Online Appendix

Intraday Arbitrage Between ETFs and Their Underlying Portfolios

TRAVIS BOX, RYAN DAVIS, RICHARD EVANS, and ANDREW LYNCH*

* *Travis Box is an Assistant Professor of Finance at Clemson University, Ryan Davis is an Assistant Professor of Finance in the Collat School of Business at the University of Alabama at Birmingham, Richard Evans is an Associate Professor of Business Administration and the Donald McLean Wilkinson Research Chair in the Darden School of Business at the University of Virginia, and Andrew Lynch is an Assistant Professor of Finance at the University of Arkansas. Send correspondence to Travis Box at Clemson University: tsbox@clemson.edu, Address: 730 Greenville One, 1 North Main St. Greenville, SC 29601.*

A. Replicating ETF portfolios

The CRSP Survivor-Bias-Free US Mutual Fund Database provides a list of historical holdings and total net asset values for a number of publicly traded open-end mutual funds. For our purposes, however, the database also reports the monthly holdings for 2,007 unique ETFs between 2006 and 2015, 1,459 of which are passively managed index funds.[1] From summary statistics reported in Box, Davis and Fuller (2019), these funds should approximate the universe of all U.S.-listed ETFs during this time period. Our goal is to estimate the intraday intrinsic value of an ETF based on how many shares of each security a fund holds multiplied by the national best bid and offer prices for those securities reported by the Trades and Quotes (TAQ) Database. Errors in the replication of these constituent portfolios would make it difficult to identify intraday arbitrage opportunities. Therefore, we verify the accuracy of each holding reported in the Mutual Fund Database by corroborating the value of that position using another source.

The Mutual Fund Database summarizes constituent value in three ways: the dollar value, the percent market value and the number of shares held in each position. Fund holdings are self-reported by each ETF sponsor. However, the values implied by these three measures are often inconsistent. For instance, the percentage market values often sum to a number other than 100%, and the sum of all holdings' market values rarely equals the total net asset value reported by the fund.

To first deal with inconsistencies in reported total net asset values, we calculate the implied total net asset value, $iTNAV_{fh}$, of each fund f based on the percent market value, $\%Val_{fh}$, of a position, h, and its reported dollar value, $DolVal_{fh}$:

[1] ETFs have an ETF flag recorded as "F," and pure index funds have an Index Fund flag equal to "D."

$$iTNAV_{fh} = \frac{DolVal_{fh}}{\%Val_{fh}}. \tag{A-1}$$

If an ETF consists of 300 holdings, Equation (A-1) gives us 300 estimates of the fund's total net asset value. To minimize the impact of outliers on our final estimate, we choose the median $iTNAV_{fh}$ to represent the actual total net asset value of the fund.

The majority of funds hold at least some foreign assets, derivatives or fixed income securities. To generate a sample of funds that hold only U.S. equities, we attempt to find matches for each fund's reported holdings in the CRSP Daily Stock File.[2] The Mutual Fund Database provides five identifiers, PERMNO, CUSIP, ticker, PERMCO and security name, for all ETF constituents. To ensure our sample is as accurate as possible, we merge the Mutual Fund Database and the Daily Stock File independently on all five identifiers.[3]

It is not uncommon that a single holding matches up with multiple equity securities depending on which identifier is used, so we verify that the reported value of each holding is consistent with the closing stock price recorded in the Daily Stock File. Our value comparisons are based on two different combinations of the Mutual Fund Database and the Daily Stock File. First, we calculate an estimate of each position's dollar value based on the sponsor-reported number of shares held and the closing stock price of each potential match from the Daily Stock File. Next, we estimate each position's percentage value by dividing our estimated dollar value by the fund's median $iTNAV_{fh}$. For matches based on each of the five identifiers, we compare our estimates of dollar value and percent market value with those reported by the fund sponsors. We remove all matches

[2] From our sample of ETF holdings, we also remove any observations where the reported coupon rate is greater than 0 or the maturity date is not missing. This filter should remove most fixed income securities.
[3] ETFs often hold small amounts of cash between dividend distributions, so we retain all positions in the Mutual Fund Database with the security name "USD CASH," and we assume that those securities have a closing price of $1 on the close of each reporting date.

where either estimate of value differs from its corresponding reported value by more than 10%. If any individual holdings are still matched with multiple securities in the Daily Stock File, we choose the one whose estimated dollar value is closest to what was reported by the fund.[4]

Having removed all of the positions in foreign assets, derivatives and fixed income securities, we next compare the combined value of our matched holdings to the total net asset value of the fund. Once again, we rely on each fund's median $iTNAV_{fh}$ and the percentage market values calculated from Daily Stock File closing prices. Specifically, we require that the sum of estimated percent market values falls somewhere between 97% and 103% of $iTNAV_{fh}$ during each reporting date. As a final check on our portfolio replication, we compare the total number of matched holdings to the number of holdings originally reported in the Mutual Fund Database. Here, we only retain portfolios whose matched holdings account for at least 97% of the total holdings reported. The filtering process described above leaves us with 423 passively managed U.S. equity funds that have replicable holdings for at least one reporting date during our sample period.

The monthly reporting frequency of the Mutual Fund Database prevents us from knowing the size of each position during non-reporting trading sessions. Furthermore, we observe significant variation in the absolute size of ETF holdings from one month to the next as ETF shares are created and redeemed. Fortunately, constraining the sample to passively managed funds ensures that the relative allocations between each position do not change from one day to the next as long as index membership is held constant. The stability of portfolio composition allows us to interpolate the number of shares held in each security throughout the month.

[4] It is not uncommon for funds to report multiple positions in the same security. In these cases, we only retain the position with the highest reported market value.

The Mutual Fund Database provides closing daily net asset values for each ETF, f, in our sample. The sponsor calculates this measure by dividing the fund's total net asset value by the number of ETF shares outstanding, $ShrOut_f$:

$$DNAV_f = \frac{\sum_h Price_h \times Shares_{fh}}{ShrOut_f}, \tag{A-2}$$

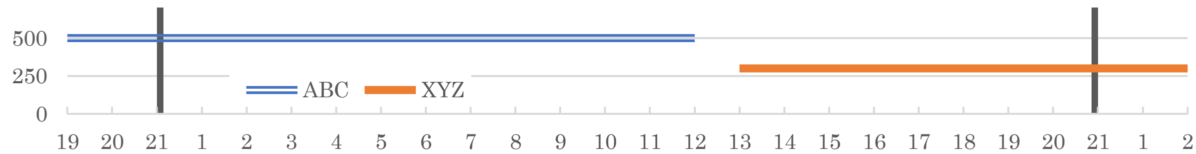
where $Price_h$ is the daily closing price of each position, h, and $Shares_{fh}$ is the size of each position within the fund. To verify the accuracy of share interpolation, we create our own proxy for daily net asset value. Each time that ETF shares are created or redeemed, the denominator on the right side of Equation (A-2) adjusts accordingly. The total net asset value, the numerator of Equation (A-2), also changes when the number of ETF shares outstanding changes between reporting dates. However, our estimates of total net asset value are based on share counts for individual positions that are fixed between reporting dates, which may not line up with the dates of ETF share creation or redemption.

Our solution to this problem is based on the observation that, while market weightings for individual portfolio holdings change constantly, the share weights of those holdings never change within a trading day. With this in mind, we multiply both sides of the Equation (A-2) by the ratio of ETF shares outstanding and the sum of constituent shares:

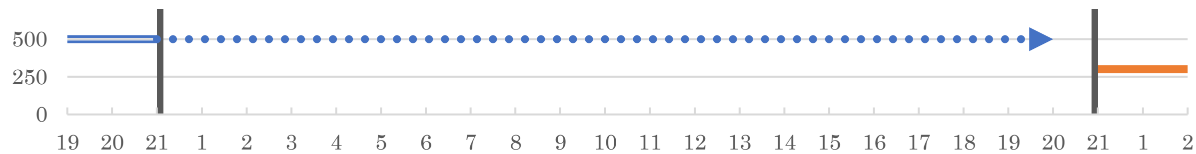
$$DNAV_f \times \frac{ShrOut_f}{\sum_h Shares_{fh}} = \sum_h \frac{Price_h \times Shares_{fh}}{\sum_h Shares_{fh}}. \tag{A-3}$$

For an ETF whose relative composition is constant between monthly reporting dates, the relationship between $ShrOut_f$ and $\sum_p Shares_{fh}$ will also remain constant across trading days. Therefore, changes in daily net asset value, $DNAV_f$, should be perfectly correlated with changes in the share-weighted price of underlying holdings, $\sum_h \left[\left(Price_h \times Shares_{fh} \right) / \sum_h Shares_{fh} \right]$.

Panel A: Actual share count



Panel B: Share count from most recent holdings report



Panel C: Share count that maximizes correlation between daily net asset value and share-weighted price

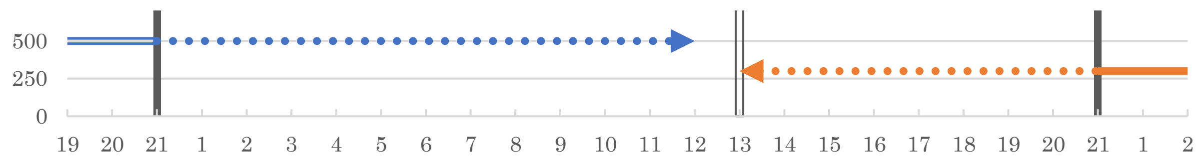


Figure A-1. Share count interpolation example

Even though changes in the size of an ETF portfolio preserve the correlation between daily net asset value and share-weighted price, adjustments to the relative composition of an ETF portfolio can spoil this relation. While changes in composition are infrequent for passively managed funds, their holdings must be adjusted anytime their underlying benchmarks are reconstituted. In Panel A of Figure A-1 we illustrate a hypothetical ETF portfolio that exchanges 500 shares of stock ABC for 300 shares of XYZ thirteen trading days after they last reported holdings. Panel B describes a common approach to share interpolation in the mutual fund literature (see, e.g., Wermers, 2000; Cremers and Petajisto, 2009), whereby share counts are assumed to remain constant until the following report date. In this case, we should not expect perfect correlation between reported daily net asset value and share-weighted price because the former is based on the holdings in Panel A, whereas the latter is based on the composition assumed in Panel B.

To account for changes in portfolio holdings between reporting dates, we maximize the correlation between daily net asset value, $DNAV_f$, and share-weighted price, $\sum_h \left[\left(Price_h \times Shares_{fh} \right) / \sum_h Shares_{fh} \right]$, by varying the date for which we assume these portfolio changes occurred. Thus, for a month containing twenty-one trading days, we compute twenty-one different month-long series of share-weighted price. If the portfolio's composition changes on day thirteen, as in Figure A-1, the share-weighted price series with day thirteen as the assumed reallocation date will have the highest correlation with daily net asset value.

To calculate correlations, we require the ETFs in our sample to have consecutive months of replicable portfolio holdings. To ensure that our replicated portfolios accurately represent the underlying holdings of each fund, we only retain fund-months where the correlation between reported daily net asset value and estimated share-weighted price is greater than 99.5%. The final sample contains 12,170 fund-months and preserves at least one month of holdings for all 423 ETFs.

B. TAQ filters

To identify arbitrage opportunities, we compare the quoted price of an ETF to the price of its underlying portfolio inside of each trading day. From the portfolio replication and interpolation processes described in Appendix 0, we are given a list of share-weights for 241,810 fund-days during our sample period. We combine these holdings with intraday prices from TAQ to compute the end-of-minute bid and ask quotes for each ETF portfolio.

Most analyses of TAQ data rely on a fairly standard set of filters (see Huang and Stoll, 1996; Bessembinder, 1999; 2003a; 2003b) to mitigate the impact of erroneous quotes and outliers. For our study, however, it is critical that we simulate the experience of a potential arbitrage trader that is interacting with the market in real-time. We worry that excessive filtering could lead to optimistic appraisals of available liquidity. To be sure that an arbitrage trader could, for instance, profitably buy an ETF while selling its underlying portfolio, we retain the original quotes whenever possible and document their removal in extreme cases where filtering becomes necessary.

We begin by removing quotes with conditions 1, 2, 6, 10, 12, and 23. Next, we identify the subset of best bid and ask prices whose quoted spreads, defined as the difference between each price scaled by their midpoint, are greater than zero but less than 50%. From this subset, we calculate the daily median midquote for every ETF and underlying stock that appears in our sample. These median midquotes represent the "typical" trade price for a security within each trading session. To identify quotes that diverge dramatically from typical prices, we flag the bid or ask as missing if the quote exceeds the daily median midquote by more than a factor of five.

For quotes with spreads that are less than zero or greater than 50%, we attempt to document an explanation for the quoted spread's irregularity. For instance, we classify either the bid or ask as missing when no price is recorded in TAQ. When both prices are available, we classify the ask as missing if the quote is at least 25% larger than the previous midquote. Likewise, we flag the bid as missing when the quote is 25% smaller than the previous midquote. Whenever the quoted spread is negative, we record both the bid and ask as missing.

For all quotes that are flagged as missing, the most recent satisfactory quote is recorded in its place. However, the quote is not replaced entirely, unless both the bid and ask prices are classified as missing. Unlike removing these observations entirely, our approach allows for the preservation of a reasonable bid quote in situations where the accompanying ask quote has deviated, and vice versa. If the replacement quote results in a spread that also violates our zero or 50% bounds, we flag both the bid and ask as missing and replace them with the most recent satisfactory observations.

Panel A: Minute-by-minute percentiles for $MisBid_t^{Und}$ and averages for $MisBid_t^{Und}$

Panel B: Minute-by-minute percentiles for $MisAsk_t^{Und}$ and averages for $MisAsk_t^{Und}$

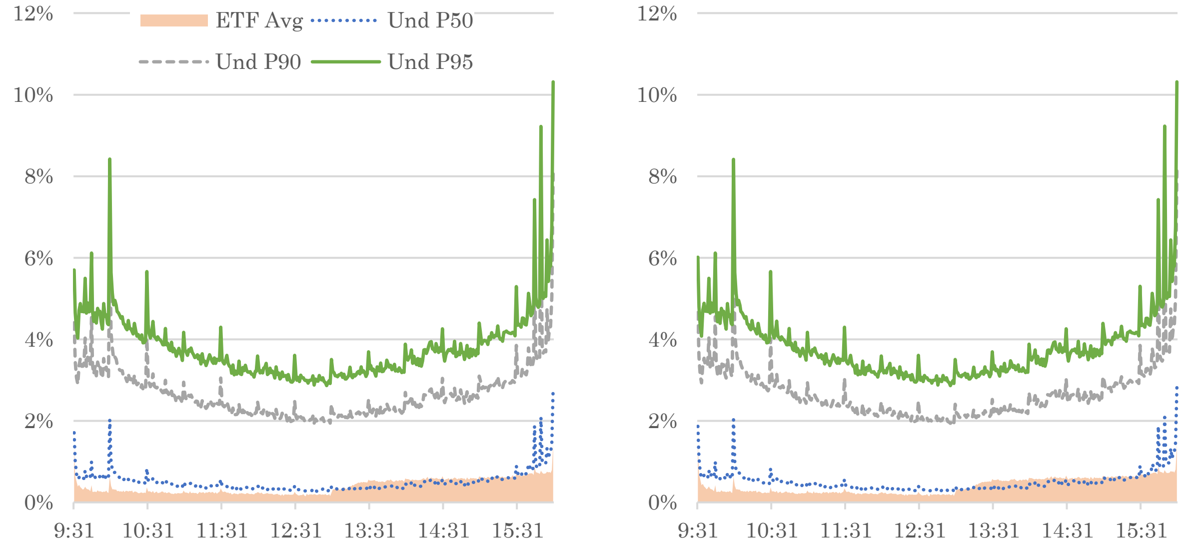


Figure B-1. Percentage of missing NBBO quotes during the trading day

This figure describes the 50th, 90th and 95th percentile of $MisAsk_t^{Und}$ and $MisBid_t^{Und}$, as well as the average of $MisBid_t^{ETF}$ and $MisAsk_t^{ETF}$, during each minute of the trading day. $MisBid_t^{Und}$ ($MisBid_t^{ETF}$) and $MisAsk_t^{Und}$ ($MisAsk_t^{ETF}$) are the weighted average proportion of missing or inferred bid, and ask, quotes in the underlying portfolio (ETF), respectively, for each fund-day and minute t in the sample.

Across our sample period, one-minute best bid and ask quotes are flagged as missing 0.519% and 0.515% of the time, respectively, for our entire sample of ETFs and underlying securities. For observations where either the bid or the ask is missing, 85.89% are missing because their quoted spread is negative or greater than 50% of the quoted midpoint.[5] 12.72% are missing because one of the quotes did not have a recorded price in TAQ. Only 2.48% are flagged because they exceed the daily median midquote by a factor of five.

We calculate the weighted average proportion of missing or inferred bid, $MisBid_t^{Und}$, and ask, $MisAsk_t^{Und}$, quotes in the underlying portfolio for each fund-day and minute t in our sample.

[5] Even allowing for a relatively wide range of acceptable quotes, we still observe a missing bid or ask in approximately 1 out of 100 minutes. To get a sense of how this magnitude might affect our results, consider an ETF with 100 holdings. If the likelihood that quotes are missing is not correlated across securities, an admittedly bold assumption, then the probability that none of the securities had a missing ask, for instance, would be approximately $0.99^{100} = 36.6\%$. Thus, at least one stock in the portfolio would have a missing ask observation $1 - 0.366 = 63.4\%$ of the time. This implies that arbitrage traders would be unable to purchase the entire underlying portfolio most of the time.

Similarly, the binary variables $MisBid_t^{ETF}$ and $MisAsk_t^{ETF}$ equal one whenever the fund's bid and ask prices are either missing or inferred. Panels A and B of Figure B-1 describe the 50th, 90th and 95th percentile of $MisAsk_t^{Und}$ and $MisBid_t^{Und}$, as well as the average of $MisBid_t^{ETF}$ and $MisAsk_t^{ETF}$, during each minute of the trading day. According to Figure B-1, the proportion of missing or inferred quotes is highest at market open and close, rising above 10% for some underlying portfolios by the end of the day.

Typical TAQ filters eliminate all observations where quoted spreads exceed some threshold, resulting in a smooth sequence of prices that are always within certain bounds. In reality, these quotes may or may not represent liquidity that was actually available in the market. Our process of replacing divergent quotes still produces a dataset where average spreads are artificially low and prices are less volatile than those that would be observable in real time. However, flagging these quotes allows us to acknowledge them in our statistical inferences.

C. Intraday quotes

With an accurate record of daily underlying share-weightings for each ETF, along with minute-by-minute bid and ask prices for each holding, we can calculate quoted prices for each portfolio. In most contexts, estimating the value of an ETF portfolio would simply require calculating the closing market-weighted average price for each of the individual constituents. However, variation in intraday prices cause the relative market-weights of these constituents to change constantly throughout the day.

Once again, our solution to this problem is based on the relationship described by Equation (A-3). Even if relative market-weights adjust with quoted prices, variation in a fund's net asset value is perfectly correlated with changes in the share-weighted price of the underlying holdings.

Therefore, the intraday underlying bid, Und_f^{Bid}, and ask, Und_f^{Bid}, price for fund f can be represented as follows:

$$
\begin{aligned}
Und_f^{Bid} \times \frac{ShrOut_f}{\sum_h Shares_{fh}} &= \sum_h \frac{Bid_h \times Shares_{fh}}{\sum_h Shares_{fh}} \\
Und_f^{Ask} \times \frac{ShrOut_f}{\sum_h Shares_{fh}} &= \sum_h \frac{Ask_h \times Shares_{fh}}{\sum_h Shares_{fh}},
\end{aligned}
\tag{A-4}
$$

where Bid_h and Ask_h are the quoted bid and ask prices of each position h. However, to compare Und_f^{Bid} and Und_f^{Ask} with quoted ETF prices, we also need the ratio of fund shares outstanding to the sum of constituent shares, $ShrOut_f / \sum_h Shares_{fh}$.

The ratio of shares outstanding to the sum of constituent shares is difficult to estimate because of intramonth variations in $\sum_h Shares_{fh}$. Here, we rely on the equilibrium relationship between an ETF's quoted price and its intrinsic value. In the absence of tradeable arbitrage opportunities, all three of the following conditions should hold:

$$
\begin{aligned}
&1.\quad ETF_f^{Ask} \geq Adj_{Max} \times \sum_h \frac{Bid_h \times Shares_{fh}}{\sum_h Shares_{fh}} \\
&2.\quad ETF_f^{Bid} \leq Adj_{Min} \times \sum_h \frac{Ask_h \times Shares_{fh}}{\sum_h Shares_{fh}} \\
&3.\quad \frac{ETF_f^{Bid}}{\sum_h \frac{Ask_h \times Shares_{fh}}{\sum_h Shares_{fh}}} = Adj_{Min} \leq \left(\frac{ShrOut_f}{\sum_h Shares_{fh}}\right)^{-1} \leq Adj_{Max} = \frac{ETF_f^{Ask}}{\sum_h \frac{Bid_h \times Shares_{fh}}{\sum_h Shares_{fh}}},
\end{aligned}
\tag{A-5}
$$

where Adj_{Max} and Adj_{Min} represent the upper and lower bounds for the inverse of $ShrOut_f / \sum_h Shares_{fh}$.

At the end of every fund-day-minute in our sample, we calculate the midpoint between both bounds. Next, we find the median daily midpoint, Adj_{Med}, for each fund. Finally, we multiply this median midpoint by the intraday share-weighted bid and ask prices to generate estimates of Und_f^{Bid} and Und_f^{Ask}. To ensure that underlying quotes are comparable to ETF bids and asks, Und_f^{Bid} and Und_f^{Ask} are also rounded to the nearest penny. The daily median midpoint, Adj_{Med}, provides an adjustment factor for share-weighted price that minimizes the frequency of perceived arbitrage opportunities. Thus, our approach makes us more likely to underestimate the total number of tradeable price discrepancies in our sample.

We also attempt to estimate Und_f^{Bid} and Und_f^{Ask} by comparing the market closing prices of each underlying security to the daily net asset value, $DNAV_f$, reported by the fund sponsor. Rearranging Equation (A-3), gives:

$$\frac{ShrOut_f}{\sum_h Shares_{fh}} = \frac{\sum_h Price_h \times \frac{Shares_{fh}}{\sum_h Shares_{fh}}}{DNAV_f}. \qquad \text{(A-5)}$$

All of the inputs on the right-hand side remain fixed within each trading session because creations and redemptions only occur after the market has closed. Therefore, the ratio of $DNAV_f$ to share-weighted closing price can be used to calculate Und_f^{Bid} and Und_f^{Ask} from share-weighted bid and ask prices.

Unfortunately, the procedure for calculating $DNAV_f$ can vary across funds. For instance, some funds may choose the prevailing market price at 4:00pm, while others base security prices on the closing auction. Due to narrow spreads in the ETF and underlying portfolios, even small errors in the estimation of underlying bid and ask prices can lead to a dramatic increase in the misclassification of arbitrage opportunities. When $DNAV_f$ is used to calculate Und_f^{Bid} and Und_f^{Ask}, we classify 13.15% of fund-day-minutes as tradeable arbitrage opportunities. When median midpoints, Adj_{Med}, are used instead, we only recognize mispricing in 1.66% of our observations.

D. Raw return and volume results

As discussed in Section 2, we symmetrically log returns and order imbalances for our intraday analyses. We assess the impact of this transformation by repeating our PVAR analyses using untransformed variables below. Figure D-1 shows the impulse response function using our 5-minute intraday sample with raw, untransformed returns, $RetRaw^{Und}$ and $RetRaw^{ETF}$, and order imbalances, $VolDiffRaw^{Und}$ and $VolDiffRaw^{ETF}$, analogous to Figure 1. Table D-1 shows the forecast error variance decomposition results, analogous to Table II. Comparing the FEVD results to Table II and the IRF results to Figure 1, the basic result that ETF returns and order flows have no statistically or economically significant effect on underlying returns is confirmed. However, consistent with the impact of outliers, the results appear to have less precision. In the IRF, for example, the impact of underlying raw return and underlying raw order flow on raw ETF returns is never statistically different from zero. Similarly, the IRF of raw ETF return on all four response variables is never statistically different from zero.

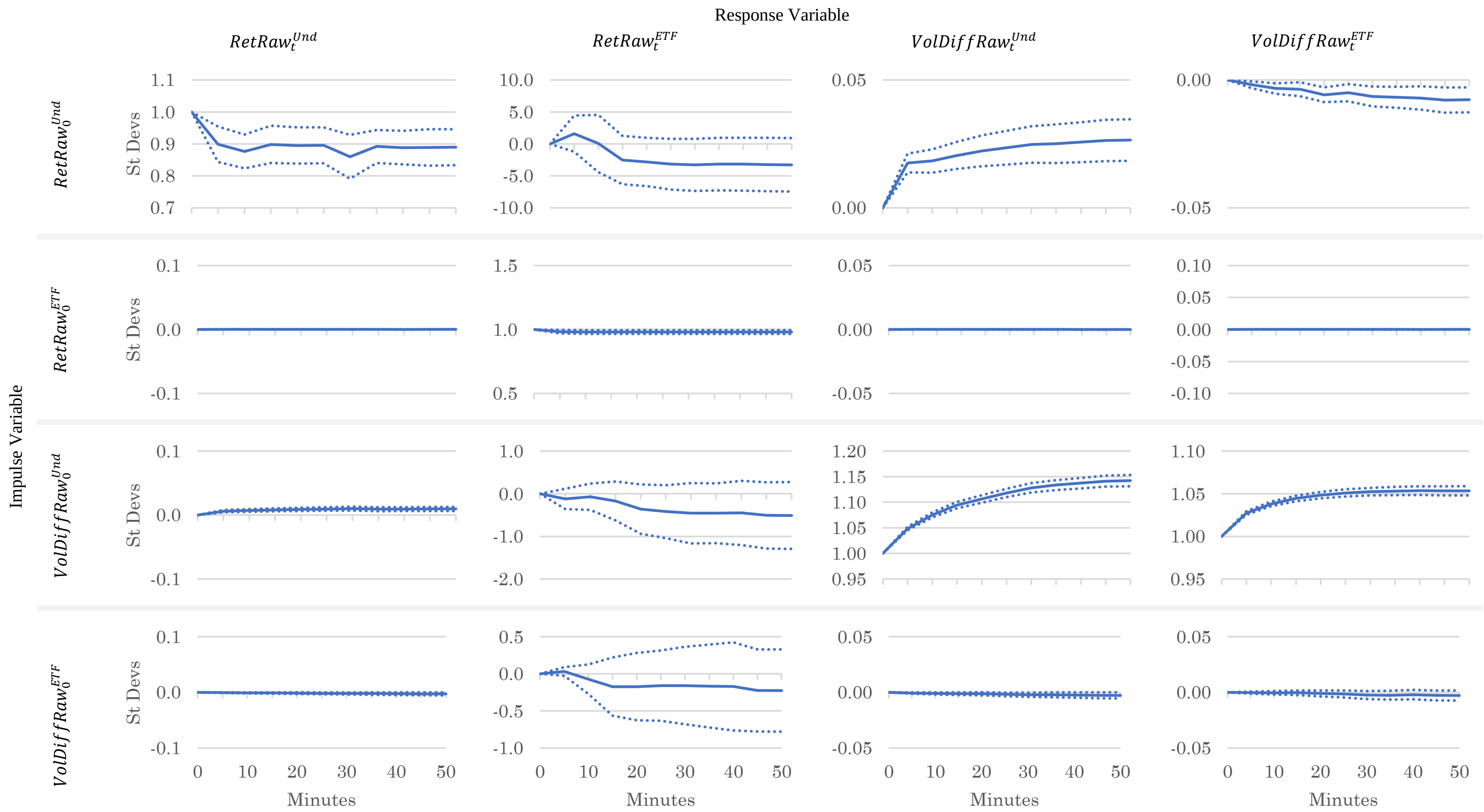


Figure D-1. 5-minute raw cumulative impulse response functions

This figure depicts the cumulative impulse response functions (IRF) derived from the estimated parameters from Equation (4). These IRFs describe each dependent variable's evolution following a one standard deviation shock in the associated impulse variable. The four variables included are $RetRaw_t^{Und}$ and $RetRaw_t^{ETF}$ and $VolDiffRaw_t^{Und}$ and $VolDiffRaw_t^{ETF}$ measured over 5-minute intervals. Confidence intervals, 97.5% and 2.5%, are denoted by dotted lines.

Table D-1
5-Minute Raw Intraday Panel Vector Autoregression Forecast-Error Variance Decomposition

This table presents the fraction of forecasted error variance explained by exogenous shocks to impulse variables after 10 periods. The four variables included are $RetRaw_t^{Und}$ and $RetRaw_t^{ETF}$ and $VolDiffRaw_t^{Und}$ and $VolDiffRaw_t^{ETF}$, measured over 5-minute intervals. Shocks are orthogonalized from top to bottom in the order presented.

Panel B: 5-Minute Window

		Response Variable			
		$RetRaw_{10}^{Und}$	$RetRaw_{10}^{ETF}$	$VolDiffRaw_{10}^{Und}$	$VolDiffRaw_{10}^{ETF}$
Impulse Variable	$RetRaw_0^{Und}$	99.993%	0.000%	0.007%	0.000%
	$RetRaw_0^{ETF}$	0.009%	99.991%	0.000%	0.000%
	$VolDiffRaw_0^{Und}$	0.156%	0.000%	99.844%	0.000%
	$VolDiffRaw_0^{ETF}$	0.001%	0.000%	0.000%	99.999%
	Observations	15,309,936			

E. Correlated alternatives to trading in constituent shares

Studies which have suggested that ETFs might introduce nonfundamental shocks into the prices of underlying securities have presumed that such noise is transferred though an arbitrage mechanism (see Broman and Shum, 2018; Ben-David, Franzoni and Moussawi, 2018; Charupat and Miu, 2011; Israeli, Lee and Sridharan, 2017; Malamud, 2016). Following a spurious shift in demand for ETF shares, arbitrage traders would take opposing positions in the fund and its portfolio, thereby pushing constituent prices away from fundamental values. However, instead of trading directly in the fund's holdings, arbitrageurs could take an opposing position in some other highly correlated security.

One obvious possibility would be to trade in the shares of another ETF that tracks a similar benchmark. Similar to our strategy for classifying funds according to their relative bid-ask spreads, we also group ETFs by their degree of comovement with other funds. For each day, we calculate the minute-by-minute intraday return correlation between all funds in our sample over the prior 200 trading days. Next, we determine which other ETF has comoved most closely to each fund and divide our 5-minute sample into quintiles based on the intensity of that highest correlation.

Following a nonfundamental price shock, opposing positions between an ETF and its best match are likely to converge quickly if the pair's comovement is high. Thus, arbitrageurs might be less inclined to trade in the underlying portfolio, and transfer noise into constituent prices, when a highly correlated alternative is available.

We tabulate the results from FEVD analysis in Table E-1 after estimating the parameters of Equation (4) independently for each correlation quintile. Just as in the previous analyses, we are unable to isolate a subset of ETFs where shocks to the demand or pricing of a fund affect its holdings after the shock. Even for ETFs with the most uncorrelated alternatives, the contributions of Ret_t^{ETF} and $VolDiff_t^{ETF}$ to the total forecast-error variance of portfolio returns are 0.01% and 0.00%, respectively. Much like our FEVD analysis of liquidity subsamples, we find considerable variation in the share of ETF returns' forecast-error variance that is attributable to underlying price changes. Funds without a strongly correlated alternative to trading directly in the underlying securities respond less efficiently to innovations in portfolio values. Yet, the absence of suitable alternatives to direct arbitrage still does not seem to encourage the transmission of price shocks from the ETF to its holdings.[6]

[6] While locating a strongly correlated alternative is necessary for a successful arbitrage strategy, realizing a profit requires that round-trip transaction costs do not exceed the magnitude of mispricing. In untabulated results, we augment our system of equations with the returns and order imbalances from portfolios consisting of, up to, three ETF alternatives that offer lower transaction costs and higher correlation with a fund than its own holdings. Ultimately, we find no evidence that nonfundamental demand shocks from the portfolio of correlated alternatives are associated with subsequent innovations in constituent prices.

Table E-1
Intraday 5-Minute Panel Vector Autoregression Forecast-Error Variance Decomposition Across ETF Correlation Match Quintiles

This table presents the fraction of forecasted error variance explained by exogenous shocks to impulse variables after 10 periods. The four variables included are $Ret_t^{Und}$and Ret_t^{ETF}, as defined in Equation (3) measured over 5-minute intervals. Shocks are orthogonalized from top to bottom in the order presented. The PVAR specification is estimated separately on five subsamples divided daily on the highest minute-by-minute return correlation a fund has with any other fund over the prior 200 days.

		ETF Correlation Match Quintile				
		Lowest Correlation ⟵		⟶ Highest Correlation		
Impulse Variable	Response Variable	1	2	3	4	5
Ret_0^{Und}	Ret_{10}^{Und}	99.99%	99.87%	99.77%	99.91%	99.93%
	Ret_{10}^{ETF}	36.73%	46.34%	56.34%	50.94%	55.15%
	$VolDiff_{10}^{Und}$	0.53%	0.55%	0.48%	0.52%	0.72%
	$VolDiff_{10}^{ETF}$	0.01%	0.00%	0.01%	0.01%	0.01%
Ret_0^{ETF}	Ret_{10}^{Und}	0.01%	0.13%	0.23%	0.09%	0.07%
	Ret_{10}^{ETF}	63.27%	53.66%	43.65%	49.05%	44.85%
	$VolDiff_{10}^{Und}$	0.00%	0.00%	0.00%	0.00%	0.01%
	$VolDiff_{10}^{ETF}$	0.01%	0.01%	0.00%	0.00%	0.00%
$VolDiff_0^{Und}$	Ret_{10}^{Und}	0.00%	0.00%	0.00%	0.00%	0.00%
	Ret_{10}^{ETF}	0.00%	0.00%	0.00%	0.00%	0.00%
	$VolDiff_{10}^{Und}$	99.47%	99.45%	99.52%	99.47%	99.27%
	$VolDiff_{10}^{ETF}$	0.00%	0.00%	0.00%	0.00%	0.00%
$VolDiff_0^{ETF}$	Ret_{10}^{Und}	0.00%	0.00%	0.00%	0.00%	0.00%
	Ret_{10}^{ETF}	0.00%	0.00%	0.00%	0.00%	0.00%
	$VolDiff_{10}^{Und}$	0.00%	0.00%	0.00%	0.00%	0.00%
	$VolDiff_{10}^{ETF}$	99.98%	99.99%	99.99%	99.99%	99.99%

F. Stochastic jump identification

A discontinuity in a Gaussian process is commonly referred to as a stochastic jump. As observable returns are discrete, the identification of a return discontinuity is non-trivial, requiring a probabilistic estimate that the observed return is inconsistent with a diffusion process. Lee and Mykland (2008) propose a jump detection measure which compares a return at time t to the bipower variation in returns over the prior k time periods. The use of bipower variation in jump detection is common across multiple techniques (see Barndorff-Nielsen and Shephard, 2006). More formally, their jump detection measure for a return series is defined as:

$$L_t = \frac{log\left({}^{S_t}/_{S_{t-1}}\right)}{\widehat{\sigma_t}}, \tag{A-6}$$

where

$$\widehat{\sigma_t}^2 = \left(\frac{1}{k-2}\right) \sum_{i=t-k+2}^{t-1} \left|log\left({}^{S_i}/_{S_{i-1}}\right)\right| \left|log\left({}^{S_{i-1}}/_{S_{i-2}}\right)\right|, \tag{A-7}$$

Lee and Myland (2008) demonstrate that in the presence of Gaussian diffusion process the distribution of $\widehat{L_t}$ is normal with an expected value conditional on the frequency of discrete observation of the series. The presumption of a distribution of this test statistic becomes the null hypothesis over which realized values of L_t can be compared to reject a continuous smooth diffusion process. The derived threshold for null rejection is:

$$\theta = \beta^* S_n + C_n, \tag{A-8}$$

where

$$\beta^* = -\log(-\log(\alpha)), \tag{A-9}$$

and

$$C_n = \frac{\left(2log(n)\right)^{1/2}}{c} - \frac{log(\pi) + log\left(log(n)\right)}{2c\left(2log(n)\right)^{1/2}}, \tag{A-10}$$

and

$$S_n = \frac{1}{c\left(2log(n)\right)^{1/2}}, \tag{A-11}$$

where α is the confidence level for the test, n is the number of observations in the return time series, and c is a constant equal to 0.7979. When $|L_t| > \theta$ the magnitude of the return at time t is considered too large for a diffusion process, implying a stochastic jump in the process. Simulations using a 5 percent confidence level have a combined misclassification (both Type I and II errors) of actual jumps less than 0.01 (0.08) percent using 15-minute (1-hour) returns (the measure becomes more accurate at higher frequencies).

The only points of judgement in the application of this jump detection process are the choice of the confidence level, α, and k, the number of lags of the return time series used to estimate the bipower variation at time t. Lee and Mykland (2008) demonstrate that k needs to be within the range of $\sqrt{252 * \Delta t}$ and $252 * \Delta t$, where Δt is the frequency of return observations each day, though they note that little accuracy is gained using values of k above the minimum required. This implies we should use between 313 and 98,280 prior minutes. Though our results in Section 3 are robust to alternative choices, we utilize an α of 5 percent and a k of 5,850 minutes (the prior 15 days' worth of intraday returns).

References

Barndorff-Nielsen, O., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. Journal of Financial Econometrics 4, 1-30.

Ben-David, I., Franzoni, F., Moussawi, R., 2018. Do ETFs increase volatility? Journal of Finance 73, 2471-2535.

Bessembinder, H., 1999. Trade execution costs on NASDAQ and the NYSE: A post-reform comparison. Journal of Financial and Quantitative Analysis 34, 387-407.

Bessembinder, H., 2003a. Issues in assessing trade execution costs. Journal of Financial Markets 6, 233–257.

Bessembinder, H., 2003b. Quote-based competition and trade exectuion costs in NYSE-listed stocks. Journal of Financial Economics 70, 385-422.

Box, T., Davis, R., Fuller, P., 2019. ETF competition and market quality. Financial Management 48, 873-916.

Broman, M., Shum, P., 2018. Relative liquidity, fund flows and short-term demand: Evidence from exchange-traded funds. Financial Review 53, 87-115.

Charupat, N., Miu, P., 2011. The pricing and performance of leveraged exchange-traded funds. Journal of Banking & Finance 35, 966-977.

Cremers, K., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. The Review of Financial Studies 22, 3329-3365.

Huang, R., Stoll, H., 1996. Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE. Journal of Financial Economics 41, 313-357.

Israeli, D., Lee, C., Sridharan, S., 2017. Is there a dark side to exchange traded funds (ETFs)? An information perspective. Review of Accounting Studies 22, 1048-1083.

Lee, S., Mykland, P., 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. The Review of Financial Studies 21, 2535–2563.

Malamud, S., 2016. A dynamic equilibrium model of ETFs. Unpublished working paper. Swiss Federal Institute of Technology, Lausanne.

Wermers, R., 2000. Mutual fund performance: an empirical decomposition into stock-picking talent, style, transactions costs, and expenses. Journal of Finance 55, 1655-1695.