ETF Trading and the Bifurcation of Liquidity*

Jonathan Brogaard
David Eccles School of Business
University of Utah

Davidson Heath
David Eccles School of Business
University of Utah

Da Huang
David Eccles School of Business
University of Utah

January 14, 2020

.

*We thank Rajesh Aggarwal, Ben Connault, Pamela Moulton, Matt Ringgenberg, Elaine Wah, and participants at the IEX Market Structure Conference for comments. We thank Shaun Davies for sharing code and data. Email: davidson.heath@eccles.utah.edu.

.

ETF Trading and the Bifurcation of Liquidity

Abstract

Passively managed exchange traded funds (ETFs) are a financial technology that has risen dramatically in the last two decades. Over the same period liquid stocks have become more liquid while illiquid stocks have not experienced a similar improvement. We model investors shifting from trading individual stocks to trading ETFs and generate predictions consistent with the documented bifurcation in liquidity. Using daily ETF creation and redemption activity, we provide empirical evidence that closely matches the model's predictions. The results show that the effects of ETFs on underlying asset markets are driven by their index implementation strategy.

JEL: G11, G12, G20

In recent years the structure of U.S. equity trading has dramatically changed. In 1998 there was \$6.8 billion of assets under management (AUM) in exchange traded funds (ETFs); by 2018 there was \$3.4 trillion of AUM in ETFs. At the same time ETFs went from less than 2% to over 31% of total U.S. equity trading volume.[1] During this same period the U.S. stock market has become more liquid on average (Angel, Harris, & Spatt, 2015). However, the improvement in liquidity has not been uniform. Liquid stocks have become more liquid while illiquid stocks have not experienced similar gains.[2] Figure 1 shows the evolution of liquidity in U.S. equities since 2000. In this paper we examine whether the rise of ETF trading can help explain recent trends in liquidity in U.S. equity markets.

Insert Figure 1 About Here

We develop a model that predicts that ETF trading has a differential, bifurcating effect on asset liquidity. The key insight is that ETFs that passively track an index can, and frequently do, deviate from the weights of the index they track.[3] The model predicts that ETFs will systematically underweight or omit index components that are less liquid *ex ante*. Thus, trading activity is not simply substituted one for one by the ETF but is tilted towards the stocks the ETF chooses to sample. Consequently, there is a bifurcation in asset liquidity as noise trading flows preferentially into more liquid index assets.

Empirically, we find that the model's predictions are borne out in the data. ETFs are more likely to omit less liquid index stocks from their holdings, and ETF primary flows (*i.e.* share creations and redemptions) have a differential effect on stock liquidity. As the magnitude of ETF primary flows increases – in either direction – *ex ante* liquid stocks

[1] All numbers are calculated from the CRSP monthly security files. The ETF numbers are for all entries with $shrcd = 73$; total equity trading volume also includes all common stocks ($shrcd = 10, 11$).

[2] Hendershott, Jones, and Menkveld (2011); Jones (2013); Haslag and Ringgenberg (2016).

[3] The aspect of index replication that we focus on is distinct from the observation of Easley, Michayluk, O'Hara, and Putnins (2019) that many ETFs diverge from the value-weighted market portfolio.

become more liquid, while illiquid stocks become less liquid. This effect is separate from the previously documented effects of market fragmentation and algorithmic trading.

While straightforward in principle, replicating a target index involves tradeoffs in its implementation. Models of index investing usually assume that index funds simply replicate the weights of their target index. In practice ETFs have considerable discretion in defining their creation/redemption baskets (Lettau & Madhavan, 2018a) and many ETFs' holdings diverge from the weights of their target index. Six of the ten largest ETFs as of 2018 state in their prospectus that they statistically replicate their target index by investing in a basket of representative securities.[4]

We model the fundamental tradeoff faced by a passive fund: To simultaneously minimize expected tracking error and expected transactions costs. The model applies to both ETF providers (that set the creation and redemption basket for authorized participants) and traditional open-ended index funds (that rebalance their portfolio after inflows and outflows). For fund holdings the model predicts that the optimal weighting of index assets is driven by their transaction costs and their correlation with other index assets. Assets that are more expensive to trade and less correlated with the index are more likely to be underweighted or omitted. We empirically test these predictions and find that they are borne out in the fund holdings data.

We next turn to the effects of ETF trading on underlying asset markets. The main prediction of the model is that ETF trading activity due to primary flows (*i.e.* the creation

[4]One example is the Vanguard Total Stock Market ETF (VTI), the third largest ETF by assets as of December 2018. The fund's prospectus states, "The Fund invests by sampling the Index, meaning that it holds a broadly diversified collection of securities that, in the aggregate, approximates the full Index in terms of key characteristics." Sampling or "optimized" index replication is not only popular with funds that track relatively illiquid indexes, but is also used by ETFs that track more liquid indexes. For example, the second largest ETF as of 2018, the iShares Core S&P 500 ETF (IVV), states in its prospectus that BlackRock "uses a representative index sampling strategy to manage the Fund."

and redemption of ETF shares in exchange for the posted basket of index assets) amplifies preexisting differences in liquidity. Thus, the ETF's optimal implementation strategy moves noise trading out of illiquid index assets and into liquid index assets. We test these predictions and find support for them in the stock-level daily data. On days with more ETF primary flows – in either direction – we find that liquidity is higher for index stocks that were more liquid *ex ante* and lower for index stocks that were illiquid *ex ante*. Depending on the measure of primary flow, we found a 12% to 16% increase *per annum* associated with primary flow in the liquidity gap between the most and least liquid stocks.

These results are potentially confounded by other market forces that drive both ETF flows and asset liquidity. However, a unique prediction of the model is that the effects of ETF flows on index asset liquidity are determined by the fund's implementation strategy. Thus, the effects are different for ETFs that follow a statistically-sampling strategy compared to ETFs that follow a fully-replicating strategy.[5] Comparing the effects of samplers versus full replicators breaks the potential simultaneity between ETF flows and asset liquidity. Using a matched sample of ETFs, we find that the bifurcation in stock liquidity is *only* driven by primary flows in sampling ETFs.

Other factors such as the arrival of market-moving news could also drive ETF primary flows and asset liquidity. We examine this possibility using two different approaches. First, we control directly for daily market movements. Second, in an additional analysis we restrict the sample to trading days when no market-moving news arrived. In both cases the differential effects of ETF activity on asset liquidity are unchanged or even larger, inconsistent with market-wide news explaining the results.

[5]Four of the ten largest ETFs as of 2018 state in their prospectus that they follow a fully replicating strategy. Two examples are SPY and QQQ, which are the State Street S&P 500 ETF and the Invesco Nasdaq 100 ETF, respectively.

Finally, we consider two additional alternative explanations for differential liquidity effects in recent years: Market fragmentation and algorithmic trading. Regulation National Market System (Reg NMS) was established in 2005 and had significant effects on quoted spreads, market fragmentation, and market quality that differ by stock market capitalization (Haslag & Ringgenberg, 2016). Algorithmic and high frequency trading have risen significantly for a segment of the stock market (Weller, 2017). We control for these effects directly, and find that stock-by-day measures of market fragmentation and algorithmic trading activity do not explain the differential effects of ETF activity on asset liquidity.

This paper makes contributions to both theoretical and empirical research on passive investing. Theoretical studies of passive investing assume that passive funds replicate the weights of their benchmark index *pro rata*. We demonstrate that models of the impact of passive investing on asset prices can be made richer and more realistic by taking into account the implementation strategy of passive funds. Moreover, empirical studies of passive investing often implicitly take full replication as given by using index assignment to study the treatment effects of index investing. We show that the institutional details of how index funds interact with the market means that full replication is not a given and therefore the intensity of index investing on specific underlying assets may be mismeasured.

This paper relates to two main strands of the literature. First, a large literature has investigated the growth of passive and ETF investing and its effects on individual stocks. Greenwood (2007) finds that a higher index weight leads a stock to co-move more with the index and less with stocks that are not in the index. Da and Shive (2018) attribute increased co-movement to ETF arbitrage activity. Ben-David, Franzoni, and Moussawi (2018) find that increased ETF ownership leads to higher stock volatility, due to arbitrage trading between ETFs' market price and net asset value (NAV). Israeli, Lee, and Sridharan (2017) find that increased ETF ownership leads to lower price efficiency, higher return synchronicity, and lower analyst coverage of the securities in the underlying basket. Evans, Moussawi, Pagano, and Sedunov (2019) find that increased ETF ownership widens the intraday bid-ask spreads of the underlying stocks. Sağlam, Tuzun, and Wermers (2019) find that increased ETF ownership makes underlying index stocks more liquid. Agarwal, Hanouna, Moussawi, and Stahel (2018) find that increased ETF ownership increases the commonality in liquidity of the underlying stocks.

We add to this literature by focusing on the implications of ETFs' index implementation strategy. We show that many ETFs sample a subset of liquid index assets, and underweight or omit less liquid index assets, and that this strategy amplifies preexisting differences in asset liquidity: Liquid assets become more liquid while illiquid assets become less liquid. We examine the create/redeem mechanism as a channel through which ETFs affect the liquidity of the underlying assets in the index.

Second, this paper relates to theoretical work on the impact of ETFs on underlying asset markets. Carpenter (2000) and Basak, Pavlova, and Shapiro (2007) show that fund flows tilt the portfolio toward stock that belong to the benchmark because of fund managers' risk aversion. Malamud (2016) constructs a general equilibrium model in which ETF creation/redemption serves as a shock propagation mechanism; Pan and Zeng (2019) construct a model in which a liquid ETF tracks a single illiquid asset, and they analyze the effects of authorized participants' market making activity on the asset's liquidity. These models all assume the ETF replicates the underlying index pro rata. By contrast, we construct a model of optimal index replication and show how these effects depend crucially on ETFs' index implementation strategy. We characterize the optimal differences between the basket and the index and the resulting differential effects on asset markets.

I. Model

An exchange-traded fund (ETF) is a fund that tracks an index of underlying securities. ETFs are in many ways similar to open-ended mutual funds except that ETF shares are listed on an exchange and traded throughout the day. An ETF tracks its underlying index because of the arbitrage activity of authorized participants (APs), which are usually large market making firms. An AP has access to the creation and redemption mechanism, which allows them to exchange ETF shares for the basket of underlying securities. If the ETF's shares trade sufficiently above the fund's net asset value, which the ETF provider publishes in real time, the AP sells ETF shares and buys the underlying basket of securities and vice versa. Authorized participants are thus *de facto* market makers for the ETF, in that they provide a liquid two-sided market for its shares and facilitate secondary market trading (Lettau & Madhavan, 2018a; Evans et al., 2019). The net creation and redemption of ETF shares, through which investor dollars move in and out of the ETF provider, is referred to as ETF primary flow.

We construct a simple one-period model that captures the fundamental tradeoff faced by any passive index fund: To simultaneously minimize expected tracking error and expected transaction costs. The model is written from the point of view of an ETF provider; we show in Appendix B that the same decision problem and solution applies to traditional open-ended funds as well.

A. Setup

There are N assets in the market with a vector of prices $\boldsymbol{p}$, and one-period excess returns $\tilde{\boldsymbol{r}}$ which are normally distributed with expectation 0 and covariance matrix Σ.

There are three types of agents: ETF providers, authorized participants, and investors. We consider the market for ETFs that track a specified index such as the S&P 500, the Russell 2000, or the CRSP value-weighted U.S. market index. The index is a vector of weights $\boldsymbol{v}$ that add to 1 and are exogenous and fixed.

An ETF provider enters the market by publishing a basket, which is a vector of weights $\boldsymbol{w}$ that add to 1. She agrees to create or redeem shares of the ETF in exchange for that basket of individual assets.[6] The net asset value (NAV) of one ETF share is $\text{NAV} = \boldsymbol{w}'\boldsymbol{p}$. The provider incurs administrative costs and collects a management fee. There is free entry, so in equilibrium ETF providers' fees equal their costs.

The ETF provider nominates one or more authorized participants (APs) who have access to the creation and redemption mechanism. The AP is risk neutral and makes the market for ETF shares by posting a bid price and an ask price around the NAV. The quotes are executed against by order flows from index investors. At the end of the period the AP nets the buy and sell orders that arrived within the period. She then closes out her netted position by trading in the individual index assets in the basket and making basket-ETF exchange with the provider.

It follows that the bid and offer prices relative to the NAV are pinned down by the transaction costs in the individual asset markets. The AP's expected profit from posting the offer $\text{NAV} + b$ and being lifted is:

$$E[(\text{NAV} + b) - (\boldsymbol{w}'\boldsymbol{p} + C(\boldsymbol{w}))] = b - \text{E}[C(\boldsymbol{w})],$$

[6]Some ETF providers also allow the authorized participants to create or redeem ETF shares in exchange for cash. After netting the daily primary flow from its authorized participants, the provider may transact in derivatives or the underlying asset markets to zero out its residual position. As long as the ETF provider faces nonzero transaction costs of doing so, the model is the same.

where b is the spread and $\mathrm{E}[C(\boldsymbol{w})]$ is the expected transaction cost incurred by the netted order flow. As long as the provider nominates at least two authorized participants, if one posts an offer that is above the zero-profit bound, the other will undercut them. That is, competition between authorized participants is Bertrand. (Lettau and Madhavan (2018b) cite that a sample of large ETFs had an average of 38 authorized participants apiece). Thus, the bid and offer prices that investors face to trade the ETF shares are NAV $\pm\, \mathrm{E}[C(\boldsymbol{w})]$.

In the competition among ETFs, all investors prefer an ETF that has a lower bid/ask spread and a lower tracking error. Thus, in equilibrium the ETF that captures the market is the one that minimizes

$$U = \mathrm{E}[C(\boldsymbol{w})] + \lambda(\boldsymbol{w} - \boldsymbol{v})'\Sigma(\boldsymbol{w} - \boldsymbol{v}), \tag{1}$$

where λ is the shadow price that investors attach to a higher tracking error (term 2) relative to a higher bid/ask spread (term 1).[7]

B. Fund Weights

The first order condition for the optimal weight of the fund in asset i is:

$$0 = \frac{\partial \mathrm{E}[C(w)]}{\partial w_i} + 2\lambda \sum_j (w_j - v_j)\sigma_i\sigma_j\rho_{ij}$$

$$\frac{Cov(\tilde{r}_i, \tilde{r}_{ETF} - \tilde{r}_{Index})}{\partial \mathrm{E}[C(w)]/\partial w_i} = -1/2\lambda.$$

[7] We assume that all investors have the same preference λ. Relaxing this assumption would result in a frontier of ETFs that express different tradeoffs between tracking error and transaction costs and cater to investors with different preferences. In practice, the multiplicity of funds tracking the same index seems to be driven by investor search costs and not by different investor preferences (Hortaçsu & Syverson, 2004).

That is, the optimality condition for each asset is that the marginal increase in trading costs equals the marginal decrease in the expected tracking error, which is a product of the covariance of $\tilde{r}_i$ with the other index assets.

To solve for $\boldsymbol{w}$ explicitly, we specify the trading cost as linear and additively separable [8] and assume noise index investors have independently normally distributed exogenous flow $\tilde{f} \sim N(0, \sigma_f^2)$:

$$\mathrm{E}[C(\boldsymbol{w})] = 2\mathrm{E}\left[\left|\tilde{f}\right|\right] \sum_i c_i w_i = 2\sqrt{\frac{2}{\pi}} \sum_i c_i w_i,$$

where c_i measures the the trading cost of stock i.

It follows that:

$$w_i^* = \left(v_i - \frac{c_i}{\lambda^{ETF} \sigma_i^2}\right) + \sum_{j \neq i} (v_j - w_j) \beta_{j,i}, \quad (2)$$

where $\lambda^{ETF} = \sqrt{\pi/2}$. Intuitively, the first term says that in general, index holdings are underweighted relative to their index weight v_i. The optimal weight balances transaction costs against the direct contribution to tracking error (hence, c_i over $\lambda \sigma_i^2$). The second term captures the indirect second-order effects on tracking error: An asset that covaries positively with other index assets that are underweighted has a higher optimal weight, and vice versa.

In general we expect the first term, which captures the direct effects on trading costs and tracking error, to dominate. However, there are exceptions such as index futures and redundant assets. Appendix B analyzes these cases in detail.

The model predicts that w_i^* is decreasing in c_i (a more illiquid stock has a lower optimal

[8] Almgren, Thum, Hauptmann, and Li (2005) propose that trading cost is exponential in order size for each asset, and using a large set of execution data they estimate the exponent to be 1.375. The model's predictions remain the same irrespective of the choice of the exponent.

weight) and increasing in ρ_{ij} (an asset which has higher correlation with other index assets has a higher optimal weight). The optimal weight is ambiguous in σ_i, as a higher volatility both increases the covariance with other index assets and increases the tracking error.

C. Effects on Asset Liquidity

The model predicts that ETFs, and passive index funds in general, should underweight or omit index assets with relatively high trading costs. We now explore the effects of ETF trading on the liquidity of underlying asset markets.

We consider two assets with the same weight in the index, $v_A = v_B = v$ and the same volatility $\sigma_A = \sigma_B = \sigma$. For simplicity we assume both assets are uncorrelated with the rest of the index. There are a mass of N_A noise traders in asset A and a mass of $N_B < N_A$ noise traders in asset B. Noise traders have exogenous flows that are independently normally distributed, $\tilde{f}_i \sim N(0, \sigma_f^2)$. The market in each individual asset is made by a single market maker as in Kyle (1985). Thus, the trading cost of the individual assets is:

$$c_i = \frac{\sigma}{N_i \sigma_f}, \quad c_A < c_B. \tag{3}$$

In other words, the two assets are otherwise identical except that asset A is more liquid *ex ante*. We now introduce a mass of N_I index investors. The index investors are rational: when they trade the index, they solve the optimization problem in equation (1) as well. With the index investors included, the expected trading volume in the underlying assets is:

$$c_i = \frac{\sigma}{(N_i + N_I w_i)\sigma_f}. \tag{4}$$

Note that for each asset we have two equations ((2) and (4)) and two unknowns (c_i and

w_i). This pins down the market structure before the ETF is introduced.

We now introduce the ETF. Compared to the pre-ETF period when each individual submits their demand directly to the market maker, now the authorized participant collects all the index investors' order flow, submits the netted flow to the market maker, and creates or redeems shares with the ETF provider. This has two effects on the market.

First, as the individual index investors' flows are netted by the AP, the market maker's expected trading volume falls, and the liquidity in the individual assets becomes:

$$c_i' = \frac{\sigma}{(N_i + \sqrt{N_I} w_i^{ETF})\sigma_f}.$$

This netting effect strictly worsens liquidity in the underlying assets. However it is welfare-improving for the index investors, because they equally share the reduced trading cost incurred by their netted flow. The effective trading cost they face is:

$$c_{i,net}' = \frac{\sigma}{N_I(N_i + \sqrt{N_I} w_i^{ETF})\sigma_f}.$$

This relation makes clear why index investors substitute into the ETF as $\sum c_{i,net}' w_i^{ETF} < \sum c_i w_i$.

We now turn to the second effect of the ETF's introduction. Because the ETF is much cheaper to trade, demand for investing in the index increases. This is the demand effect. That is, a new mass N_I' of index investors enter the market. This effect increases the expected order flow faced by the market maker, and changes the trading costs in the individual assets:

$$c_i'' = \frac{\sigma}{(N_i + \sqrt{N_I + N_I'} w_i^{ETF})\sigma_f}. \tag{5}$$

The model predicts that the effects of ETF trading on the liquidity of index assets is ambiguous, depending on the relative strength of the netting effect and the demand effect. For intuition, we next provide a numerical example. Details of the calibration are in Appendix section C.

We characterize the change in the liquidity gap between the two assets as:

$$\Delta \text{Gap} = (c''_B - c''_A) - (c_B - c_A).$$

When ΔGap is positive, the preexisting liquidity differential between asset A and B becomes wider, and vice versa. Figure 2 plots ΔGap over different values for the primary flow – the creation and redemption activity – of the ETF. The blue line shows the effects on the liquidity gap from an ETF that strategically samples the assets in its underlying index i.e. following the index implementation strategy in equation (2). We see that for low and moderate levels of index investing the netting effect dominates, and the ETF widens the liquidity gap. For very high levels of index investing, the demand effect dominates, and the ETF reduces the liquidity gap between assets.

Insert Figure 2 About Here

Alternatively, an ETF can choose not to strategically sample, but instead commit to fully replicate their index weights. The red line in Figure 2 shows the effect on asset liquidity from an ETF that is fully replicating. We see that the fully replicating ETF has a much lower point beyond which the demand effect on liquidity dominates, and the ETF reduces the preexisting liquidity gap. This prediction is intuitive, because the replicating fund overweights the less liquid asset compared to both the individual index investors and the sampling ETF. Thus the model predicts that sampling ETFs always have a larger (more positive) differential effect on asset liquidity than replicating ETFs.

II. Data

The data covers all stocks in the Russell 3000 index from 2009 to 2018. We focus on Russell 3000 index stocks to make sure all the stocks in the sample are constituents of indices that are tracked by ETFs with meaningful assets under management and trading volume. We obtain quarterly fund holdings from the union of the CRSP mutual fund holdings database and the Thompson-Reuters S12 database as both databases have gaps in their coverage. Returns, trading volume, and other market data for both stocks and ETFs are from the CRSP daily file. We obtain index membership and their weights by month directly from Russell Investments for the Russell 1000 (large-cap) and Russell 2000 (small-cap) indexes.

Table 1 Panel A reports the summary statistics from the sample. Note that the Russell 3000 excludes micro-cap and foreign stocks. The market capitalization measure has a mean of $6.7 billion and a median of $1.2 billion. The effective spread captures trading costs, measured as the difference between the price in which a market order executes and the mid-quote on the market the instant before. The lower the effective spread, the more liquid the stock. The effective spread in our sample has a mean of 0.0043 and a median 0.0016, showing that a few illiquid stocks skew the effective spread distribution to the right. Stock volatility has a slightly right-skewed distribution with a mean of 0.025 and a median of 0.021. The precise definition of all variables can be found in Appendix Section A.

Insert Table 1 About Here

Table 1 Panel B shows summary statistics for the ETFs in the sample, which is all U.S.

equity ETFs from 2009 to 2018 in the CRSP mutual funds database with at least $10 million in assets under management. Figure 3 Panel A shows how the total assets under management of U.S. equity ETFs has evolved over time.

Insert Figure 3 About Here

We split the sample ETFs on the basis of their index implementation strategy. Equity ETFs mainly adopt two implementation strategies, full replication and statistical sampling. We categorize ETFs that state in their prospectus that they hold all or substantially all stocks in the index as Replicator ETFs. We categorize ETFs that state in their prospectuses that they statistically sample a subset of index stocks as Sampler ETFs. 7% of ETFs in the sample use a different implementation strategy such as derivative contracts to track the underlying indices, which we refer to as Other ETFs. These are typically leveraged ETFs or inverse ETFs.

We obtain implementation strategy information from Bloomberg's equity fund characteristics and verify via manual checks through ETFs' prospectuses.[9] On average Sampler ETFs have slightly more assets under management (AUM), with a mean of $2.8 billion, than replicator ETFs, with a mean of $1.9 billion. There is almost complete overlap (common support) across the distributions of the AUMs of the two types of ETFs. The expense ratios, reported in CRSP Mutual Fund Database, are also similar between the two types of ETFs. Replicator ETFs have a mean expense ratio of 45 basis points (bps) while Sampler ETFs have an expense ratio of 44 bps.

[9]ETF prospectuses are obtained from each ETF's own official website by the ETF provider. Note that we constructed the sample using CRSP, which is survivorship free, and then obtained the implementation strategy information according to the list of ETFs we have. This addresses concerns of potential survivorship bias that may arise due to some inactive ETFs' data not being available in Bloomberg or ETF's website alone.

The main difference between Replicator and Sampler ETFs is in their annual turnover ratio, measuring the portion of the portfolio that is replaced compared to the previous year. Replicator ETFs have a mean turnover ratio of 41% while Sampler ETFs have a mean turnover ratio of 28%. This is natural, as Replicator ETFs' portfolio weights are strictly dictated by the index weights and so need to be adjusted more frequently. Sampler ETFs on the other hand have discretion in their trading decisions and can delay or forego adjusting their portfolio weights.

We drop ETFs in the Other implementation strategy category in all subsequent tests, as their index replication strategy is entirely different. Reflecting this, their summary statistics are quite different from the Replicator and Sampler ETFs. The Other ETFs on average have much smaller AUMs, higher fees, and higher turnover.

III. Empirical Tests: ETF Trading and Stock Liquidity

This section tests the model's predictions. We first test whether ETFs underweight or completely omit stocks that are *ex ante* illiquid. Next we test how ETFs' weighting decision changes the liquidity of the underlying index assets. Finally, we separately test the effect by samplers and replicators. The empirical results in all tests are consistent with the model predictions.

A. Do Funds Omit Illiquid Assets?

The model predicts that ETF's underlying basket underweights illiquid stocks due to their higher trading costs. ETF provider can choose to omit certain stocks altogether if their liquidity is low. To test this predictions, we use the quarterly holdings of ETFs that track the Russell 1000 (large-cap) and the Russell 2000 (small-cap) index. We focus on ETFs that track Russell Index here because historical index member list and their weights are not publicly available in general. We obtain directly from FTSE Russell the index weights as well as the ETFs that tracks Russell Index. 22 ETFs with unique CRSP portno track Russell 1000 Index while 23 track Russell 2000 Index. Table 2 tests how funds' holdings deviated from their target index. The unit of observation is at the quarter t, ETF j, and index stock i level. The dependent variable is a dummy variable, $\mathbb{1}_{Omitted_{i,j,t}}$, that equals 1 when the stock is in the fund's target index, but is not in the fund's holdings, and zero otherwise. That is, the dependent variable is turned on when the fund did not hold that stock, although the stock was a member of its target index. The independent variable of interest is stock liquidity, *ESpread*, measured by the dollar-weighted percentage effective spread and winsorized at 1% and 99% level to make sure the results are not driven by outliers. Too rule out alternative explanations, stock-month level controls and different fixed effects are added. We use the Linear Probability Model instead of Probit or Logit because fixed effects lead to biased estimates in non-linear models (Greene, Han, and Schmidt (2002)):

$$\mathbb{1}_{Omitted_{i,j,t}} = \beta ESpread_{i,t} + \chi X_{i,t} + \kappa + \epsilon_{i,j,t} \ , \tag{6}$$

Insert Table 2 About Here

where $X_{i,t}$ is the control variables including stock return volatility, $Volatility_{i,t}$, measured by standard deviation of daily stock return of stock i in the last month in quarter t, and stock's correlation with the index, $Correlation\_w\_Index_{i,t}$, measured by the the correlation between daily return of stock i and daily return of the index (either Russell 1000 Index or Russell 2000 Index) that includes stock i in the last month in quarter t. We control for volatility and correlation with the index because the model predicts that they affect ETF's basket weighting decision as well and the variable of interest is stock liquidity. We also control for the index weight of stock i in month t, $IndexWeight_{i,t}$, so that the regression compares the stocks that are equally important to the index. All the control variables are also winsorized at 1% and 99% level.

κ is fix effects. In Table 2 Column 1, we include fund fixed effects and quarter fixed effects. It is possible that there is a general trend affecting ETFs' weighting decision throughout time. For example, as ETF market grew larger and more sophisticated and investors became more demanding, ETFs choose to put more emphasize on minimizing tracking errors in recent years than before, which could bias the results. Adding quarter fixed effects rules out this possibility. Similarly, fund fixed effects sweep out potential time-invariant fund-specific omitted variables, such as the heterogeneity in ETFs' preference over lower track error and lower trading costs. We are also concerned about potential time-varying fund-specific omitted variables. For example, each fund can hire new fund managers and/or change their preference in cost-error tradeoff at any point in the sample period, therefore driving the weighting decision. Column 2 addresses this issue by adding fund-by-quarter fixed effects, and thus compares stocks only within each fund's quarterly holdings snapshot.

Finally, because Russell index weights are float-adjusted, the index weight for each stock covaries strongly with its liquidity. For example, highly liquid stocks in general have a higher float-adjusted market capitalization, therefore mechanically have a higher weight in Russell Index. To address this, in Column 3 we assign each index stock in each quarter to one of 300 index buckets, sorted by their index weight. Therefore each bucket contains 10 index stocks that had similar size and liquidity. Adding bucket fixed effects thus compares each stock with a small set of peer stocks, in the same index, that were equally important from the index's perspective. Unsurprisingly, after adding bucket fixed effects, $IndexWeight_{i,t}$ is no longer statistically significant.

In all cases a stock's liquidity, measured by its effective spread, positively predicts omission by ETFs. That is, index stocks that were less liquid *ex ante* were more likely to be omitted from funds' holdings. The coefficient of $ESpread_{i,t}$ of 0.015 suggests the relationship is economically sizeable, as a 10 basis points increase in a stock's percentage effective spread is associated with a 15% increase in its probability of being omitted by ETFs that track the index. The holdings data show that ETF holdings systematically deviate from their target index in ways that are consistent with the model. We test how ETFs' basket weighting decision affects the liquidity of underlying stocks over time in the next section.

B. Does ETF Primary Flow Affect Liquidity?

The model predicts that as illiquid stocks are underweighted or omitted in ETF holdings compared to their target index, ETFs generate more trading activity in underlying index assets that are more liquid *ex ante*, and vice versa, widening the liquidity gap. We test this hypothesis in this section. The objective is not to isolate the direct effects of variation in ETF trading as in e.g., Ben-David et al. (2018). Rather, the hypothesis is that in equilibrium as investors move from trading individual index assets to trading ETFs, trading volume and liquidity will decline in relative terms for index assets that are more illiquid *ex ante*. Specifically, the model predicts that the effects of ETF trading on asset markets are driven by the fund's implementation strategy, via the authorized participants' trading activities.

The key measure is primary flow, the net ETF share creation and redemption. To calculate primary flow, we obtain ETF's daily shares outstanding from Bloomberg.[10] Each

[10]Brown, Davies, and Ringgenberg (2019) point out that compared to CRSP, Bloomberg's data on daily

ETF j's primary flow on day t is calculated as the change in shares outstanding on day t compared to day $t-1$ times the closing price on day t:

$$PrimaryFlow_{j,t} = (Shrout_{j,t} - Shrout_{j,t-1}) \times p_{j,t} \ .$$

We focus on primary flow instead of total trading volume for two reasons. First, much of the trading activity in ETF shares is bilateral between secondary market investors and does not involve an authorized participant on either side. For such trades there is no accompanying trading in the underlying assets. Second, large block trades may be executed directly through an authorized participant but do not appear in the exchange volume (Lettau & Madhavan, 2018b). The primary flow measure address these two issues by capturing the net inflow or outflow to the ETF each day. This measure directly corresponds to the end-of-day netted order flow in the model.

As a measure of fund flows, ETF primary flow has two main differences with the literature on mutual fund flows (i.e. Coval and Stafford (2007); Goldstein, Li, and Yang (2013)). First, in the mutual fund flows setting the funds are open-ended mutual funds and money flows directly between investors and the fund. Second, in that setting the mutual funds are almost all actively managed, and thus they have discretion in adjusting their holdings. By contrast, in our setting ETF primary flows occur through authorized participants, while the weights on the underlying assets are determined by the ETF provider.

In order to test the model prediction that the primary flow widens the *ex ante* liquidity gap, we construct an aggregate primary flow measure on daily level. We aggregate the the individual primary flow of all the ETFs in the sample, which is all U.S. equity ETFs from 2009 to 2018 in the CRSP Mutual Fund Database with at least \$10 million assets under

shares outstanding is more accurate.

management. This covers over 99% of the assets under management and trading activities of all U.S. equity ETFs:

$$PrimaryFlow_t = \sum_j PrimaryFlow_{j,t} \ .$$

This measure captures the intensity of all authorized participants' trading in the underlying assets as a result of making the market for all the ETFs in the sample at day t. The model predicts that this measure drives the *ex ante* liquidity gap between index stocks to go up. We focus this measure as the main independent variable of interest and refer to this aggregate measure when we use "primary flow" in all subsequent tests. [11]

Insert Figure 4 About Here

For each month, we sort the sample stocks into quintiles based on their average effective spread in previous month. The first quintile contains the most liquid stocks and the fifth quintile contains the most illiquid stocks *ex ante*. We choose one-month window for two reasons. First, stock liquidity exhibits strong short-term reversal. Thus, if we sort stocks based on their effective spread on day $t-1$, the relative liquidity change on day t will

[11] Ideally, we would want to construct a more refined primary flow measure on stock-day level:

$$PrimaryFlow_{i,t} = \sum_j PrimaryFlow_{j,t} \times w_{i,j,t} \quad \forall j \ s.t. \text{ stock } i \in \text{ETF } j \ .$$

where w_i, j, t is stock i's basket weight in ETF j on day t. That is, we captures the exact portion of the primary flow that goes into each stock. This measure directly corresponds to the $\sqrt{N_I + N_I'} w_i^{ETF}$ term in equation (5). However, this requires us to know all the ETFs' basket weights on daily level. The basket weight is considered proprietary knowledge of ETF provider and is difficult to obtain. To circumvent this, we tried focusing only on the 45 Russell ETFs used in previous section and use the index weights (which we obtained from FTSE Russell) to proxy for ETFs' basket weights. This allows us to come up with a stock-day level primary flow but severely limited the scope of ETFs in the study. All the results using this refined primary flow measure are qualitatively the same compared to the results in the paper. We decide to use the less refined but broader primary flow measure to draw more general conclusions.

be largely mechanical. That is, stocks in most liquid quintile will become less liquid, and vice versa, dictated by liquidity's short-term reversal pattern. Second, when ETF providers observe stock liquidity to choose their basket weight, it is unlikely that they will decide what stocks are liquid and illiquid based solely on stocks' liquidity yesterday. In other words, they will use a relatively longer window to form a more informative view on stock liquidity to drive their weighting decisions. Therefore, albeit somewhat arbitrary, one month is a natural choice for the sorting window. The results are qualitatively the same when we use two-month and three-month window for lagged liquidity sorting.

Figure 4 plots the relation between ETF primary flow and asset liquidity. The x-axis is the daily primary flow, and the y-axis is the daily percentage change in stock liquidity, measured as $\log(ESpread_{i,t}) - \log(ESpread_{i,t-1})$. Panel A shows that there is no significant association between the daily total ETF primary flow and daily changes in stock liquidity. That is, ETF primary flows in either direction have no effect on the liquidity of *ex ante* liquid stocks.

By contrast, Figure 4 panel B shows that for illiquid stocks there is a strongly upward-sloping relation in both directions. On days with larger magnitude of ETF primary flows, either positive or negative, the effective spread of illiquid stocks goes up significantly. That is, ETF primary flows, in both directions, are strongly associated with a reduction in the liquidity of illiquid stocks. This pattern is consistent with the model, as illiquid stocks are underweighted or omitted in the ETF's underlying basket, therefore in the trading activity of authorized participants.

To formally examine the relation between ETF primary flows and stock liquidity, we regress the daily change in effective spread for each index stock i on the magnitude of ETF primary flow that day, separately for each lagged liquidity quintile by interacting primary flow magnitude with $Liquid^q_{i,t}$, a dummy variable that equals 1 if stock i is in the liquidity quintile q for the last month as of day t. We also add stock-level lagged controls and fixed effects by stock and date:

$$\Delta\, ESpread_{i,t} = \sum_{q=1}^{5} \beta^q \times |PrimaryFlow_t| \times Liquid_{i,q,t} + Liquid_{i,q,t} + \chi X_{i,t-1} + \gamma_i + \kappa_t + \epsilon_{i,t}\ , \quad (7)$$

where $X_{i,t-1}$ is stock-level lagged controls including lagged stock liquidity, measured by stock i's log turnover, $log(Turnover_{i,t-1})$, and lagged stock market capitalization. This makes sure the main coefficients of interests, β^q, captures the $PrimaryFlow_t$ effect on change in stock liquidity among stocks that are equally important to the index. We use stock turnover instead of effective spread to control for lagged liquidity here because we want to avoid any mechanical relationship due to having effective spread on both sides of the equation. All control variables are winsorized at 1% and 99% level to make sure outliers don't drive the results.

We controlled for quintile main effects by including $Liquid^q_{i,t}$. This sweeps out potential omitted variables that differentially affect the liquidity of stocks in different quintiles but are orthogonal to primary flow. The fixed effects by date, κ_t, sweeps out all observed or unobserved factors, for each day, that change stock liquidity in the same direction across all stocks. In other words, the specification isolates daily changes in stock liquidity in relative terms across the quintiles. Notice that one dummy variable $Liquid^q_{i,t}$ must be excluded, as including all five is collinear with the daily fixed effects. In the estimates the omitted dummy variable is always for quintile 3, $Liquid^3_{i,t}$. This convention lets us estimate the relative effects of daily ETF primary flows on the most-liquid and least-liquid stocks. Stock fixed effects, γ_i, is also added to rule out potential time-invariant stock-specific omitted variables that affect stock liquidity. The results are reported in Table 3.

Insert Table 3 About Here

Table 3 Column 1 shows a clear differential relation of individual stock liquidity with daily ETF primary flows. On days with larger ETF primary flows, in either direction, liquidity for liquid stocks improved while liquidity for illiquid stocks worsened. The difference between the coefficients of the most liquid vs. most illiquid stocks is strongly statistically significant (and for all other columns in the table as well), with an F-statistic of 90.1. The difference is also economically significant. The average annual absolute primary flow in the sample is \$1.03 trillion, which corresponds to an 11.9% ($\exp(0.019 \times 1.03) - 1$) increase in the ratio of the effective spreads of least liquid versus most liquid stocks.

As Figure 3 panel B shows, the magnitude of ETF primary flows has increased steadily throughout the sample period; hence, we scale the daily dollar amount so that that the independent variable has a stationary distribution. Table 3 Column 2 repeats the estimates, scaling the daily ETF primary flows by the total AUM of the ETFs in the sample as of the previous trading day. The results are still both statistically and economically significant. The average annual absolute primary flow as a percentage of total AUM in the sample is 124%, which corresponds to a 16.0% ($\exp(12.00 \times 124\%/100) - 1$) increase in the ratio of the effective spreads of least liquid versus most liquid stocks.

The model predicts that primary flow in either direction have the effect documented in Columns 1 and 2, because both positive and negative primary flow will trigger authorized participants' trading activities in the underlying assets. To test this prediction, in Table 3 Columns 3 and 4 we break the sample into days when the total ETF primary flow is positive (Column 3) versus days when it is negative (Column 4). As shown by the number of observations in the table, there are slightly more net creation days (58%) than redemption days (42%). We see that the larger the magnitude of ETF primary flow – in either direction – the larger the bifurcation of individual stocks' liquidity.

The effect on the most liquid stocks of ETF redemptions is not statistically significant. This is likely due to authorized participants' existing operational short positions (Evans et al., 2019). When redeeming, authorized participants receive a basket of underlying assets, the most liquid ones in which can be netted with their operational short therefore there is no trading in the underlying market and the effect is muted. Least liquid assets get traded regardless because operational shorts only take place in most liquid stocks. On the other hand, in creation days, the most liquid stocks still need to be bought in underlying market to be delivered to the provider because of authorized participants' leverage constraint.

For trading volume, measured by share turnover, $Volume_{i,t}/Shrout_{i,t}$, in individual stock, the pattern is in the opposite direction. Table 3 Columns 5 through 8 repeat the analyses using the daily percentage change in share turnover as the dependent variable and show that regardless of the unit of measurement, and symmetrically in both directions, a larger magnitude of ETF primary flow is associated with a widening liquidity gap between *ex ante* liquid and illiquid stocks.

In sum, across a variety of measures of ETF primary flow, and in both positive and negative directions (i.e. net creation and net redemption), we find differential effects on liquidity and trading volume for individual stocks that are consistent with the predictions of the model. Next, we test how Replicator and Sampler ETFs affect stock liquidity differently.

C. Replicator versus Sampler ETFs and Stock Liquidity

A key prediction that is unique to the model is that Sampler ETFs should cause a larger (more positive) differential effect on asset markets than Replicator ETFs. Moreover, from an investor's point of view there is no salient difference between an ETF that samples and an ETF that fully replicates. Because an ETF's implementation strategy is set at inception, and because investors are indifferent whether they are trading a replicator or sampler, confounding factors such as news arrival, algorithmic trading, or investor behavior predict no difference in the relation for replicator versus sampler funds. Thus, comparing the effects of primary flows between the two types of index implementation strategy is a empirically clean test of the model.

One concern with such a comparison is that replicator and sampler funds differ on other dimensions. For example, funds that track an index of large liquid assets such as the S&P500 are more likely to be fully replicating while funds that track an index containing small and illiquid assets such as the Russell 2000 are more likely to be samplers. This fact is again consistent with the main hypothesis, but it potentially distorts the comparison between replicator and sampler ETFs' primary flow effects.

To address this concern we construct a matched set of replicator and sampler ETFs. For each fund-year in the data that is a sampler (i.e. replicates their target index statistically and is not fully replicating), we attempt to match it to a replicator fund in the same year. Matched fund-year pairs must have the same detailed four-character CRSP objective code and assets under management (AUM) at the beginning of the year that is within 25% of each other. When there are multiple matches we pick the closest match in terms of AUM. For example, the ALPS Dividend Dogs ETF (ticker SDOG, AUM $1.86B, replicator), which tracks a subset of the S&P500 index consisting of the five firms in each sector with the highest dividend yield, is matched with the WisdomTree U.S. Large-Cap Dividend Fund (ticker DLN, AUM $1.88B, sampler) which tracks the 300 largest companies ranked by market capitalization from the WisdomTree U.S. Dividend Index. Thus, these are two large-cap dividend funds which began that year (2016) with almost identical assets under management.

In all, we construct 599 matched fund-year pairs, for an average of 60 matched pairs per year. The average total AUM per year in 2018 dollars is $114B for the sampler funds and $118B for the replicator funds, so the average sampler (replicator) fund has $1.90B ($1.97B) in assets under management. Figure 5 compares the distributions of fund AUM, expense ratio and turnover between the matched samples. By construction, the distributions of fund AUM are nearly identical. The replicator funds have slightly higher expense ratios and turnover on average, but overall the distributions of those fund characteristics are also very similar.

Insert Figure 5 About Here

Table 4 compares the effects of daily ETF primary flows between the matched set of replicator ETFs versus sampler ETFs.

Insert Table 4 About Here

We see that primary flows for both types of ETF affect trading activity (share turnover) in a consistent direction (Columns 2 and 4). However, the effects of ETF primary flow on liquidity are of opposite sign between the two types of ETF. Primary flows for replicator ETFs (Column 1) are associated with a shrinking liquidity gap while, by contrast, primary flows for sampler ETFs (Column 3) are associated with a widening liquidity gap. The magnitude of sampler ETFs' differential effect ($\beta^5 - \beta^1$) is much smaller compared to the results with pooled primary flow in table 3 Column 2. This is because in the matching process we disproportionately lose more AUM than dollar amount primary flow due to lack of common support in ETFs with exotic fund objectives, which usually have muted creation and redemption activities. This inflates the magnitude of the independent variable, dollar amount primary flow scaled by ETFs AUM, and attenuates the relation.

The result is consistent with the model's specific prediction that sampler ETFs' effect on liquidity is always larger (more positive) than that of the replicators. Empirically, we see that the sampler ETFs have a positive gap-widening effect while replicators have a negative gap-shrinking effect. This empirical pattern is difficult to explain via other confounding factors because (i) other confounding variables are swept out via the stock-day controls and fixed effects, (ii) the matched fund pairs follow similar indices and have similar assets under management, and (iii) traders and investors are indifferent whether a given fund is a replicator or a sampler.

IV. Alternative explanations

One concern with the results is that market dynamics change over time, and could covary with both ETF primary flow and stock turnover and liquidity. For example, the arrival of index-relevant information could drive increased ETF trading activity, and also cause market makers to reduce their trading activity and widen their bid/ask spreads, particularly in less liquid stocks. We examine these potential confounds in two ways.

First, it could be that when there is market-moving news (or the risk of market-moving news arriving), market makers reduce liquidity more in stocks that were less liquid *ex ante*.

To examine this possibility, we first add the magnitude (absolute value) of the CRSP value-weighted U.S. market index as an additional explanatory variable, interacted with each stock's lagged liquidity quintile. Table 5 Columns 1 and 2 show that the relationship between ETF primary flows and asset liquidity is effectively unchanged when we add the daily market return as an additional explanatory factor (compare with Table 3 Column 3 and 7).

Insert Table 5 About Here

Second, we drop from the sample any days in which the U.S. stock market had a return, measured by CRSP value-weighted U.S. market index return, outside the range [-0.5%, +0.5%]. This filter leaves us with a subsample of 1,291 trading days on which the market return was nearly unchanged. Table 5 Columns 3 and 4 show the results. The differential relationship between ETF primary flows and stock liquidity is even stronger than in the full sample, particularly for the least liquid set of underlying stocks, and the relationship with stock-level trading activity is again apparent. Thus, on quiet market days, the relationship between ETF primary flows and stock liquidity is actually stronger than on days when the market moved a lot.

The third way that we examine other market factors is to control directly for time-varying factors such as high frequency trading activity and market fragmentation that are well known to have great impact on liquidity. We use the SEC MIDAS data to construct measures of both high frequency trading activity and market fragmentation for each stock by month individually from 2012 to 2018. For high frequency trading activity, we construct the trade-to-order ratio following the same procedure in (Weller, 2017). For market fragmentation, we construct an HHI of trading volumes across market venues:

$$HHI_{i,t} = \sum_j \left(\frac{\text{Trading volume}_{i,j,t}}{\sum_j \text{Trading volume}_{i,j,t}} \right)^2$$

where i denotes stock, j denotes market venue, and t denotes month. The results when we add those stock-level measures as controls are shown in Table 6.

Insert Table 6 About Here

We see that the differential relations of daily total ETF primary flow with stock liquidity and turnover are still present, both across all days in the more recent sample (Columns 1 and 2) and when we condition down to days on which the market did not move (Columns 3 and 4). We conclude that recent trends in high frequency trading activity and market fragmentation do not explain the results.

V. Conclusion

An ETF's objective is to closely track a target index at a low cost. In this paper we document the direct consequences of ETF implementation strategy, most prominently a bifurcation effect on the liquidity of the underlying index assets. Liquid stocks become more liquid, and illiquid stocks become more illiquid, due to ETF trading activity.

We construct a stylized model to characterize the trade-off between tracking error and trading cost. The model predicts that for stocks that are illiquid and expensive to trade, index funds and ETF providers are better off underweighting or omitting these stocks. The model further predicts that the effects of ETF trading on underlying assets markets is determined by their index replication strategy. This is an important point for the academic literature on the effects of the rise of index investing.

Our contribution is two-fold. First, although comovement in stock liquidity is well studied over the past two decades, the widening liquidity gap in the U.S. stock market since 2006 has received little attention. The theory and results in our paper help explain this fact and further predict that as long as ETF trading activity continues to increase, the liquidity gap is predicted to widen even further. Second, we point out a treatment effect that has been ignored by empirical studies in this field – index constituents being systematically underweighted or omitted from the ETF basket. Mistakenly classifying omitted stocks as treated stocks (as will happen using either index weights or index assignment to proxy for fund ownership) can result in biased estimates of the effects of index investing. Thus, the results should inform future empirical research as well.

References

Agarwal, V., Hanouna, P., Moussawi, R., & Stahel, C. W. (2018). Do ETFs increase the commonality in liquidity of underlying stocks? *Working Paper*.

Angel, J. J., Harris, L. E., & Spatt, C. S. (2015). Equity trading in the 21st century: An update. *Quarterly Journal of Finance*, *5*(1), 1–39.

Basak, S., Pavlova, A., & Shapiro, A. (2007). Optimal asset allocation and risk shifting in money management. *Review of Financial Studies*, *20*(5), 1583–1621.

Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do ETFs increase volatility? *Journal of Finance*, *73*(6), 2471–2535.

Blocher, J., & Whaley, R. E. (2016). Two-sided markets in asset management: Exchange-traded funds and securities lending. *Working Paper*.

Brown, D. C., Davies, S., & Ringgenberg, M. (2019). Etf flows, non-fundamental demand, and return predictability. *Working Paper*.

Carpenter, J. N. (2000). Does option compensation increase managerial risk appetite? *Journal of Finance*, *55*(5), 2311–2331.

Coval, J., & Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, *86*(2), 479–512.

Da, Z., & Shive, S. (2018). Exchange traded funds and asset return correlations. *European*

Financial Management, *24*(1), 136–168.

Easley, D., Michayluk, D., O'Hara, M., & Putnins, T. J. (2019). The Active World of Passive Investing. *Working Paper*.

Evans, R. B., Moussawi, R., Pagano, M. S., & Sedunov, J. (2019). ETF short interest and failures-to-deliver: Naked short-selling or operational shorting? *Working Paper*.

Goldstein, I., Li, Y., & Yang, L. (2013). Speculation and hedging in segmented markets. *Review of Financial Studies*, *27*, 881-922.

Greene, W., Han, C., & Schmidt, P. (2002). The bias of the fixed effects estimator in nonlinear models. *Unpublished Manuscript, Stern School of Business, NYU*, *29*.

Greenwood, R. (2007). Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights. *Review of Financial Studies*, *21*(3), 1153–1186.

Haslag, P. H., & Ringgenberg, M. (2016). The causal impact of market fragmentation on liquidity. *Working Paper*.

Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *Journal of Finance*, *66*(1), 1–33.

Holden, C. W., & Jacobsen, S. (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance*, *69*(4), 1747–1785.

Hortaçsu, A., & Syverson, C. (2004). Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A case study of S&P 500 index funds. *Quarterly Journal of Economics*, *119*(2), 403-456.

Israeli, D., Lee, C. M., & Sridharan, S. A. (2017). Is there a dark side to exchange traded funds? An information perspective. *Review of Accounting Studies*, *22*(3), 1048–1083.

Jones, C. M. (2013). What do we know about high-frequency trading? *Working Paper*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1335.

Lee, C. M., & Ready, M. J. (1991). Inferring trade direction from intraday data. *Journal of Finance*, *46*(2), 733–746.

Lettau, M., & Madhavan, A. (2018a). Exchange-traded funds 101 for economists. *Journal of Economic Perspectives*, *32*(1), 135–54.

Lettau, M., & Madhavan, A. (2018b). Exchange-traded funds 101 for economists. *Journal of Economic Perspectives*, *32*(1), 135–54.

Malamud, S. (2016). A dynamic equilibrium model of ETFs. *Working Paper*.

Pan, K., & Zeng, Y. (2019). ETF arbitrage under liquidity mismatch. *Working Paper*.

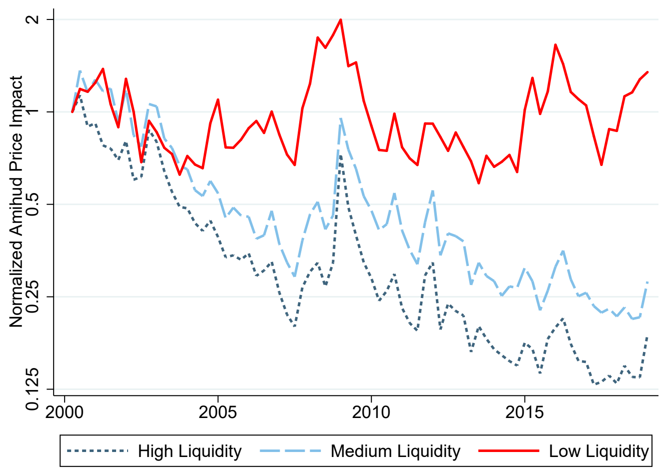
Sağlam, M., Tuzun, T., & Wermers, R. (2019). Do ETFs increase liquidity? *Working Paper*.

Weller, B. M. (2017). Does algorithmic trading reduce information acquisition? *Review of*

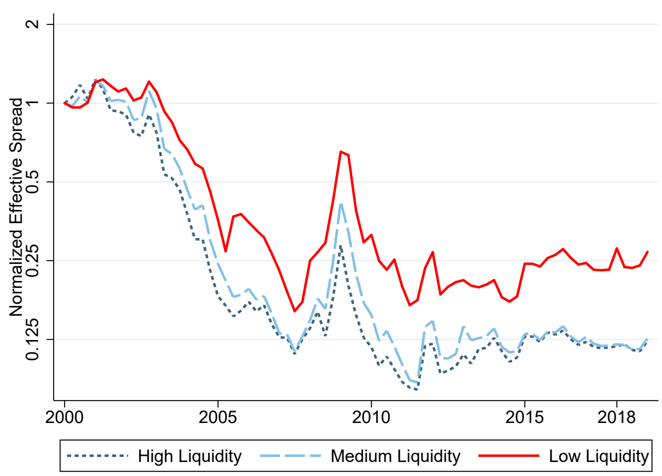
Financial Studies, *31*(6), 2184–2226.

Figure 1. The evolution of U.S. stock liquidity

(a) Amihud measure

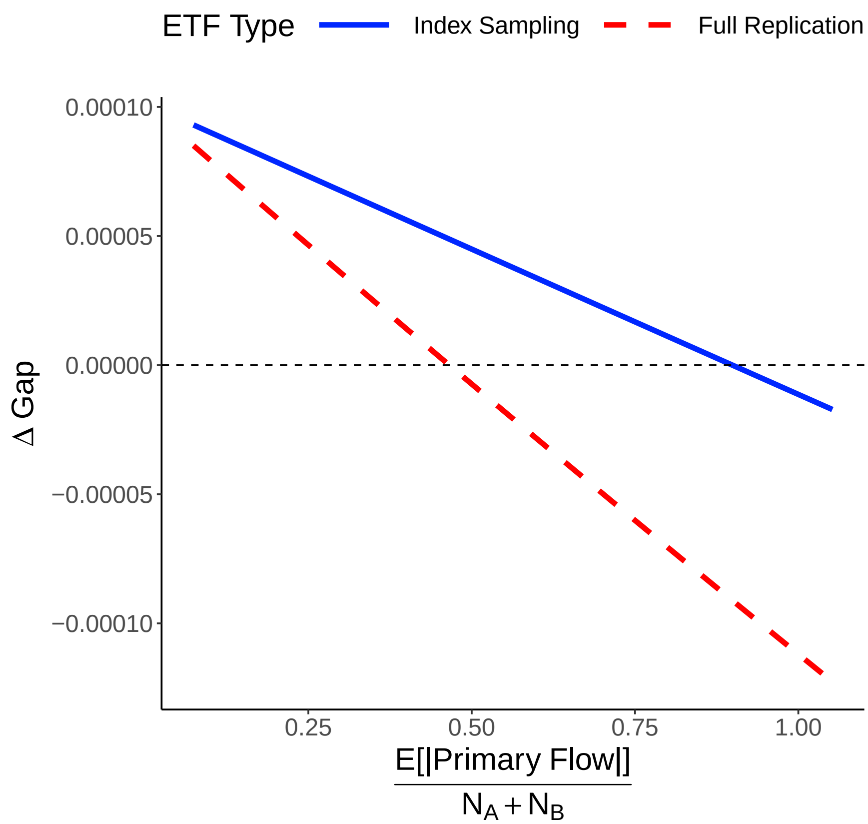


(b) Effective Spread



The figure plots the distribution of liquidity across U.S. common stocks over time. Stocks are sorted into terciles on the basis of their liquidity each quarter. Panel A plots the average quarterly Amihud measure from 2000 to 2018 for each tercile, while panel B plots the average quarterly effective spreads from 2000 to 2018 for each tercile. Both measures are normalized to their 2000 level. The sample includes all U.S. common stocks with market capitalization greater than $300M in 2018 dollars.

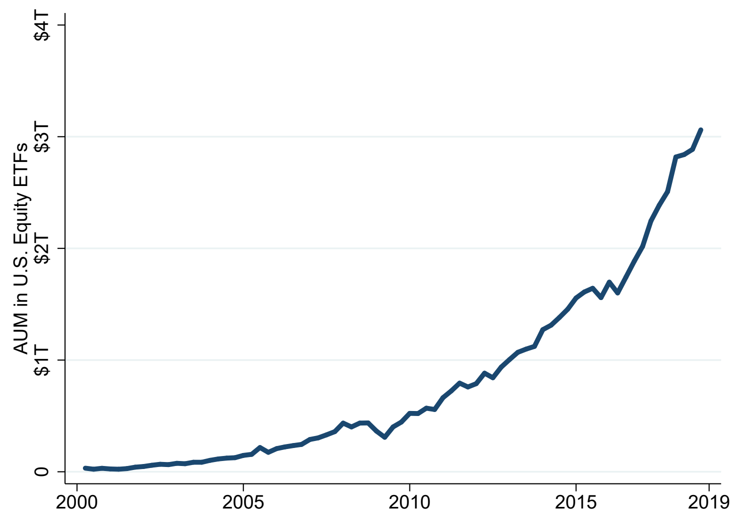
Figure 2. The relative change in liquidity between underlying assets



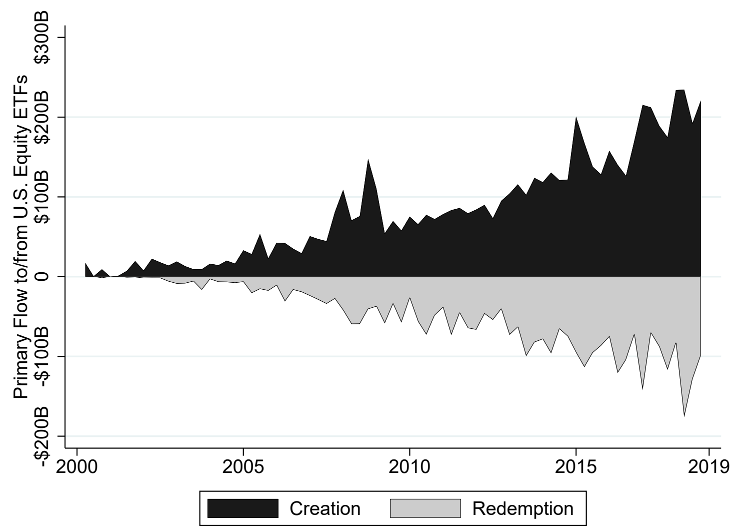
The figure plots the model implied relative change in liquidity, from pre- to post-ETF, between the underlying index assets as a function of ETF primary flow (net creation / redemption activity), scaled by trading volume in the individual assets. A positive value of ΔGap means the liquidity gap widens (bifurcating effect), while a negative value means the liquidity gap is reduced.

Figure 3. ETF assets under management and primary flows

(a) ETF assets under management



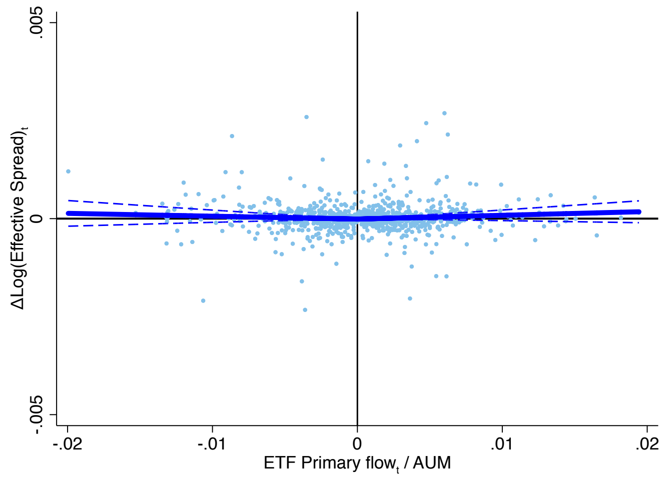
(b) ETF primary flows



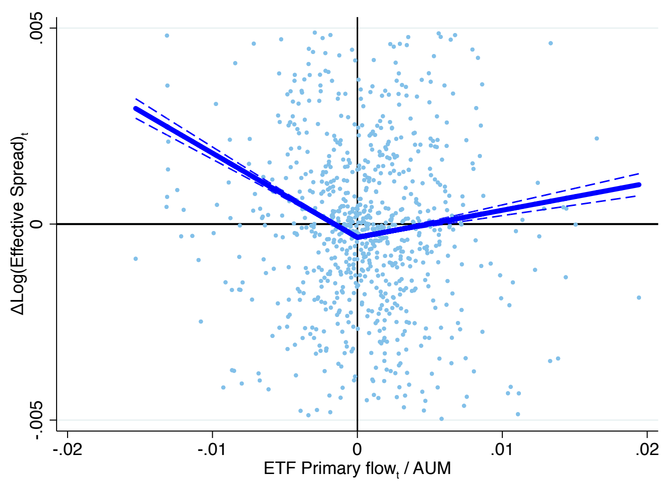
Panel A plots total assets under management and panel B plots total positive and negative primary flows (total dollar create and redeem activity, respectively) in U.S. equity exchange traded funds (ETFs), quarterly from 2000 to 2018.

Figure 4. ETF primary flow and changes in asset liquidity

(a) Quintile 1 (Most liquid *ex ante*)



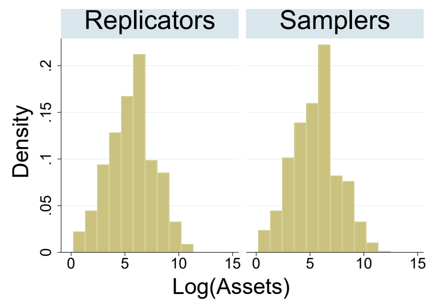
(b) Quintile 5 (Least liquid *ex ante*)



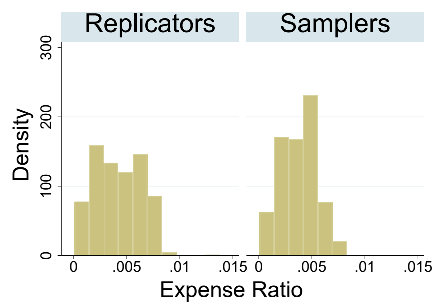
The figure plots daily changes in effective spreads for individual stocks against daily total primary flows (total dollar creations minus redemptions scaled by AUM) for U.S. equity ETFs. Panels A and B plot the relation for stocks that were in the first (most liquid) and fifth quintile (least liquid) respectively sorted by effective spread as of the previous month (the most liquid and least liquid stocks *ex ante*). The blue lines show the linear best-fit line, separately estimated for positive and negative ETF primary flows. The dashed lines show 95% confidence intervals.

Figure 5. Fund characteristics of replicators and samplers in matched sample

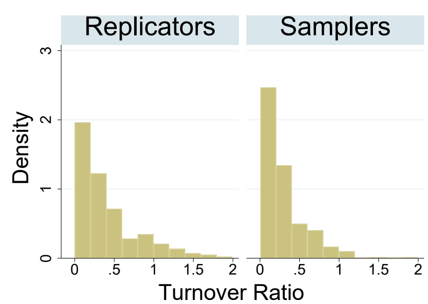
(a) Fund AUM



(b) Fund expense ratio



(c) Fund turnover ratio



The figure plots histograms of log assets under management (a), expense ratios (b), and yearly turnover ratios (c), comparing the matched samples of replicator ETFs (left) and sampler ETFs (right).

Table 1
Summary statistics

Panel A presents summary statistics of the sample stocks, which consist of all Russell 3000 members (including both the Russell 1000 large-cap index and the Russell 2000 small-cap index) monthly from 2009-2018. The sample contains 5,743 unique stocks. Panel B displays summary statistics of ETFs in the sample, which consists of all U.S. equity ETFs in the CRSP mutual fund database with at least $10 million assets under management from 2009-2018. Panel B is split on the basis of each fund's implementation strategy. Variable definitions can be found in Appendix section A.

Panel A: Index Stocks

	Mean	StDev	P10	Median	P90
Market capitalization ($ Millions)	6,700	25,267	198	1,198	12,672
Effective spread	0.0043	0.0287	0.0004	0.0016	0.0082
Volatility	0.025	0.014	0.011	0.021	0.042

Panel B: Exchange Traded Funds

	Observations	Mean	StDev	P10	Median	P90
	Full Replication					
AUM ($ Millions)	4,233	1,921	9,239	23	203	3,902
Expense Ratio (%)	4,059	0.45	0.22	0.14	0.48	0.70
Turnover	4,043	0.41	0.55	0.06	0.25	0.94
	Sampling / Optimized					
AUM ($ Millions)	1,279	2,802	9,894	22	282	4,725
Expense Ratio (%)	1,226	0.44	0.17	0.20	0.48	0.63
Turnover	1,220	0.28	0.28	0.05	0.20	0.61
	Other					
AUM ($ Millions)	425	268	538	15	60	816
Expense Ratio (%)	397	0.81	0.27	0.45	0.95	0.99
Turnover	386	1.00	2.03	0.07	0.45	2.66

Table 2
Stock characteristics and ETF holdings

The table presents regressions of quarterly fund holdings by ETFs on liquidity of index stocks:

$$\mathbb{1}_{Omitted_{i,j,t}} = \beta ESpread_{i,t} + \chi X_{i,t} + \kappa + \epsilon_{i,j,t} \ ,$$

where the dependent variable $\mathbb{1}_{Omitted_{ijt}}$ is a dummy variable that equals 1 if fund j omitted stock i in its holdings in quarter t, and 0 if fund j held any number of shares of stock i in quarter t. The independent variable is effective spread, $ESpread_{i,t}$, measured as the average percentage effective spread of stock i in quarter t. $X_{i,t}$ is stock level controls including stock volatility, correlation with stock return and index return, and stock's index weight. Precise definitions can be found in Appendix section A. κ is fixed effects. The sample unit is stock-fund-quarter. The sample includes all stocks that were in the Russell 1000 or 2000 index and all Russell 1000 and 2000 ETFs that reported their holdings in quarter t, from 2009 to 2018. Standard errors are clustered by stock. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)
$ESpread_{i,t}$	0.014***	0.014***	0.015**
	(0.005)	(0.005)	(0.006)
$Volatility_{i,t}$	0.016***	0.017***	0.016***
	(0.004)	(0.004)	(0.004)
$Correlation\_w\_Index_{i,t}$	-0.007	-0.005	-0.006
	(0.012)	(0.013)	(0.013)
$IndexWeight_{i,t}$	1.683	1.663	-0.683
	(2.131)	(2.130)	(2.531)
Fund FE	Yes	No	No
Year-Quarter FE	Yes	No	No
Fund x Year-Quarter FE	No	Yes	Yes
Index Bucket FE	No	No	Yes
Observations	1,005,002	1,005,002	1,005,002
R-squared	0.22	0.23	0.24

Table 3
ETF primary flows and asset liquidity

The table presents regressions of the daily percent changes in effective spreads $\%\Delta ESpread_{i,t}$ and turnover $\%\Delta Turnover_{i,t}$ of individual stocks on the magnitude of daily ETF primary flow i.e. creation and redemption activity, $|PrimaryFlow_t|$. The regression estimate the Primary Flow effect separately for each lagged liquidity quintile:

$$\%\Delta Y_{i,t} = \sum_{q=1}^{5} \beta^q \times |PrimaryFlow_t| \times Liquid^q_{i,t} + Liquid^q_{i,t} + \chi X_{i,t-1} + \gamma_i + \kappa_t + \epsilon_{i,t} \ ,$$

The dependent variable, daily percentage change in effective spreads and turnover, is calculated as $\log(Y_{i,t}) - \log(Y_{i,t-1})$. The independent variable of interest is the magnitude of aggregate primary flow of all U.S. equity ETFs with at least $10 million assets under management. $Liquid^q_{i,t}$ represents lagged liquidity quintile and is a dummy variable that equals one if stock i is in the liquidity quintile q for the last month as of day t and zero otherwise. Quintile 1 contains the most liquid stocks while quintile 5 contains the least liquid stocks. $Liquid^3_{i,t}$ is always omitted to avoid multicollinearity with stock fixed effects. $X_{i,t-1}$ is stock-level controls including market capitalization and liquidity lagged by one day (measured as $Turnover_{i,t-1}$ for Columns 1-4 and $ESpread_{i,t-1}$ for Columns 5-8 to avoid mechanical relationship). The precise definitions can be found in Appendix section A. The bottom of the table reports the difference of the primary flow effects on liquidity of stocks in the most and the least liquid quintile in previous month. The sample unit is stock by day. The sample consists of all Russell 3000 member stocks daily from 2009 to 2018. Standard errors are clustered by stock. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\%\Delta ESpread_{i,t}$	$\%\Delta ESpread_{i,t}$	$\%\Delta ESpread_{i,t}$	$\%\Delta ESpread_{i,t}$	$\%\Delta Turnover_{i,t}$	$\%\Delta Turnover_{i,t}$	$\%\Delta Turnover_{i,t}$	$\%\Delta Turnover_{i,t}$
$\|PrimaryFlow_t\| \times Liquid^1_{i,t}$ (Most Liquid)	-0.018***	-2.11***	-3.73***	0.07	-0.005	-0.34	-1.43**	0.75
	(0.003)	(0.24)	(0.37)	(0.39)	(0.005)	(0.42)	(0.66)	(0.75)
$\|PrimaryFlow_t\| \times Liquid^2_{i,t}$	-0.007***	-1.47***	-1.93***	-0.82*	0.002	0.88*	0.62	1.04
	(0.003)	(0.27)	(0.44)	(0.45)	(0.005)	(0.47)	(0.73)	(0.84)
$\|PrimaryFlow_t\| \times Liquid^3_{i,t}$	-	-	-	-	-	-	-	-
$\|PrimaryFlow_t\| \times Liquid^4_{i,t}$	0.006*	1.49***	2.49***	0.12	-0.011**	-1.02**	-1.35*	-0.65
	(0.004)	(0.35)	(0.55)	(0.58)	(0.005)	(0.47)	(0.76)	(0.89)
$\|PrimaryFlow_t\| \times Liquid^5_{i,t}$ (Least Liquid)	0.085***	9.91***	8.63***	11.12***	-0.042***	-3.10***	-2.86***	-3.63***
	(0.011)	(1.47)	(1.51)	(3.23)	(0.007)	(0.51)	(0.79)	(0.87)
Quintile Main Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock-level Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,219,418	7,219,418	4,207,821	3,011,594	7,219,418	7,219,418	4,207,821	3,011,594
R-squared	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02
PrimaryFlow Units	$Trillion	($/AUM)/100	($/AUM)/100	($/AUM)/100	$Trillion	($/AUM)/100	($/AUM)/100	($/AUM)/100
Sample Dates	All	All	PrimaryFlow>0	PrimaryFlow<0	All	All	PrimaryFlow>0	PrimaryFlow<0
$\beta^5 - \beta^1$	0.109***	12.00***	12.35***	11.01***	-0.035***	-2.76***	-1.43**	-4.37***
F-stat	90.1	68.6	69.8	11.8	32.8	42.8	5.1	39.8

Table 4
ETF primary flows and asset liquidity: Replicator vs sampler ETFs

The table repeats the analyses in table 3 with different way of measuring primary flow. The table presents regressions of the daily percent changes in effective spreads $\%\Delta ESpread_{i,t}$ and turnover $\%\Delta Turnover_{i,t}$ of individual stocks on the magnitude of daily ETF primary flow i.e. creation and redemption activity, $|PrimaryFlow_t|$. The daily ETF primary flows are calculated using a matched sample of pairs of replicator and sampler fund-years that have the same Lipper objective code and similar assets under management (AUM). We divide sample stocks each month into quintiles on the basis of their liquidity in the previous month: *Liquid* 1 contains the most liquid stocks, while *Liquid* 5 contains the least liquid stocks. Standard errors are clustered by stock. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) $\%\Delta ESpread_{i,t}$	(2) $\%\Delta Turnover_{i,t}$	(3) $\%\Delta ESpread_{i,t}$	(4) $\%\Delta Turnover_{i,t}$
$\lvert PrimaryFlow_t\rvert \times Liquid^1_{i,t}$ (Most Liquid)	0.21***	0.40***	-1.47***	-0.15
	(0.08)	(0.13)	(0.33)	(0.30)
$\lvert PrimaryFlow_t\rvert \times Liquid^2_{i,t}$	0.34***	0.33**	-0.24	0.36
	(0.09)	(0.16)	(0.38)	(0.35)
$\lvert PrimaryFlow_t\rvert \times Liquid^3_{i,t}$	-	-	-	-
$\lvert PrimaryFlow_t\rvert \times Liquid^4_{i,t}$	-0.41***	-0.39**	0.31	-1.53***
	(0.13)	(0.19)	(0.44)	(0.56)
$\lvert PrimaryFlow_t\rvert \times Liquid^5_{i,t}$ (Least Liquid)	-0.63**	-0.55***	3.15***	-2.96***
	(0.32)	(0.17)	(0.90)	(0.42)
ETF Type	Replicators	Replicators	Samplers	Samplers
Quintile Main Effects	Yes	Yes	Yes	Yes
Stock-level Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Daily FE	Yes	Yes	Yes	Yes
Observations	7,219,418	7,219,418	7,219,418	7,219,418
R-squared	0.01	0.02	0.01	0.02
PrimaryFlow Units	($/AUM)/100	($/AUM)/100	($/AUM)/100	($/AUM)/100
$\beta^5 - \beta^1$	-0.85**	-0.95***	4.63***	-2.81***
F-stat	7.2	43.9	28.4	57.7

Table 5
ETF primary flows and asset liquidity: Controlling for market-wide news

The table repeats the analyses in table 3 with additional controls and different sample. The table presents regressions of the daily percent changes in effective spreads $\%\Delta ESpread_{i,t}$ and turnover $\%\Delta Turnover_{i,t}$ of individual stocks on the magnitude of daily ETF primary flow i.e. creation and redemption activity, $|PrimaryFlow_t|$. Columns 1 and 2 interact the liquidity quintiles with the magnitude of market return, measured as CRSP value-weighted U.S. market index return. The No News sample in columns 3 and 4 consists only of days on which the market return was smaller than +/- 50 basis points. We divide sample stocks each month into quintiles on the basis of their liquidity in the previous month: *Liquid* 1 contains the most liquid stocks, while *Liquid* 5 contains the least liquid stocks. Standard errors are clustered by stock. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) $\%\Delta ESpread_{i,t}$	(2) $\%\Delta Turnover_{i,t}$	(3) $\%\Delta ESpread_{i,t}$	(4) $\%\Delta Turnover_{i,t}$
$\lvert PrimaryFlow_t\rvert \times Liquid^1_{i,t}$ (Most Liquid)	0.21***	0.40***	-1.47***	-0.15
	(0.08)	(0.13)	(0.33)	(0.30)
$\lvert PrimaryFlow_t\rvert \times Liquid^2_{i,t}$	-1.52***	-0.65	-3.61***	0.38
	(0.31)	(0.54)	(0.70)	(1.10)
$\lvert PrimaryFlow_t\rvert \times Liquid^3_{i,t}$	-	-	-	-
$\lvert PrimaryFlow_t\rvert \times Liquid^4_{i,t}$	0.42	-0.82	1.98**	-0.97
	(0.41)	(0.54)	(0.90)	(1.12)
$\lvert PrimaryFlow_t\rvert \times Liquid^5_{i,t}$ (Least Liquid)	9.06***	-1.51***	17.03***	-3.11***
	(2.15)	(0.57)	(4.07)	(1.14)
$\lvert MarketRet_t\rvert \times Liquid^1_{i,t}$	-0.00*	0.01***	-0.01*	0.02**
	(0.00)	(0.00)	(0.01)	(0.01)
$\lvert MarketRet_t\rvert \times Liquid^2_{i,t}$	0.00	0.01***	-0.00	0.03***
	(0.00)	(0.00)	(0.00)	(0.01)
$\lvert MarketRet_t\rvert \times Liquid^3_{i,t}$	-	-	-	-
$\lvert MarketRet_t\rvert \times Liquid^4_{i,t}$	0.01***	-0.00	0.02***	0.02***
	(0.00)	(0.00)	(0.01)	(0.01)
$\lvert MarketRet_t\rvert \times Liquid^5_{i,t}$	0.01	-0.01***	0.02	0.02***
	(0.01)	(0.00)	(0.01)	(0.01)
Sample Dates	All	All	No News	No News
Quintile Main Effects	Yes	Yes	Yes	Yes
Stock-level Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Daily FE	Yes	Yes	Yes	Yes
Observations	7,219,418	7,219,418	3,834,554	3,834,554
R-squared	0.01	0.02	0.01	0.02
PrimaryFlow Units	($/AUM)/100	($/AUM)/100	($/AUM)/100	($/AUM)/100
$\beta^5 - \beta^1$	10.95***	-0.20	22.02***	-2.03**
F-stat	26.2	0.2	29.6	5.3

Table 6
ETF primary flows and asset liquidity: Controlling for market fragmentation and HFT activity

The table repeats the analyses in table 3 with additional controls. The table presents regressions of the daily percent changes in effective spreads $\%\Delta ESpread_{i,t}$ and turnover $\%\Delta Turnover_{i,t}$ of individual stocks on the magnitude of daily ETF primary flow i.e. creation and redemption activity, $|PrimaryFlow_t|$. The Herfindahl of trading volume across venues, $HHI_{i,t}$, measures market fragmentation. The trade-to-order ratio, $TOR_{i,t}$, measures high frequency trading activity. The precise definitions can be found in Appendix section A. The results are similar using other measures such as Odd-lot ratio, average trade size, and cancel-to-trade ratio. We divide sample stocks each month into quintiles on the basis of their liquidity in the previous month: *Liquid* 1 contains the most liquid stocks, while *Liquid* 5 contains the least liquid stocks. Standard errors are clustered by stock. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) $\%\Delta ESpread_{i,t}$	(2) $\%\Delta Turnover_{i,t}$	(3) $\%\Delta ESpread_{i,t}$	(4) $\%\Delta Turnover_{i,t}$
$\|PrimaryFlow_t\| \times Liquid^1_{i,t}$ (Most Liquid)	-3.90***	1.68*	-7.97***	2.03
	(0.68)	(0.93)	(1.04)	(1.53)
$\|PrimaryFlow_t\| \times Liquid^2_{i,t}$	-1.58**	2.17**	-4.42***	3.20**
	(0.77)	(0.97)	(1.35)	(1.57)
$\|PrimaryFlow_t\| \times Liquid^3_{i,t}$	-	-	-	-
$\|PrimaryFlow_t\| \times Liquid^4_{i,t}$	-1.86**	-3.50***	0.17	-2.71*
	(0.92)	(1.01)	(1.67)	(1.53)
$\|PrimaryFlow_t\| \times Liquid^5_{i,t}$ (Least Liquid)	6.57***	-10.43***	11.71***	-11.20***
	(2.17)	(1.14)	(3.87)	(1.68)
$HHI_{i,t}$	0.11***	0.08***	0.13***	0.09***
	(0.03)	(0.01)	(0.04)	(0.01)
$TOR_{i,t}$	-1.29***	11.56***	-1.20***	11.83***
	(0.05)	(0.20)	(0.06)	(0.27)
Sample Dates	All	All	No News	No News
Control for Market Returns	Yes	Yes	Yes	Yes
Quintile Main Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Daily FE	Yes	Yes	Yes	Yes
Observations	4,527,432	4,527,432	2,646,165	2,646,165
R-squared	0.02	0.04	0.02	0.04
PrimaryFlow Units	($/AUM)/100	($/AUM)/100	($/AUM)/100	($/AUM)/100
$\beta^5 - \beta^1$	10.47***	-12.11***	19.68***	-13.24***
F-stat	24.9	127.5	27.4	70.1

Appendix

A. Variable Definitions

Table A1: Variable Definitions

Variable Names	Description
$\%\Delta \text{ESpread}_{i,t}$	The percentage change in effective spread of stock i on day t is calculated as following: $$\%\Delta \text{ESpread}_{i,t} = \log(\text{ESpread}_{i,t}) - \log(\text{ESpread}_{i,t-1})$$
$\%\Delta \text{Turnover}_{i,t}$	The percentage change in turnover of stock i on day t is calculated as following: $$\%\Delta \text{Turnover}_{i,t} = \log(\text{Turnover}_{i,t}) - \log(\text{Turnover}_{i,t-1})$$
$\mathbb{1}_{\text{Omitted};\ i,j,t}$	Dummy variable that equals 1 if ETF j omitted stock i in quarter t.
$\text{Amihud}_{i,t}$	Absolute value of monthly return of stock i divided by monthly dollar trading volume in month t. Monthly dollar trading volume is calculated as monthly share trading volume times end of month closing price.
$\text{AUM}_{i,t}$ (ETF)	Price of ETF i times shares outstanding of ETF i on day t.
$\text{Correlation\_w\_Index}_{i,t}$	The correlation between daily returns of stock i and the daily . returns of the index that includes stock i in quarter t

Continued on next page

Table A1 – continued from previous page

Variable Definitions	Description
$\text{ESpread}_{i,t}$	Let i denote stock, t denote day, and s denote intraday time. The dollar-weighted percentage effective spread of stock i on day t is calculated as following: $$\sum_s \frac{\lvert \log(\text{price}_{its}) - \log(\text{midpoint}_{its}) \rvert \cdot \text{price}_{its} \cdot \text{size}_{its}}{\sum_s \text{price}_{its} \cdot \text{size}_{its}}$$ Buy-sell indicator are created using method in Lee and Ready (1991) and quotes prior to 2015 are interpolated using method in Holden and Jacobsen (2014).
$\text{Expense Ratio}_{i,t}$	Expense ratio of ETF i in year t from CRSP Mutual Fund Database.
$\text{HHI}_{i,t}$	Let i denote stock, j denote exchange, and t denote month. Herfindahl-Hirschman Index is calculated monthly as following: $$\text{HHI}_{i,t} = \sum_j \left(\frac{\text{Trading Volume}_{ijt}}{\sum_j \text{Trading Volume}_{ijt}} \right)^2$$
$\text{IndexWeight}_{i,t}$	Stock i's weight in index j at month t, from Russell proprietary data.
$\text{Market Capitalization}_{i,t}$	Closing price of stock i times shares outstanding of stock i on day t.
MarketRet_t	Return of CRSP value-weighted U.S. market index on day t.
PrimaryFlow_t, dollar unit	The sum of all individual ETF's primary flow, $\text{PrimaryFlow}_{j,t}$, on day t.
PrimaryFlow_t, percent unit	The dollar unit aggregate primary flow on day t divided by the total ETF AUM on day t.

Continued on next page

Table A1 – continued from previous page

Variable Definitions	Description
$\text{PrimaryFlow}_{j,t}$, dollar unit	The change of the shares outstanding of ETF j from day *t-1* to day t times closing price of ETF j on day t.
$\text{PrimaryFlow}_{j,t}$, percent unit	The dollar unit primary flow of ETF j on day t divided by the AUM of ETF j on day t.
$\text{Shrout}_{i,t}$ ($\text{Shrout}_{j,t}$)	The shares outstanding of stock i (ETF j) on day t.
$\text{TOR}_{i,t}$	Trade to order ratio, calculated as the trading volume of stock i on day t divided by the order volume of stock i on day t.
$\text{Turnover}_{i,t}$	Trading volume of stock i divided by shares outstanding of stock i on day t.
$\text{Volatility}_{i,t}$	Standard deviation of daily return of stock i in month t.
$\text{Volume}_{i,t}$	Trading volume of stock i on day t.

B. Further discussion of the model

B.1. Open-ended index funds

As noted before, the model also applies to traditional open-ended index funds that rebalance their portfolios after inflows and outflows. The index fund manager chooses their implementation strategy to attract investors, who again have the same utility function over expected tracking error and expected trading costs (which investors now bear directly via the management fee, instead of indirectly via the bid/ask spread):

$$U = \boldsymbol{C}(\boldsymbol{w}) + \lambda(\boldsymbol{w} - \boldsymbol{v})'\Sigma(\boldsymbol{w} - \boldsymbol{v})$$

The rest of the solution and all comparative statics are the same.

B.2. Trading in index futures

Recall that

$$w_i^* = \frac{1}{1 + c_i/\lambda\sigma_i^2} v_i + \frac{1}{1 + c_i/\lambda\sigma_i^2} \sum_{j \neq i} (v_j - w_j)\beta_{j,i}$$

Consider an index for which there is a liquid futures contract. The futures contract has $v_i = 0$, so the first term is zero. That is, the futures contract is not an index constituent, and for such assets we would think w_i^* should also be zero. But the futures contract is perfectly correlated with the weighted return of the index constituents, and has very low trading costs c_i. In this case, the optimal weights w_j on all the index constituents themselves are close to zero and the optimal weight on the futures contract is close to one.

B.3. Redundant assets

Recall that

$$w_i^* = \frac{1}{1 + c_i/\lambda\sigma_i^2} v_i + \frac{1}{1 + c_i/\lambda\sigma_i^2} \sum_{j \neq i} (v_j - w_j)\beta_{j,i}$$

Consider two perfectly substitutable assets ($\rho_{1,2} = 1$) with different relative trading costs

$(c_1/\sigma_1^2 \neq c_2/\sigma_2^2)$. Assume for simplicity that $\rho_{1,j}, \rho_{2,j} = 0, \forall j \neq 1, 2$.[12] We have a system of equations with two equations and two unknowns:

$$w_1^* = \kappa_1 v_1 + \kappa_1 v_2 \frac{\sigma_2}{\sigma_1} - \kappa_1 w_2 \frac{\sigma_2}{\sigma_1}$$
$$w_2^* = \kappa_2 v_2 + \kappa_2 v_1 \frac{\sigma_1}{\sigma_2} - \kappa_2 w_1 \frac{\sigma_1}{\sigma_2}$$

where

$$\kappa_i = \frac{1}{1 + c_i/\lambda\sigma_i^2}$$

Solve for:

$$w_1^* = \frac{\kappa_1(1-\kappa_2)}{1-\kappa_1\kappa_2}\left(v_1 + \frac{\sigma_2}{\sigma_1} v_2\right)$$

Note that, if security one is almost costless to trade, then $\kappa_1 \to 1$, and $w_1^* = v_1 + v_2(\sigma_2/\sigma_1)$. In other words, security one completely takes security two's place in the basket. Alternatively, if security one is infinitely expensive to trade, then $\kappa_1 \to 0$ and its optimal weight is zero. Between the two corner solutions, the optimal weights tilt in favor of holding the asset that is relatively cheaper to trade. The cheaper asset does not completely take over because of the quadratic trading cost.

[12] More general case with arbitrary ρ gives the same qualitative result.

B.4. Short selling revenues

On top of the utility function used in the paper and discussed above in appendix, fund managers may have additional incentives to hold certain assets if they collect the lending fees from offering their shares to short sellers (Blocher & Whaley, 2016). Short borrow fees will shift the optimal holdings, but has no impact on the other directional predictions. To see this, consider the modified utility function:

$$U = \boldsymbol{C(w)} - \boldsymbol{S'w} + \lambda(\boldsymbol{w} - \boldsymbol{v})'\Sigma(\boldsymbol{w} - \boldsymbol{v})$$

where $\boldsymbol{S}$ is the expected short borrow fee per asset, which offsets the expected trading costs. Solving for asset i, we have:

$$w_i^* = \frac{1}{1 + c_i/\lambda\sigma_i^2} v_i + \frac{1}{1 + c_i/\lambda\sigma_i^2} \sum_{j \neq i} (v_j - w_j)\beta_{j,i} + \frac{1}{c_i + \lambda\sigma_i^2} s_i$$

Notice that the optimal weights are higher than before, but the comparative statics with respect to $c_i, \rho_{i,j}$ and σ_i^2 are unchanged.

C. Numerical example

We standardize the distributions of the asset payoffs and the noise flow, $\sigma = \sigma_f = 1$. The mass of noise traders who trade asset A and B are set to be 100 and 70, $N_A = 100, N_B = 70$. Both assets have 1% index weight, $v_A = v_B = 0.01$, and their payoffs have zero correlation with the rest of the index. [13] The mass of noise traders who previously homemake index are 200, $N_I = 250$ and their preference parameter over cost-error trade-off is 10, $\lambda = 3$. Note that except for asset A's noise mass has to be greater than that of asset B, which makes asset A more liquid, all values can be chosen arbitrarily, and the results are robust to the choice of numerical values of those parameters.

For the period before the introduction of the ETF, plugging in the numbers and solve for 2 and 4, we have:

$$c_A = 0.0098,\ c_B = 0.0140, \text{ and } w_A = 0.0067,\ w_B = 0.0053$$

which characterize the market structure before the ETF is introduced. When ETF is introduced, its cost-error preference is endogenously determined as $\lambda^{ETF} = \sqrt{\pi/2}\lambda = 12.53$. The netting effect is captured by the square root in the denominator in 5, and the demand effect is captured by N_I', which is exogenously determined. We leave it open and plug in different values to repeatedly solve for the system of equations of (2) and (5). A1 plots the equilibrium solution given different level of new mass being drawn in.

As shown in A1, the introduction of ETF strictly increases the optimal weights for both assets. This initial jump is due to the fact that ETF has a higher preference for low tracking error compared to homemade indexer ($\lambda^{ETF} = \sqrt{\pi/2}\lambda$). As more noise traders trading ETF, the ETF underweights both assets less. This is because the variance of the aggregate end-of-period flow goes up, which increases the observed noise volatility by the market maker and decreases the spread she posts.

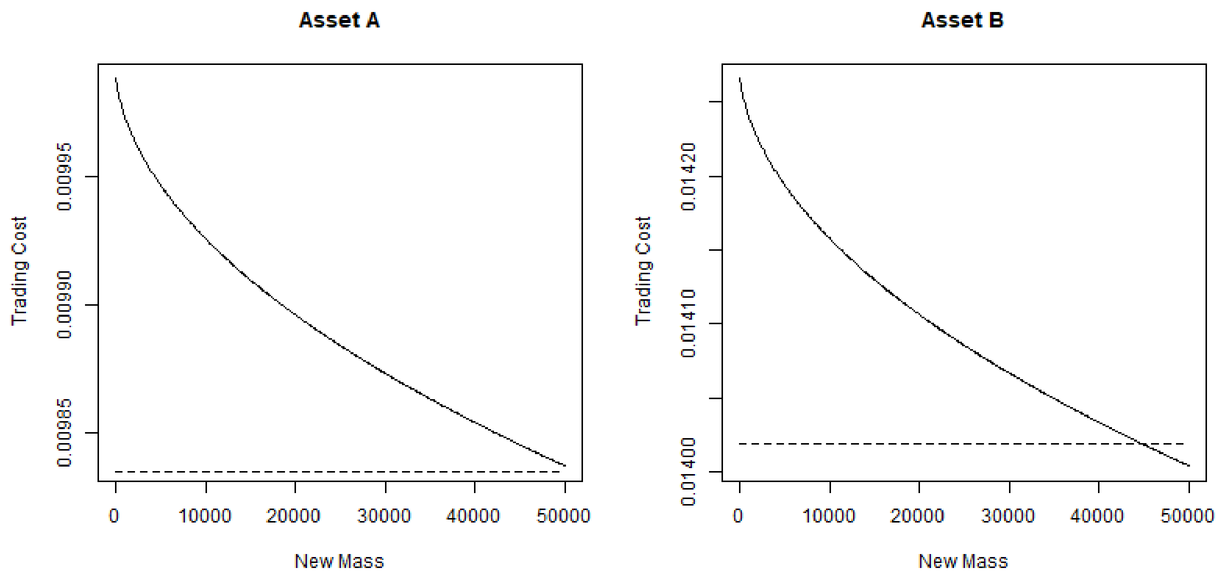
[13] This is a simplifying assumption and forces the optimal basket to always underweight assets as there is no tracking-error mitigating benefit from holding at-full or even overweighting. Relaxing this assumption will produce optimal weights that are fully flexible. Therefore, in this numerical example, if an asset is underweighted relatively less (more), it should be interpreted as being overweighted (underweighted) in a more generalized setting.

Trading costs for both assets initially go up with the introduction of the ETF as the pure netting effect is strictly liquidity-impairing. With the demand effect growing stronger, the trading costs decreases as it is strictly liquidity-improving. When the new noise traders pass a critical mass, the demand effect dominates the netting effect and ETF becomes overall liquidity-improving. However, it's worth noting that it takes roughly 39100 new noise traders for asset A and roughly 38700 for asset B for demand effect to start to dominate, while the initial noise trader mass is only 200. This suggests that ETF is likely overall liquidity-impairing as the critical mass is too high to achieve.

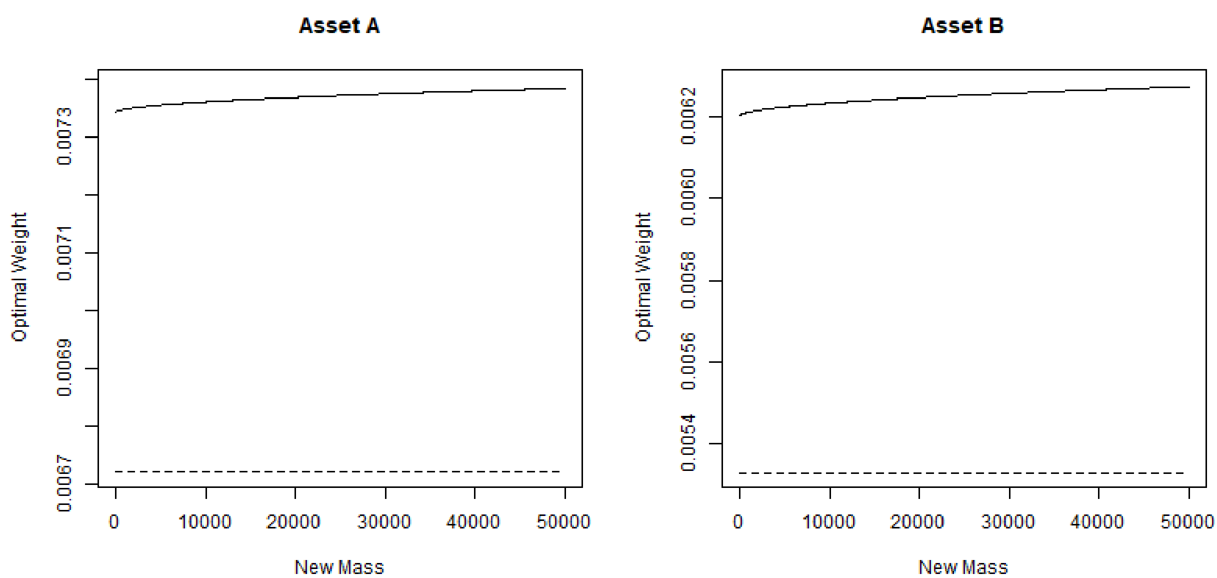
However, as demonstrated before, even though ETF worsens the liquidity unless the new noise mass is sufficiently high, noise traders who trade ETF is strictly better off because they now share the trading cost incurred by the one aggregate flow as opposed to bearing the cost incurred by their own index-homemaking activity. A2 captures the welfare implications for different types of investors. The welfare here is measured as the trading cost before ETF introduction minus that after (such that decreasing trading cost is welfare-improving).

In addition, we can alternatively plot the x-axis as $E[|\text{Primary Flow}|]$ instead of the new mass N_I', as N_I' is a sufficient statistic that determines $E[|\text{Primary Flow}|]$:

$$E[|\text{Primary Flow}|] = \sqrt{N_I + N_I'}\sqrt{\frac{2}{\pi}}$$



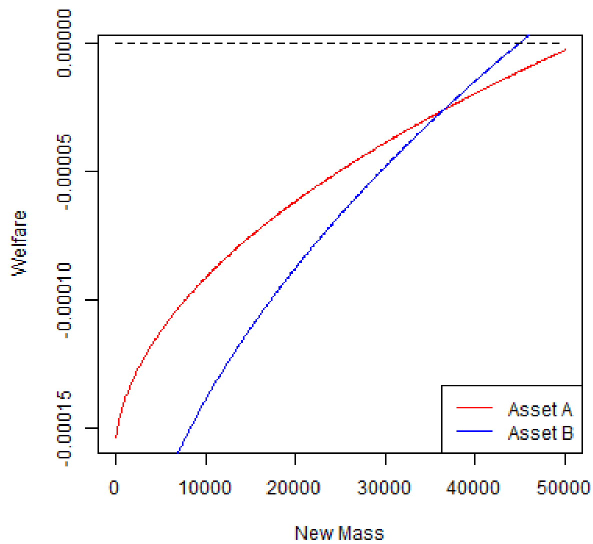
(a) Trading cost



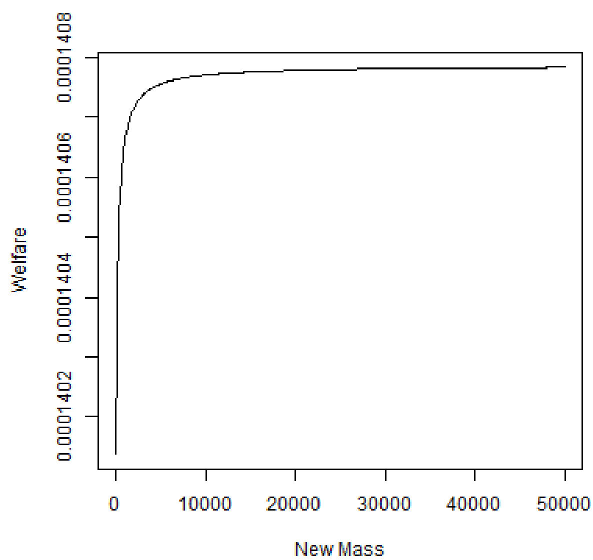
(b) Optimal weight

Figure A1. Asset trading cost and optimal weight pre- and post-ETF

Panel A plots the trading cost and panel B plots the optimal basket weight of both assets. The solid line represents the post-ETF value and the dashed line represents the pre-ETF value.



(a) Underlying assets



(b) ETF

Figure A2. Welfare implication for different traders

The figure plots the difference in trading costs pre- and post-ETF as a measure of investor's welfare. Panel A plots the welfare of investors in the underlying assets market and panel B plots the welfare of investors who trade the ETF.