Optimal Index-Linked Rebalancing with Anticipatory Trading*

Stefano Pegoraro† Marco Sammon‡ John J. Shim§

August 27, 2026

Abstract

Using a model of index-linked rebalancing around reconstitution events, we show front-runners provide liquidity to index investors and benefit them. Index investors trade off execution costs against tracking-error concerns, while speculators maximize trading profit. In competitive markets, even loose index trackers optimally rebalance at index reconstitution, rationalizing concentrated reconstitution-day trading with little contemporaneous price impact. Empirically, index-linked investors bear small rebalancing costs relative to standard transaction-cost estimates and traders short sell shares to provide liquidity at reconstitution. We show fast-track additions increase IPO issue prices but impose costs on index investors when the newly-listed stocks are illiquid and hard to short.

Keywords: index rebalancing, index funds, passive investing, benchmark tracking, tracking error, front running, anticipatory trading, liquidity provision, price impact, fast-track additions, IPO.

*We thank Adem Atmaz, Rob Battalio, Patrick Blonien, Thomas Eisenbach, Miklos Farkas, Alessandro Melone, Taisiya Sikorskaya, Adi Sunderam, and audiences at the Bristol Financial Markets Conference, Penn State Happy Valley Finance Conference, University of Washington Summer Finance Conference, Wabash River Finance Conference, Vanguard, Chicago Fed, and Washington University in St. Louis for helpful comments.

†Mendoza College of Business, University of Notre Dame, United States of America. Email: s.pegoraro@nd.edu

‡Harvard Business School, Harvard University, United States of America. Email: mcsammon@gmail.com

§Mendoza College of Business, University of Notre Dame, United States of America. Email: jshim2@nd.edu

1 INTRODUCTION

As of the end of 2025, investors had delegated $19 trillion to funds whose sole mandate is to track an index. These index funds hold 19% of total U.S. equity market capitalization.[1] These figures likely understate the true scale of index-linked investing: pension funds, active funds, and other institutional investors also benchmark their portfolios to indexes (Pavlova and Sikorskaya, 2023). Because index providers regularly change the composition of their indexes, index-linked investors must trade to keep their portfolios aligned with their benchmark. The growth of index-linked investing has made these rebalancing trades enormous: on the exact date when S&P and Russell index changes are implemented, index trackers often trade more than 20% of a stock's shares outstanding in the closing auction of that day (Chinco and Sammon, 2024). The sheer scale and predictability of these rebalancing trades have attracted speculators looking to earn easy profits.

On the surface, the speculators' strategy is simple: if index-linked investors buy index additions and sell index deletions on the day when changes go into effect, buy/sell these stocks in anticipation of the predictable demand. Even if the speculators' trading affects prices, they will be able to offload their positions at the new prices to the index-linked investors. That is, index investors will rebalance by buying high and selling low. And with the growth of index investing, a sizable mass of speculators has emerged alongside.[2] This type of trading, sometimes viewed as predatory and colloquially referred to as "front-running," gives the impression that benchmark-linked investors unwittingly bear a cost of rebalancing. For example, Harvey et al. (2025) estimate that institutional investors following an asset allocation strategy with 60% in stocks and 40% in bonds incur a cost of $16 billion per year from mechanically and predictably rebalancing.

Why are index-linked investors willing to pay these costs? If the perceived costs are large enough, even index funds that strictly track an index and meticulously manage their tracking error should be willing to bear some tracking error in favor of reducing costs by, say, smoothing out their rebalancing trades. What is perhaps more puzzling is that even flexible index-linked investors seem to trade on index implementation day and account for as much volume on those days as the index funds that have relatively strict mandates (Chinco and Sammon, 2024). Surely fund managers are aware of speculators – why do they continue to trade in such a predictable fashion?

In this paper, we build a model of optimal index-linked rebalancing to better under-

[1] See the 2026 Investment Company Institute Factbook for details.
[2] For example, see articles from Bloomberg and Business Insider.

stand the trade-offs faced by an index-linked fund, the role of speculators, and under what conditions the fund is made worse off by speculators. We find that the fund is almost always made better off with the presence of competitive speculation – the only exception is when the fund has very little concern over tracking error with respect to the index (i.e., the most flexible index-linked investors). Our model clarifies that we can interpret speculators as liquidity providers, and the fund can coordinate this liquidity by committing to trading on the day when index changes come into effect. That is, there are gains from trade: since speculators do not have tracking error concerns, they can reduce transaction costs by spreading out trading over time and handing over nearly all the necessary shares to the fund on the rebalancing date to reduce tracking error. And while a fund with minimal tracking error concerns is made worse off in a world with speculators, the fund's optimal strategy in this world is still to concentrate all of its trading on the index rebalancing date.

Before describing the details of the model, we highlight and update two facts from the literature that, when taken together, already suggest that index-implementation day trading involves intermediation. First, index-rebalancing trading volume is concentrated on the day when index changes become effective and can be as much as 25 times these stocks' past average daily volume (ADV) (Chinco and Sammon, 2024). Second, despite this extraordinary volume, average implementation-day returns for direct S&P 500 additions were -9 bps between 2010 and 2024 (Greenwood and Sammon, 2025). These two facts show that index rebalancing can be concentrated on a single day with seemingly no concession required by counterparties. To the extent that there is a price concession, it comes from pre-implementation day returns. For both S&P and Russell index rebalancing events, stocks eventually added (dropped) exhibit increasing (decreasing) prices leading up to the implementation day. At the very least, this pattern suggests that the index rebalancing events are predictable by market participants.

The model features an index-tracking fund that must sell a stock that is leaving the index and buy a stock that is being added to the index on or in the days around the index implementation date. The fund can split its order across periods, but incurs a tracking-error cost if it trades outside the implementation date. Strategic speculators trade to maximize profits. By trading, all agents generate transitory and persistent price impact, which they internalize. We solve the model under two equilibrium concepts. Under no commitment, the fund chooses its schedule while taking speculators' positions as given. Under full commitment, the fund can credibly pre-announce its trading schedule and speculators respond to that announcement.

The fund faces a trade-off between avoiding tracking error and reducing transaction

costs, and the latter includes the rents that accrue to speculators. When tracking error concerns are sufficiently strong, the fund's unique optimum is to rebalance entirely on the implementation day. Furthermore, if a fund can credibly commit to a trading schedule, concentrated trading on the implementation day is optimal under even weaker tracking-error cost conditions. By committing to a concentrated trading schedule on the implementation day, the fund can coordinate the provision of liquidity on that day, thereby minimizing tracking error while also reducing price impact. In contrast, if a fund cannot credibly commit, the temptation to rebalance outside the reconstitution date crowds out speculator trading (i.e., intermediary liquidity provision) at reconstitution, raising the costs for the fund. Through this lens, speculator activity is a form of intermediation or liquidity provision rather than a purely predatory activity that harms index-linked investors.

For any level of tracking-error concerns, the fund increasingly trades on the implementation date as competition between speculators intensifies. In particular, if a fund can credibly commit to a trading schedule, we show that, for *any* degree of tracking-error concern, concentrated trading on the implementation date becomes optimal when competition between intermediaries is sufficiently intense. Intuitively, as speculators compete more fiercely, the rents they can extract fall to zero. By committing to this concentrated trading schedule, the fund enjoys both zero tracking error and cheap liquidity provided by competitive intermediaries.

We then extend the model to allow for exogenous stochastic flows at and around the implementation date and free entry by speculators. The extension allows us to study how a relatively small investor with some tracking-error concerns trades when large order flows from "strict" index investors are expected to arrive to the market on the implementation date. We show that, as exogenous order flows scale up, even weak index trackers with no commitment power fully rebalance at implementation. With free entry, as exogenous order flows increase, more speculators enter, placing a bound on the potential gains an index tracker could obtain by actively trading before or after implementation. However, as flows scale up, the risk of accruing substantial tracking error also scales. When tracking error costs outweigh the gains from active trading, the fund shifts to implementation-date rebalancing.

The model delivers a set of results that jointly rationalize the observed separation between the timing of returns and quantities around an index implementation date. As speculators gradually pre-position shares and accumulate inventory, prices increase ahead of the implementation date. When they sell to the fund on the implementation date, prices change only modestly despite the extraordinarily large trading volume, consistent with

empirical evidence in Greenwood and Sammon (2025) and Chinco and Sammon (2024).

We use our model to evaluate the welfare implications of fast-tracking the addition of a stock to an index shortly after the initial public offering (IPO). Fast-track additions are particularly costly for index investors when the stock is illiquid and short selling is constrained. Both conditions are likely to apply shortly after an IPO. With less time to build up inventory, speculators incur large transaction costs, intermediating a lower share of the volume and forcing index investors to suffer steeper price reversals. In contrast, fast-track additions are beneficial for issuers and initial stock owners, especially when speculators are highly competitive. To build inventory in a short time period, competitive speculators bid up the stock's price as early as possible, boosting IPO prices.[3]

The model predicts that speculators acquire shares over multiple periods to minimize their price impact. As a part of this strategy, speculators also acquire shares *after* the implementation day. That is, they short-sell on the implementation day and cover the short in the subsequent periods. This prediction clarifies a way to empirically isolate speculator activity: short interest and securities lending data. Short-selling data allow us to isolate genuine intermediation activity from generic asset sales (such as those by mutual funds and active investors attempting to "sell high" to generate alpha). We find that short interest rises on implementation dates and then falls afterwards as speculators cover, consistent with the model's predictions. We also document the reverse pattern for negative index-induced demand shocks.

In an extension of the model, we examine how increasing the number of periods impacts trading volume around the implementation date. When speculators have fewer trading periods to pre-position ahead of implementation, trading volume becomes concentrated on days closer to implementation. We test this prediction using migrations between the Russell 1000 and 2000 indexes. Russell uses a transparent rank formula and a predetermined rank cutoff to determine which stocks will switch indexes and which will remain in their existing index. Some stocks are well beyond the cutoff and the market will correctly infer that these stocks are certain to switch, even before the official rank date (i.e., announcement date). Other stocks are closer to the cutoff, leaving open the possibility that a return shock for that stock or other stocks close to the cutoff might still affect the index they will be assigned to on the rank date.

[3]This analysis also at least partially applies to rebalancing a 60/40 stock/bond portfolio. As described in Harvey et al. (2025), when these portfolios become skewed towards stocks or bonds from differences in returns, they will need to rebalance to get back towards the 60/40 target. In this scenario, competitive speculation may push prices back towards the 60/40 target and prevent rebalancing in the first place. This limits speculator inventory accumulation, similar to how greater trading costs and a shorter pre-positioning period impact speculator activity in fast-track index additions.

We use a measure of the pre-rank-date predictability as a way of capturing the number of periods available for speculators to pre-position shares for implementation-day trading. That is, stocks far from the cutoff are available for pre-positioning well in advance of the rank date. For stocks near the cutoff, speculators are much more likely to wait until closer to the rank date or even the rank date itself to pre-position shares. We find that less predictable migrations have more trading between the rank and implementation dates, and in the period after the implementation date, consistent with speculators having less time to pre-position and therefore needing to intermediate over more constrained time frames.

Finally, the model provides a simple measure of index-linked investor costs: the post-implementation price reversal. Intuitively, price reversals measure the *avoidable* transaction cost borne by index investors. Empirically, we show that post-implementation reversals for S&P 500 direct additions have fallen sharply over time, and that the decline is even more pronounced when reversals are scaled by the size of the index demand shock, as the model prescribes. Strikingly, despite trading enormous quantities, index funds trade very cheaply – a back-of-the-envelope calculation suggests that the cost of trading 2,500% of average daily volume is around 1%. For comparison, Frazzini et al. (2012) estimate the cost of trading 11% of average daily volume at 0.3%.

Our model and empirical results highlight the role of speculators as intermediaries. To assess the welfare of index-linked investors, we must also consider a counterfactual world without speculators. Index funds and benchmark-linked investors may incur even greater costs without speculators, either in terms of greater tracking error or higher trading costs.

RELATED LITERATURE. This paper sits at the intersection of research on the index effect, benchmark-driven demand, and strategic liquidity provision.

A first related literature studies price pressure around index additions and deletions. Early work in this area documents significant abnormal returns around reconstitution dates (Harris and Gurel, 1986; Shleifer, 1986; Petajisto, 2011). More recent evidence shows that, in major indices, announcement-day and implementation-day effects have weakened while the price response has shifted earlier (Greenwood and Sammon, 2025). Our contribution is to emphasize why the timing of *prices* and *quantities* has diverged and to offer a mechanism for that divergence. In our model, benchmark-linked investors continue to concentrate their trading on the implementation date, but intermediaries spread their own trading across the surrounding window by building inventory in advance and unwinding short positions afterward. This allows the intermediary to effectively act as

an agent for the index fund, smoothing the trades – and thus the returns – over time.

A second literature studies execution and anticipatory trading. Standard models predict gradual trading if the informed investor is expected to have price impact (Kyle, 1985; Almgren and Chriss, 2001), while theories of concentrated trading emphasize coordination during periods of deep liquidity (Admati and Pfleiderer, 1988). Related work highlights the benefits of commitment power in execution (Shim and Wang, 2025), and that predictable trading can invite strategic responses by rent-extracting intermediaries (Brunnermeier and Pedersen, 2005). We bring these ideas together in an index-reconstitution setting where benchmark-tracking concerns create a motive to trade on the implementation date, and anticipatory traders endogenously provide liquidity by pre-positioning for expected demand. In our model, synchronized execution at implementation is an equilibrium outcome because off-event fund trading crowds out intermediary liquidity provision at the close.

A closely related paper is Harvey et al. (2025), who study predictable rebalancing by institutional investors that aim to maintain target stock-bond allocations. They estimate that these predictable cross-asset trades impose costs of $16 billion per year on investors. Their setting, however, differs from ours in two important ways. First, their rebalancing occurs *across* asset classes, whereas ours occurs *within* equities. This distinction matters because the price impact of cross-asset rebalancing is likely governed by a macro-elasticity, while within-equity index rebalancing is likely governed by a micro-elasticity (Gabaix and Koijen, 2021). Because intermediating aggregate stock-market exposure is riskier than intermediating relative demand across individual stocks, macro elasticities are likely smaller than micro elasticities, making expected price impact larger in their setting. Further, their $16 billion estimate captures the total cost of predictable cross-asset rebalancing, including price pressure generated by the rebalancing investors' own trades. By contrast, in our setting, much of the (avoidable) price impact is generated by intermediaries who trade in anticipation of index demand. Our model therefore tells us to focus on the avoidable component of implementation shortfall: the part of price pressure that reverses after the implementation date.

A third literature studies the rise of passive and benchmarked investing and the extent of inelastic demand (Koijen and Yogo, 2019; Coles et al., 2022; Pavlova and Sikorskaya, 2023). Benchmark-driven demand has grown substantially, making it important to understand how markets have adapted to large, predictable rebalancing trades. Consistent with an ecosystem that has adapted to these predictable events (Lo, 2004), trading volume is extraordinarily concentrated on the reconstitution date itself (Chinco and Sammon, 2024). Our model helps rationalize this concentration by emphasizing a coordina-

tion channel that is specific to pre-announced and predictable index events. Specifically, index providers like Russell and S&P communicate rules, calendars, and implementation details in advance. This transparency supports a commitment-like environment in which intermediaries can build inventory ahead of the implementation date and supply liquidity when index funds trade to minimize tracking error.

Finally, our paper can inform index design, especially in settings where pre-positioning is constrained. Our mechanism suggests that concentrated rebalancing is less costly when index changes are predictable and intermediaries have time to build inventory before implementation. By contrast, short implementation windows can limit pre-positioning and concentrate price pressure on the implementation date. Consistent with this logic, Murray and Sammon (2026) show that fast-track IPO inclusions, which are costly to short and do not have many periods between IPO and index inclusion, amplify implementation-day price impact and the subsequent reversals relative to a world with fewer frictions for intermediaries.

2 Data and Variables

This subsection describes the data sources, sample construction, and the event-time datasets used in our analyses of index front-running and rebalancing around S&P and Russell index reconstitution. We construct the data separately for (i) S&P direct additions and deletions, (ii) S&P tier migrations (e.g., between large-, mid-, and small-cap indices), (iii) Russell migrations between the Russell 1000 and Russell 2000, and (iv) Russell direct additions and deletions from outside the Russell universe. Constructing the data separately reduces the risk of overlap, as some stocks may be added to or dropped from multiple indexes within the same event window.

We also clarify definitions and the language used in this paper. We use the term *reconstitution* to indicate the change in the composition of a certain index. In the context of the S&P 500 index, a reconstitution involves the *addition* of a stock to the index and the *deletion* of a stock. Many additions and deletions, particularly in the context of the Russell 1000 and 2000 indexes, are *migrations* between indexes.

We use the term *implementation date* to refer to the trading day when the index provider implements the reconstitution and, hence, the updated index becomes effective. We use the term *rebalancing* to specifically indicate the trading activity of index-tracking investors who trade on or near the implementation date in order to align their portfolio holdings with the updated index.

2.1 DATA

S&P 500 RECONSTITUTION EVENTS. S&P 500 additions and deletions are based on index-change lists provided by the index provider and include both announcement and effective dates. We also identify a set of unusual cases that are not typical index changes because they are driven by corporate events such as mergers and associated delistings. In the baseline analysis we remove these unusual cases. In some analyses we additionally separate S&P changes associated with regularly scheduled quarterly rebalances from ad hoc changes that occur when a slot opens up due to a merger or other corporate event.

The S&P event data also indicate whether an event reflects a movement across S&P size tiers. In our baseline analyses, we treat migrations as movements between the S&P 500 and the mid-cap 400 or small-cap 600. In many analyses, we separate these from stocks that enter from or exit to outside the S&P 1500 index family.

RUSSELL 1000/2000 MIGRATIONS. Russell migration events are based on index-change lists provided by the index provider and Russell reconstitution files that provide the annual rank date (when membership is determined) and the effective reconstitution date (when index constituent changes take effect). We classify each event as either a move into the Russell 2000 from the Russell 1000, or the reverse move into the Russell 1000 from the Russell 2000. We separately analyze Russell direct events. These are cases where firms enter or exit the Russell 1000 or Russell 2000 from outside the Russell universe.

INDEX RETURNS For return-based outcomes, we benchmark each event against the return of the relevant index. For S&P events, we use the S&P index return that matches the event (e.g., S&P 500 for S&P 500 additions/deletions and S&P tier returns for tier migrations). For Russell events, we use the Russell 1000 return for Russell 1000 additions and the Russell 2000 return for Russell 2000 additions, and analogously benchmark migration events against the destination/source index as appropriate. We obtain all daily total index return data from Datastream.

STOCK-LEVEL DATA AND INSTITUTIONAL HOLDINGS Our daily stock-level outcomes come from CRSP and are organized at the security-trading-day level.

2.2 MAIN DATA SET

We build a single main data set at the security-trading-day level that combines CRSP data with the reconstitution event-time datasets. Unless stated otherwise, the baseline event-

time panels require a complete event window for each stock and event. Specifically, we require the stock to have trading-day data for all $\tau \in [-120, +20]$ relative to the implementation date. This mechanically excludes firms that list recently or otherwise do not have a full pre-reconstitution history, such as some spinoffs and corporate events associated with mergers and related index changes.

TRADING CALENDAR AND DATE ALIGNMENT Index dates (announcement, rank, effective) can fall on weekends or market holidays. Our convention is to map any non-trading event date to the next trading day (e.g., Saturday to Monday). We implement this using a CRSP-based trading calendar and a next-trading-day map from calendar dates to valid CRSP trading dates. This alignment is applied before constructing event windows.

EVENT-TIME PANEL CONSTRUCTION For each reconstitution event group we construct a long event-time panel that indexes observations by event, by the implementation date, and by event time (days relative to the anchor). We use the same procedure for each group. First, we clean the raw event information and align announcement, rank, and implementation dates to trading days. Second, we assign a unique event identifier and create separate versions of each event anchored on the relevant dates (announcement versus implementation for S&P, and rank versus implementation for Russell). Third, we expand each anchored event into a fixed event window and map event time to trading dates using the trading calendar. Lastly, we merge in daily outcomes from the main data set.

SAMPLE SELECTION AND TARGET EVENT BUCKETS Our baseline figures and tables group events into four conceptually distinct categories. First, for the S&P 500 we study direct additions and direct deletions. Second, we study S&P tier migrations between the S&P 500 and adjacent size tiers. Third, for Russell we study direct additions and direct deletions into the Russell 1000 and into the Russell 2000, meaning entry from or exit to outside the Russell universe. Fourth, we study Russell migrations between the Russell 1000 and Russell 2000. For S&P analyses, we exclude unusual cases (e.g., those driven by listings, delistings, mergers, acquisitions, etc.) by default, and conduct robustness checks also excluding ad-hoc additions/deletions (i.e., additions/deletions which do not occur at the regular quarterly rebalancing events).

2.3 EVENT-TIME OUTCOME VARIABLES

This subsection describes the construction of the main event-time outcomes used in the analysis. These include scaled trading volume and cumulative abnormal returns.

SCALED VOLUME Scaled volume is defined as shares traded divided by shares outstanding, or

$$scaledvolume_{it} = \frac{vol_{it}}{shrout_{it} \cdot 1000}.$$

CUMULATIVE ABNORMAL RETURNS (CAR) For a given event, define the stock cumulative return from the anchor date to τ as

$$R^{\text{stock}}(\tau) = \prod_{s=0}^{\tau}(1 + r_s^{\text{stock}}) - 1,$$

and define the cumulative return on the appropriate index benchmark over the same days as

$$R^{\text{bench}}(\tau) = \prod_{s=0}^{\tau}(1 + r_s^{\text{bench}}) - 1.$$

Cumulative abnormal return is then

$$\text{CAR}(\tau) = R^{\text{stock}}(\tau) - R^{\text{bench}}(\tau).$$

A key implementation detail is that we cumulate stock and benchmark returns separately and subtract *after* cumulation.

BENCHMARKS AND ANCHORS We use index-specific benchmarks that match the event. S&P 500 events are benchmarked to the S&P 500 return. Russell 1000 direct additions and deletions are benchmarked to the Russell 1000 return, and Russell 2000 direct additions and deletions are benchmarked to the Russell 2000 return. For migration events, we benchmark to the index most directly tied to the event. For S&P events, we anchor event time on the announcement date. For Russell events, we anchor event time on the effective reconstitution date.

WINSORIZATION To reduce the influence of extreme outliers, we winsorize the outcome variables within each event time and decade. Our baseline approach uses one-sided winsorization at the 99th percentile.

3 Motivating Facts

In this section, we provide background and institutional details on indexes that are commonly and explicitly tracked by index funds and ETFs, and implicitly tracked by mutual funds and institutional investors. We then outline several facts on index investing that directly inform our model.

3.1 Background and Institutional Details

Indexes are a common way to track segments of the market across and within asset classes, as well as serve as benchmarks for asset managers. Over the last thirty years, they have grown as a way to construct investment portfolios. Among the most common indexes used for investing are the S&P 500 index and the Russell 2000 index.

Indexes add and drop constituents from the indexes, usually at regular intervals. The ways that indexes reconstitute vary, ranging from complete transparency following a published formula (FTSE Russell indexes) to using subjective decision-making with limited information about selection criteria (S&P indexes). Depending on the policy, market participants find it simple or difficult to predict what changes will be made to an index in advance. Most indexes have regularly scheduled dates when the index is adjusted, where the scheduled reconstitution is announced at some point before the implementation date of the index reconstitution. There are some exceptions to this regular schedule, usually where index adjustments are made because of certain types of market events (e.g., a sudden merger, acquisition, or bankruptcy).

There are many types of investors or investment vehicles that follow indexes. We will broadly refer to these investors as *index trackers*. The most strict index trackers are index funds and ETFs. Most of these investment vehicles replicate the index with very little tracking error, and specifically cite an explicit goal to minimize tracking error. For index funds and ETFs that track indexes with relatively liquid underlying securities (e.g., U.S. equities), the tracking error is relatively small (Sammon and Shim, 2026), but can be larger due to selective replication when the underlying securities are illiquid (e.g., high-yield corporate bonds).

There are other index trackers that are not as strict. Many investors now engage in direct replication, where an investor or institution mimics an index but with some flexibility on several dimensions. These dimensions include holding only a representative subset of the index, strategically trading securities to engage in tax-loss harvesting, avoiding illiquid securities, and timing the purchase and sale of index securities. The reasons for market timing vary, from economizing on trading costs and index-provider fees, or

because mandates, risk-systems, and compensation are tied to benchmarks (Pavlova and Sikorskaya, 2023). These index trackers may differ in how closely they follow the index, but are still considered index trackers in that, at any given point in time, their portfolios will largely be similar to the tracked index.

The last group of index trackers we highlight are so-called "closet indexers." These investors do not disclose that they are indexers and may even state that they are active investors, selecting securities they believe will outperform an index or benchmark. However, their portfolio holdings appear to closely replicate some index (and often the index they are benchmarked against). We think of closet indexers as also varying how closely they track the index, perhaps based on the degree of manager risk aversion or agency issues – if these factors are large, they may track the index as closely as index funds or ETFs.

More broadly, predictable rule-based rebalancing is not limited to stock market indexes. For example, many institutional investors maintain stock-bond portfolios with target weights, such as 60% invested in stocks and 40% invested in bonds, and rebalance at regular calendar dates or when portfolio weights deviate significantly from predetermined targets (Harvey et al., 2025). While these investors do indeed follow a predictable rule-based rebalancing policy, our main analysis focuses on index-linked rebalancing within equities, where investors substitute between individual stocks rather than between aggregate asset classes.

Each index typically specifies index-eligible shares – a type of adjusted shares outstanding – for each constituent stock. Index-eligible shares combined with prices are used to compute the index price level and index returns. Prices are almost always drawn from official closing prices when they are available. For U.S. equities, these prices come from the closing auction that marks the end of regular-hours trading for each stock. For more illiquid securities without official closing prices, indexes can use prices that come from independent pricing services or market snapshots.

The price that indexes use is relevant for index trackers that attempt to exactly replicate the index. Because these strict index trackers minimize tracking error, they attempt to buy and sell securities at the same price that the index uses. For example, these index trackers will often attempt to buy adds and sell drops for U.S. equity indexes at the closing auction in the exact units they need for rebalancing to incur as close to zero tracking error as possible.

3.2 Motivating Facts

In this subsection, we provide two facts on trading volume and prices around implementation days of index reconstitutions. The first fact is that there is a huge burst of trading volume on the implementation day itself, with much less trading before or after the event. The second fact is that, despite all this volume, prices move very little. Collectively, these two facts present a puzzle that motivates our model.

Fact 1: Index trackers trade on the index implementation day.

The first fact is that there is a huge amount of trading volume on the exact day index reconstitutions take place, which we refer to as the implementation day. When an index adds or drops a constituent, index trackers tend to trade in the closing auction of the implementation day. This allows index trackers to buy and sell securities at the same price that the index uses in constructing the index price level, and therefore minimizes tracking error.

Figure 1 shows volume and cumulative abnormal returns around implementation days for S&P 500 and Russell 1000/2000 index events. The S&P 500 events are separated into direct additions and deletions, which enter or exit the S&P 500 from outside the S&P index universe, and migration additions and deletions, which move between the S&P 500 and another S&P size index (the S&P 400 midcap or the S&P 600 smallcap). For Russell events, we focus on migrations between the Russell 1000 and Russell 2000. For both sets of events, we focus on reconstitutions taking place between 2010 and 2024.

The figure shows that, in every case, there is an extremely large spike in volume on the implementation day. For S&P 500 direct additions, nearly 20% of shares outstanding trade on the implementation day. For S&P 500 deletions and migrations, implementation-day volume is also extremely large, ranging from roughly 18% to 26% of shares outstanding. Russell migrations have smaller levels of implementation-day volume, but still show a clear and large spike relative to surrounding days.

Surprisingly, there is not a comparable run-up in volume before implementation, nor is there a comparable continuation of volume after implementation. This is not surprising for index funds and ETFs, i.e., strict index trackers. Most of these trackers have very small tracking error budgets. If the index uses the closing price on the implementation day, then these trackers naturally want to trade at that same closing price.

What is more surprising is that more flexible index trackers and closet indexers also seem to trade on implementation day. Chinco and Sammon (2024) document that the volume in the closing auction on implementation days is double what we would expect

from strict index trackers alone. This suggests that implementation-day trading is not limited to mechanical index funds and ETFs. It also appears to include other benchmark-linked investors that have at least some flexibility in when to trade.

FACT 2: THERE IS LITTLE TO NO RETURN ON INDEX IMPLEMENTATION DAY, DESPITE ENORMOUS TRADING VOLUME.

The second fact is that implementation-day returns are small despite this extraordinary volume. Given the significant demand for newly added index constituents and the large amount of trading on the implementation day, it seems sensible to believe that the price movement would have to be large to clear the market. In other words, if index trackers buy or sell up to 20% of all shares outstanding for a stock, and do so on a single day, we would expect significant price impact for such extreme demand. For context, average daily trading volume for S&P 500 stocks is on the order of 1% of shares outstanding.

Historically, there were large returns around reconstitution in the early years of index-linked investing. Shleifer (1986) and Harris and Gurel (1986) document that in the 1970s and 1980s, the index addition price effect for S&P 500 stocks was large, with an abnormal one-day return of around 3%. However, as the dollars devoted to indexing have grown, the price movements on announcement and implementation days themselves have become *smaller*, not larger (Greenwood and Sammon, 2025).

Figure 1 shows this fact in the modern, post-2010 sample. The bars show the implementation-day volume spike, while the black line shows cumulative abnormal returns around the same event. For S&P 500 direct additions, the average abnormal return on the implementation day is only -0.09%. For S&P 500 migration additions, it is -0.24%. For Russell migrations, the average implementation-day abnormal returns are similarly small: 0.10% for stocks moving from the Russell 1000 to the Russell 2000 and 0.11% for stocks moving from the Russell 2000 to the Russell 1000.

The S&P 500 deletion panels are somewhat different, but they point to the same tension. Direct deletions have a larger implementation-day abnormal return of -1.48%, and they also show a large negative pre-event return. These are unusual cases, however, as firms directly deleted from the S&P 500 are firms that historically were large, but became so small they are no longer eligible to be included even in the S&P 600 small cap. Migration deletions, by contrast, have an implementation-day abnormal return of only -0.45%. Thus, even in the deletion panels, where the events are less clean than additions, the implementation-day return is modest relative to the amount of trading.

The figure also shows that returns are not completely flat before implementation. For example, S&P 500 direct additions have some positive run-up before implementation,

while S&P 500 direct deletions have a pronounced decline. These patterns are not surprising. For S&P events, there can be announcement effects before implementation. For Russell events, the ranking date is more predictable, and pre-rank returns mechanically affect index assignment. A stock that moves from the Russell 1000 to the Russell 2000 has typically performed poorly relative to the cutoff, while a stock that moves from the Russell 2000 to the Russell 1000 has typically performed well. Thus, some of the pre-event price movement likely reflects selection into the index event itself, rather than only anticipation of index-tracker demand.

The key fact is therefore not that prices don't move at *any* point before implementation. The key fact is that, on the implementation day itself, prices move very little relative to the amount of trading volume. The implementation day is when index trackers appear to trade, yet it is not when we see the kind of price movement that would normally be associated with trading 20% of shares outstanding.

This is especially striking when compared with estimates of price impact for institutional investors. For example, Frazzini et al. (2012) show that average price impact from trading 6% of average daily volume is around 30 basis points, and trading 11% of average daily volume is around 40 basis points. The index trades in our sample are orders of magnitude larger: implementation-day trading can be roughly 20 to 25 *times* average daily volume. The largest volume bin in Frazzini et al. (2012) is 11% of average daily volume, while these index events involve trading that is closer to 2,000% to 2,500% of average daily volume. Extrapolating from the estimates in Frazzini et al. (2012), we might expect to see price impact of over 6,800 basis points. Instead, we see little to no implementation-day return.

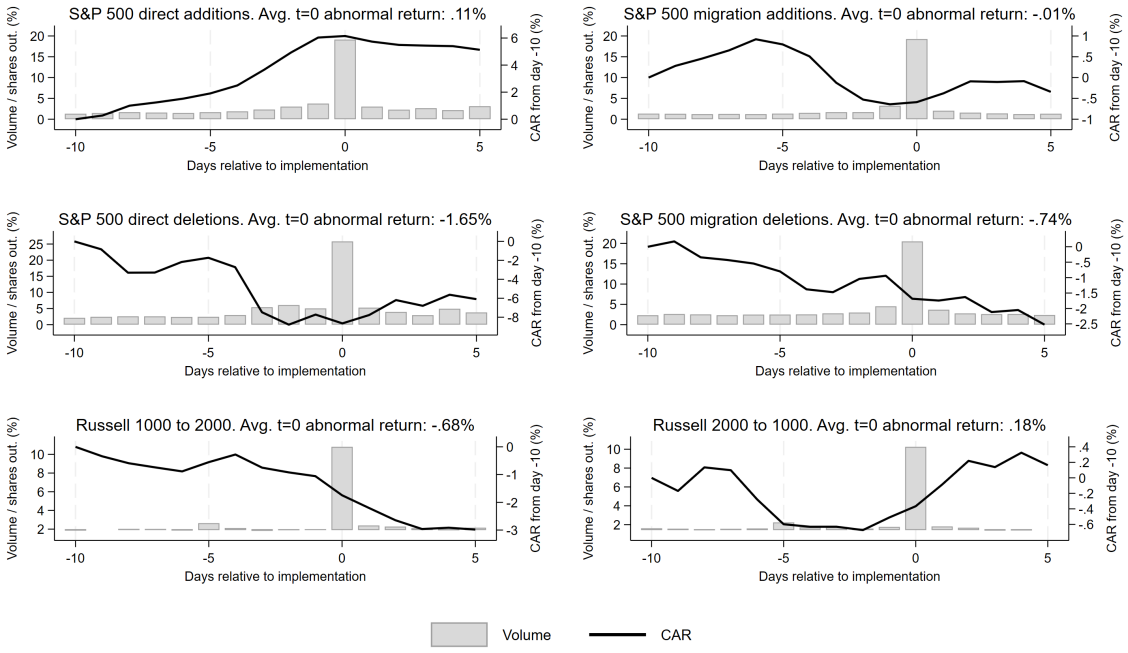
This does not mean that index trackers face zero trading costs. The implementation-day return is not the same thing as the total cost of trading. If speculators acquire shares in advance and sell them to index trackers at implementation, then the cost to index trackers may be embedded in the pre-event price rather than in the implementation-day return itself. Measuring the actual trading cost therefore requires a model of who trades before implementation, who supplies liquidity at implementation, and what the counterfactual price path would have been without anticipatory trading.

3.3 Discussion

The two indexing facts together present a puzzle. Index trackers – including loose trackers who have flexibility in when to trade – all seem to trade on the index implementation day. But prices do not jump on implementation day despite the significant volume.

So, the natural question is: why do prices not move? The amount of shares traded on

Figure 1: Volume and Returns Around Index Implementation Days



Notes. This figure shows average daily volume and cumulative abnormal returns around index implementation days. The gray bars show daily volume scaled by shares outstanding, in percent, and correspond to the left axis. The black line shows cumulative abnormal returns from day -10, in percent, and corresponds to the right axis. We separately show S&P 500 direct additions, S&P 500 migration additions, S&P 500 direct deletions, S&P 500 migration deletions, Russell 1000 to Russell 2000 migrations, and Russell 2000 to Russell 1000 migrations. S&P 500 direct additions and deletions enter or exit the S&P 500 from outside the S&P index universe. S&P 500 migration additions and deletions move between the S&P 500 and another S&P size index. Russell migrations move between the Russell 1000 and Russell 2000. The figure uses index events from 2010 to 2024.

the implementation day represents a demand shock that is, by any measure, extremely large and inelastic. Since the shock is widely anticipated, it is possible that the shares are provided with little additional price concession on the actual implementation day. If the price movement occurs before implementation, this implies that index trackers may buy at high prices and sell at low prices, which can be interpreted as a transaction cost or implementation shortfall. Still, it is striking to see 5% to 20% of a stock's shares outstanding changing hands with little contemporaneous price movement.

This in turn suggests a natural second question: why do index trackers trade on implementation day instead of in the days before? If prices rise leading up to an addition, it seems like they would benefit from trading before the implementation day itself. This seems especially valuable for loose index trackers and/or "enhanced" indexers that have less strict tracking error constraints and have flexibility to pay, what seems like, a slightly

lower price by trading slightly early or late (Arnott et al., 2023).

Given the volume we observe on implementation dates, however, a significant proportion of index trackers do not appear to trade outside of the implementation date. If anyone trades in the days or months before the implementation day, it is likely speculators, who provide the shares to the index trackers on the implementation day itself. The final question, therefore, is whether speculator activity is sustainable for index trackers in equilibrium, and whether this activity represents an excessive cost to investors.

At a glance, it seems as though speculators are imposing a cost on index trackers because their behavior resembles front running (or predatory trading more broadly) (Brunnermeier and Pedersen, 2005). On the other hand, we do not know the counterfactual, which is what price impact and trading costs would be for index trackers if speculators did not acquire the shares in advance and provide shares on the implementation day, or if index trackers trading became less predictable.

The model that follows is designed to discipline this counterfactual. It asks why an index tracker may rationally concentrate trading at implementation even when prices appear to move beforehand, and how speculators who trade in advance can simultaneously earn profits, reduce implementation-day price impact, and reduce the total cost borne by index investors.

4 MODEL

We consider a model with two types of agents: an index fund that needs to rebalance its portfolio in response to an index change, and a set of speculators that can trade in anticipation of the fund's demand. We study the fund's optimal order execution when it has a price impact and speculators can trade in anticipation of the fund's demand. All proofs are in the Internet Appendix IA.1, which is available upon request.

4.1 MODEL SETUP

The model spans five periods, $t \in \{0, 1, 2, 3, 4\}$. Period 0 is the announcement date: the index change is publicly announced, but no trading occurs and all positions are at their long-run fundamental values. Period 1 is the pre-implementation trading phase, during which the fund and speculators may establish positions in anticipation of the upcoming index change. Period 2 is the implementation date, on which the index change takes effect and the fund executes its mandatory rebalancing trade. Period 3 is the post-implementation trading phase, in which remaining positions may be adjusted. Period 4 is a terminal payoff period: no trading occurs, but each security pays its fundamental value

$V_{J,4}$ per share to its holders. Because the terminal payoff equals the fundamental value rather than the prevailing market price, any price impact incurred during periods $\{1, 2, 3\}$ represents a real cost that is not recovered at the terminal date.

Two assets trade in the market. The *original security* (O) is the incumbent index constituent, and the *new security* (N) is the incoming constituent, added to the index in period 2. In period 2, the index substitutes N for O: the target shifts from Q shares of O and zero shares of N to zero shares of O and Q shares of N. Accordingly, the fund must sell Q shares of O and purchase Q shares of N. We require that the fund's portfolio be fully rebalanced by the end of period 3: $\sum_{t=1}^{3} x_t = Q$.

The market contains a *fund* (F) and $L \geq 1$ *identical speculators*. The fund minimizes the sum of net execution cost and tracking-error cost. It may trade both securities in any period $t \in \{1, 2, 3\}$: let x_t^N denote N-shares purchased and x_t^O denote O-shares traded (negative for sales) in period t. We focus on the symmetric case where the fund trades in lockstep; that is, $x_t \equiv x_t^N = -x_t^O$. So the fund buys x_t shares of N and sells x_t shares of O in each period.

Each speculator $l \in \{1, \ldots, L\}$ maximizes trading profit net of execution and holding costs. Speculators may trade both securities in any period $t \in \{1, 2, 3\}$: let $s_{N,t,l}$ and $s_{O,t,l}$ denote speculator l's N-side and O-side trades in period t, respectively. Speculators are required to close all positions by the end of period 3, so that cumulative holdings satisfy $S_{N,3,l} = \sum_{\tau=1}^{3} s_{N,\tau,l} = 0$ and $S_{O,3,l} = \sum_{\tau=1}^{3} s_{O,\tau,l} = 0$ for every speculator l. Speculators compete in *Cournot* fashion, taking the quantities of other speculators as given. We characterize the symmetric equilibrium in which all speculators adopt identical strategies, $s_{N,t,l} = s_{N,t}$ and $s_{O,t,l} = s_{O,t}$ for all l, so that aggregate speculator order flows are $S_{N,t} = L\, s_{N,t}$ and $S_{O,t} = L\, s_{O,t}$.

4.1.1 Fundamental Values and Prices

Each security has a fundamental (long-run) value that evolves as a martingale with IID innovations,

$$V_{J,t} = V_{J,t-1} + \varepsilon_{J,t}\,, \qquad J \in \{N, O\}\,, \tag{1}$$

where $\{\varepsilon_{J,t}\}$ are IID across time with $\mathbb{E}[\varepsilon_{J,t}] = 0$ and finite variance; innovations need not be independent across securities. We write $M_{J,t} \equiv \sum_{\tau=1}^{t} \varepsilon_{J,\tau}$ for the cumulative martingale innovation, so that $V_{J,t} = \bar{v} + M_{J,t}$. Let $P_{J,t}$ denote the transaction price of security $J \in \{N, O\}$ at time t. All transaction prices and fundamental values are initialized at the common level $\bar{v}$ at $t = 0$:

$$V_{N,0} = V_{O,0} = P_{N,0} = P_{O,0} = \bar{v}\,. \tag{2}$$

A key parameter is the standard deviation of the difference in fundamental innovations between N and O,

$$\nu \;\equiv\; \mathrm{sd}\big(\varepsilon_{N,t} - \varepsilon_{O,t}\big) > 0\,, \tag{3}$$

which is strictly positive whenever $\varepsilon_{N,t} \neq \varepsilon_{O,t}$ with positive probability, the natural case for distinct securities. In the terminal period 4, each security pays its fundamental value $V_{J,4} = \bar{v} + M_{J,4}$ to its holders.

INTERMEDIATED TRADING. For each security J, the market contains n_J competitive risk bearers and $\mathcal{D}_J \geq 3$ identical dealers. The fund and speculators submit aggregate signed customer flow $z_{J,t} \equiv x_t^J + \sum_{l=1}^{L} s_{J,t,l}$ in period t, and $Z_{J,t} \equiv \sum_{\tau=1}^{t} z_{J,\tau}$ is cumulative customer flow. In the spirit of Ho and Stoll (1981), Stoll (1978), Amihud and Mendelson (1980), Madhavan and Smidt (1993), O'Hara and Oldfield (1986), and Grossman and Miller (1988), dealers intermediate these orders between customers and risk bearers. Let $h_{r,J,t}$ denote risk bearer r's signed inventory. Given the transfer price $\widetilde{P}_{J,t}$, each risk bearer solves

$$\max_{h_{r,J,t}\in\mathbb{R}} \left\{ (V_{J,t} - \widetilde{P}_{J,t}) h_{r,J,t} - \frac{\omega_J}{2} h_{r,J,t}^2 \right\}, \qquad r = 1, \ldots, n_J,$$

where $\omega_J > 0$ summarizes risk aversion, payoff risk, and balance-sheet costs. Risk-bearer market clearing requires $\sum_{r=1}^{n_J} h_{r,J,t} = -Z_{J,t}$. With this specification, risk bearers generate a downward-sloping residual demand as in models by Shleifer (1986), Harris and Gurel (1986), and Greenwood (2005).

Let $u_{i,J,t} \in \mathbb{R}$ denote dealer i's signed allocation of the current order: $u_{i,J,t} > 0$ means that it sells to customers and $u_{i,J,t} < 0$ means that it buys from customers. Conditional on the clearing allocation and price, dealer i chooses its quote slope to maximize

$$\Pi_{i,J,t}^{D} = (P_{J,t} - \widetilde{P}_{J,t}) u_{i,J,t} - \theta_J u_{i,J,t}^2,$$

where $\theta_J > 0$ is its intermediation-cost parameter.

After customer orders arrive, dealers purchase shares from, or sell shares to, risk-bearing investors at the common transfer price $\widetilde{P}_{J,t}$ implied by their aggregate inventory position. Dealer i submits the linear marginal quote

$$\widetilde{P}_{J,t} + a_{i,J} u, \qquad u \in \mathbb{R},$$

where $a_{i,J} > 0$ is its strategic choice. The allocation mechanism selects signed dealer

allocations $u_{i,J,t}$ and a common customer-facing transaction price $P_{J,t}$ such that

$$\widetilde{P}_{J,t} + a_{i,J} u_{i,J,t} = P_{J,t} \quad \text{for every active dealer}, \qquad \sum_i u_{i,J,t} = z_{J,t}.$$

Thus all customer units transact at $P_{J,t}$, whereas dealers transact with risk-bearing investors at $\widetilde{P}_{J,t}$.

Lemma 1 (Price-impact function)**.** *Under the dealer-market structure above, the transaction price is*

$$P_{J,t} = V_{J,t} + \lambda_J Z_{J,t} + \rho_J z_{J,t}, \tag{4}$$

where

$$\lambda_J = \frac{\omega_J}{n_J}, \qquad \rho_J = \frac{2\theta_J(\mathcal{D}_J - 1)}{\mathcal{D}_J(\mathcal{D}_J - 2)}.$$

The equilibrium price-impact function corresponds to the functions used in models of optimal trading execution (Almgren and Chriss, 2001; Bertsimas and Lo, 1998).

We refer to λ_J and ρ_J as the persistent and temporary price-impact coefficients, respectively. For algebraic convenience, we define

$$\mu_J \equiv \lambda_J + \rho_J$$

which measures the total within-period price impact of flows. We allow $\lambda_O \leq \lambda_N$ and $\rho_O \leq \rho_N$, so that $\mu_O \leq \mu_N$, reflecting the possibility that the original security is a more liquid position with lower total per-unit price impact than the new constituent. When order flow is negative (net selling), the transaction price falls accordingly.

When the fund trades both securities in lockstep, $x_t^N = x_t$ and $x_t^O = -x_t$, the effective within-period impact is $\mu_N + \mu_O$. No trading occurs in period 4. For later use, define

$$\bar{P}_N \equiv \bar{v} + \lambda_N Q, \tag{5}$$

the N-security price after all trading has taken place and cumulative N-side net order flow equals Q.

4.1.2 COSTS AND OBJECTIVES

Fund's objective. The fund chooses the trading schedule (x_1, x_2, x_3) to minimize the sum of transaction cost, TC^F, and tracking-error cost, $\kappa \sum_{t=1}^{3} \mathrm{TE}_t$:

$$C^F = \mathrm{TC}^F + \kappa \sum_{t=1}^{3} \mathrm{TE}_t, \tag{6}$$

where $\kappa \geq 0$ is a tracking-error penalty parameter that measures the fund's aversion to deviations from the index.

The fund's *transaction cost* TC^F is the total execution cost of buying Q shares of N and selling Q shares of O across periods $\{1, 2, 3\}$, measured relative to the expected fundamental value of the security. Because the terminal payoff equals the fundamental value, any price impact incurred during trading represents a real cost not recovered at the terminal date. For period t trading of x_t shares, the cost is the price-impact component of the current transaction price: $\lambda_N \, x_t \, Z_{N,t} + \rho_N \, x_t \, z_{N,t}$ on the buy side and $-\lambda_O \, x_t \, Z_{O,t} - \rho_O \, x_t \, z_{O,t}$ on the sell side. The total transaction cost is

$$\mathrm{TC}^F = \sum_{t=1}^{3} \Big[\lambda_N \, x_t \, Z_{N,t} + \rho_N \, x_t \, z_{N,t} - \lambda_O \, x_t \, Z_{O,t} - \rho_O \, x_t \, z_{O,t} \Big], \tag{7}$$

with $Z_{N,0} = Z_{O,0} = 0$.

The fund's *tracking error* in period t measures the risk of unexpected deviations relative to the underlying index. Write $\Delta\Lambda_t \equiv (P_{N,t} - P_{N,t-1}) - (P_{O,t} - P_{O,t-1})$ for the index-relative return over period t, computed from (4) *with every price-impact term retained*. Let $h_{N,t}$ denote the fund's holdings of N at the end of period t, with initial position $h_{N,0} = 0$ and an initial holding of Q shares of O; the index holds Q shares of O for $t < 2$ and Q shares of N from period 2 onward, and $d_{t-1} \equiv h_{N,t-1} - h^{\mathrm{idx}}_{N,t-1}$ denotes the fund's deviation from the index holding. The per-period tracking error is

$$\mathrm{TE}_t \;=\; \mathrm{sd}_{t-1}\big(d_{t-1} \, \Delta\Lambda_t\big) \;=\; \hat{\nu}_t \, |d_{t-1}|, \qquad d_{t-1} \equiv h_{N,t-1} - h^{\mathrm{idx}}_{N,t-1}, \tag{8}$$

where $\hat{\nu}_t \equiv \mathrm{sd}_{t-1}(\Delta\Lambda_t)$ is the standard deviation of the index-relative return, and sd_{t-1} denotes the standard deviation conditional on the information available at the end of period $t-1$.

At the end of period 0 the fund and the index hold identical portfolios, so $\mathrm{TE}_1 = 0$; entering period 2 the fund holds x_1 excess shares of N, so $\mathrm{TE}_2 = \hat{\nu}_2 \, x_1$; and entering period 3 it holds $Q - x_1 - x_2 = x_3$ shares of O in excess of the index, so $\mathrm{TE}_3 = \hat{\nu}_3 \, x_3$. Tracking error

is thus linear in the fund's off-implementation trades, and the total tracking-error cost is

$$\kappa \sum_{t=1}^{3} \mathrm{TE}_t \;=\; \kappa\,(\hat{\nu}_2\, x_1 + \hat{\nu}_3\, x_3), \tag{9}$$

where $\kappa > 0$ is the tracking-error penalty parameter.

If all order flows are deterministic, every price-impact term in $\Delta\Lambda_t$ is measurable at the end of period $t-1$ and is removed by the conditional expectation, leaving $\Delta\Lambda_t - \mathbb{E}_{t-1}\Delta\Lambda_t = \varepsilon_{N,t} - \varepsilon_{O,t}$. Hence $\hat{\nu}_2 = \hat{\nu}_3 = \nu$ and (9) reduces to

$$\kappa \sum_{t=1}^{3} \mathrm{TE}_t \;=\; \kappa\,\nu\,(x_1 + x_3). \tag{10}$$

We maintain the assumption of deterministic order flows until Section 4.6, where we relax it to allow for stochastic order flows.

Speculator's objective. Each speculator l maximizes trading profit net of execution and holding costs. The execution cost from N-side trades of $s_{N,t,l}$ shares in period t is $\lambda_N\, s_{N,t,l}\, Z_{N,t} + \rho_N\, s_{N,t,l}\, z_{N,t}$, with the O-side defined analogously. The compact coefficient $\mu_N = \lambda_N + \rho_N$ is used in the algebra below where convenient. In addition, each speculator bears a holding cost proportional to the square of cumulative positions,

$$\mathrm{HC}_l^S = c \sum_{t=1}^{3} \big[(S_{N,t,l})^2 + (S_{O,t,l})^2\big], \tag{11}$$

where $S_{N,t,l} = \sum_{\tau=1}^{t} s_{N,\tau,l}$ and $S_{O,t,l} = \sum_{\tau=1}^{t} s_{O,\tau,l}$ are cumulative positions, and $c \geq 0$ captures financing, margin, and inventory risk. Because speculators must close all positions by the end of period 3, the terminal payoff terms that would otherwise appear in the profit expression vanish identically. The per-speculator profit across both securities therefore reduces to the negative of total execution cost minus holding cost:

$$\Pi_l^S = -\sum_{t=1}^{3} \big[P_{N,t}\, s_{N,t,l} + P_{O,t}\, s_{O,t,l}\big] - \mathrm{HC}_l^S, \tag{12}$$

where $P_{J,t}$ is the execution price from (4). The speculator profits when the period-2 execution gain (from selling into the fund's demand at temporarily elevated prices) exceeds the combined execution and holding costs from periods 1 and 3.

4.1.3 Strategies and Payoffs

In the symmetric equilibrium, each speculator employs a *three-period strategy* that closes all positions by the end of period 3. On the N-side, the strategy is parameterized by $s_1 \equiv s_{N,1}$, the pre-positioning purchase in period 1, and $r \equiv s_{N,3}$, the covering trade in period 3; the implied period-2 trade $-(s_1 + r)$ sells the accumulated position into the fund's implementation demand. The O-side is symmetric, with $s_{O,1}$ denoting the period-1 short and $r_O \equiv s_{O,3}$ the period-3 re-short. Aggregate positions across L speculators are

$$S_1 \equiv L\, s_1, \quad R \equiv L\, r \quad \text{(N-side)}; \qquad S_{O,1} \equiv L\, s_{O,1}, \quad R_O \equiv L\, r_O \quad \text{(O-side)}. \tag{13}$$

By construction, cumulative N-side holdings per speculator are s_1 at the end of period 1, $-r$ at the end of period 2, and 0 at the end of period 3; the O-side is symmetric. All positions close by design, confirming that the terminal-payoff terms in (12) vanish and that the holding cost per speculator is $c\,(s_1^2 + r^2)$ on the N-side plus $c\,(s_{O,1}^2 + r_O^2)$ on the O-side.

The net order flows in each period are

$$\begin{aligned}
&\text{Period 1:} \quad z_{N,1} = x_1 + S_1, \qquad z_{O,1} = -x_1 + S_{O,1};\\
&\text{Period 2:} \quad z_{N,2} = x_2 - S_1 - R, \qquad z_{O,2} = -x_2 - S_{O,1} - R_O;\\
&\text{Period 3:} \quad z_{N,3} = x_3 + R, \qquad z_{O,3} = -x_3 + R_O.
\end{aligned}$$

Execution prices follow from the pricing function (4). Period 4 involves no trading; holders receive the fundamental value $V_{J,4}$.

4.2 Equilibrium Analysis

We now characterize equilibrium. We begin by deriving the symmetric Cournot best-response functions for $L \geq 1$ speculators on each side, taking the fund's schedule as given. Then, we characterize the fund's optimal trading schedule under no commitment and full commitment, respectively. In both cases, if the tracking-error penalty is sufficiently high, the fund optimally concentrates all trading at implementation. The full-commitment threshold is below its no-commitment counterpart: by publicly announcing its schedule, the fund can secure the implementation-day outcome under a strictly broader set of parameters.

4.2.1 SPECULATOR'S GENERAL THREE-PERIOD STRATEGY

Next, we characterize the speculators' symmetric best responses to an arbitrary trading schedule implemented by the fund.

Proposition 1 (L-speculator symmetric Cournot equilibrium)**.** *Suppose $\rho_N > 0$ (and correspondingly $\rho_O > 0$ on the O-side). With $L \geq 1$ identical speculators in symmetric Cournot competition and an arbitrary fund schedule (x_1, x_3) (with $x_2 = Q - x_1 - x_3$):*

1. Equilibrium positions at the implementation-date rebalancing schedule. *At $(x_1, x_3) = (0, 0)$, the aggregate equilibrium positions are*

$$S_1(L) = L\, s_1(L), \qquad R(L) = L\, r(L), \tag{14}$$

with $s_1(L) > 0$ for all $L \geq 1$. Closed-form expressions for $s_1(L)$ and $r(L)$ are given in (IA.7) *in the proof.*

2. Affine best-response structure. *For a general schedule (x_1, x_3), the per-speculator best responses are affine corrections to the implementation-date rebalancing positions:*

$$s_1^*(x_1, x_3) = s_1(L) + \frac{\alpha_{s,1}\, x_1 + \alpha_{s,3}\, x_3}{\Delta_N(L)}\,, \tag{15}$$

$$r^*(x_1, x_3) = r(L) + \frac{\alpha_{r,1}\, x_1 + \alpha_{r,3}\, x_3}{\Delta_N(L)}\,., \tag{16}$$

where $\Delta_N(L) > 0$ is defined in (IA.6) *in the proof. All correction coefficients are non-positive except $\alpha_{r,1}$, which has ambiguous sign; $\alpha_{s,1}$ and $\alpha_{r,3}$ are strictly negative, and $\alpha_{s,3} = 0$ exactly when $\lambda_N = c = 0$ (see the proof for explicit expressions). Hence fund trades weakly reduce speculator equilibrium positions on both the accumulation and close-out margins, and strictly reduce their sum.*

3. Crowding out: aggregate liquidity provision decreases when the fund trades outside of the implementation date. *The aggregate N-side liquidity provision on the implementation date $\Sigma_N^*(x_1, x_3) \equiv S_1^*(x_1, x_3) + R^*(x_1, x_3)$ is strictly decreasing in both x_1 and x_3:*

$$\frac{\partial \Sigma_N^*}{\partial x_j} = \frac{L\,(\alpha_{s,j} + \alpha_{r,j})}{\Delta_N(L)} < 0, \qquad j \in \{1, 3\}. \tag{17}$$

That is, the more the fund trades outside of the implementation date, the less total speculative liquidity is available at implementation.

4. Fund trades increase net order flow. *Each aggregate derivative is bounded below by* -1*:*

$$\frac{\partial S_1^*}{\partial x_j},\ \frac{\partial R^*}{\partial x_j},\ \frac{\partial \Sigma_N^*}{\partial x_j}\ >\ -1, \qquad j \in \{1,3\}. \tag{18}$$

Hence, the net order flow $z_{N,t}$ *is strictly increasing in fund trades outside of the implementation date.*

The O-side positions and best-response functions are given by the same expressions with $(\rho_O, \lambda_O, \mu_O)$ *replacing* $(\rho_N, \lambda_N, \mu_N)$*.*

The crowding-out result established in Part 3 is the central economic mechanism linking the two equilibrium concepts. When the fund trades before or after implementation ($x_1 > 0$ or $x_3 > 0$), it compresses speculator margins and induces speculators to scale back their positions. Because speculators supply the liquidity that absorbs the fund's period-2 demand x_2, reduced speculative participation raises the fund's implementation-date execution cost. Committing to the implementation-date rebalancing schedule $(x_1, x_3) = (0, 0)$ therefore maximizes speculative liquidity at implementation, thus reducing transaction costs.

The condition $\rho_N > 0$ (and correspondingly $\rho_O > 0$ on the O-side) is necessary for the aggregate crowding-out derivative in Part 3 to be strictly negative and for the bounds in Part 4 to be strict, and is maintained throughout the analysis.

4.2.2 Equilibrium with No Commitment

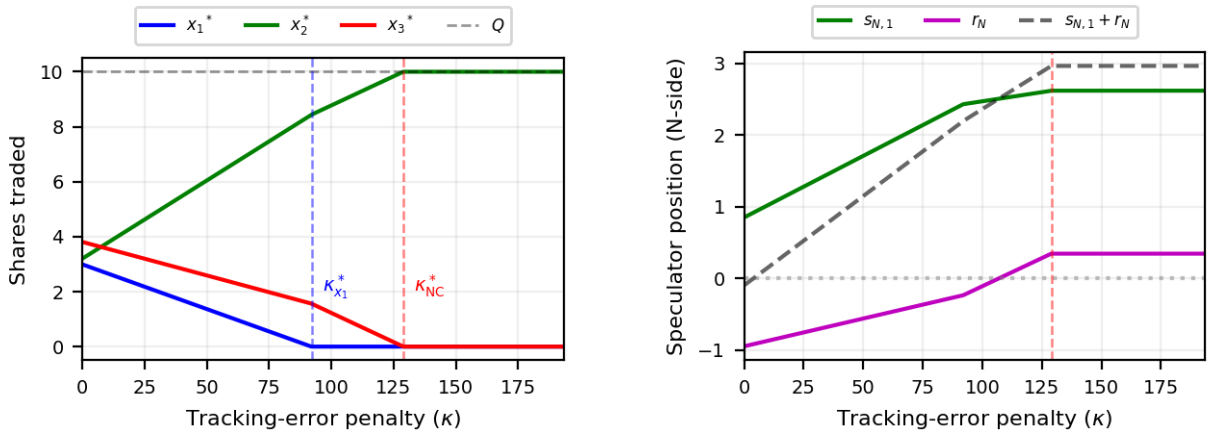
We formally define an equilibrium with no commitment as follows.

Definition 1 (Equilibrium with no commitment)**.** *An* equilibrium with no commitment *is a strategy profile* $(x_1^*, x_2^*, x_3^*; (s_{1,l}^*, r_l^*)_{l=1}^L)$ *such that: (i) each speculator's strategy* $(s_{1,l}^*, r_l^*)$ *is a best response to the fund's strategy and the other speculators' strategies; and (ii) the fund's strategy* (x_1^*, x_2^*, x_3^*) *is a best response to the aggregate speculator positions, taking speculator behavior as given. Neither party commits ex ante; strategies are chosen simultaneously.*

That is, with no commitment, the fund takes speculator positions as given when choosing its trading schedule. The equilibrium is determined by the mutual best responses of all players simultaneously. The fund's optimal schedule is therefore a fixed point of the best-response functions, and the speculators' equilibrium positions are determined by the fund's schedule through the best-response functions derived in Proposition 1. In this case, the equilibrium is characterized by the following result.

Figure 2: No-commitment equilibrium as a function of the tracking-error penalty κ.

(a) Fund trading schedule **(b)** Speculator positions



Notes. Panel 2(a) shows fund trade sizes x_1^*, x_2^*, and x_3^*. For $\kappa \geq \kappa_{\text{NC}}^*$ (red dashed line) implementation-date rebalancing obtains ($x_2^* = Q$). In the intermediate regime $\kappa_{x_1}^* \leq \kappa < \kappa_{\text{NC}}^*$ (blue dashed line), $x_1^* = 0$ and $x_3^* > 0$; below $\kappa_{x_1}^*$ the solution is fully interior. Panel 2(b) shows N-side speculator positions at the no-commitment equilibrium (accumulation $s_{N,1}$ and close-out trade r_N) together with the total period-2 sell $s_{N,1} + r_N$. As κ rises, the fund concentrates trading in period 2, increasing speculators' front-running incentives. Parameters: $Q = 10$, $\lambda_N = 0.5$, $\rho_N = 0.55$, $\lambda_O = 0.2$, $\rho_O = 0.2$, $c = 0.3$, $\nu = 0.15$, $L = 1$.

Proposition 2 (Fund's optimal rebalancing schedule)**.** *There exist a no-commitment threshold $\kappa_{\text{NC}}^* > 0$ (see* (IA.14)*) and a second threshold $\kappa_{x_1}^* \in (0, \kappa_{\text{NC}}^*]$, both derived explicitly in the proof, such that the equilibrium has three regimes:*

1. $\kappa \geq \kappa_{\text{NC}}^*$ ***(implementation-date rebalancing):*** *The fund's unique optimal strategy is $(x_1^*, x_3^*) = (0, 0)$, and hence, $x_2^* = Q$. That is, the fund trades its entire demand in period 2 and incurs exactly zero tracking error.*

2. $\kappa_{x_1}^* \leq \kappa < \kappa_{\text{NC}}^*$ ***(intermediate regime):*** *$x_1^* = 0$ and $x_3^* > 0$.*

3. $\kappa < \kappa_{x_1}^*$ ***(interior regime):*** *$x_1^* > 0$ and $x_3^* > 0$. That is, the fund spreads trades across all three periods, with $x_3^* \geq x_1^*$.*

Both $\kappa_{x_1}^* < \kappa_{\text{NC}}^*$ in the statement of the proposition and $x_3^* > x_1^*$ in Part 3 hold strictly if and only if permanent impact is present on at least one side, $\max\{\lambda_N, \lambda_O\} > 0$. When $\lambda_N = \lambda_O = 0$ the two thresholds coincide, the intermediate regime is empty, and the interior solution is symmetric, $x_1^* = x_3^*$.

Figure 2 illustrates the three-regime structure as a function of κ: the left panel traces equilibrium fund trade sizes (x_1^*, x_2^*, x_3^*) across the three regimes, and the right panel shows the corresponding equilibrium speculator positions on the N-side.

The fund's marginal cost of shifting one share from period 2 to period 3 is

$$\left.\frac{\partial C^F}{\partial x_3}\right|_{(0,0)} = \underbrace{\kappa \left.\frac{\partial \,\mathrm{TE}}{\partial x_3}\right|_{(0,0)}}_{\kappa\,\nu} + \underbrace{\left.\frac{\partial \,\mathrm{TC}^F}{\partial x_3}\right|_{(0,0)}}_{-\iota(L)}, \tag{19}$$

a trade-off between the marginal tracking-error penalty $\kappa\,\nu$ and $\iota(L)$, the marginal transaction-cost saving from thinning the period-2 block. Implementation-date rebalancing is optimal once the penalty dominates the saving, $\kappa \geq \iota(L)/\nu = \kappa^*_{\mathrm{NC}}$; because speculators' period-2 sell flow already lowers the fund's execution cost, their presence compresses $\iota(L)$ and hence the threshold. Yet under no commitment the fund takes speculator positions as given, so it never internalizes the liquidity it would destroy by trading outside period 2, and κ^*_{NC} remains strictly positive.

The second threshold $\kappa^*_{x_1}$ and the back-loading pattern $x_3^* \geq x_1^*$ reflect a further asymmetry, between trading before versus after implementation. A period-1 purchase permanently displaces the N-security price *before* the fund's own period-2 block, raising execution cost on the remaining x_2 shares; a period-3 purchase arrives after the block and leaves that price unaffected. Since the tracking-error penalty is the same in either direction, the marginal cost of shifting to period 1 never falls below that of shifting to period 3:

$$\left.\frac{\partial C^F}{\partial x_1}\right|_{(0,0)} = \kappa\,\nu - \iota(L) + \underbrace{A}_{\geq\, 0}, \tag{20}$$

where the wedge A is strictly positive whenever permanent impact is present on at least one side, $\max\{\lambda_N, \lambda_O\} > 0$. Consequently, as κ falls, the fund shifts trades to period 3 first, and only below $\kappa^*_{x_1}$ does x_1^* also become positive.

4.2.3 Equilibrium with Full Commitment

Full commitment is the more natural assumption for large fund managers, and a repeated-game perspective shows why it can be sustained endogenously. Index reconstitutions recur on a predictable calendar, so the fund and speculators interact repeatedly: speculators observe the fund's realized execution pattern after each event and update their pre-positioning strategies accordingly. In such a setting, the implementation-date rebalancing schedule $(x_1, x_3) = (0, 0)$ can be self-enforcing as a reputation equilibrium: if the fund deviates by trading outside the implementation window, speculators interpret this as a signal of opportunistic behavior, scale back their pre-positioning in subsequent

reconstitutions, and thereby raise the fund's future execution costs. Under standard folk-theorem arguments, a sufficiently patient fund therefore finds it optimal to maintain the committed schedule in every period, rendering commitment credible without any external enforcement mechanism. We therefore define the full-commitment equilibrium as follows.

Definition 2 (Equilibrium with full commitment)**.** *An* equilibrium with full commitment *is a strategy profile* $(x_1^*, x_2^*, x_3^*; (s_{1,l}^*, r_l^*)_{l=1}^L)$ *such that: (i) the fund publicly commits to its trading schedule* (x_1^*, x_2^*, x_3^*) *to minimize total cost, anticipating the speculators' best responses; and (ii) each speculator's strategy* $(s_{1,l}^*, r_l^*)$ *is a best response to the committed schedule and the other speculators' strategies. The fund is bound by its stated schedule.*

That is, with full commitment, the fund moves first: it selects and publicly commits to a cost-minimizing schedule (x_1^*, x_2^*, x_3^*), taking into account how speculators will respond to that commitment. Each speculator then observes the complete schedule and chooses trades to maximize profit, competing in Cournot fashion with the remaining $L-1$ speculators. This notion of equilibrium differs from no commitment in that the fund internalizes the effect of its committed schedule on speculator behavior, namely the crowding-out effect described in Proposition 1, rather than treating speculator positions as exogenous. The fund's optimal commitment schedule is characterized by the following result.

Proposition 3 (Fund's optimal rebalancing schedule under full commitment)**.** *There exists a full-commitment threshold* κ_{FC}^* *satisfying* $0 < \kappa_{\mathrm{FC}}^* \leq \kappa_{\mathrm{NC}}^*$*, with strict inequality under the parameter conditions of Remark 4.2.1. The following hold.*

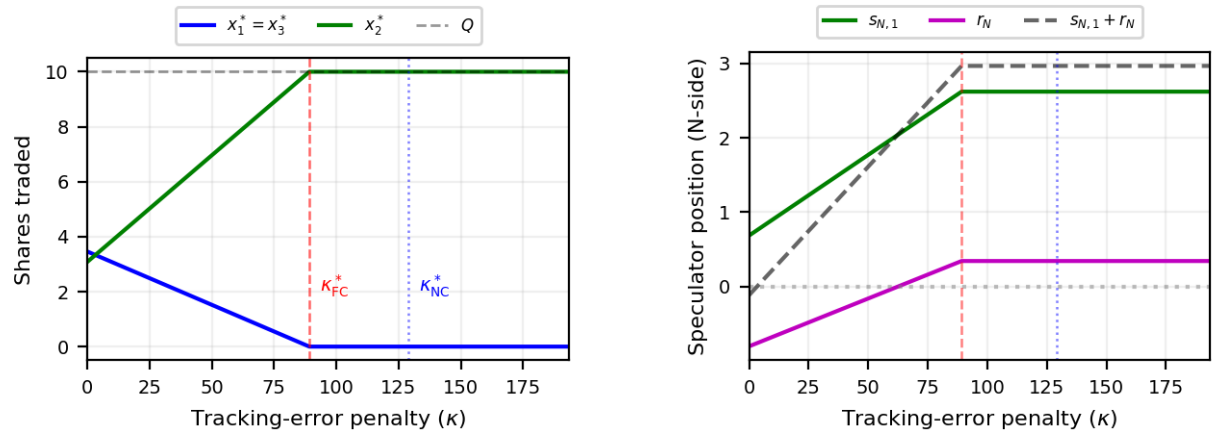
1. $\kappa \geq \kappa_{\mathrm{FC}}^*$ ***(implementation-date rebalancing):*** *The fund's unique optimal commitment is* $(x_1^*, x_3^*) = (0, 0)$*, and hence* $x_2^* = Q$*. The fund trades its entire demand in period* 2 *and incurs exactly zero tracking error. The implementation-day schedule also maximizes aggregate speculator liquidity provision:* $\Sigma_N^* = S_1^* + R^*$ *and* $|S_{O,1}^*| + |R_O^*|$ *are both strictly decreasing in* x_1 *and* x_3*.*

2. $\kappa < \kappa_{\mathrm{FC}}^*$***:*** *The fund's unique optimal commitment satisfies* $x_1^* = x_3^* > 0$*: trades are spread* equally *across periods* 1 *and* 3*, with the remainder concentrated in period* 2*. Speculator positions on both sides are strictly reduced relative to the implementation-day benchmark.*

Figure 3 illustrates these features: the left panel shows the fund's optimal committed schedule, with the interior solution satisfying $x_1^* = x_3^*$ and the switch to implementation-date rebalancing at κ_{FC}^* (strictly below the no-commitment threshold κ_{NC}^*); the right panel shows the corresponding reduction in speculator positions relative to the no-commitment benchmark.

Figure 3: Full-commitment equilibrium as a function of the tracking-error penalty κ.

(a) Fund trading schedule **(b)** Speculator positions



Notes. Panel 3(a) shows the fund's optimal trading schedule. By Proposition 3, the interior solution satisfies $x_1^* = x_3^*$ (symmetric around implementation); implementation-date rebalancing ($x_1^* = x_3^* = 0$) obtains at the full-commitment threshold κ^*_{FC} (red dashed), strictly below the no-commitment threshold κ^*_{NC} (blue dotted). Panel 3(b) shows N-side speculator positions at the full-commitment equilibrium. When the fund spreads trading ($x_1^* = x_3^* > 0$), speculators reduce their accumulation $s_{N,1}$ and covering r_N, reflecting the strategic-effect terms SE_N and SE_O in eq. (IA.55). Parameters: $Q = 10$, $\lambda_N = 0.5$, $\rho_N = 0.55$, $\lambda_O = 0.2$, $\rho_O = 0.2$, $c = 0.3$, $\nu = 0.15$, $L = 1$.

The optimality of the threshold κ^*_{FC} admits a marginal-cost interpretation analogous to (19). Starting from $(x_1, x_3) = (0, 0)$, the committed fund's total marginal cost of shifting one share to period 3 is

$$\left.\frac{d\tilde{C}^F}{dx_3}\right|_{(0,0)} = \underbrace{\kappa \left.\frac{d\,\mathrm{TE}}{dx_3}\right|_{(0,0)}}_{\kappa\,\nu} + \underbrace{\left.\frac{\partial\,\mathrm{TC}^F}{\partial x_3}\right|_{(0,0)}}_{-\iota(L)} + \underbrace{\frac{\partial\,\mathrm{TC}^F}{\partial \mathbf{S}^*} \cdot \left.\frac{d\mathbf{S}^*}{dx_3}\right|_{(0,0)}}_{\mathrm{SE}(L)\,>\,0} \tag{21}$$

where the first two terms are identical to the no-commitment gradient (19) and the third term is the indirect effect on transaction costs through the speculator best responses $\mathbf{S}^*(x_1, x_3) = (S_1^*, R^*, S_{O,1}^*, R_O^*)$, which we label as the strategic effect, $\mathrm{SE}(L) > 0$. Implementation-date rebalancing is optimal when $\kappa \geq (\iota - \mathrm{SE}(L))/\nu = \kappa^*_{\mathrm{FC}}$.

The threshold κ^*_{FC} is strictly below its no-commitment counterpart κ^*_{NC}. The difference arises because the committed fund internalizes the strategic effect $\mathrm{SE}(L)$: by announcing $(x_1, x_3) = (0, 0)$, it preserves speculator margins and thereby secures the maximum speculative liquidity at implementation. An uncommitted fund ignores this externality and therefore perceives a smaller benefit from concentrating all trades at period 2.

Moreover, below the full-commitment threshold, the optimal schedule satisfies $x_1^* = x_3^*$: pre- and post-implementation trades are equal in size. This symmetry reflects the

structure of the transaction cost under commitment. When the fund accounts for the speculator best-response functions (15)–(16), the crowding-out effect is symmetric between period 1 and period 3: announcing one share of pre-implementation trading compresses speculator accumulation by the same amount that announcing one share of post-implementation trading compresses speculator covering. Because the crowding-out derivatives enter the fund's effective cost symmetrically, and the tracking-error penalty $\kappa\,\nu$ is identical for deviations in either direction, the cost-minimizing committed schedule equates x_1 and x_3. This contrasts sharply with the no-commitment case, where the asymmetry in permanent price impact (period-1 purchases displace prices ahead of the implementation block, whereas period-3 purchases do not) generates back-loading ($x_3^* > x_1^*$) at every interior solution.

4.3 Equilibrium in the Competitive Limit

Next, we study the optimality of implementation-date rebalancing as the number of speculators grows. Under no commitment, the threshold $\kappa^*_{\mathrm{NC}}(L)$ remains bounded away from zero even as $L \to \infty$: even with a fully competitive speculator sector, the fund requires a strictly positive tracking-error penalty to prefer implementation-date rebalancing. Under full commitment, by contrast, $\kappa^*_{\mathrm{FC}}(L) \to 0$: once speculators are sufficiently numerous and the fund can credibly commit to its schedule, *any* fund will optimally concentrate all trading at implementation.

Proposition 4 (Competitive limit: $L \to \infty$)**.** *As the number of identical speculators $L \to \infty$:*

1. ***No-commitment threshold:*** *$\kappa^*_{\mathrm{NC}}(L)$ converges to a strictly positive limit:*

$$\kappa^*_{\mathrm{NC}}(\infty) = \frac{(\mu_N + \mu_O)\,Q}{\nu} > 0. \tag{22}$$

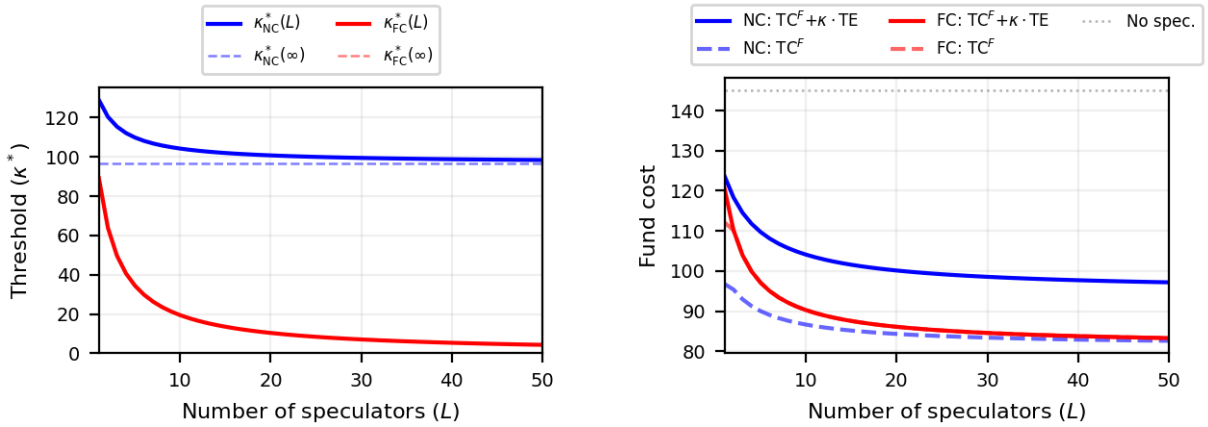
   When $c = 0$, $\kappa^*_{\mathrm{NC}}(L)$ is strictly decreasing in L.

2. ***Full-commitment threshold:*** *$\kappa^*_{\mathrm{FC}}(L)$ is strictly decreasing in L and $\kappa^*_{\mathrm{FC}}(L) \to 0$ as $L \to \infty$. Consequently, for any $\kappa > 0$ there exists $L^*(\kappa)$ such that for all $L \geq L^*(\kappa)$, implementation-date rebalancing ($x_1^*{=}x_3^*{=}0$) is the unique full-commitment equilibrium.*

Figure 4 illustrates both results: the left panel plots $\kappa^*_{\mathrm{NC}}(L)$ (blue) and $\kappa^*_{\mathrm{FC}}(L)$ (red) as functions of L, with dashed horizontal asymptotes; the right panel shows equilibrium fund cost at a fixed κ, confirming that the committed fund always achieves lower total cost and switches to implementation-date rebalancing at a smaller L.

Figure 4: Effect of the number of speculators L on equilibrium outcomes.

(a) Implementation-date rebalancing thresholds **(b)** Equilibrium fund cost



Notes. Panel 4(a) shows implementation-date rebalancing thresholds $\kappa^*_{\mathrm{NC}}(L)$ (blue) and $\kappa^*_{\mathrm{FC}}(L)$ (red) as functions of L, with dashed lines showing their limits as $L \to \infty$. Both thresholds decrease in L at these parameters; strict monotonicity in L is established in general for $\kappa^*_{\mathrm{FC}}(L)$ and, when $c = 0$, for $\kappa^*_{\mathrm{NC}}(L)$; the no-commitment limit is strictly positive ($\kappa^*_{\mathrm{NC}}(\infty) = (\mu_N + \mu_O)Q/\nu$) while the full-commitment threshold converges to zero (Proposition 4). Panel 4(b) shows fund equilibrium cost at fixed $\kappa = 80$. Solid lines show total cost $\mathrm{TC}^F + \kappa \cdot \mathrm{TE}$; dashed lines show TC^F alone. The FC total cost (solid red) is always below the NC total cost (solid blue), confirming that commitment benefits the fund. When $\kappa^*_{\mathrm{FC}}(L)$ drops below 80 for large L, the FC fund switches to implementation-date rebalancing (zero TE, higher TC^F), whereas the NC fund continues to split trades. Parameters: $Q = 10$, $\lambda_N = 0.5$, $\rho_N = 0.55$, $\lambda_O = 0.2$, $\rho_O = 0.2$, $c = 0.3$, $\nu = 0.15$.

When evaluating a marginal deviation from implementation-date rebalancing, a non-committed fund internalizes the impact saving on its *entire* residual order, a benefit proportional to Q. Hence, the threshold under no commitment remains strictly positive regardless of how many speculators enter.

Because the full-commitment threshold $\kappa^*_{\mathrm{FC}}(L) \to 0$, any fund operating in a sufficiently competitive market and able to credibly pre-commit to its schedule will, in equilibrium, concentrate *all* rebalancing at the implementation date. This provides a direct explanation for the empirical regularity that passive funds almost universally execute at implementation rather than spreading trades across the announcement window: credible commitment, backed by reputation and disclosure requirements, aligns the fund's incentives with the strategy that minimizes total cost, and competition among liquidity providers supports the optimality of the strategy.

4.3.1 Heterogeneous Effects of Competitive Speculators on Fund Investors' Welfare

To isolate how competition affects the fund investors' welfare, define

$$\Delta C^F_{\mathcal{E}}(L;\kappa) \equiv C^F_{\mathcal{E}}(L;\kappa) - C^F_{\mathcal{E}}(0;\kappa), \qquad \mathcal{E} \in \{\mathrm{NC}, \mathrm{FC}\},$$

where $C^F_{\mathcal{E}}(L;\kappa)$ is the equilibrium fund total cost (transaction cost plus tracking-error cost) under regime $\mathcal{E}$ with L speculators. Positive ΔC^F means that speculator participation raises the fund's total cost relative to the no-speculator benchmark; negative ΔC^F means that it lowers total cost.

Figure 5 plots ΔC^F as a function of κ for $L = 1$ and $L \to \infty$, separately for no commitment and full commitment. In both regimes, the effect on fund investors' welfare is heterogeneous across mandates: for low tracking-error penalties, speculator participation increases total cost relative to $L = 0$, while for sufficiently high penalties it reduces total cost. The gain from competition is also heterogeneous in magnitude: moving from $L = 1$ to $L \to \infty$ amplifies both tails, producing larger cost increases in the low-κ region and larger cost reductions in the high-κ region.

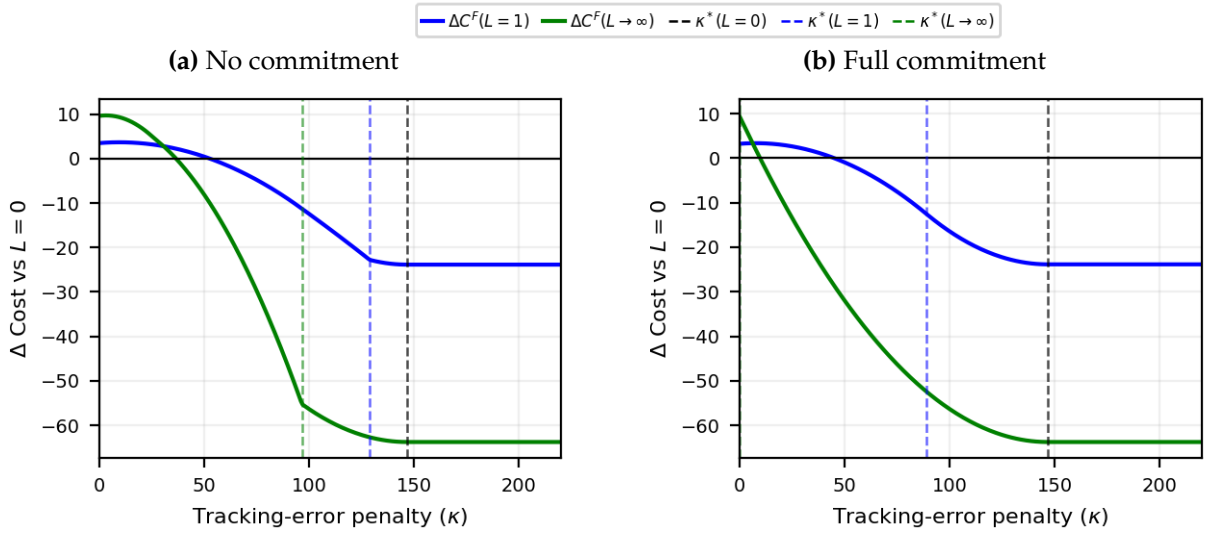
The dashed vertical lines in each panel mark the corresponding implementation-date rebalancing thresholds at $L = 0$, $L = 1$, and $L \to \infty$. Under full commitment, the competitive-limit threshold is essentially at the origin ($\kappa^*_{\mathrm{FC}}(\infty) \approx 0$), consistent with Proposition 4: even very weak tracking-error mandates are sufficient for implementation-date rebalancing once competition is deep and commitment is credible.

4.4 Optimal Arrangement

In the market equilibrium, the fund and the speculators interact non-cooperatively through a market mechanism. We now characterize the fund's *first-best* outcome. Specifically, we study the lowest cost the fund could attain if, instead of facing speculators as strategic counterparties, it could *direct* their trades and settle their accounts through side payments. Formally, suppose the fund chooses its own schedule $(x_1, x_3) \in \mathcal{X}$ together with every speculator's position, and pays each speculator a lump-sum transfer that offsets the speculators' trading costs, so that every speculator exactly breaks even. The fund then internalizes the entire cost of the arrangement, yielding the total cost

$$\mathcal{J} \equiv \underbrace{C^F}_{\text{fund's own cost}} + \underbrace{\left(-\textstyle\sum_{l=1}^{L} \Pi^S_l\right)}_{\text{side payments to speculators}}.$$

Figure 5: Heterogeneous effects of speculator competition on fund investors' welfare across commitment regimes.



Notes. Each panel plots $\Delta C^F(L;\kappa) \equiv C^F(L;\kappa) - C^F(0;\kappa)$, where C^F includes transaction cost and tracking-error cost $\kappa \cdot \text{TE}$. The fund's trading schedule is re-optimized for each κ and each L under the corresponding equilibrium concept; the implementation-date trade is not fixed at $x_2 = Q$. Blue lines show $L = 1$, green lines show $L \to \infty$, and the black horizontal line is the $L = 0$ benchmark. In each panel, dashed vertical lines mark the corresponding implementation-date rebalancing thresholds for $L = 0$ (black), $L = 1$ (blue), and $L \to \infty$ (green). Parameters: $Q = 10$, $\lambda_N = 0.5$, $\rho_N = 0.55$, $\lambda_O = 0.2$, $\rho_O = 0.2$, $c = 0.3$, $\nu = 0.15$.

The total cost includes the fund's own trading cost C^F (inclusive of tracking-error costs) plus the side payments $-\sum_l \Pi_l^S$ that compensate the L speculators for the positions the fund directs them to take. Minimizing $\mathcal{J}$ is therefore equivalent to the cooperative benchmark with binding side payments. Let $\mathcal{J}^\circ(L;\kappa)$ denote the optimized cost under this optimal arrangement.

Proposition 5 (Fund's optimal arrangement)**.** *Suppose the fund directs the speculators' trades and compensates them through side payments that leave each speculator breaking even, choosing the trading schedule and directed speculator positions to minimize its total cost and attain $\mathcal{J}^\circ(L;\kappa)$. Then:*

1. ***Symmetric first best.*** *The fund's unique first-best schedule is symmetric around implementation: $x_1^\circ = x_3^\circ$. There is a threshold $\kappa^\circ(L) \geq 0$ such that the fund rebalances entirely on the reconstitution date ($x_1^\circ = x_3^\circ = 0$) if and only if $\kappa \geq \kappa^\circ(L)$. Below the threshold, it spreads its own trading across all three periods, allocating equally to the pre- and post-implementation dates.*

2. ***Holding costs govern the trade-off.*** *The threshold $\kappa^\circ(L)$ is strictly decreasing in L and $\kappa^\circ(L) \to 0$ as $L \to \infty$. When speculators bear no holding cost ($c = 0$) the threshold*

vanishes; that is, $\kappa^\circ(L) = 0$ for every *$L \geq 1$. In this case the fund fully delegates all pre-positioning and covering to the speculators, who hold their positions costlessly, and simply compensates them through side payments; concentrating its own execution at the reconstitution date is optimal no matter how few speculators operate.*

3. ***First-best versus equilibrium.*** *The first-best threshold satisfies $\kappa^\circ(L) < \kappa^*_{\mathrm{FC}}(L) \leq \kappa^*_{\mathrm{NC}}(L)$. Moreover, provided permanent impact is present on at least one side ($\max\{\lambda_N, \lambda_O\} > 0$), the first-best total cost $\mathcal{J}^\circ(L;\kappa)$ is strictly lower than that of either equilibrium in every regime, with $\mathcal{J}^\circ(L;\kappa) < C^F_{\mathrm{FC}}(L;\kappa) \leq C^F_{\mathrm{NC}}(L;\kappa)$.*

4. ***Permanent impact creates the residual inefficiency.*** *Let*

$$\Delta\mathcal{J}_{\mathrm{FC}}(\infty;\kappa) \;\equiv\; \lim_{L\to\infty}\big[C^F_{\mathrm{FC}}(L;\kappa) - \mathcal{J}^\circ(L;\kappa)\big]. \tag{23}$$

For any fixed $\kappa > 0$, the fund investors' welfare gap $\Delta\mathcal{J}_{\mathrm{FC}}(\infty;\kappa)$ is non-negative. It is zero if and only if $\lambda_N = \lambda_O = 0$; hence competitive speculation implements the first best only in the absence of permanent impact.

The limiting fund investors' welfare gap is driven entirely by permanent price impact: under competitive speculation ($L \to \infty$), speculators absorb the optimal amount of temporary-impact cost, but permanent impact creates an externality that non-cooperative behavior cannot internalize. Only a binding arrangement, one in which the fund directs the speculators' trades and settles their books through side payments, eliminates this residual inefficiency.

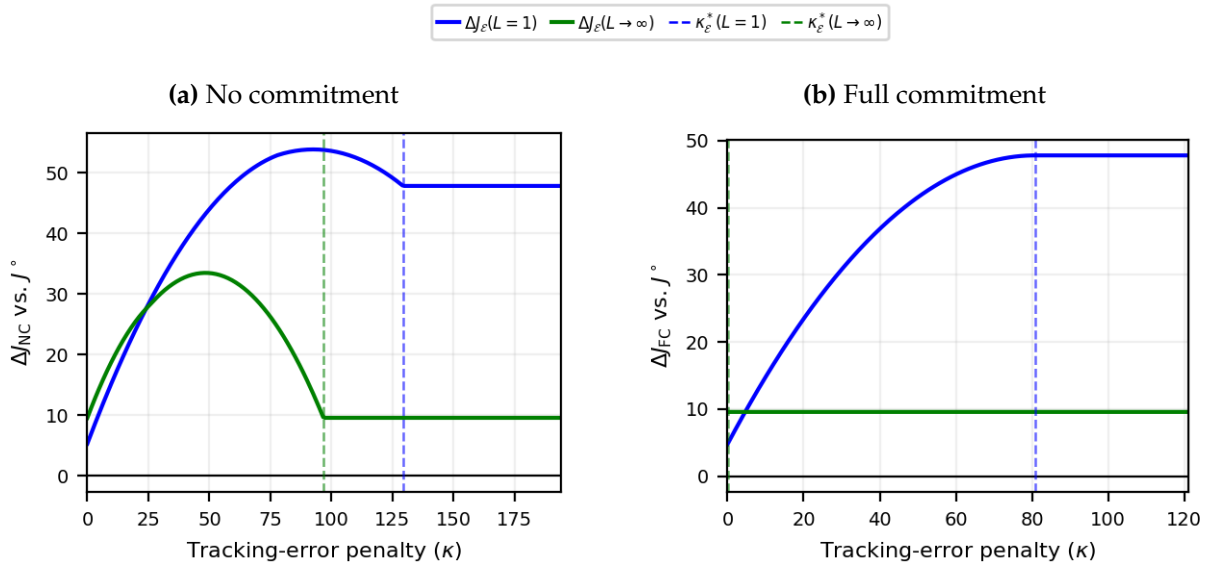
To quantify this coordination loss across mandates, define

$$\Delta\mathcal{J}_{\mathcal{E}}(L;\kappa) \equiv C^F_{\mathcal{E}}(L;\kappa) - \mathcal{J}^\circ(L;\kappa), \qquad \mathcal{E} \in \{\mathrm{NC}, \mathrm{FC}\},$$

where $C^F_{\mathcal{E}}(L;\kappa)$ is the equilibrium fund total cost under concept $\mathcal{E}$ and $\mathcal{J}^\circ(L;\kappa)$ is the first-best benchmark at the same L. Positive $\Delta\mathcal{J}_{\mathcal{E}}$ means that lack of coordination increases the total cost borne by fund investors relative to the optimal arrangement. Figure 6 plots these coordination gaps as functions of κ under no commitment and full commitment.

Figure 7 illustrates the convergence of the full-commitment fund cost to the first-best as L grows. The gap remains strictly positive for all finite L when $\lambda_N > 0$ or $\lambda_O > 0$, and converges to the closed-form limit in (IA.76).

Figure 6: Coordination-loss gap relative to the optimal arrangement across commitment regimes.



Notes. Each panel plots $\Delta\mathcal{J}_\mathcal{E}(L;\kappa) \equiv C^F_\mathcal{E}(L;\kappa) - \mathcal{J}^\circ(L;\kappa)$, where $\mathcal{E} \in \{\text{NC}, \text{FC}\}$. For each curve, the equilibrium and first-best terms use the same number of speculators L. Blue lines show $L = 1$ and green lines show $L \to \infty$. Dashed vertical lines mark the corresponding implementation-date rebalancing thresholds for each equilibrium concept at $L = 1$ (blue) and $L \to \infty$ (green). Positive values indicate higher total investor cost due to lack of coordination between the fund and speculators. Parameters: $Q = 10$, $\lambda_N = 0.5$, $\rho_N = 0.55$, $\lambda_O = 0.2$, $\rho_O = 0.2$, $c = 0$, $\nu = 0.15$.

4.5 Price and Quantity-Based Measure of Speculative Competition

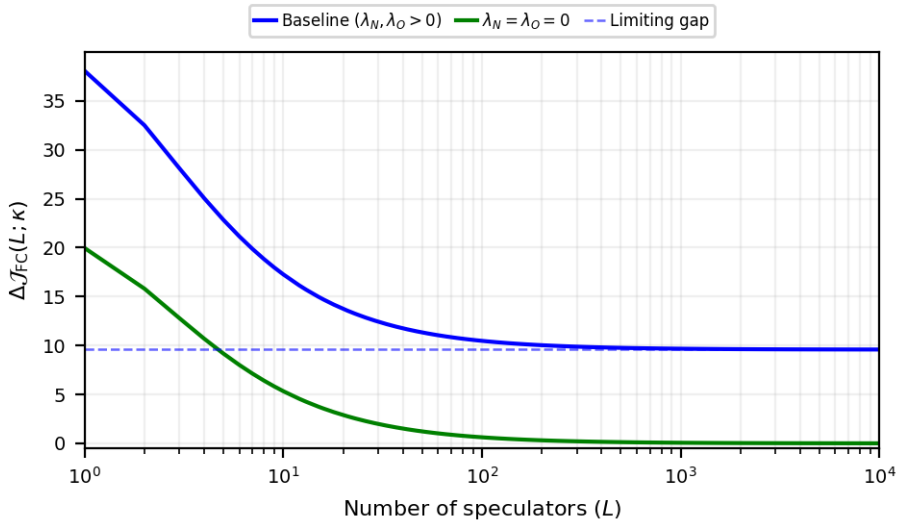
All structural quantities derived above depend on parameters that are not directly observable. We now show that speculative competition leaves a clean signature in transaction prices alone. Define the *post-implementation reversal* as

$$\Gamma(L) \equiv \frac{\mathbb{E}\left[\bar{P}_N - P_{N,2}\right]}{P_{N,0}\, Q}, \tag{24}$$

the gap between the post-trading price $\bar{P}_N = \bar{v} + \lambda_N Q$ and the expected implementation-date transaction price, scaled by the initial N-security transaction price $P_{N,0}$ and the rebalancing quantity Q; taking expectations removes the mean-zero martingale innovations, isolating price-impact components.

Because all positions are closed by period 3, cumulative N-side net order flow satisfies $Z_{N,3} = Q$ regardless of L, so the post-trading price $\bar{P}_N = \bar{v} + \lambda_N Q$ is pinned by fundamentals and the rebalancing size, independent of speculator competition. Using the

Figure 7: Residual coordination loss under full commitment as speculative competition deepens.



Notes. The figure shows how far the full-commitment outcome remains from the optimal arrangement as the number of speculators L grows, holding the tracking-error penalty fixed. The vertical axis is the coordination loss $\Delta\mathcal{J}_{\mathrm{FC}}(L;\kappa)$, the cost borne by fund investors in excess of the first best at the same L; the horizontal axis is on a log scale. The blue line uses the baseline parameters, under which both securities carry permanent impact; the green line switches permanent impact off on both sides, leaving temporary impact unchanged. Deeper competition shrinks the loss in both cases, but it disappears only when permanent impact is absent: otherwise the loss flattens out at the strictly positive limit of Proposition 5(iv), marked by the dashed horizontal line. Competitive speculators end up absorbing the right amount of temporary impact, but no depth of competition leads them to internalize the permanent-impact externality. Parameters: $Q = 10$, $\lambda_N = 0.5$, $\rho_N = 0.55$, $\lambda_O = 0.2$, $\rho_O = 0.2$, $c = 0.3$, $\nu = 0.15$, $\kappa = 80$; the green line sets $\lambda_N = \lambda_O = 0$.

execution-price formula (4) with $Z_{N,1} = S_1(L)$ under implementation-date rebalancing:

$$\Gamma(L) = \frac{\mu_N\, R(L) - \rho_N(Q - S_1(L))}{P_{N,0}\, Q}\,. \tag{25}$$

The second term in the numerator captures the temporary concession (at rate $\rho_N > 0$) on the residual block not pre-positioned by speculators, which depresses the implementation-date price below $\bar{P}_N$; the first term reflects the price impact of period-3 covering flow that has not yet been priced in at period 2 and partially offsets the concession.

Proposition 6 (Post-implementation reversal)**.** *For $\rho_N > 0$, the post-implementation reversal $\Gamma(L)$ defined in* (24) *satisfies:*

1. ***Competitive effects.*** *$\Gamma(L) < 0$ for all $L \geq 1$, and Γ is strictly increasing in L for all sufficiently large L. If $c = 0$, then Γ is strictly increasing in L for every $L \geq 1$.*

2. ***Competitive limit.*** *$\Gamma(L)$ converges to a strictly negative limit,*

$$\Gamma(\infty) \equiv \lim_{L\to\infty} \Gamma(L) = -\frac{\rho_N^3}{P_{N,0}\big(3(\rho_N + \frac{\lambda_N}{2})^2 + \frac{\lambda_N^2}{4}\big)} < 0, \tag{26}$$

where $\Gamma(\infty) \to 0$ as $\rho_N \to 0$.

We complement the price-based measure Γ with a *volume-based* measure of speculative competition. Define the *intermediated volume*

$$\mathcal{V}(L) \equiv \frac{S_1(L) + R(L)}{Q}, \tag{27}$$

the fraction of the fund's rebalancing demand that is absorbed by speculators rather than by the market maker at implementation. Under implementation-date rebalancing, the period-2 net order flow is $z_{N,2} = Q - S_1 - R$, so $\mathcal{V}$ measures how much of Q is intermediated.

Proposition 7 (Intermediated volume)**.** *Under $\rho_N > 0$ and $c \geq 0$, the intermediated volume $\mathcal{V}(L)$ satisfies:*

1. ***Strict monotonicity.*** *$\mathcal{V}(L)$ is strictly increasing in L for all $L \geq 1$.*

2. ***Competitive limit.*** *$\mathcal{V}(L)$ converges to*

$$\mathcal{V}(\infty) \equiv \lim_{L\to\infty} \mathcal{V}(L) = \frac{2(\rho_N + \frac{\lambda_N}{2})^2 + \frac{\lambda_N^2}{2}}{3(\rho_N + \frac{\lambda_N}{2})^2 + \frac{\lambda_N^2}{4}} \in (0, 1), \tag{28}$$

where the displayed ratio converges to 1 as $\rho_N \to 0$.

While Γ captures the price signature of competition, $\mathcal{V}$ captures its quantity signature: as more speculators enter, a larger share of the fund's demand is pre-positioned and covered by intermediaries, reducing the residual order flow that hits the market maker at implementation. The strict upper bound $\mathcal{V}(\infty) < 1$ reflects the same economic force that sustains $\Gamma(\infty) < 0$: temporary price impact creates a wedge that prevents full intermediation even under perfect competition.

4.6 Large Reconstitution Events

The analysis so far treats the fund as the entire reconstitution event: the fund's own demand Q is the flow that speculators anticipate, and the number of speculators is a parameter. Actual reconstitutions are different in two respects. The flow that executes on the

implementation date is the aggregate demand of many index-tracking vehicles, so any one fund is a small part of a large and imperfectly predictable event; and the intermediaries who supply liquidity against that flow enter because the event is large enough to cover their costs, so their number is set by the event rather than given. This subsection embeds a fund of fixed size q in an event of size Q, lets the number of speculators be determined by free entry, and asks what happens to the fund's rebalancing decision as the event grows.

We focus on funds with no commitment power for two reasons. First, in this section we are primarily interested in understanding the optimal strategy of relatively small active funds, who could strategically exploit predictable price patterns to minimize their transaction cost. Second, Proposition 4 shows that competition alone does not fully remove the fund's incentive to trade away from the implementation date. Therefore, any convergence to implementation-date rebalancing in this extension can be attributed to the interaction between event size and endogenous speculator entry. We will show that, as Q grows, the no-commitment threshold converges to zero, so that a fund with *any* tracking-error penalty eventually rebalances entirely on the implementation date.

SETTING. An exogenous index-tracking flow buys ζ_1 shares of N in period 1, $Q + \zeta_2$ shares in period 2, and ζ_3 shares in period 3, selling the same quantities of O. The announced block Q is known at $t = 0$; the shocks ζ_t are realized and publicly observed at date t, before that date's trading, and are independent of each other and of the fundamental innovations, with

$$\mathbb{E}[\zeta_t] = 0, \qquad \mathrm{sd}(\zeta_t) = \chi\, Q, \qquad t \in \{1, 2, 3\}, \tag{29}$$

and bounded support. The fundamental innovations $\{\varepsilon_{J,t}\}$ of (1) are independent across dates and independent of the shocks. The parameter $\chi > 0$ is the standard deviation of the mis-estimate of the event's flow, as a fraction of the event: ζ_1 is anticipatory tracking flow ahead of the date, ζ_2 is uncertainty about how much tracking capital actually executes on the date, and ζ_3 is the unpredictable post-event flow.

The fund needs to rebalance a position of size q and, under no commitment, chooses its trades as the shocks arrive: x_1 may depend on ζ_1, x_3 on (ζ_1, ζ_2), and $x_2 = q - x_1 - x_3$, to minimize $\mathbb{E}[\mathrm{TC}^F] + \kappa \sum_t \mathrm{TE}_t$. Speculators do the same: each speculator's period-1 trade may depend on ζ_1 and its period-2 trade on (ζ_1, ζ_2), with the period-3 trade closing the position. They compete in Cournot fashion as before. The fund takes the speculators' *shock-contingent policies* as given, with strategies now functions of the observed flow, and

an equilibrium is a Nash equilibrium in policies. At $t = 0$, a speculator pays a sunk cost K to participate on both sides and enters if the expected profit of the resulting equilibrium covers K; we denote the resulting number of speculators on each side by $L^{\mathrm{FE}}(Q)$. Tracking error is measured as in Section 4.1: the dispersion of the fund's index-relative return over each period, so that flow-driven price movements enter through the model's own impact coefficients.

The fund trades off transaction costs against tracking error. As the event Q grows, the fund leans further toward minimizing tracking error. As the event gets larger, more speculators pay K and enter, and each entrant takes on flow until it earns just enough to cover that cost. The fund therefore competes for the same price pressure against a crowd that grows with the event. Hence, the gain it can capture by trading outside the implementation date remains bounded as the event grows. However, the risk of unexpected deviations from the index's return grows with the event, because flow uncertainty scales with the size of the event. As a result, past some event size, the fund abandons early and late trading and rebalances entirely on the implementation date, however small its tracking-error penalty, in order to minimize the risk of large unexpected deviations from the index. We formalize this intuition in the following proposition.

Proposition 8 (Large events under no commitment)**.** *Fix $q > 0$, $K > 0$, $\chi > 0$ and $c = 0$, and let $L = L^{\mathrm{FE}}(Q)$.*

1. *There is a threshold $\kappa^*_{\mathrm{NC}}(Q; q)$ such that implementation-date rebalancing in every state, together with the speculators' state-contingent best responses to it, is an equilibrium with no commitment if and only if $\kappa \geq \kappa^*_{\mathrm{NC}}(Q; q)$; for $\kappa > \kappa^*_{\mathrm{NC}}(Q; q)$ it is the fund's unique best response among all state-contingent policies.*

2. ***Large-event limit:*** *as the size of the event grows,*

$$\lim_{Q \to \infty} \kappa^*_{\mathrm{NC}}(Q; q) = 0. \tag{30}$$

   Consequently, for any $\kappa > 0$ there exists $\bar{Q}(\kappa)$ such that for all $Q \geq \bar{Q}(\kappa)$, implementation-date rebalancing in every state is the fund's unique best response and the profile of item 1 is an equilibrium with no commitment.

The proposition should be interpreted against Proposition 4 and against free entry without flow uncertainty. In the competitive limit the event was the fund's own demand Q and the number of speculators grew without bound; the threshold converged to $(\mu_N + \mu_O)Q/\nu > 0$ because the fund's gain from thinning its block is infra-marginal

and proportional to Q, while the tracking scale ν is a constant. Here the fund's own block is q, so its infra-marginal gain is proportional to q and bounded; and free entry, unlike the competitive limit, also leaves the run-up and the reversal at a positive level, so that with a perfectly predictable flow the threshold would stay bounded away from zero, with the strictly positive limiting floor of Corollary IA.1. With an imperfectly predictable flow the event enters the tracking scale, which grows with Q, while the intermediaries' capacity to absorb the event's surprises grows with the event through entry.

Proposition 8 relies on three key conditions. *(i)* The fund is small relative to the event: q is held fixed as Q grows, so the fund's own price impact stays bounded while the event's does not. *(ii)* χ is bounded away from zero. At $\chi = 0$, there are no flow shocks and, as Q increases, the risk of unexpected deviations from the benchmark remains fixed. As a result, the trade-off between profitable deviations from the benchmark and tracking-error risk remains constant and $\kappa^*_{\mathrm{NC}}(Q; q)$ does not converge to 0. *(iii)* The shocks have bounded support. With unbounded shocks a large enough realized flow always justifies some trade, so no finite κ makes implementation-date rebalancing optimal in *every* state. However, as we show in the next proposition, even with unbounded shocks, implementation-date rebalancing becomes optimal in almost all states as Q grows large.

Proposition 9 (Unbounded flow shocks: execution of the mandate)**.** *Maintain the setting above but drop the assumption that the shocks are bounded: let $\zeta_t = \chi Q\, u_t$ with u_t i.i.d., mean zero, $\mathrm{sd}(u_t) = 1$, independent of the fundamental innovations, with support not necessarily bounded and not necessarily symmetric. Fix $q, K, \chi, \kappa > 0$ and $c = 0$. Then, in every equilibrium with no commitment, the fund executes almost all of its trades on the implementation date as the event grows:*

$$\lim_{Q\to\infty} \frac{\mathbb{E}[x_2^e]}{q} = 1. \tag{31}$$

Consequently the fund's pre- and post-event trading converges to zero in probability.

With bounded shocks the statement is deterministic and state by state: beyond a finite threshold the fund rebalances at implementation in *every* realization of the flow. With unbounded shocks, a sufficiently extreme realization always makes some off-implementation trading worthwhile. However, such extreme shocks become increasingly unattractive to exploit for the fund as Q grows: free entry and competition from speculators bound the fund's gains from such trades, while the risk of accruing tracking error grows in Q. Therefore, even weak index trackers are very likely to rebalance at implementation when they face large order flows on the implementation date.

5 IMPLICATIONS: FAST-TRACK INCLUSIONS AND LIQUIDITY COMPLEMENTARITY

In this section, we apply the theoretical framework of Section 4 in two directions. We first derive the model's implications for the price path, trading volume, and price discovery around index-reconstitution events. We then study fast-track index additions after an initial public offering (IPO), evaluating the execution costs they impose on index funds and their effect on IPO prices. To map the model to the data more directly, we allow for multiple trading periods. Throughout the section, we take the fund's rebalancing schedule as exogenously fixed at implementation-date rebalancing. We also set the number of speculators to an exogenous quantity L and do comparative statics as it changes.[4] We then study the speculators' equilibrium response and the implied prices and trading volume.

5.1 MULTI-PERIOD SETUP

We denote by T the implementation date. There are $T-1$ pre-implementation periods and H post-implementation periods, for a total of $T+H$ trading periods, all indexed by $t \in \{1, \ldots, T+H\}$. As in the base model, period 0 represents the initial state and a terminal period at $T+H+1$ represents the realization of the fundamental value. Setting $T=2$ and $H=1$ recovers the base model exactly.

The fund's schedule is fixed at implementation-date rebalancing: $x_T = Q$ and $x_t = 0$ for $t \neq T$. Each speculator l chooses trades $(s_{N,1,l}, \ldots, s_{N,T+H,l})$, with cumulative positions $S_{N,t,l} = \sum_{\tau=1}^{t} s_{N,\tau,l}$, subject to the zero-net-position constraint $S_{N,T+H,l} = 0$, and analogously on the O-side. In every trading period, transaction prices are given by the price-impact function (4), and each speculator maximizes the multi-period extension of the profit (12), with the holding cost (11) accruing on cumulative positions in every period. Because speculator profit remains a quadratic function of trades, the symmetric Cournot best response is obtained by numerically solving a $(2(T+H)) \times (2(T+H))$ linear system; with the fund's schedule fixed, these best responses pin down order flows, prices, and reversals.

[4] An equivalent comparative static is to consider endogenous speculator entry and vary the entry cost K. Because the final equilibrium implications depend on the number of active speculators and not the entry cost itself, we do comparative statics with respect to the number of active speculators to emphasize the economic force, namely competition, behind the results.

5.2 Price Run-Up, Trading Volume, and the Role of Speculation

The model's most direct empirical implications concern the shape of the price path and trading volume around implementation. A set of predictions emerges that corresponds to empirical regularities documented in the literature: the secular decline in implementation-day abnormal returns, the migration of price discovery into the pre-event window, and the concentration of index-driven volume in the closing auction on the effective date.

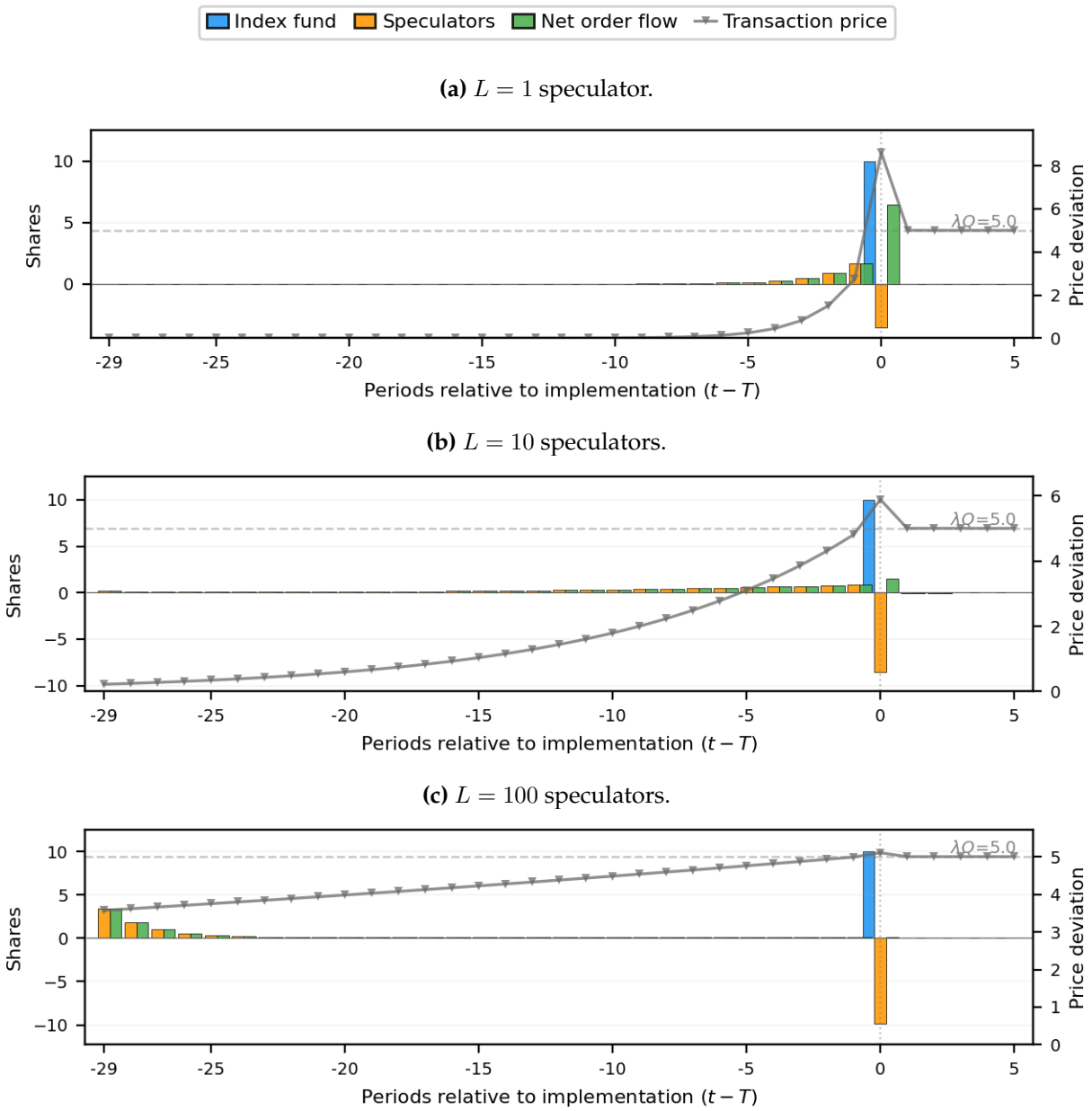
Figure 8 traces the predicted order flows and execution-price path for the added security when the fund rebalances entirely at period $T = 30$.

With $L = 1$, speculators accumulate few shares pre-T, keeping pre-implementation prices low. As speculators sell their limited inventory at implementation against the large demand of the fund, prices spike at implementation, far above the post-trading price $\bar{P}_N = \bar{v} + \lambda_N Q$ of (5), thus creating a large reversal. At $L = 10$, greater competition drives pre-T accumulation higher, thus smoothing the price spike at implementation and subsequent reversal. At $L = 100$, speculators are highly competitive, bid aggressively for the stock before implementation, and, as a result, push prices before and at implementation near the post-trading price $\bar{P}_N$. Therefore, the model predicts a cross-sectional pattern: securities added to a competitive market should exhibit immediate price changes at the announcement and a smaller implementation-date spike than securities added to uncompetitive markets.

Speculators provide a form of price discovery by impounding the known demand shock into transaction prices as soon as information is available and they can trade on it. The same competition mechanism offers a time-series rationalization for the disappearing index effect documented by Greenwood and Sammon (2025): competition among speculators intensified and price discovery migrated progressively earlier in event time. The model predicts that this process leaves total cumulative price adjustment unchanged: it equals the permanent price impact $\lambda_N Q$ of the rebalancing quantity, independent of L. However, competitive speculators redistribute this price impact from the implementation date to the pre-event period, precisely the pattern documented by Greenwood and Sammon (2025).

Moreover, these results provide a demand-side rationalization for the disproportionate concentration of index-driven trading in the closing auction on effective dates documented by Chinco and Sammon (2024): the closing auction is the economically optimal execution venue for any benchmarked investor whose tracking-error penalty exceeds the threshold κ^* (Proposition 2), not merely a focal convention. As the intermediation sector has grown and speculator competition has intensified, the threshold has declined toward

Figure 8: Order flows and transaction price under implementation-date rebalancing.



Notes. $x_T = Q$, $x_t = 0$ for $t \neq T$; periods 1–35 shown (activity negligible thereafter). Bars: fund trade (blue), aggregate speculator flow (orange), net order flow (green). Gray line: transaction price, net of fundamental value (right axis); dashed line: $\lambda_N Q = 5$. Greater speculator competition shifts price pressure from the implementation spike to a pre-T run-up. Parameters: $T = 30$, $Q = 10$, $\lambda_N = 0.5$, $\rho_N = 0.55$, $c = 0.3$, $H = 29$.

zero (Proposition 4), expanding the set of fund types for which implementation-date rebalancing is individually optimal.

5.3 Fast-Track Inclusions

We now evaluate the cost index trackers incur when a stock is included in the index shortly after its IPO. Newly listed stocks tend to be illiquid (high ρ_N) and, by fast-tracking their inclusion (lowering T), index providers force speculators to concentrate their inventory accumulation over a short period of time, exacerbating costs for index trackers. Speculators can make up for the short pre-implementation window by shorting the stock at implementation and closing their short positions afterward. However, newly issued stocks are hard to short, as the supply of lendable shares is severely limited. To capture this friction, we compare the baseline with a no-short-selling variant in which each speculator's cumulative position in the added security is constrained to be non-negative, $S_{N,t,l} \geq 0$ for every t.

Fast-track additions are costly for index investors. We evaluate the cost for index trackers using the model-implied measure of avoidable transaction costs: the post-implementation reversal Γ of Proposition 6, where a more negative Γ means that the implementation-date price sits further above the post-trading price $\bar{P}_N$ and the subsequent reversal is larger. The left column of Figure 9 shows that, in general, index trackers incur larger price reversals when the stock is illiquid (high ρ_N) and speculators are not competitive (low L). However, as long as short selling is possible and speculation is sufficiently competitive, speculators can intermediate the demand shock by shorting at implementation and covering afterward, so index trackers incur only a marginally worse reversal when the stock is fast-tracked after its IPO (low T).

In contrast, the right column of Figure 9 shows that when short selling is prohibited, the reversal is much higher for fast-track inclusions, especially when the stock is illiquid. Even in highly competitive markets (high L), index trackers incur substantial reversal costs in illiquid stocks because speculators cannot intermediate the demand shock by shorting at implementation and covering afterward.

Fast-track additions increase IPO prices. Whereas index trackers suffer from fast-track inclusions of illiquid, hard-to-short stocks, issuers greatly benefit from them. The left column of Figure 10 shows that when short selling is unconstrained, the first-trading-day price is only moderately higher for fast-track inclusions, even when the stock is illiquid. In contrast, the right column of Figure 10 shows that when short selling is

prohibited, the first-trading-day price is much higher for fast-track inclusions, especially when the stock is illiquid. Because speculators are forced to build inventory before implementation with no option to short at implementation, they bid up the price prior to implementation, leading to substantially higher prices on the first trading day for fast-track inclusions of illiquid, hard-to-short stocks.

Interestingly, issuers and initial stock owners benefit more when markets are highly competitive. As discussed earlier, competitive speculators trade aggressively and immediately bid up prices. In the context of an IPO, competitive speculators will therefore bid fiercely at the IPO stage or when the stock begins trading, ultimately transferring value to the issuer and initial stock owners.

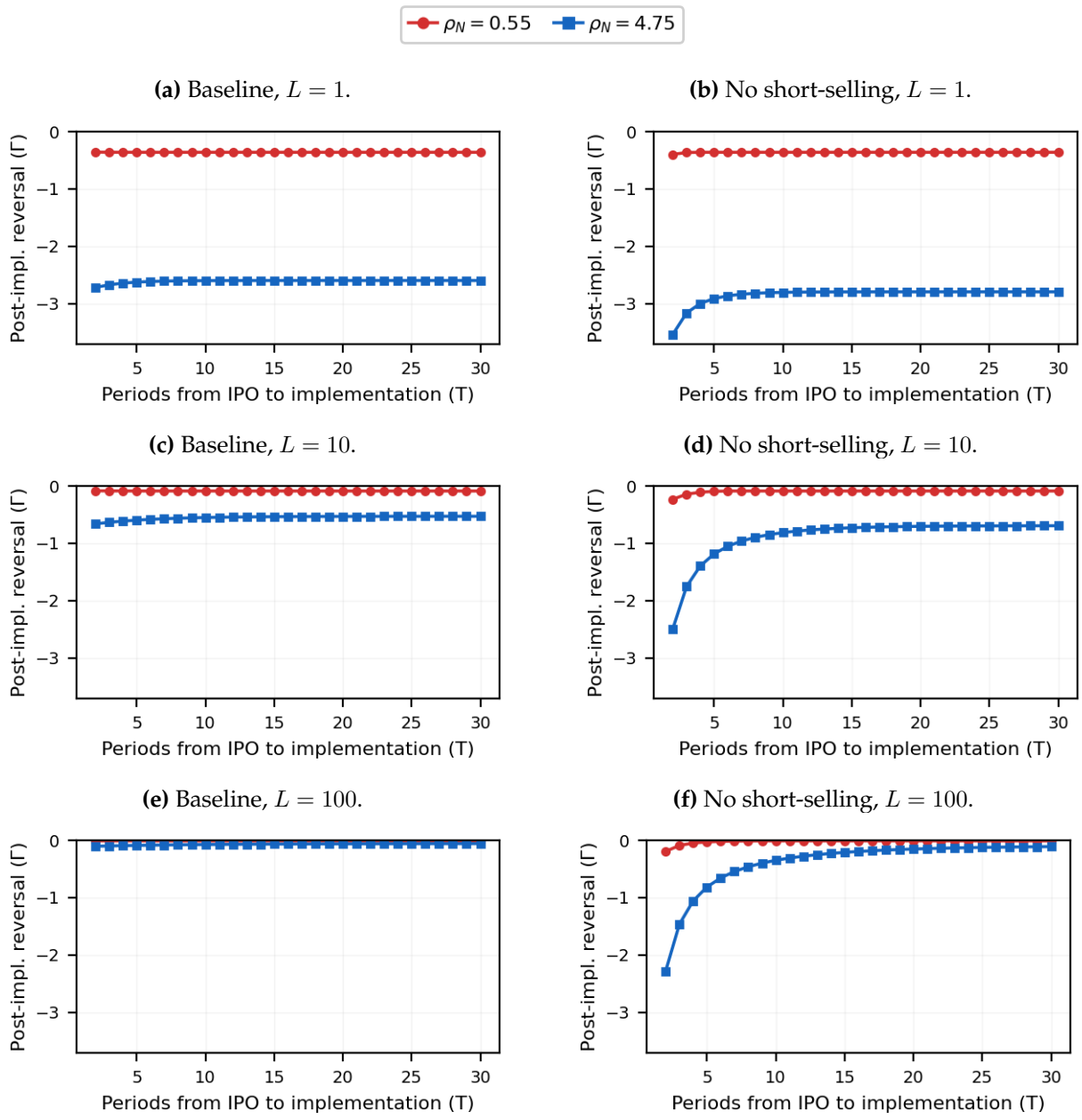
Overall, these predictions are consistent with the findings in Murray and Sammon (2026). Our model provides a framework to rationalize and qualify their empirical findings. First, the model highlights that index investors suffer costs from fast-track additions only as long as markets are illiquid (high ρ_N) and short selling is constrained. Both conditions are likely after an IPO. In contrast, the surprise addition of a liquid and established stock to an index could be performed at a low cost to investors even with a short pre-implementation trading period, as trades are spread over the post-implementation period by short selling. Second, issuers and initial stock owners especially benefit from fast-tracking in highly competitive markets. Conversely, in the less competitive markets of the past, speculators would have retained a larger share of total profits.

Figure 11 illustrates the inventory and price dynamics behind these patterns. In the baseline, speculators accumulate a moderate long inventory before implementation, sell more than that inventory into the fund's demand, and cover the resulting short position afterward; intermediation is spread over both sides of the implementation date, so shortening the pre-implementation window (low T) matters little. When short selling is prohibited, all intermediation must run through inventory accumulated before implementation, which is expensive to build quickly in an illiquid stock: speculators bid up pre-implementation prices, index trackers face a larger reversal, and issuers can expect high stock prices on their first trading day.

6 Interpreting Index Events Through the Model

This section uses the model to reinterpret several empirical facts around index reconstitution. The first fact examines implementation-day trading volume relative to demand from strict index trackers, which speaks to apparent coordination among benchmark-linked investors. The second fact uses post-implementation day reversals to measure the

Figure 9: Post-implementation reversal Γ as a function of the pre-implementation horizon T: baseline versus no short-selling.



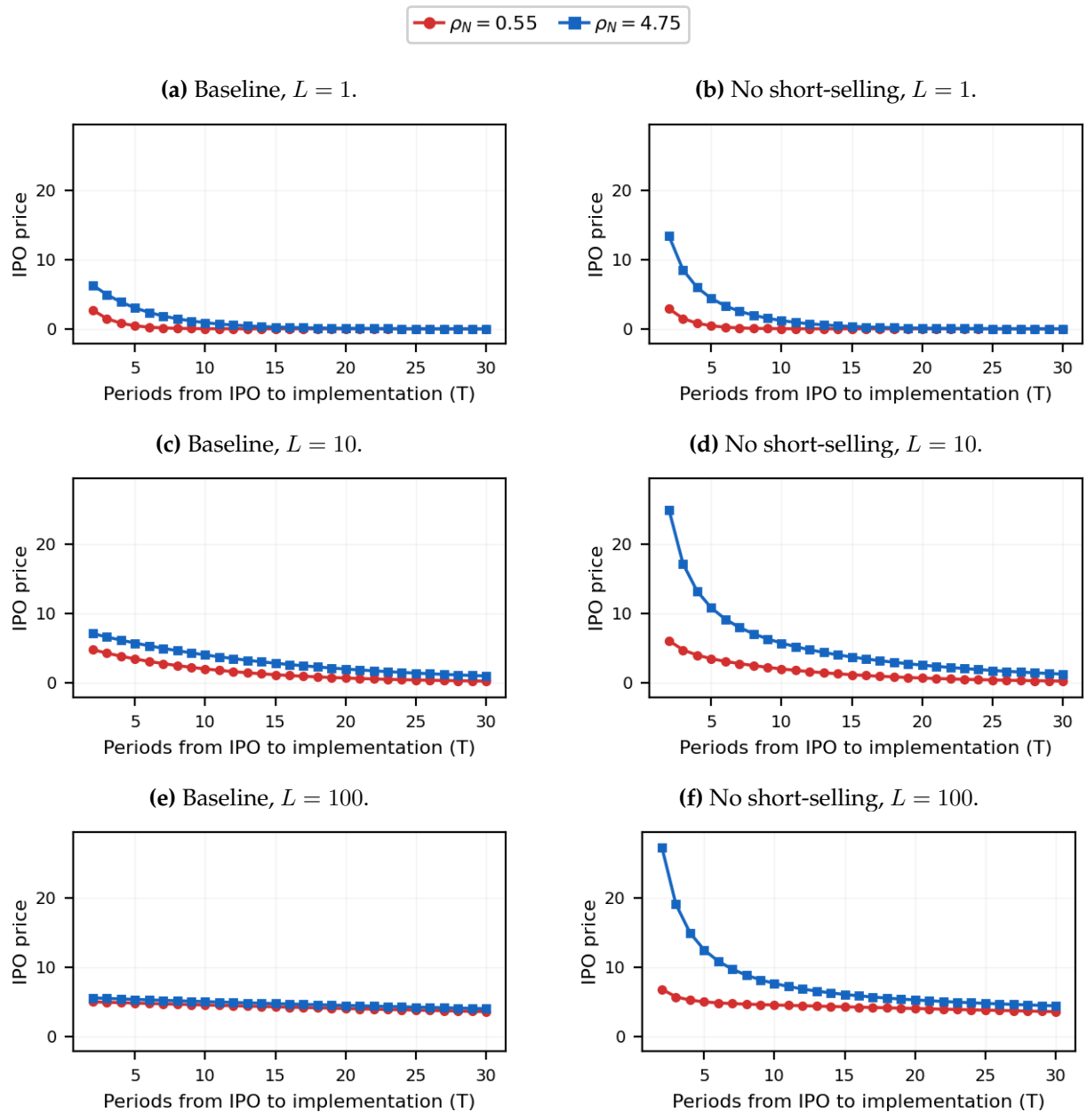
(a) Baseline, $L = 1$. **(b)** No short-selling, $L = 1$.

(c) Baseline, $L = 10$. **(d)** No short-selling, $L = 10$.

(e) Baseline, $L = 100$. **(f)** No short-selling, $L = 100$.

Notes. Red and blue lines denote $\rho_N = 0.55$ and $\rho_N = 4.75$, respectively. In the left column, short selling is unrestricted. In the right column, short selling is not allowed. Rows show outcomes with $L = 1$ speculator (first row), $L = 10$ speculators (second row), and $L = 100$ speculators (third row). Parameters: $Q = 10$, $\lambda_N = 0.5$, $c = 0.3$, $H = 29$.

Figure 10: IPO price (first-trading-day transaction price at $t = 1$, net of fundamental value) as a function of the pre-implementation horizon T: baseline versus no short-selling.



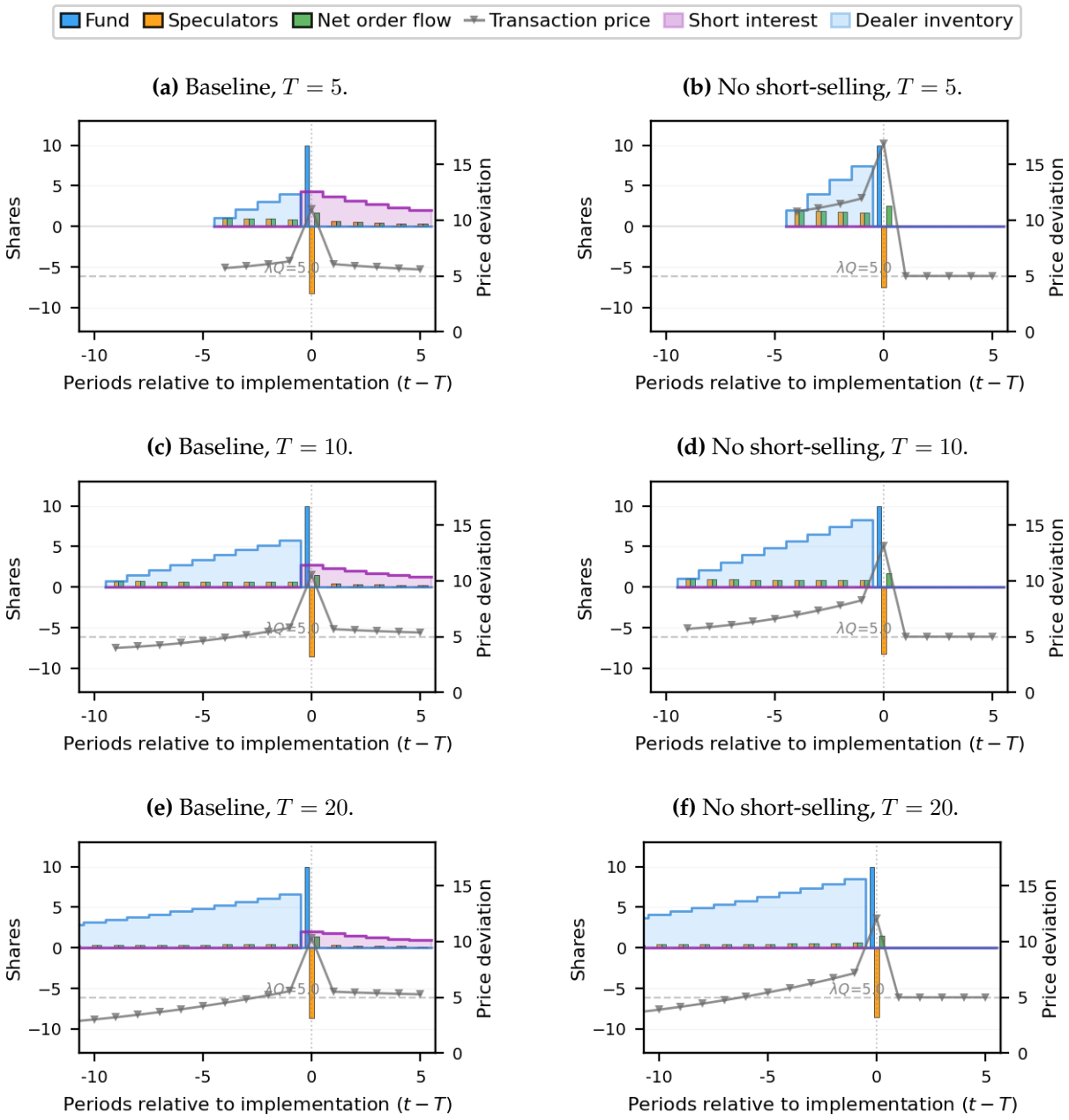
(a) Baseline, $L = 1$. **(b)** No short-selling, $L = 1$.

(c) Baseline, $L = 10$. **(d)** No short-selling, $L = 10$.

(e) Baseline, $L = 100$. **(f)** No short-selling, $L = 100$.

Notes. Red and blue lines denote $\rho_N = 0.55$ and $\rho_N = 4.75$, respectively. In the left column, short selling is unrestricted. In the right column, short selling is not allowed. Rows show outcomes with $L = 1$ speculator (first row), $L = 10$ speculators (second row), and $L = 100$ speculators (third row). Parameters: $Q = 10$, $\lambda_N = 0.5$, $c = 0.3$, $H = 29$.

Figure 11: Order flows, short interest, and speculators' inventory under high temporary impact ($\rho_N = 4.75$): baseline versus no short-selling.



Notes. Rows set $T = 5$ (first row), $T = 10$ (second row), and $T = 20$ (third row); columns compare unrestricted short selling (left) and no short selling (right). Bars: fund trade (blue), aggregate speculator flow (orange), net order flow (green). Purple shading: speculator short interest $\max(-L\,S_{N,t,l}, 0)$; light blue shading: speculators' inventory $\max(L\,S_{N,t,l}, 0)$, where $S_{N,t,l}$ is the symmetric cumulative speculator position. Gray line: transaction price, net of fundamental value (right axis); dashed line: $\lambda_N Q = 5$. Parameters: $Q = 10$, $\rho_N = 4.75$, $\lambda_N = 0.5$, $c = 0.3$, $L = 10$, $H = 29$.

avoidable component of price impact. Third, we use short interest to identify how intermediaries build and unwind positions around index reconstitutions. The fourth fact studies the link between variation in how predictable index reconstitutions are and abnormal trading volume. Finally, we combine reversals with the size of index-linked demand to provide a back-of-the-envelope calculation for the effective cost borne by index investors.

The main empirical setting is S&P 500 direct additions, events that involve stocks entering the S&P 500 from outside the S&P index universe (i.e., from outside the S&P 400 mid cap and S&P 600 small cap). These are useful in our setting, as we do not need to estimate the size of possibly offsetting demand shocks (as would be the case for e.g., migrations from the S&P 400 to the S&P 500). For some tests, we also use Russell index migrations to study cases where the expected index demand shock is negative (e.g., when stocks switch from the Russell 2000 to the Russell 1000). In addition, given the mechanical nature of Russell reconstitution, we use Russell indexes to isolate when stocks have anticipated and less-anticipated demand shocks.[5]

6.1 Coordinated Trading on the Implementation Day

The first fact builds on Chinco and Sammon (2024), showing that implementation-day trading is too large to be explained by the demand of strict index trackers (i.e., index funds and ETFs) and loose index trackers (identified using the method in Behmaram and Davis (2026)) alone. Quantifying trading from strict and loose trackers is important because this will be needed to quantify the avoidable component of trading costs for index fund investors.

To measure demand from strict index trackers, we identify plain-vanilla index mutual funds and ETFs tracking the S&P 500. We use CRSP mutual fund classifications, fund names, active share and realized tracking error to classify funds. For example, we identify S&P 500 funds using the CRSP/Lipper S&P 500 flag when available, as well as name-based matches for funds whose names explicitly reference the S&P 500. For this exercise, we exclude style, sector, factor, ESG, non-value weighted, and enhanced-index products. We also drop funds whose holdings counts or realized tracking errors are inconsistent with tracking the S&P 500 (e.g., stocks with a stated benchmark to the S&P 500 but hold less than 400 securities). This procedure gives us a conservative measure of strict index-fund demand for each index.

[5]We do not use Russell tests when examining reversals or, in general, when tests involve prices. The mechanical Russell reconstitution involves ranking stocks based on market capitalization, which is directly affected by prices. That is, returns in the days and weeks leading up to the rank date can directly affect anticipated index-linked demand.

We then supplement this measure using the broader shadow-indexing measure from Behmaram and Davis (2026), who expand on Pavlova and Sikorskaya (2023) by constructing a *holdings*-based measure of realized passive ownership adjusted for Active Share (Cremers and Petajisto, 2009). This broader measure is useful because many benchmark-linked investors do not explicitly identify themselves as index funds, but still hold portfolios that closely resemble an index and therefore may trade around index events to minimize tracking error.

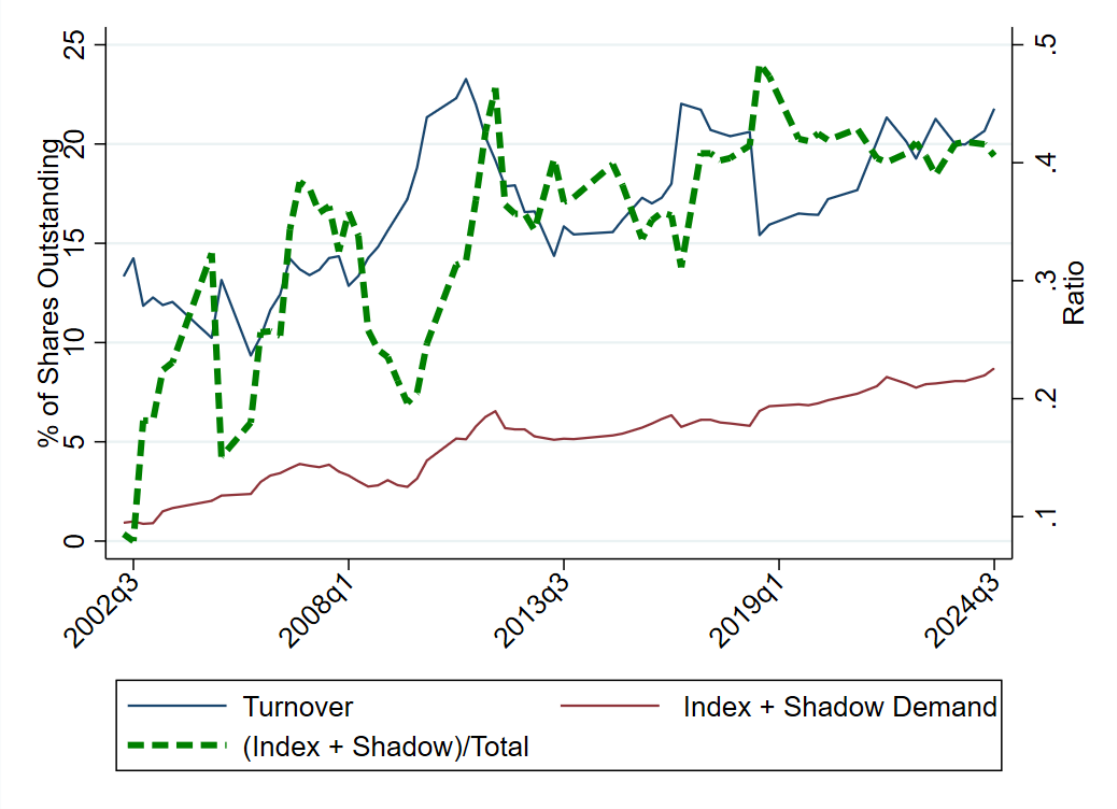
Figure 12 shows the comparison of S&P 500 direct addition trading activity and predicted index-linked demand. The blue solid line shows total implementation-day turnover. The red solid line shows estimated demand from strict S&P 500 index funds plus shadow S&P 500 index trackers. The dashed green line shows the ratio of index-plus-shadow demand to total turnover. Two facts stand out. First, total turnover on the implementation day is extremely large, often exceeding 15% to 20% of shares outstanding. Second, even after including shadow-index demand, benchmark-linked demand explains only part of total implementation-day trading (less than 50%). The red line shows that, if all strict and shadow index trackers traded on implementation day, we would expect turnover of less than 10% of shares outstanding. The ratio of index demand to implementation-day turnover, plotted in the dashed green line, shows this more clearly: the ratio rises over time but remains well below one and is closer to 0.4.

Our interpretation of these patterns is that implementation-day trading is not limited to strict index funds. Of course, one might argue that some of the additional volume is driven by intermediation, but given that most trading on these days takes place in the closing auction, or at trades linked to the closing auction price (Chinco and Sammon, 2024), we believe this is unlikely to be the main driver of the difference. Our model provides an interpretation for the "missing" passive investors: many non-strict index trackers and shadow indexers not captured by the Behmaram and Davis (2026) measure, as well as other benchmark-linked investors, choose to trade on the same day as strict trackers. This is exactly the type of coordination emphasized by the model. When tracking-error concerns are sufficiently high, or when intermediary competition is sufficiently strong, benchmark-linked investors optimally concentrate trading on the implementation day.

6.2 Post-Implementation Reversals and the Avoidable Component of Price Impact

The second fact is that post-implementation price reversals have fallen over time. In standard execution models, we typically find that larger trades generate larger tempo-

Figure 12: Implementation-Day Turnover and Benchmark-Linked Demand



Notes. This figure shows implementation-day trading and demand from strict and shadow index trackers for S&P 500 direct additions. The blue line shows total implementation-day turnover, measured as trading volume scaled by shares outstanding. The red line shows estimated index-plus-shadow demand, also scaled by shares outstanding. The dashed green line shows the ratio of index-plus-shadow demand to total implementation-day turnover. Strict index trackers are identified using CRSP mutual fund classifications, fund names, holdings counts, and realized tracking error. Shadow-index demand is based on the holdings-based approach in Behmaram and Davis (2026). The figure presents five-year moving averages.

rary price impact and therefore larger reversals. Our model predicts, however, that the relationship between trading volume and reversals depends on the degree of intermediation. As competition among speculators increases – i.e., as L increases – more of the implementation-day demand is pre-positioned. This spreads price impact over time and reduces the temporary component of price impact that reverts – all of the shares demanded by index trackers is provided with little to no price impact on the implementation day.

The model also clarifies the right empirical object to quantify the amount and cost of intermediation taking place. While the raw reversal (i.e., return) is informative, the

model's analogue is the reversal scaled by the size of the index demand shock:

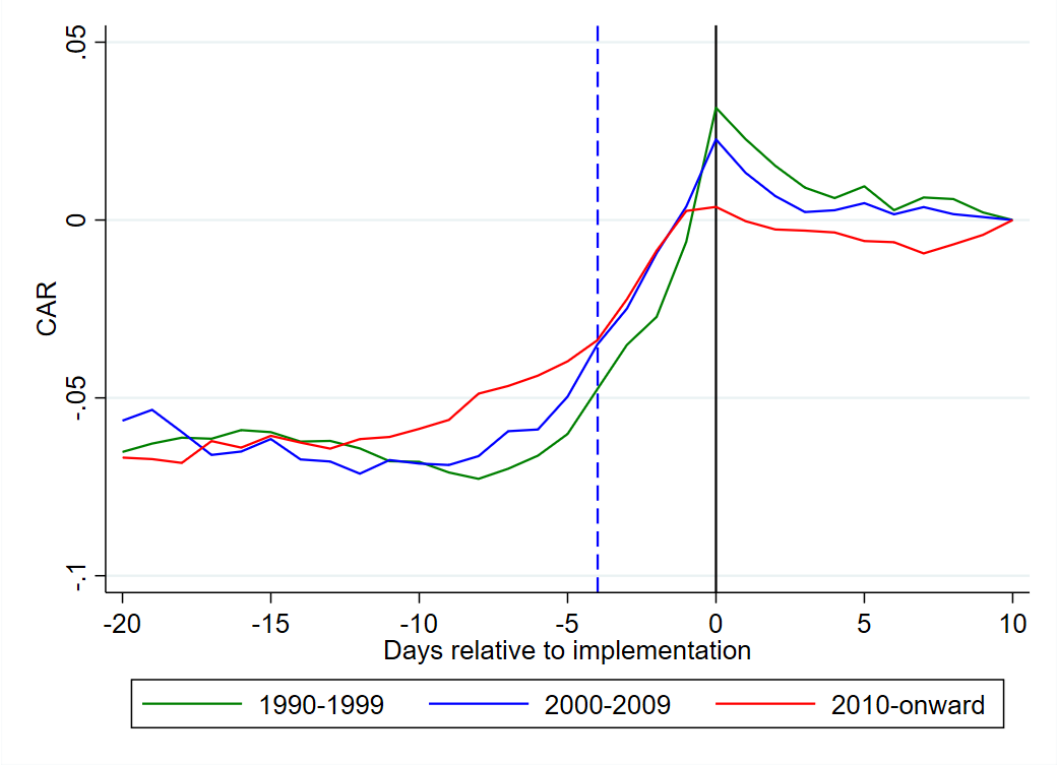
$$\Gamma(L) \equiv \frac{P_{\text{post}} - P_{\text{implementation}}}{Q}.$$

This object measures the avoidable component of price impact per unit of benchmark-linked demand, not the full implementation shortfall faced by index investors. If index demand grows over time while reversals shrink, then $\Gamma(L)$ must have fallen substantially. The decline in reversals *despite* growing index-linked demand therefore points to a large increase in intermediation capacity or competition among intermediaries.

Figure 13 shows basic return evidence for S&P 500 direct additions. Specifically, we plot cumulative abnormal returns around implementation by decade, normalizing prices so that the post-event price is the same 10 days after the implementation day ($t = +10$). This normalization makes it easier to see how the reversal has changed over time. In the 1990s, prices rise into implementation and then reverse noticeably afterward. In the 2000s, the reversal is smaller even though index-linked demand is larger. In fact, the 2000s is when many index ETFs were introduced and saw rapid growth as an investment vehicle. In the 2010-and-onward period, the post-implementation reversal is essentially zero, despite still larger index-linked demand. Importantly, the *total* price movement is roughly similar across time – the starting point at $t = -20$ is almost the same across decades even though returns have been normalized to zero at $t = +10$. The major change over decades is how much of the implementation-day price effect takes place before implementation. For example, in the 2010 and onward sample, a larger share of the return happens before the *announcement* itself.

Figure 14 shows the time-series version of the same fact. Panel A plots five-year moving averages of post-event reversals for S&P 500 direct additions, using five-day and ten-day post-implementation-day windows. In the 1980s, there was continuation, while starting in the 1990s, there was post-implementation reversal. Reversals started to shrink around 2000 and are nearly zero by the end of our sample. Panel B divides the same reversal by the size of the index-linked demand shock estimated in Figure 12. We focus on the post-2002 period for this sample based on data availability for the size of the index-linked demand shock. Panel B shows that the decline in reversals is even more striking after this normalization. Index-fund demand has increased substantially, yet the reversal per unit of demand has fallen to almost zero. In addition, we know from Section 6.1 that our measure of index-linked demand accounts for a little less than half of all demand as measured by implementation-day turnover, most of which comes from the closing auction. Accounting for the hidden index-linked demand would push the reversal per unit

Figure 13: Price Paths Around S&P 500 Direct Additions by Decade



Notes. This figure shows cumulative abnormal returns around S&P 500 direct additions, separated by decade. Event time is measured relative to the implementation day of the index reconstitution. The solid vertical line marks the implementation day, and the dashed vertical line marks the average announcement day. Cumulative abnormal returns are normalized so that each decade's average price path equals zero at day $t = +10$.

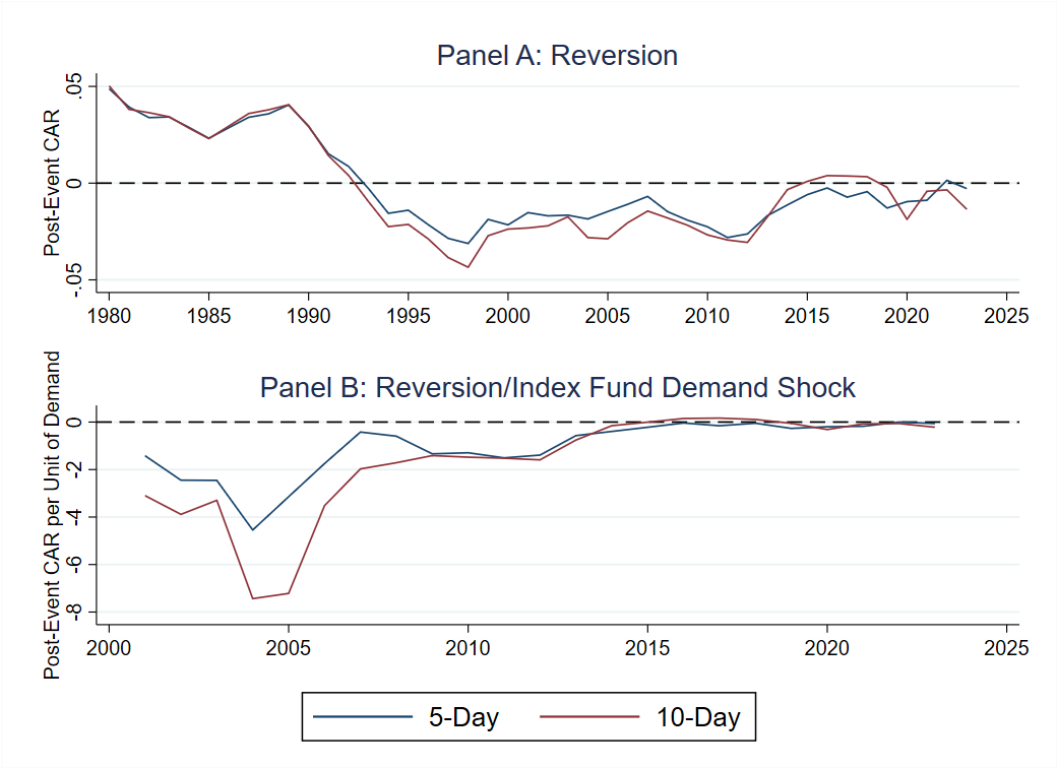
of demand even closer to zero for every period.

This joint pattern is difficult to reconcile with the view that larger index demand mechanically creates larger temporary price distortions. It is instead exactly what the model predicts when intermediary competition increases. As L rises, speculators compete more aggressively to supply liquidity at implementation. This pushes the temporary component of price impact down, even as the quantity traded at implementation rises. Thus, in light of the model, the empirical fact that quantities have increased while reversals have decreased suggests that the market for intermediating index events has become much more competitive over time.

6.3 Short Interest Dynamics

The evidence on reversals suggests that intermediation has become more competitive over time. In this subsection, we examine securities-lending data to study how that inter-

Figure 14: Post-Implementation Reversals Around S&P 500 Direct Additions



Notes. This figure shows post-implementation reversals for S&P 500 direct additions. Panel A plots five-year moving averages of post-event cumulative abnormal returns measured over five- and ten-day windows after implementation. Panel B plots the same reversals divided by the estimated index-fund demand shock estimated in Figure12. The model implies that the demand-normalized reversal is the empirical analogue of the avoidable component of price impact per unit of index demand. A value closer to zero indicates less temporary price impact.

mediation occurs. In the model, spreading trades across more periods is valuable because it reduces temporary price impact. One way speculators do this is by acquiring positions outside of the implementation day and selling into benchmark-linked demand at implementation. One way to conduct this trade is through short selling on the implementation day and unwinding afterward. Specifically, for index additions, speculators can borrow shares on the implementation day, sell those shares to index trackers buying at implementation, and then buy shares back over time to cover the short position. In this sense, the post-implementation period gives intermediaries additional time to spread out their trades. In other words, rather than all intermediation occurring before the event, some of it can occur after the event through short covering. This mechanism is also useful to examine for data purposes – speculator purchasers will be mixed in with other investors' purchases, but speculator sales will likely show as short sales, distinguishing them from "long sales" from investors. In this sense, securities lending and short interest data can

provide insight into speculator activity more specifically.

For this purpose, we use Markit data from 2010 to 2023. The data include short interest, borrowing fees, and lendable quantity. We also compute utilization, defined as short interest divided by lendable quantity. This last measure is useful because index inclusion can itself increase lendable supply: once index funds and ETFs own the stock, some of those shares may become available for lending. Thus, both the numerator and denominator of short-interest measures can move around when there are index reconstitutions. Because shorting settles at T+2 trading days, we align shorting with event timing by shifting the Markit data backward by two trading days.

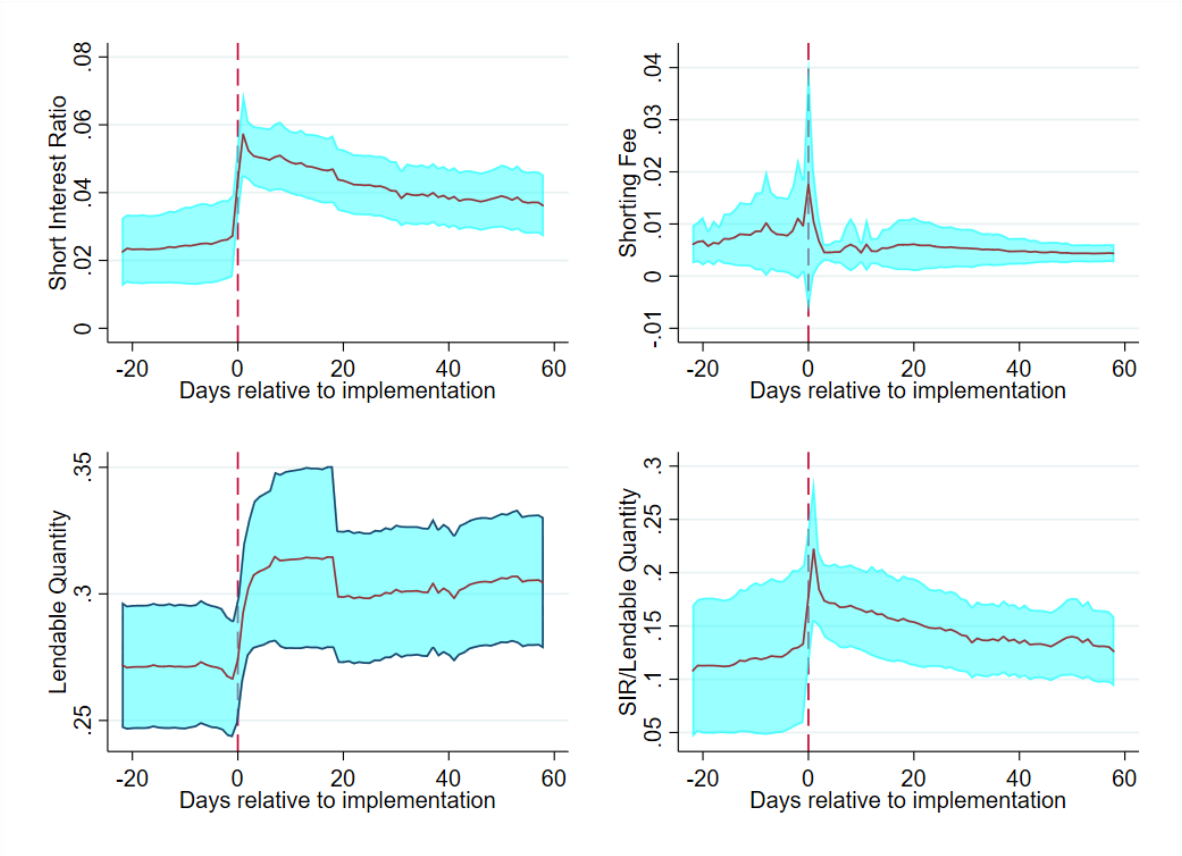
Figure 15 shows the pattern for S&P 500 direct additions. Short interest rises sharply around implementation and then declines gradually afterward, consistent with speculators entering into short positions to satisfy index-linked implementation demand. Lendable quantity also increases after implementation, consistent with the idea that index funds are willing to lend out their shares. Borrow fees spike around implementation and then revert, consistent with the increased supply of shares from passive funds driving down the fee for shorting. However, shorting increases even relative to this increase in supply. Therefore, the overall patterns in Figure 15 are consistent with intermediaries borrowing shares, selling them into benchmark-linked demand, and then covering over time.

Russell migrations provide an additional useful test because the sign of the net demand shock can differ across migration types. Stocks moving from the Russell 1000 to the Russell 2000 experience a positive demand shock because more dollars track the Russell 2000 than the Russell 1000 relative to the total capitalization of each index. Figure 16 shows that short interest rises around implementation and then declines afterward, similar to the evidence in Figure 15, which are also positive net index-tracker demand shocks.

The reverse migration is especially useful. Stocks moving from the Russell 2000 to the Russell 1000 experience a negative index-linked demand shock. In this case, the model predicts the opposite pattern: rather than borrowing and shorting more shares to supply positive demand at implementation, intermediaries should reduce short positions at implementation. Figure 17 shows this is exactly the case empirically. Short interest falls around implementation. This symmetry is important because it shows that the short-interest evidence is not simply a generic feature of any index reconstitution. Instead, it captures speculator activity and moves in the direction predicted by the sign of benchmark-linked demand.

Taken together, the securities-lending evidence supports the model's liquidity provi-

Figure 15: Short Interest and Lending Markets for S&P 500 Direct Additions



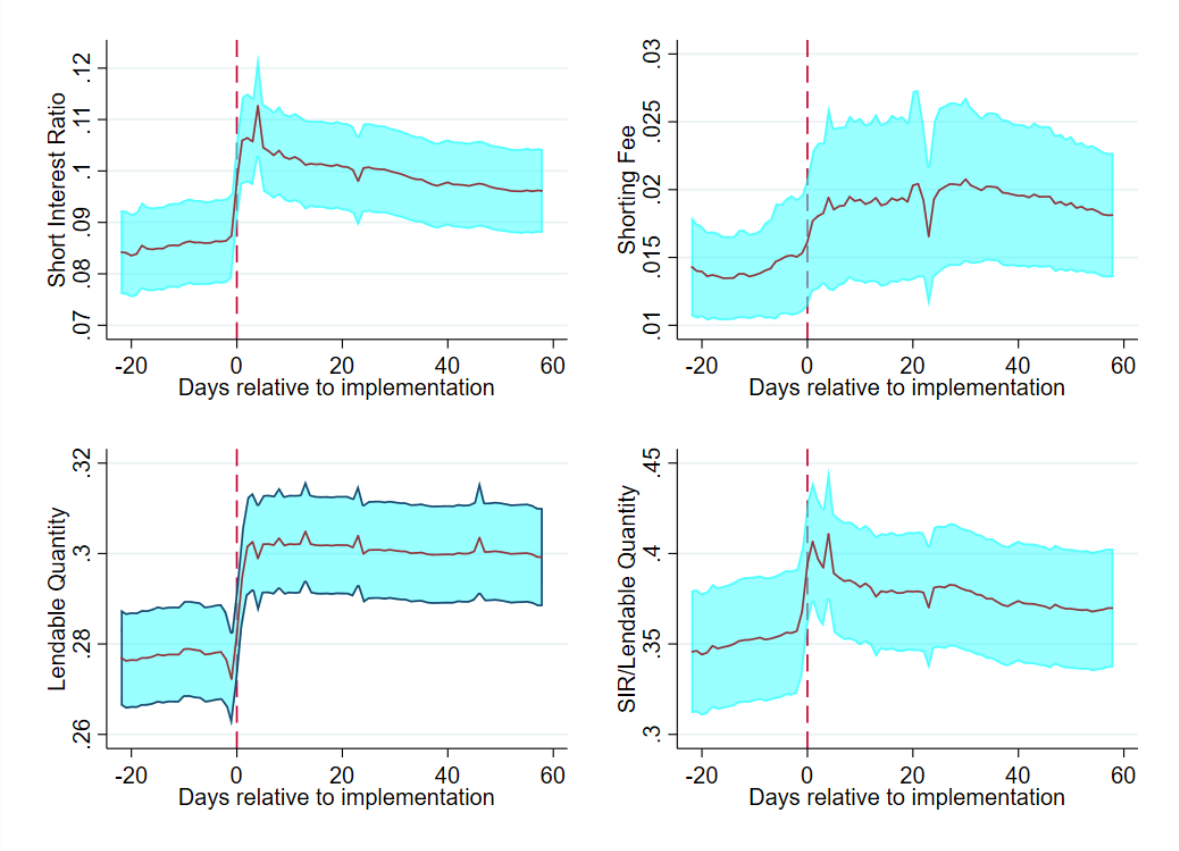
Notes. This figure shows securities-lending variables for S&P 500 direct additions around implementation days of index reconstitutions. The four panels show the short-interest ratio, indicative borrow (shorting) fee, the lendable quantity, and short interest divided by the lendable quantity. Event time is measured relative to the implementation day. The red line shows the average event-time path, and the shaded region shows confidence intervals. The dashed vertical line marks the implementation day. The sample uses Markit securities-lending data from 2010 to 2023.

sion mechanism. The model highlights that the post-implementation period is also useful for intermediation. Specifically, these are another set of trading periods that intermediaries can use to spread price impact. Intermediaries spread price impact by trading before and after the event, with shorting serving as a measure of intermediary activity.

6.4 Predictability and Last-Minute Intermediation Around Russell Migrations

The extended model with multiple pre-implementation periods emphasizes the importance of predictability. When the demand shock is known far in advance, intermediaries have more time to pre-position. When there is little time to pre-position, intermediation

Figure 16: Short Interest and Lending Markets Around Russell 1000 to Russell 2000 Migrations

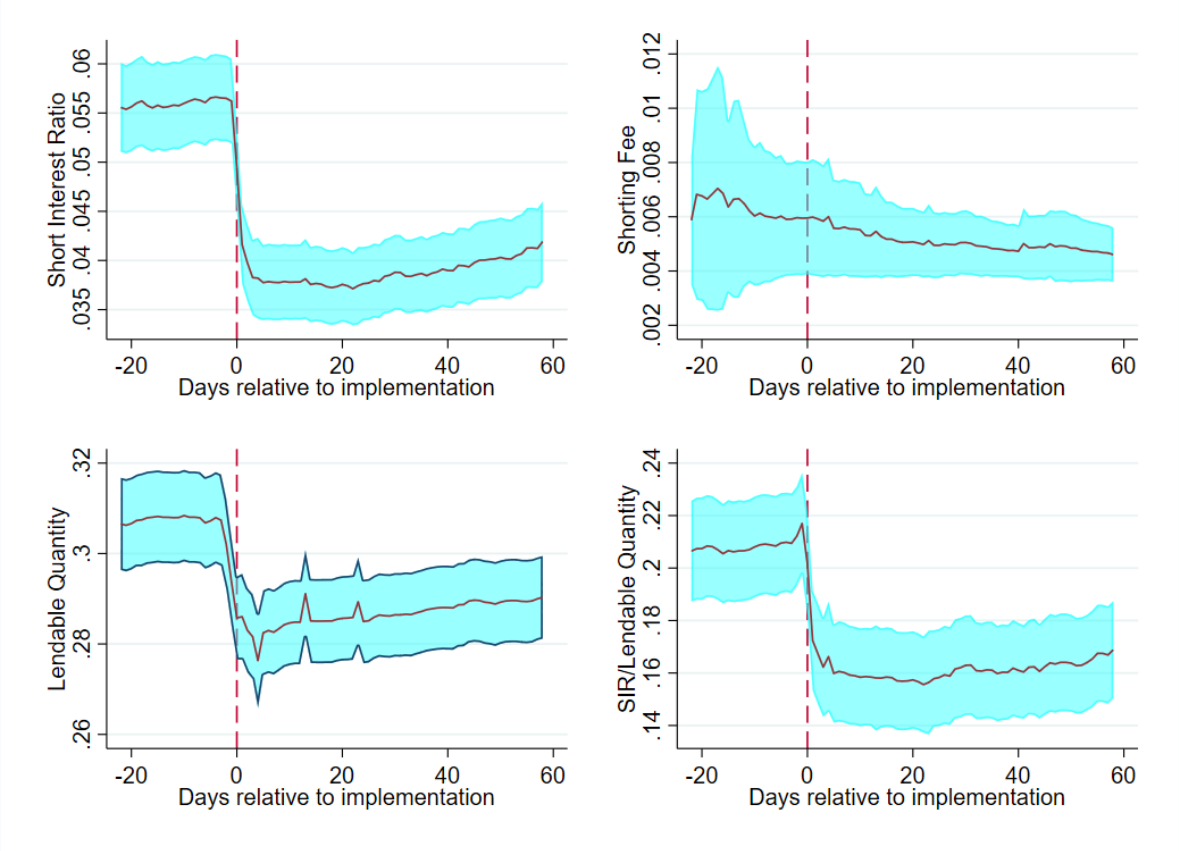


Notes. This figure shows securities-lending variables for Russell 1000 to Russell 2000 migrations around implementation days of index reconstitutions. These migrations generally correspond to positive net benchmark-linked demand because more assets track the Russell 2000 than the Russell 1000. The four panels show the short-interest ratio, indicative borrow (shorting) fee, the lendable quantity, and short interest divided by lendable quantity. Event time is measured relative to the implementation day. The red line shows the average event-time path, and the shaded region shows confidence intervals. The dashed vertical line marks the implementation day. The sample uses Markit securities-lending data from 2010 to 2023.

must occur in a compressed time frame – effectively, there are fewer pre-implementation periods to take advantage of. Based on this logic, we expect less predictable events to generate more trading concentrated between the reconstitution announcement and the implementation date, as well as more post-event trading (given the shorting and post-event covering mechanism described above).

Russell migrations are useful for studying this idea because index assignment is rule-based and depends on market capitalization ranks. Some migrations are "slam dunks" by the rank date: the stock is far enough over the cutoff that the migration was likely anticipated far in advance. Other migrations are close calls. For these stocks, changes in returns for one or more stocks could reshuffle rankings and change whether stocks stay

Figure 17: Short Interest and Lending Markets Around Russell 2000 to Russell 1000 Migrations



Notes. This figure shows securities-lending variables for Russell 2000 to Russell 1000 migrations around implementation days of index reconstitutions. These migrations generally correspond to negative net benchmark-linked demand because more assets track the Russell 2000 than the Russell 1000. The four panels show the short-interest ratio, indicative borrow (shorting) fee, the lendable quantity, and short interest divided by lendable quantity. Event time is measured relative to the implementation day. The red line shows the average event-time path, and the shaded region shows confidence intervals. The dashed vertical line marks the implementation day. The sample uses Markit securities-lending data from 2010 to 2023.

in their current index or move to a different index. As a result, the market learns relatively late (and closer to or on the actual rank date) whether the stock will switch indexes. We expect that close-call switchers should require more last-minute intermediation than slam-dunk switchers, and therefore have more trading volume just before and after the implementation date.

Table 1 tests this idea. The sample includes Russell migration candidates from 2013 to 2024, focusing on years in which the net demand shock exceeds 1% of shares outstanding. We compare three groups: close switchers within 50 ranks of the cutoff, close non-switchers within 50 ranks of the cutoff, and slam-dunk switchers more than 50 ranks

away from the cutoff.[6] We estimate a regression that examines turnover for these three groups. The dependent variables are turnover and abnormal turnover in three windows: from the rank date to the day before the implementation date, the implementation day itself, and the 40 days after the implementation date. Abnormal turnover is defined as the residual from an OLS regression of turnover on the stock's pre-rank return, lagged Russell index membership, and year fixed effects. We regress each turnover variable and in each of the three periods on dummies to capture group membership, as well as pre-rank turnover as a control. The omitted group is the slam dunk switchers.

The main result is that close switchers have significantly higher turnover and abnormal turnover before and after the implementation date (relative to the slam dunk switchers), though the post implementation raw turnover is not statistically significant. The implementation-day coefficient is small, which is a useful placebo test: close-call switchers do not just (all else equal) have more trading volume at every point in time. Rather, Table 1 suggests that when index assignment is resolved late, the intermediation process is more compressed and leads to more trading around the implementation day (both before and after). However, the implementation day trading is very similar, consistent with the idea that the implementation day volume is driven by index-linked demand, which is the same for switchers (as a percentage of shares outstanding) regardless of whether the stock was close or far from the rank cutoff.[7]

Table 1 also shows that the stocks that do not end up switching have much smaller turnover on all days surrounding implementation, including the implementation day itself. Since these stocks stay in their existing index, there is essentially no index-linked demand shock. This means that there is no demand shock to drive up turnover on the implementation day, and no interest from speculators in the days surrounding implementation.

6.5 How Large Are the Costs to Index Investors?

The previous facts show that intermediation for index reconstitutions has become more competitive and that the temporary component of price impact has fallen. We now use the model to translate these reversals into a simple measure of the cost of intermediation for index fund investors.

[6] For more details on how these cutoffs are computed in the "banding" regime, see, e.g., Coles et al. (2022).

[7] The Russell indexes (as well as almost all value-weighted stock indexes) simply hold a constant fraction of (float-adjusted) shares outstanding, regardless of rank. While higher ranked stocks will have larger weights, each stock in an index-tracker holds the same fraction of shares outstanding for each stock in the index (float adjustments tweak this slightly). See Sammon and Shim (2026) for more details.

Table 1: Turnover Around Close-Call Russell Migrations

	Rank Day to Implemen. Day -1		Implemen. Day		Implemen. Day $+1$ to $+40$ Days	
	Turnover	Abnormal Turnover	Turnover	Abnormal Turnover	Turnover	Abnormal Turnover
Within 50 Ranks + Switch	**0.138****	**0.214*****	5.13E-04	0.00437	0.200	0.312***
	(0.059)	**(0.054)**	(0.005)	(0.006)	(0.111)	(0.095)
Within 50 Ranks + No Switch	-0.302*	-0.245*	-0.0265	-0.0236	-0.313**	-0.230*
	(0.140)	(0.135)	(0.020)	(0.021)	(0.103)	(0.112)
Pre-Rank Turnover	0.299***	0.251***	0.00963***	0.00743***	0.542***	0.477***
	(0.053)	(0.055)	(0.001)	(0.001)	(0.042)	(0.040)
Within 50 Ranks + Switch $\times$ Pre-Rank Turnover	-0.0611	-0.0993**	-0.000328	-0.00215	-0.200***	-0.253***
	(0.041)	(0.039)	(0.002)	(0.002)	(0.036)	(0.043)
Within 50 Ranks + No Switch $\times$ Pre-Rank Turnover	0.169	0.180**	-0.00549	-0.00509	0.121	0.132
	(0.098)	(0.076)	(0.006)	(0.006)	(0.085)	(0.080)
Observations	368	368	368	368	368	368
R-squared	0.696	0.553	0.794	0.780	0.689	0.609
FE				Year		

Notes. This table studies turnover around Russell migration candidates close to the Russell 1000/Russell 2000 cutoff. "Within 50 Ranks + Switch" is an indicator for stocks within 50 ranks of the cutoff on the rank date that switch indexes. "Within 50 Ranks + No Switch" is an indicator for stocks within 50 ranks of the cutoff that do not switch. The omitted group is "slam-dunk" switchers, or stocks that are more than 50 ranks away from the cutoff (and thus are certain to switch). The dependent variables are turnover and abnormal turnover measured from the rank date to the day before the implementation date, on the implementation day, and from one to 40 trading days after the implementation date. Abnormal turnover is residual turnover after controlling for pre-rank returns, lagged Russell index membership, and year fixed effects. Standard errors are clustered by year. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

As mentioned above, the model implies that the post-implementation reversal captures the avoidable component of the trading cost. Permanent price impact is not avoidable: the market must ultimately absorb the benchmark-linked demand shock, and this will move prices. By contrast, temporary price impact is the part that reverts after implementation and could be avoided through smoother trading. A simple empirical analogue of the avoidable cost is therefore

$$\text{Avoidable Cost}_e \approx \text{Benchmark-linked dollars traded}_e \times \text{Post-event reversal}_e.$$

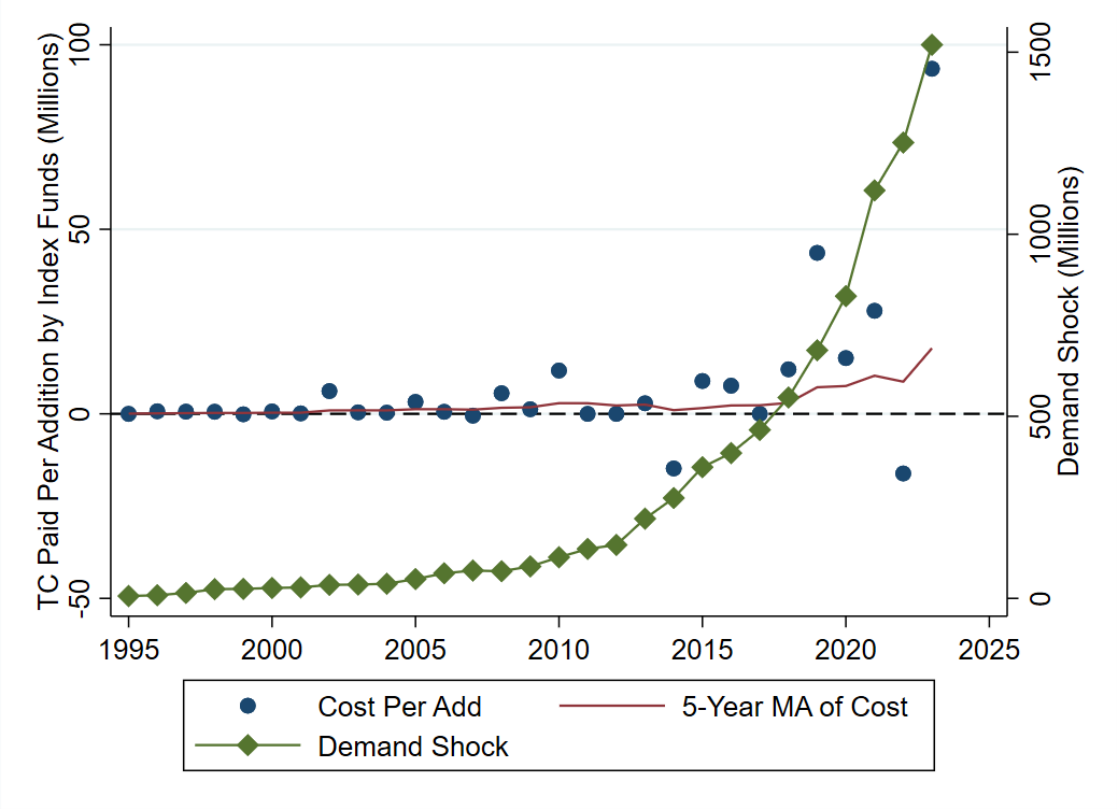
Equivalently, in the notation of the model,

$$\text{Avoidable Cost}_e \approx Q_e \times \left(P_{\text{implementation},e} - P_{\text{post},e}\right).$$

This measure is not the total implementation shortfall. If speculators acquire shares in advance and sell them to index trackers at implementation, some of the total cost may be embedded in the pre-event price. The point is instead that the reversal measures the temporary and avoidable part of price impact associated with concentrated implementation-day trading.

Figure 18 plots the resulting cost estimate for S&P 500 direct additions over time. The green line shows the index-linked demand shock, which has grown dramatically as in-

Figure 18: Estimated Avoidable Trading Costs for S&P 500 Direct Additions



Notes. This figure shows a back-of-the-envelope estimate of avoidable trading costs for S&P 500 direct additions. The green line shows the index-linked demand shock, measured in millions of dollars and plotted on the right axis. The blue points show the estimated cost paid per addition by index funds, measured in millions of dollars and plotted on the left axis. The red line is a 5-year moving average of the cost estimate. The cost is computed as benchmark-linked dollars traded multiplied by the post-implementation reversal.

dexing has expanded. The blue points show the estimated trading cost paid per addition by index funds, and the red line shows the 5-year moving average of this cost. The key fact is that benchmark-linked demand has grown enormously, but the estimated cost has not grown proportionally. In recent years, the cost is roughly 1% of the benchmark-linked trade.

Our estimated cost per trade – especially given the size of these trades – is extremely small. For example, Frazzini et al. (2012) find that price impact is roughly 40 bps for a trade about 10% of ADV. By contrast, S&P 500 additions can involve benchmark-linked trades over 2500% of ADV, and yet the price impact is only about 100 bps. In this sense, the surprising fact is not that index investors face trading costs – any trade must generate costs in some form – it is that index investors can trade such enormous quantities at a cost that appears very low relative to standard benchmarks.

Again, the model is useful because it can resolve this seeming puzzle. As index demand has grown, the incentive to intermediate index events has also grown. More intermediaries enter the trade, competition increases, and the temporary component of price impact falls. This is why rising Q and falling reversals are jointly informative. In a model without increasing intermediary competition, larger demand shocks should *mechanically* generate larger reversals. In our model, the fact that reversals fall while Q rises indicates that the market has become substantially more efficient at intermediating index demand.

6.6 Implications for Index Design

Collectively, the empirical evidence implies that predictable index demand is potentially costly to index fund investors, but predictability also allows intermediaries to prepare and reduce the temporary component of price impact. The relevant counterfactual is therefore not a world in which index funds trade for free. The relevant counterfactual is a world in which index trackers must trade large quantities without intermediaries providing shares to satisfy demand.

This has implications for index design – and we suspect that index providers are aware of these channels. When index changes are predictable and the implementation window is long enough, intermediaries can pre-position and supply liquidity at the close. This can reduce implementation-day price impact and lower the avoidable component of trading costs for index investors. In this sense, transparency and commitment can be valuable because they coordinate liquidity. For example, starting in 1989, S&P started pre-announcing index changes. Similarly, Russell has increased the gap between the ranking date and the implementation date, another way to give intermediaries more time to acquire shares and spread out trading over more days.

At the same time, this logic does not imply that increased predictability is always good. If the pre-positioning window is too short, intermediaries may not have enough time to acquire inventory and smooth price impact. In those cases, implementation-day price impact and post-event reversals may reappear, as explained in Section 5.3 and shown in Murray and Sammon (2026) for post-IPO fast-track additions to CRSP indexes. The reason is that there simply are not enough periods after the IPO day and before the implementation day, and shorting is restricted due to IPO lock-ups, preventing speculators from smoothing trading over the days after implementation. Similarly, if predictability attracts informed trading or creates adverse selection, index providers may face additional costs. The point of the model is to clarify this tradeoff. Concentrated implementation-day trading can be optimal when tracking-error concerns are important and intermediaries are competitive, but the benefits depend on giving intermediaries

enough time and certainty to supply liquidity.

7 Conclusions

We show that apparent front-running around index reconstitution events is a form of liquidity provision that benefits index investors. Our model also rationalizes the empirical disconnect between trading volume and price changes around such events, whereby large reconstitution-day trading volumes, with over 20% of shares outstanding trading, occur with small implementation-day returns.

The model delivers four main takeaways. First, speculators act as endogenous liquidity providers rather than pure predators. They pre-position before implementation, trade against benchmark-linked demand at reconstitution, and unwind short positions after implementation. This mechanism can generate price pressure before the event, but reduces implementation-day price impact and post-implementation reversal. Second, competition between front-runners plays a central role. As competition among speculators increases, index trackers—including those that only loosely track the index—optimally concentrate rebalancing on the implementation date. Third, the model identifies post-implementation price reversal as the measure of the avoidable cost borne by index-linked investors. Fourth, we establish a liquidity complementarity between strict and weak index trackers. In highly competitive markets, loose index trackers increasingly concentrate their rebalancing at implementation when the size of the passive rebalancing volume increases.

We use the model to interpret several empirical facts. Implementation-day volume is much larger than strict index-fund demand alone, suggesting that the implementation date serves as a coordination point for strict index trackers, shadow indexers and benchmarked investors. Moreover, post-implementation reversals have fallen sharply over time for S&P 500 direct additions, even as benchmark-linked demand has grown. While the raw reversal is interesting in its own right, the model implies that the relevant object is the reversal scaled by the size of the demand shock, and this demand-normalized reversal has also fallen substantially over the past two decades. Under the lens of the model, this suggests competition among speculators has increased and that speculators, as a result, effectively provide liquidity at a lower cost.

Additionally, securities-lending data provide direct evidence on the way speculators are able to spread their trades to reduce price impact. Around positive-demand additions, short interest and borrow fees rise around implementation and then decline as speculators cover their shorts. For negative-demand Russell migrations, the pattern reverses,

with short interest falling around implementation. Finally, close-call Russell switchers have more trading before and after implementation than slam-dunk switchers, consistent with the idea that less predictable demand gives speculators less time to pre-position and leads to more concentrated trading.

These results have practical implications for how to interpret, and potentially reduce, the costs of index rebalancing. If speculators acquire shares in advance and sell them to index trackers at implementation, some of the effective trading cost is embedded in the pre-event price path. The model implies that post-event reversals are the right way to measure the *avoidable* component of price impact. Using this logic, the striking fact is not that index investors face trading costs. Any large trade must generate transaction costs in some form. The striking fact is that, in recent years, index funds appear able to trade enormous quantities – in some cases more than 20 times average daily volume – with costs of roughly 1% of the trade. Relative to standard price-impact estimates for institutional trades, this is surprisingly *cheap*, not surprisingly expensive. Apparent front-running can improve outcomes for index fund investors if competition between speculators is high enough, because it shifts risk to speculators that are willing to supply immediacy at implementation and allows index investors to minimize tracking error without incurring substantial price impact.

Using our framework, we also shed light on index design and the welfare implications of fast-track index additions after initial public offerings (IPOs). Index investors can rebalance their portfolio at small costs when the incoming constituent is very liquid, or when competitive speculators can intermediate the required shares by building inventory before implementation or shorting shares at implementation, covering the short positions afterward. In the case of fast-track additions, the incoming constituents are likely illiquid and hard to short. If the time from the IPO to the index addition is short, index investors will suffer from implementation-day price impact and post-event reversals, as empirically shown by Murray and Sammon (2026). In contrast, issuers and original owners benefit from fast-track additions, especially when competitive speculators bid up IPO prices.

REFERENCES

Admati, A. R. and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies*, 1(1):3–40.

Almgren, R. and Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40.

Amihud, Y. and Mendelson, H. (1980). Dealership market: Market-making with inventory. *Journal of Financial Economics*, 8(1):31–53.

Arnott, R. D., Brightman, C., Kalesnik, V., and Wu, L. (2023). Earning alpha by avoiding the index rebalancing crowd. *Financial Analysts Journal*, 79(2):76–97.

Behmaram, A. and Davis, C. (2026). Indexing and the elasticity of stock demand.

Bertsimas, D. and Lo, A. W. (1998). Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50.

Brunnermeier, M. K. and Pedersen, L. H. (2005). Predatory trading. *Journal of Finance*, 60(4):1825–1863.

Chinco, A. and Sammon, M. (2024). The passive ownership share is double what you think it is. *Journal of Financial Economics*, 157:103860.

Coles, J. L., Heath, D., and Ringgenberg, M. C. (2022). On index investing. *Journal of Financial Economics*, 145(3):665–683.

Cremers, K. M. and Petajisto, A. (2009). How active is your fund manager? a new measure that predicts performance. *Review of Financial Studies*, 22(9):3329–3365.

Frazzini, A., Israel, R., and Moskowitz, T. J. (2012). Trading costs of asset pricing anomalies. *Fama-Miller Working Paper, Chicago Booth Research Paper*, (14-05).

Gabaix, X. and Koijen, R. S. (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 Sammon, M. (2025). The disappearing index effect. *Journal of Finance*, 80(2):657–698.

Grossman, S. J. and Miller, M. H. (1988). Liquidity and market structure. *Journal of Finance*, 43(3):617–633.

Harris, L. and Gurel, E. (1986). Price and volume effects associated with changes in the

S&P 500 list: New evidence for the existence of price pressures. *Journal of Finance*, 41(4):815–829.

Harvey, C. R., Mazzoleni, M. G., and Melone, A. (2025). The unintended consequences of rebalancing. Technical report, National Bureau of Economic Research.

Ho, T. and Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1):47–73.

Koijen, R. S. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, pages 1315–1335.

Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management, Forthcoming*.

Madhavan, A. and Smidt, S. (1993). An analysis of changes in specialist inventories and quotations. *Journal of Finance*, 48(5):1595–1628.

Murray, C. and Sammon, M. (2026). Primary capital market transactions and index funds. *Review of Asset Pricing Studies*, page raaf013.

O'Hara, M. and Oldfield, G. S. (1986). The microeconomics of market making. *Journal of Financial and Quantitative Analysis*, 21(4):361–376.

Pavlova, A. and Sikorskaya, T. (2023). Benchmarking intensity. *Review of Financial Studies*, 36(3):859–903.

Petajisto, A. (2011). The index premium and its hidden cost for index funds. *Journal of Empirical Finance*, 18(2):271–288.

Sammon, M. and Shim, J. J. (2026). Index rebalancing and stock market composition: Do indexes time the market? *Journal of Financial Economics*, 177:104229.

Shim, J. J. and Wang, Y. (2025). Commitment power and trading frequency in optimal execution. *Available at SSRN 4978304*.

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

Stoll, H. R. (1978). The supply of dealer services in securities markets. *Journal of Finance*, 33(4):1133–1151.