t V\left(\frac{\sum_{i=1}^{I_t}\Delta\beta_{it}}{I_t}\right) + \sum_t w_t V\left(\frac{\sum_{j=1}^{J_t}\Delta\beta_{jt}}{J_t}\right) \tag{22}$$
$$= \sum_t w_t \frac{1}{I_t} V\left(\Delta\beta_{it}\right) + \sum_t w_t \frac{1}{J_t} V\left(\Delta\beta_{jt}\right) \tag{23}$$
$$= \sum_t w_t \left(\frac{V\left(\Delta\beta_{it}\right)}{I_t} + \frac{V\left(\Delta\beta_{jt}\right)}{J_t}\right). \tag{24}$$

Here, we allow $V\left(\Delta\hat{\beta}_{it}\right)$ and $V\left(\Delta\hat{\beta}_{jt}\right)$ to be different from each other, and to differ across inclusion rounds.