Internet Appendix to:
The Value of ETF Liquidity

August 18, 2021

Appendix A. Data cleaning procedures

We apply the following minor filters in the sample selection (in addition to those described in the data section of the main paper) :

- Two ETFs (SPLG, SPSM) are removed as they changed the index they track. Prior to November 16, 2017, SPLG tracked the Russell 1000 Index and traded under the ticker ONEK. Between November 16, 2017 and January 24, 2020 the SPLG fund tracked SSgA Large Cap Index. Prior to November 16, 2017, the SPSM tracked the Russell 2000 Index and traded under the ticker TWOK. Between November 16, 2017 and January 24, 2020 the SPSM fund tracked the SSGA Small Cap Index.

- We manually check that several ETFs are recorded twice in ETF Global database. We remove the instances of those ETFs that have incomplete or inaccurate data, and leave the ones with data entries (e.g., MER) that align with alternative data sources such as ETF.com. These ETFs are IVV, IJH, IJR.

- We remove several erroneous ETF comps matches that refer to ETFs that track fundamentally opposite versions (e.g., long vs short) of the same index. These ETFs are SCIJ, SCIX.

- We remove two ETFs that are erroneously matched with their group although they track a different index than the rest of ETFs in their group (S&P Mid Cap 400 Pure Growth vs S&P Mid Cap 400 Growth). These ETFs are RFG and RFV.

- We remove ETFs by the same issuer that have non-overlapping portfolios (for example Gold Miners large cap vs Junior Gold Miners small cap portfolios). These ETFs are GDX and GDXJ.

We apply the following data cleaning procedures in computing the variables used in the main regression analysis:

- If daily relative spread is less than or equal to 0, it is replaced with a missing value;

- If daily relative spread is above 10,000 bps, it is replaced with missing value;

- For each ETF, the turnover and tracking difference variables are winsorized at the 99% level;

- For 37 ETFs for which Factset does not report underlying index values, the tracking error and performance drag in the 2020 cross-section are replaced with the average value of tracking error and performance drag in the 100 remaining ETFs. In panel data, the same procedure is performed in each year-month.

Appendix B. Matching ETF characteristics in a given index

In our analysis, competing ETFs are assumed to be the same or similar with respect to structure. In the table below, we outline the strategies to ensure ETFs in a given index are effectively identical: (i) controlling for the observed ETF characteristics in regressions, and (ii) matching the same-index ETFs by observed characteristics.

Table B1: Strategies to ensure ETFs within a given index are very similar

This table provides a list of observed ETF characteristics and how we ensure those characteristics are either controlled for in regressions or matched to be the same for same-index ETFs through the sample selection. There are two types of ETFs in our sample: open-end investment funds and unit investment trusts (UITs). We introduce the UIT dummy variable to control for the UIT structure. There are only two ETFs in the sample (MDY and SPY) that are UITs. Being a UIT, these ETFs reinvest their dividends not daily, but quarterly. They also cannot lend out securities.

ETF characteristic	Control variable in regressions	Set to identical in sample selection
Currency of returns	no	yes
ETN structure	no	yes
Index exposure	yes	yes
Investment style	no	yes
Legal structure	yes	no
Performance drag	yes	no
Tax rate on capital gains	no	yes
Tax on distributions	no	yes
Tracking error	yes	no

Appendix C. Relation between ETF fees and liquidity for ETFs with identical benchmarks

This Appendix reports additional robustness tests using only ETFs that track exactly the same benchmark as at least one other ETF. The panel regressions in this table are akin to those in Table 6 of the main paper. We use panel data to increase the number of observations and the statistical power of our tests given we use a smaller sample here.

Table C1: Relation between ETF fees and liquidity in the panel of ETFs

This table reports results of panel regressions using the 61 ETFs that have the same benchmark as at least one other ETF. The dependent variable is the ETF's fee (MER or net expense ratio) in basis points. The independent variables are listed in the first column. All variables are calculated from daily data as an average per ETF-month. The period covered is 2016 – 2020. Relative Spread (in basis points) is the bid-ask spread in the ETF secondary market divided by the ETF's midpoint price. Log Dollar Volume is the log of daily ETF traded dollar volume in the secondary market. Turnover (as a fraction) is the annualized ratio of daily dollar volume in the ETF's secondary market and assets under management (AUM). Market Share (as a fraction) is the ETF's AUM divided by total AUM of all ETFs tracking the given index. Tracking Error (in basis points) is the standard deviation of daily differences in returns of the ETF and its underlying index. Performance Drag (in basis points) is the average daily difference in returns of the ETF and its underlying index. D_{UIT} is a dummy variable for ETFs that are structured as unit investment trusts. Standard errors are double clustered by year-month and by index. T-statistics are reported in parentheses. ***,**, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	MER (1)	MER (2)	MER (3)	MER (4)	MER (5)	MER (6)	MER (7)	MER (8)
$Intercept$	12.60***	9.88***	-43.77***	-39.44***	1.82***	1.63***	4.04***	2.52***
	(25.23)	(18.89)	(-33.95)	(-30.90)	(3.36)	(2.82)	(11.72)	(5.79)
$Relative Spread$	-0.49***	-0.43***						
	(-7.65)	(-7.78)						
$Log Dollar Volume$			2.87***	2.54***				
			(43.42)	(38.48)				
$Turnover$					0.61***	0.46***		
					(24.90)	(19.97)		
$Market Share$							13.56***	11.90***
							(29.74)	(28.85)
$Tracking Error$		0.17***		0.13***		0.16***		0.16***
		(19.95)		(20.26)		(19.86)		(22.02)
$Performance Drag$		0.12***		0.09***		0.12***		0.12***
		(4.94)		(5.10)		(4.39)		(5.34)
D_{UIT}		8.82***		1.67***		3.07***		7.77***
		(25.55)		(3.35)		(3.73)		(13.85)
Num obs-ns	3,807	3,807	3,807	3,807	3,807	3,807	3,807	3,807
Adjusted R^2	78%	81%	83%	84%	77%	79%	79%	82%
Fixed effects	Index	Index	Index	Index	Index	Index	Index	Index