Intraday Arbitrage Between ETFs and Their Underlying Portfolios

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

ABSTRACT

Prior research suggests that ETF arbitrage affects the market quality of underlying securities. We directly test this proposition by examining minute-by-minute returns and order imbalances, but find little evidence that trading in ETFs impacts the underlying. Panel vector autoregression shows ETF returns largely follow the underlying returns. We also find that mispricing events are preceded by underlying price and order imbalance shocks, corrected by ETF quote adjustments unrelated to order imbalance, inconsistent with an arbitrage explanation. Extending our analysis to a daily frequency also reveals little to no relation between ETFs and the market quality of their constituent securities.

JEL classification: G12, G14

Keywords: Exchange traded funds (ETFs), limits to arbitrage, liquidity

---

We appreciate the helpful comments of seminar participants at Clemson University, University of Arkansas, University of Alabama, Auburn University, University of North Texas, University of Mississippi, Missouri State University, the 2019 Financial Management Association Meeting, the 2019 Northern Finance Association Meeting, the 2019 Southwestern Finance Association Meeting, and the 2017 Eastern Finance Association Meeting. We are especially grateful to the Mississippi Center for Supercomputing Research and its director, Ben Pharr, for computational assistance.

* *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 in the Darden School of Business at the University of Virginia, and Andrew Lynch is an Assistant Professor of Finance at the University of Mississippi. Send correspondence to Travis Box at Clemson University: tsbox@clemson.edu Address: 730 Greenville One, 1 North Main St. Greenville, SC 29601.*

There is a growing literature examining the potential impact of ETF trading on the securities underlying their portfolios. While several early studies suggested that ETF trading improves price discovery for the underlying securities,[1] the more recent literature, both empirical and theoretical, suggests that increasing ETF liquidity may attract short-term traders that introduce noise into constituent security prices through the arbitrage mechanism.[2] While the mechanics of this arbitrage, buying the ETF and selling the underlying or vice-versa, provide a seemingly logical mechanism for this effect, not all ETF trading activity directly affects the underlying. The Investment Company Institute, for example, shows that from 2015 to 2018, only 10% of ETF secondary market trading activity (i.e. the purchase and sale of ETF shares) translated into authorized participant primary market or arbitrage trading activity.[3] If ETF trading does not translate into substantial creation and redemption activity in the underlying securities, a clear arbitrage channel, then how does ETF trading affect the underlying?

While much of this previous research focuses on daily correlations between the prices of ETFs and their constituent securities, it assumes but does not directly test this proposed arbitrage mechanism.[4] In this paper, we examine how ETF trading affects the underlying securities at an intraday frequency, using a sample of 423 passively managed U.S. equity ETFs from 2006 to 2015. In contrast to the existing literature we find little evidence that trading in the ETF transmits to the

[1] See Hasbrouck (2003), Yu (2005), Chen and Strother (2008), Fang and Sanger (2012), and Ivanov, Jones and Zaima (2013).
[2] See Broman and Shum (2018), Charupat and Mui (2011) and Malamud (2016). ETF ownership is associated with higher trading costs and less informative prices (Israeli, Lee and Sridharan (2017)), excess comovement (Da and Shive (2018)) and excess volatility (Ben-David, Franzoni and Moussawi (2018)) in the underlying portfolio..
[3] 2019 Investment Company Fact Book, Investment Company Institute, Figure 4.5.
[4] While previous work has not tested this arbitrage mechanism directly, several potential limitations of the ETF arbitrage process are suggested in previous studies. For instance, market frictions allow some mispricings to persist longer than a potential arbitrageurs' planned trading horizon (Ben-David, Franzoni and Moussawi (2018)). Furthermore, investors may not enforce efficient pricing due to execution risks (Malamud (2016)), transaction costs (Broman and Shum (2018)), difficulty borrowing shares, as well as varying arbitrage speeds or difficulties in observing intrinsic values, especially during times of market stress (Madhavan and Sobczyk (2016)).

underlying portfolio using simple correlations, panel regressions, GMM dynamic panel estimations, or panel vector autoregressions at intraday and daily frequencies. Conversely, we find strong evidence ETF prices follow their constituent's prices, even in cases of extreme price discrepancies or returns (i.e. stochastic jumps). Overall, rather than transmitting non-fundamental shocks to the underlying securities, our results suggest that ETFs may shield the portfolio from demand shocks by offsetting liquidity provision in the underlying securities.

Our empirical analysis begins with simple intraday correlations between ETF and underlying returns and order flows. The most basic prediction of the hypothesis that ETF trading affects the price of the underlying is a strong correlation between ETF order flow and the returns on the underlying. While we find a strong correlation between underlying order flow and ETF returns, we find no relationship between ETF order flow and returns on the underlying securities at 1-, 5-, and 10-minute intraday intervals, and only a weak relationship at a daily interval.

It is possible that failing to control for other covariates in this simple bivariate analysis is masking evidence of this arbitrage mechanism. We continue by modeling the endogenous relationship between intraday returns and marketable order imbalances of ETFs and their underlying portfolios using a panel vector autoregressive (PVAR) model. Simultaneously modeling all four relations (ETF and underlying order imbalances and returns), we estimate the impulse response function (IRF) of a shock in one variable to the other three. If arbitrage trading affects constituent securities, we would expect to find shocks in ETF order imbalances and/or returns leading underlying order imbalances and/or returns. We find no response in the underlying portfolio to shocks in ETF order imbalance. While we find a small response in underlying returns to a shock in ETF returns (6%), we find a considerably stronger effect flowing from underlying returns to ETF returns (36%) and the difference between the two is statistically significant.

We then examine the forecast error variance decomposition (FEVD) from the PVAR model to identify the fraction of each variable's variation explained by the other three. Over the 1-, 5-, or 10-minute interval window, we find that orthogonalized shocks to the ETF returns explain at most 0.26% of underlying return variance whereas orthogonalized shocks to ETF order flow explains 0.00%, to two significant figures.[5] Limiting our analysis to underlying portfolios most exposed to non-fundamental shocks in the ETF, those with low market capitalization or high spreads relative to their ETFs, we still find underlying returns unresponsive to either ETF returns or order imbalances.

While the PVAR analysis covers the entire sample of intraday data, it is possible ETF trading only impacts constituent securities when the two become mispriced or are subject to extreme price movements. To address this possibility, we revisit our 1-minute horizon intraday analysis, but focus on those periods with profitable arbitrage opportunities (as measured by the best bid and ask prices for the ETF and its underlying) in one test and periods where either the ETF or the underlying experiences a stochastic jump in a second test.

In the first test examining arbitrage opportunities, we classify each ETF into one of five relative pricing categories: overvalued (ETF bid is greater than its underlying's ask), partially overvalued (ETF midquote is greater than its underlying's ask), partially undervalued (ETF midquote is less than its underlying's bid), undervalued (ETF ask is less than its underlying's bid), and no arbitrage (ETF midquote is between its underlying's bid and ask). Focusing on these potential arbitrage opportunities, our results reveal a strong relationship between the prices quoted for the ETFs and the prices quoted for the constituent securities. For minutes where the ETF becomes overvalued (undervalued), we find that selling (buying) pressure in the underlying securities pushes the portfolio price downward (upward). In both cases, arbitrage opportunities are preceded by shocks to underlying security prices,

[5] The precise amount of underlying returns' forecast error variance attributed to ETF order imbalance is 0.00068%.

impounded by directional trading in the portfolio. Mispricing does not appear to be the result of short-term price pressure on the ETF's shares. Additionally, our results indicate that constituent price shocks do not reverse during the 30-minute period following a mispricing event, suggesting they are not transitory in nature.

During the minutes following a mispricing event, we observe a swift convergence between the bid and ask prices of the ETF and its underlying portfolio. However, our analysis also suggests that order flow continues to move in the same direction as the mispricing. After the ETF becomes overvalued (undervalued), we see additional buying (selling) pressure in the ETF and net selling (buying) pressure in the portfolio. Price discrepancies between the ETF and the underlying securities are corrected in the limit order book, but directional trading does not drive the convergence. Furthermore, marketable orders appear to be moving prices for the constituent securities and the ETF farther away from parity. If arbitrage activity were the primary motivation for directional trading, we should expect to see order flow that pushes prices closer together.

In the second test, we look at stochastic jumps in either the ETF or the underlying, as identified by the measure of Lee and Mykland (2008). Across our entire sample period, we find that that underlying (ETF) returns exhibit a jump in 0.27 (0.39) percent of observations. In those cases where the underlying experiences a stochastic jump, the underlying and ETF prices mirror each other almost perfectly. In contrast, however, those cases where the ETF experiences a stochastic jump, the ETF price change is preceded by price changes in the underlying in the same direction. Moreover and similar to the arbitrage analysis above, these stochastic jump price changes persist over the thirty minute window of the analysis, suggesting that on average, these are not transient deviations from fundamental values. Overall, the arbitrage and stochastic jump analyses are consistent with liquidity providers hedging their exposure to informed trading in the underlying shares, but we find little evidence that ETF shocks are transmitting noise to their underlying securities.

A sequence of mispricing events may lead to an imbalance in a market maker's inventory of fund and constituent shares. ETFs have a second source of liquidity, however, that mitigates much of the risk associated with market maker inventory accumulation. Specially designated market makers, known as Authorized Participants (AP), can exchange the underlying portfolio for shares of the ETF directly through the fund sponsor.[6] Because ETF shares can be created and redeemed at the fund's closing net asset value (NAV), APs do not have to worry about whether mispricings will eventually converge. Should their inventories ever become large enough, liquidity providers could simply sell their accrual of constituent securities or fund shares in the ETF's primary market, then use the creation or redemption to offset any accumulated short position.[7] We therefore repeat our PVAR analysis at the daily level with similar results, ETF variables such as order imbalance and share creation/redemption have little impact on underlying portfolios.

One possible concern with our analysis is that use of intraday data requires us to focus on a smaller ETF sample and use different econometric techniques than the ETF literature. To ensure that our smaller sample and econometric approach are not driving our observed results, we repeat the analysis of the most directly connected empirical test, namely the impact of ETF ownership on the volatility of constituent securities. We first follow the panel regression framework approach of Ben-David, Franzoni and Moussawi (2018), finding similar evidence that ETF ownership and mispricing

---

[6] Malamud (2016) argues that reductions in primary market trading costs may strengthen the shock propagation channel that allows ETFs to affect stock returns. Petajisto (2017) suggests that primary market activity may not always provide an affordable way to correct price deviations between a fund and its NAV.
[7] Another form of potential ETF arbitrage involves a pairs trading strategy, whereby investors exploit price discrepancies between correlated equivalent assets. Box, Davis and Fuller (2019) and (2020) document competition between ETFs that hold nearly identical portfolios. Marshall, Nguyen and Visaltanachoti (2013) study price discrepancies between two ETFs that track the S&P 500 and, ultimately, find evidence of intraday arbitrage. Likewise, Broman (2016) reports that unexpected shocks to ETF trading activity are correlated across funds with similar holdings.

increases volatility in S&P 500 securities.[8] We then repeat the volatility analysis incorporating insights from our intraday and daily PVAR results. In one set of specifications, we include additional lags of the relevant control variables, and in another we use a dynamic panel estimation to better account for the important interdependencies revealed in our previous analysis. Incorporating these changes has important implications for the results. In examining individual securities underlying the portfolio, we find that ETF ownership actually dampens volatility and the effect of mispricing on volatility becomes marginal. Repeating the analysis for the volatility in the underlying portfolio instead of individual securities, we also find little impact of ETF mispricing and share creation/redemption on underlying volatility the next trading day. Overall, neither primary nor secondary market activity appears to influence constituent returns the following day.

In addition to reporting how prices and order flow transmit between ETF and constituent on average and around mispricing events, we document two important empirical artifacts about ETFs that are sometimes misrepresented in the academic literature. First, across all of our 94 million fund-minutes, the quoted spread of the underlying portfolio only exceeds that of the ETF in roughly 45% of our observations. In general, ETFs do not offer superior liquidity relative to their underlying securities. Second, while we observe a strong correlation between returns and order flow in the constituents, the correlation between returns and order flow in the ETF shares is effectively zero. We find little evidence ETF prices change in response to market participants' buying and selling. If trading in the ETF does not affect the price of the fund itself, it should not be surprising that underlying security prices are also unaffected by ETF order flow.

[8] Table VIII, page 2520, Ben-David, Franzoni and Moussawi (2018).

Overall, while our findings preserve the possibility that non-fundamental shocks in the ETF may extend to the constituent securities, they suggest that arbitrage trading is not the primary channel through which these pricing pressures are transferred.

I. Data and Variable Definitions

From the CRSP Survivor-Bias-Free US Mutual Fund Database, we collect monthly holdings for 423 passively managed U.S. equity funds between 2006 and 2015. Together, these ETFs account for nearly \$1.16 trillion of the \$1.23 trillion total net assets invested in U.S. domestic equity ETFs during 2015. Due to the prevalence of errors in self-reported holdings, we verify the accuracy of each Mutual Fund Database position by corroborating the holding's value using the CRSP Daily Stock File. The details of this verification process are provided in Appendix A.

Based on best bid and ask prices collected from TAQ, we are able to calculate quotes during 94,605,900 fund-minutes for both the ETFs (ETF_t^{Bid} and ETF_t^{Ask}) and their underlying portfolios (Und_t^{Bid} and Und_t^{Ask}).[9] At each time t, we also calculate midpoint returns ($RetRaw_t$) as the percent change in midpoint price for both the ETF and its underlying portfolio. For most of the securities in our analysis, the distribution of intraday midpoint returns is highly leptokurtic. Thus, to minimize the influence of extreme price changes, we examine intraday returns (Ret_t) after performing the following adjustment:

$$Ret_t = \begin{cases} \ln(1 + RetRaw_t), & if\ RetRaw_t \geq 0 \\ -\ln(1 + |RetRaw_t|), & if\ RetRaw_t < 0 \end{cases}. \tag{1}$$

By computing midpoint return's natural log independently for positive and negative values, we preserve the distribution's natural symmetry around zero.[10]

---

[9] Please see Appendices B and C for more detail about the construction of bid and ask prices for the ETFs and their underlying portfolios.

[10] While a simple, unconditional, log transformation would dampen the influence of extreme positive observations, it would also amplify the magnitudes of negative returns and create even larger outliers on the distribution's left side. To

Next, we calculate directional order flow ($VolDiffRaw_t$), scaled by average total volume and reported in basis points, during each period t:

$$VolDiffRaw_t = \frac{BuyVol_t - SellVol_t}{\frac{1}{T} \times \sum_{t=1}^{T}(BuyVol_t + SellVol_t)} \times 10{,}000, \tag{2}$$

where $BuyVol_t$ is the dollar value of all trades occurring at prices above the prevailing midpoint, $SellVol_t$ represents the dollar value of trades occurring below the quoted midpoint and T is the total number of periods within each trading day. As with midpoint return, we correct for kurtosis in $VolDiffRaw_t$'s distribution by using a similar symmetric log transformation:

$$VolDiff_t = \begin{cases} \ln(1 + VolDiffRaw_t), & if\ VolDiffRaw_t \geq 0 \\ -\ln(1 + |VolDiffRaw_t|), & if\ VolDiffRaw_t < 0 \end{cases}. \tag{3}$$

Table I provides summary statistics for each of the four variables of interest, Ret_t^{Und}, Ret_t^{ETF}, $VolDiff_t^{Und}$ and $VolDiff_t^{ETF}$, calculated over 1-, 5- and 10-minute windows within each trading day. Regardless of frequency, constituent midpoint returns and order imbalance demonstrate minimal skewness and both have average values centered near zero.[11] For ETFs, however, Table I reports that average $VolDiff_t^{ETF}$ is positive across all three frequencies. Sellers, of any asset, should raise prices in response to strong demand and lower prices whenever demand is weak. If order imbalance captures the net demand for a security, it is not obvious why liquidity providers would choose to offer ETFs at prices low enough to encourage a disparity between buy and sell volume. Persistent order imbalances, especially over such a lengthy sample period, raise the possibility that ETF prices might not respond to buying and selling pressure in the same way as their underlying securities.

---

ensure our transformation is not affecting the results, we have repeated the analysis with the log transformation with qualitatively similar results.

[11] Despite the large discrepancy between mean and median $VolDiff_t^{Und}$ in the 5- and 10-minute samples, finer partitions of the data reveal that these distributions are still centered almost perfectly around zero. For instance, the 49th and 51st percentiles of $VolDiff_t^{Und}$ are -4.71 and 3.28, respectively, in the 5-minute sample. Given that several different securities contribute to each portfolio's order flow, at least some order imbalance is likely during each 5- or 10-minute window.

For a cursory look at how price changes and order imbalances might be associated, Table I also reports the estimated correlation coefficients between each of our four variables. Not surprisingly, we find a strong connection between ETF and underlying portfolio returns across all frequencies, with correlations becoming even stronger as measurement frequency falls. We also see that the correlation between Ret_t^{Und} and $VolDiff_t^{Und}$ is positive, implying that directional order flow usually trends in the same direction as price changes for the underlying securities. Once again, this association becomes stronger as the measurement window lengthens. For ETFs, however, Table I provides little evidence that directional order flow is positively associated with their own midpoint returns.[12] Furthermore, the estimated correlation between Ret_t^{ETF} and $VolDiff_t^{ETF}$ is even slightly negative at the 1- and 5-minute frequencies, whereas ETF returns are positively related to order imbalances within their underlying portfolios. Thus, for ETFs, price changes coincide with buying and selling pressure in their constituent securities, but not with trading in their own shares.

Prior literature presumes that order imbalance in ETF shares, possibly initiated by short-term noise traders, pushes fund prices away from fundamental values and that these fluctuations are impounded into underlying securities through an arbitrage mechanism. From Table I, it seems unlikely that ETF order imbalances are contemporaneously influencing their own returns, much less those of their underlying portfolios. For the passively managed domestic equity strategies populating our sample, persistent order imbalances, and negligible associations between trading and returns, suggest that underlying portfolio values determine ETF bid and offer prices, not buying and selling pressures in the fund's shares.

[12] This is consistent with Brogaard, Hendershott and Riordan (2019), who find that price discovery in Canadian equity markets occurs predominantly through limit orders, not through trading.

II. Panel Vector Autoregression Analysis

Even though bivariate analysis does not reveal a contemporaneous association between ETF trading and portfolio returns, a more robust empirical approach might reveal a link between fund order flow and constituent prices. Multiple confounding empirical characteristics may obscure this link. For instance, shocks to the supply and demand of underlying shares, regardless of whether those shocks are related to ETF activity, might not impact security prices contemporaneously. Also, momentum, in either order imbalance or returns, could mask a fundamental connection between these variables. Lastly, if order imbalances and price changes for ETFs and their constituents are jointly determined, measuring responses to independent shocks in one of these time series would require an approach that accounts for all potential interdependencies. Considering these interdependencies reinforces the importance expressed by the prior literature of identifying non-fundamental demand shocks when examining potentially adverse effects of ETFs on their constituent securities (Brown, Davies and Ringgenberg (2019)).

One approach to addressing the interdependencies of these variables and to identify non-fundamental demand shocks is a panel vector autoregression. Vector autoregression (VAR) models are commonly used to account for intertemporal dependence between groups of macroeconomic variables (see Hamilton (1985), Koijen, Lustig and Van Nieuwerburgh (2017), and Duffee (2018)). Unlike structural models with simultaneous equations, VAR models require little knowledge of the economic forces that are impacting each time series. Variables are usually treated as endogenous within a VAR system, but the impact of exogenous shocks to individual time series can still be modeled through an impulse response function (IRF) and forecast error variance decomposition (FEVD). Holtz-Eakin, Newey and Rosen (1988) extend the estimation of vector autoregressive systems to panel data, accounting for heterogeneity in levels, variances, and the time series correlation patterns of each panel cross-section. To do this, they apply an instrumental variables approach to the quasi-differenced auto regressive equations (Love and Zicchino (2006) and Abrigo and Love (2016)). In our setting, panel VAR allows us to simultaneously quantify the reaction of ETF and portfolio returns and order flow to independent innovations in these same variables using a large cross-section of fund-days.[13]

The following system of linear equations represents a PVAR specification of order p:

$$\boldsymbol{Y}_{fdt} = \boldsymbol{Y}_{fdt-1}\boldsymbol{A}_1 + \boldsymbol{Y}_{fdt-2}\boldsymbol{A}_2 + \cdots + \boldsymbol{Y}_{fdt-p}\boldsymbol{A}_p + \boldsymbol{\alpha}_{dt} + \boldsymbol{\gamma}_{fd} + \boldsymbol{\varepsilon}_{fdt}, \tag{4}$$

where $\boldsymbol{Y}_{fdt} =$ is a (1×4) vector of dependent variables, Ret_{fdt}^{Und}, Ret_{fdt}^{ETF}, $VolDiff_{fdt}^{Und}$ and $VolDiff_{fdt}^{ETF}$ for each fund (f), date (d) and time $(t \in \{1,2, \dots, T\})$. The vector $\boldsymbol{\alpha}_{dt}$ contains variable-specific fixed effects for each date-time combination, and $\boldsymbol{\gamma}_{fd}$ is vector of variable-specific panel effects for each fund-date. The parameters to be estimated are included in the 4×4 matrices $\boldsymbol{A}_1$, $\boldsymbol{A}_2$,…, $\boldsymbol{A}_p$, and the vector $\boldsymbol{\varepsilon}_{fdt}$ consists of idiosyncratic error terms for each dependent variable such that $E\left[\boldsymbol{\varepsilon}_{fdt}\right] = \boldsymbol{0}$, $E\left[\boldsymbol{\varepsilon}'_{fdt}\boldsymbol{\varepsilon}_{fdt}\right] = \boldsymbol{\Sigma}$ and $E\left[\boldsymbol{\varepsilon}'_{fdt}\boldsymbol{\varepsilon}_{fds}\right] = \boldsymbol{0}$ for all $t > s$. To correct for estimate biases due to panel effects with lagged dependent variables, we utilize an instrument set made up of all future observations.[14]

[13] Panel vector auto-regressions are used in a similar context by Hilscher, Pollet and Wilson (2015), Hollifield, Neklyudov and Spatt (2017), and Lee, Naranjo and Velioglu (2018).

[14] While the bias approaches zero as the total number of time periods increases, Judson and Owen (1999) find significant bias even when $T = 30$. Based on the assumption that idiosyncratic errors are serially uncorrelated, consistent GMM estimators have been proposed whereby parameter estimates are based on a first-difference transformation of the dependent variables. Thus, lagged levels and differences of $\boldsymbol{Y}_{fdt}$ become eligible instruments for the transformed dependent variables. With first-difference transformation, a second-order PVAR requires that $T_i \geq 5$ realizations are available for each subject. Here, $\Delta\boldsymbol{Y}_{fdt-1}$ is modeled as a function of $\Delta\boldsymbol{Y}_{fdt-1} = \boldsymbol{Y}_{fdt-1} - \boldsymbol{Y}_{fdt-2}$ and $\Delta\boldsymbol{Y}_{fdt-2} = \boldsymbol{Y}_{fdt-2} - \boldsymbol{Y}_{fdt-3}$, so the levels $\boldsymbol{Y}_{fdt-1}$, $\boldsymbol{Y}_{fdt-2}$ and $\boldsymbol{Y}_{fdt-3}$ are ineligible as instruments. Thus, the levels $\boldsymbol{Y}_{fdt-4}$ and $\boldsymbol{Y}_{fdt-5}$ must be available for the model to be just-identified. Instead of first-differencing, Arellano and Bover (1995) propose a forward orthogonal deviation that subtracts the average of all future observations and preserves all but the most recent realizations for the instrument set. Further, to improve efficiency, especially in unbalanced panels, Holtz-Eakin, Newey, and Rosen (1988) recommend substituting missing observations with zero.

The consistency of GMM estimation relies on optimal model and moment selection. Andrews and Lu (2001) propose criteria resembling the widely used Bayesian information criterion to select each specification's order p and how many lags q of the dependent variables form the basis of the moment conditions.[15] Once the optimal model has been chosen, IRFs, derived from the parameter estimates of $\boldsymbol{A}_1$, $\boldsymbol{A}_2$,…, $\boldsymbol{A}_p$, describe each dependent variable's evolution following some exogenous shock. These IRFs have no causal interpretation, however, because the idiosyncratic errors, $\boldsymbol{\varepsilon}_{fdt}$**,** are likely to be correlated contemporaneously.

To determine whether independent shocks to one variable, such as $VolDiff_{fdt}^{ETF}$, impact future realizations of another, like Ret_{fdt}^{Und}, we rely on FEVD analysis for most of our qualitative interpretations. Here, idiosyncratic errors are orthogonalized to isolate the contribution of exogenous shocks in one variable to the forecast-error variances of all the others. Sims (1980) proposes a Cholesky decomposition of $\boldsymbol{\Sigma}$ whereby an instantaneous "causal" ordering of each time series is specified by their arrangement within $\boldsymbol{Y}_{fdt}$.[16] For the sequence $\left[Ret_{fdt}^{Und}, Ret_{fdt}^{ETF}, VolDiff_{fdt}^{Und}, VolDiff_{fdt}^{ETF}\right]$, shocks to fund order flow are orthogonalized relative to the immediate impacts of each preceding variable. Thus, FEVD analysis can tell us how much independent shocks to the ETF order flow imbalance, $VolDiff_{fdt}^{ETF}$, at $t = 0$ influence the realizations of $VolDiff_{fdt}^{Und}$, Ret_{fdt}^{ETF} and, most importantly, Ret_{fdt}^{Und} during subsequent periods.[17] Brown, Davies, and Ringgenberg (2019) highlight the importance of identifying non-fundamental demand shocks to

---

[15] The Andrews and Lu (2001) model selection criteria suggest that the specification's order p should equal 5, 9, and 16 for the 1-, 5-, and 10-minute samples, respectively. In all cases, q is suggested to equal $p + 3$.

[16] The ordering is such that the first time series in $\boldsymbol{Y}_{fdt}$ may have an immediate impact on all other variables. The second time series may have an instantaneous impact on the remaining components of $\boldsymbol{Y}_{fdt}$, excluding the first variable, and so on. This type of causality is often referred to as Wold-causality.

[17] This ordering ensures examinations of ETF variables measure non-fundamental shocks, an important step in assessing ETF impact on underlying securities (Brown, Davies and Ringgenberg (2019)). Additionally, we estimate Equation (4) with several different orderings, however, qualitative inferences regarding the impact of fund order flow on subsequent constituent prices are unchanged by these alternative specifications.

the ETF in examining the potentially adverse effects on underlying securities and the PVAR methodology will enable us to do so.

A. Full Sample PVAR Results

To ease the computational burden of estimating Equation (4), the variables Ret_{fdt}^{Und}, Ret_{fdt}^{ETF}, $VolDiff_{fdt}^{Und}$ and $VolDiff_{fdt}^{ETF}$ are centered across date and time combinations so that variable-specific fixed effects $\boldsymbol{\alpha}_{dt}$ can be omitted from the specification.[18] Each time series is also standardized by their fund-specific intraday volatilities realized during the prior 200 days, with a minimum of 50 days required for inclusion in the analysis. After estimating the parameters from Equation (4), we derive ten-period IRFs following a unit shock to the idiosyncratic error terms of each dependent variable.

Figure 1 depicts the cumulative simple IRFs from our parameter estimates, which describe a variable's accumulated reaction in each of the proceeding ten periods. Response variables are denoted along the top of each column, and impulse variables are indicated to the left of each row. The 97.5% and 2.5% confidence intervals are designated by dotted lines above and below the average impulse response. By standardizing Ret_{fdt}^{Und}, Ret_{fdt}^{ETF}, $VolDiff_{fdt}^{Und}$ and $VolDiff_{fdt}^{ETF}$ prior to the estimation of Equation (4), the cumulative IRFs in Figure 1 describe each variable's expected reaction following a one-standard deviation shock to an impulse variable during $t=0$. While only 5-minute results are tabulated, cumulative IRFs from the 1- and 10-minute samples are qualitatively similar.

Despite their inability to generate causal inferences, cumulative IRFs can help us understand how variables within a dynamic system interact with each other on average. The IRFs along the diagonal of Figure 1, represent the cumulative response of each variable (Ret_{fdt}^{Und}, Ret_{fdt}^{ETF}, $VolDiff_{fdt}^{Und}$ and

[18] Centering variables across date and time combinations remove the systematic component of intraday returns and order flow during each period.

$VolDiff_{fdt}^{ETF}$) to a one-standard deviation shock to itself. For the underlying, we observe a small reversal of the shock (5% dissipation to 0.95 of the original shock at 50 minutes after the shock) but a strong continuation in the underlying directional order flow. For the ETF, we observe a larger reversal (32% dissipation to 0.68) and a more modest continuation in the ETF directional order flow (to 114%). Across the top row, a positive shock to underlying prices is followed by a large positive and unsurprising ETF return response (0.32 standard deviations). The directional order flow responses, however, present a more interesting pattern. The positive underlying return shock is followed by buying in the underlying shares as represented by the positive 0.08 standard deviation response, but selling in the ETF (-0.03 standard deviation response). Here, fund trading is trending in the opposite direction as price changes suggesting, once again, that ETF liquidity providers do not respond to buying and selling pressure in a conventional manner.

Moving to the second row of Figure 1, while we observe ETF price shocks generating positive responses in the underlying returns and order flow, the cumulative reactions never amount to more than a small fraction of one standard deviation. For comparison, a one-standard deviation shock to Ret_t^{ETF} only generates an accumulated 0.06 standard deviation response in underlying returns after 50 minutes, whereas, a unit shock to Ret_t^{Und} generates a 0.34 standard deviation reaction in ETF returns over the same horizon and the confidence intervals for the two show that the difference is strongly statistically significant. For the other off-diagonal figures in the two bottom rows, we see almost no evidence that demand shocks to the underlying or ETF shares are correlated with their subsequent returns. Similar to the correlation results, the IRF results are not consistent with ETF trading negatively impacting the underlying through the arbitrage mechanism.

To better understand the causal links between ETF trading and portfolio returns, we now turn to a FEVD analysis of our estimated parameters from Equation (4). The contributions of an impulse variable to the forecast-error variance of each response variable, after ten periods, are tabulated in Table II. Panels A, B and C report FEVD estimates for the 1-, 5- and 10-minute windows, respectively.

Regardless of the frequency with which the variables are measured, we find almost no evidence that independent shocks to ETF returns (Ret_t^{ETF}) or ETF order imbalance ($VolDiff_t^{ETF}$) impact future price changes or trading in the underlying securities. While fund returns may contribute a tiny fraction to their portfolio return's total forecast-error variance, 0.26% in the 1-minute sample, orthogonalized shocks to ETF order flow have no relation with future innovations in Ret_t^{Und} or $VolDiff_t^{Und}$. Conversely, Table II also suggests that future fund returns are influenced by price changes in their underlying securities, or at least by underlying and fund price changes that occur simultaneously.[19] Roughly 18.67% of Ret_t^{ETF}'s forecast-error variance can be explained by shocks to constituent prices in the 1-minute sample, rising to 43.20% in the 10-minute sample. Altogether, our results imply that, while future ETF prices respond to, potentially simultaneous, shocks in their underlying portfolios, orthogonalized ETF shocks do not affect their portfolios reciprocally during later periods.[20]

B. ETF liquidity and non-fundamental demand shocks

Across a broad sample of U.S. equity ETFs, our PVAR analysis demonstrates that independent shocks to fund prices and order imbalances have little effect on subsequent constituent returns and trading. It is possible, however, that non-fundamental demand shocks in certain types of funds impact

[19] Idiosyncratic shocks to underlying returns are not orthogonalized relative to Ret_t^{ETF}, $VolDiff_t^{Und}$ and $VolDiff_t^{ETF}$ because of the variable ordering within $\boldsymbol{Y}_{fdt}$. Statistically, it is not possible to determine whether temporally correlated shocks to the fund and portfolio are attributable to one or the other. As basket securities, however, ETF returns should be impacted instantaneously by shocks to constituent prices. This presumption is also supported by our graphical analysis.

[20] The lack of association between Ret_t^{Und} and $VolDiff_t^{Und}$ in the FEVD analysis may seem curious given their positive correlation reported in Table I.[20] FEVD only measures the contribution of independent shocks in one time series to the forecast-error variance of another. While innovations in Ret_t^{Und} and $VolDiff_t^{Und}$ are obviously related contemporaneously, FEVD analysis shows that orthogonalized shocks to order flow do not impact future realizations of returns.

the returns and order flow of their portfolios more directly. For instance, Broman and Shum (2018), Charupat and Mui (2011), Israeli, Lee and Sridharan (2017), and Malamud (2016) suggest that short-term traders, those that are most likely to introduce noise into security prices, are encouraged by the superior liquidity of ETFs. Thus, to identify cases where a fund's liquidity might be superior, relative to its underlying holdings, we divide our sample into four distinct subsets: *Small Underlying and Small ETF*, *Small Underlying and Large ETF*, *Large Underlying and Small ETF*, and *Large Underlying and Large ETF*. Funds and constituents are considered large if their mean daily market capitalization falls in the top tercile of our sample during a particular month.[21]

After generating unique parameter estimates for Equation (4) from each of the four subsamples, we tabulate results from a FEVD analysis in Table III. We expect that disparities in liquidity will be greatest whenever ETFs are large relative to their underlying holdings.[22] Even though the impact of independent fund return shocks on underlying returns, Ret_t^{Und}, is greatest for the *Small Underlying and Large ETF* subset, this particular impulse contributes only 0.32% to the portfolio return's total forecast-error variance. Thus, even in a subset where the fund's market capitalization is most likely to exceed that of its average constituent, ETF price changes have almost no influence on subsequent portfolio returns.

Where we find the most meaningful variation across the four subsamples is in the share of ETF returns' total forecast-error variance that is attributable to underlying, or potentially simultaneous, price changes. For funds that are large relative to the average market capitalization of their holdings, Table III demonstrates that future innovations in ETF prices are more strongly influenced, 51.93% of

[21] Underlying market capitalization is the dollar-weighted average of all stocks held by the ETF, based on their CRSP price and shares outstanding reported at the end of the previous year. The top tercile of ETFs (underlying portfolios) ranked by market capitalization accounts for 95.6% (78.6%) of aggregate ETF market capitalization on average.
[22] Even amongst the largest ETFs, the market capitalization of the fund is usually dwarfed by the average size of their holdings. For instance, the largest ETF, the SPDR S&P 500, has an average market value of $263 billion in 2019 according the *Wall Street Journal*. By comparison, the fund's largest holding, Microsoft, is worth more than $1 trillion.

forecast-error variance, by shocks to constituent returns. Conversely, shocks to Ret_t^{Und} contribute only 19.76% to the forecast-error variance of Ret_t^{ETF} in the *Large Underlying and Small ETF* subsample. Instead of conducting more noise into constituent security prices, the superior liquidity of relatively large ETFs only seems to facilitate a more efficient response to innovations in portfolio values.

To amplify the potential for liquidity disparities between the funds and their underlying securities, we also divide our 5-minute sample into quintiles based on the ETF's and constituents' relative bid-ask spreads. Quintile breakpoints are based on the difference between average minute-by-minute intraday quoted spreads for the fund and its market-weighted portfolio during the previous 200 trading days. The first quintile consists of ETFs with bid-ask spreads that are low relative to their holdings, whereas the fifth quintile contains funds with quotes that are wide compared to the constituents.

The results in Table IV show that independent shocks to ETF returns, Ret_t^{ETF}, and ETF order imbalance, $VolDiff_t^{ETF}$, still have almost no impact on future price changes or trading in the underlying securities, regardless of any disparities between fund and constituent quoted spreads. Furthermore, the small fraction of the portfolio return's total forecast-error variance that can be attributed to ETF returns does not seem to vary monotonically with relative liquidity. Here, 0.25% of the forecast-error variance of underlying returns is contributed by independent fund price changes in the middle subset, falling to 0.10% and 0.01% in the lowest and highest spread difference quintiles, respectively.

Once again, the most interesting differences across the subsamples are found in the response of Ret_t^{ETF} from orthogonalized shocks to Ret_t^{Und}. While variation across the quintiles is not perfectly monotonic, a larger proportion of ETF return's total forecast-error variance is contributed by portfolio price changes when the fund's quoted spread is low relative to its constituents. Altogether, the results from Table III and Table IV suggest that fund liquidity is not promoting the transfer of non-fundamental shocks from equity ETFs into their holdings. Even if a fund's liquidity encouraged short-term trading, it might also dampen the price impact of non-fundamental demand shocks. Empirically, improvements in an ETF's relative liquidity only seem to improve the efficiency of the fund's reaction to underlying price changes without promoting a transfer of noise into security prices.

C. Correlated alternatives to trading in constituent shares

Studies which have suggested that ETFs might introduce non-fundamental shocks into the prices of underlying securities have presumed that such noise is transferred though an arbitrage mechanism (Broman and Shum (2018), Ben-David, Franzoni and Moussawi (2018), Charupat and Miu (2011), Israeli, Lee and Sridharan (2017) and 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 non-fundamental 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 V 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.[23]

III. Graphical Analysis

The results from our FEVD analysis strongly suggest that independent shocks to ETF prices or order imbalances have little economic impact on the future returns and trading of their portfolio securities. However, the arbitrage mechanism, as described in prior literature, requires at least some divergence between fund and underlying prices. Perhaps, the effect of ETF trading on constituent shares is only observable when mispricing is large enough to incentivize arbitrageurs to enter opposing positions.

[23] 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 non-fundamental demand shocks from the portfolio of correlated alternatives are associated with subsequent innovations in constituent prices.

Figure 2 describes four mispricing associations between an ETF and its underlying portfolio. In all four cases, the ETF is trading at a premium or discount as measured by the percentage difference in midpoint prices. However, only Panel A and Panel D describe a tradeable arbitrage opportunity. In Panel A, for instance, the ETF bid price is above the ask price for the underlying portfolio. Therefore, the simultaneous purchase of the underlying portfolio, and sale of the ETF, would result in realized profits if prices later converged. In Panel B, the midpoint price of the ETF is still much higher than that of its underlying portfolio, but this discrepancy cannot be arbitraged away by submitting marketable orders.

For our subsequent analysis, we compare all ETF and underlying bid and ask prices at the end of every minute during our sample period. All fund-minute observations are designated as $Misp_t^{EB>UA}$, $Misp_t^{UA>EB>UM}$, $Misp_t^{UM>EA>UB}$ or $Misp_t^{UB>EA}$, in accordance with Figure 2, or classified as $NoMisp_t$ when none of these conditions is satisfied. Table VI reports the percentage of fund-minute observations that are mispriced across our entire sample. Out of 94,605,900 total fund-minute observations, 813,815 and 749,441 are classified as overvalued and undervalued, respectively. Thus, tradeable arbitrage opportunities exist between an ETF and its underlying portfolio approximately 1.66% of the time. The classifications $Misp_t^{UA>EB>UM}$ and $Misp_t^{UM>EA>UB}$ account for 3.69% and 3.63% of our observations, respectively, leaving 91.02% of fund-minutes designated as $NoMisp_t$.

While smaller ETFs are likely to have less liquidity, and ETFs that hold smaller securities are probably more difficult to value, Table VI indicates that the proportion of overvalued and undervalued fund-minutes is roughly equal across all size subsamples. To understand why, quoted spreads for the ETF and underlying portfolio are also reported in the in the far-right columns of the table. As expected, the *Large ETF* and *Large Underlying* classifications have lower spreads. However, these narrower spreads mean that relatively minor discrepancies between the ETF and its portfolio can lead to tradeable arbitrage opportunities. Thus, the margin for error shrinks with fund and security size. Furthermore, contrary to the common presumption that superior ETF liquidity could attract short-term traders and introduce noise into constituent prices (Broman and Shum (2018), Charupat and Mui (2011), Israeli, Lee and Sridharan (2017), and Malamud (2016)), quoted spreads for U.S. equity ETFs are usually wider than those of their underlying holdings. Only when ETFs are large relative to the size of their portfolios do transaction costs favor the fund.

A. Cumulative midpoint returns

After identifying all of the mispricing events in our sample, we cumulate the ETF and underlying portfolio midpoint return after each minute beginning at $t - 10$ and extending through $t + 30$.[24] Next, we calculate the average cumulative midpoint return for each event-window-minute across all observations within a particular size category.[25] When these minute-by-minute averages are depicted in Figure 3, several empirical artifacts become clear. Most notably, the figure demonstrates that mispricing events often center around the arrival of information relevant to the value of constituent securities, whose cumulative returns are represented by the dotted line. With the exception of the *Small Underlying* and *Large ETF* case, the value of the underlying portfolio experiences a persistent decline during 40-minute event windows where the ETF becomes overvalued, and a persistent increase during events where the ETF becomes undervalued. Following a price shock to the constituents during the first ten minutes, the portfolio stays at this new level for the remaining thirty minutes of the event window. While this effect is strong for extreme mispricing ($Misp_t^{EB>UA}$ and $Misp_t^{UB>EA}$), it even holds in cases of partial mispricing ($Misp_t^{UA>EB>UM}$ and $Misp_t^{UM>EA>UB}$) for

[24] We remove any observations that proceed another event by ten minutes or less. Extreme mispricing events, $Misp_t^{EB>UA}$ and $Misp_t^{UB>EA}$, are only excluded if there is another extreme mispricing event of the same direction in the prior ten minutes. Partial mispricing events, $Misp_t^{UA>EB>UM}$ and $Misp_t^{UM>EA>UB}$, are excluded if there is another partial or extreme mispricing event of the same direction in the prior ten minutes.

[25] After cumulative returns are calculated, they are Winsorized each minute at the 1st and 99th percentiles.

Small ETFs. Therefore, price adjustments to the underlying portfolio are not, on average, transient deviations from fundamental values. Conversely, when the ETF is large relative to the value of constituents, fluctuations in the price of the portfolio do not persist through the end of the event window. For these cases, which represent 17.75% of the sample's observations, arbitrage opportunities are less likely to be associated with material information about the underlying holdings. According to Table VI, this is the only size classification where transaction costs are lower in the ETF than in the underlying securities. Therefore, as suggested by Broman and Shum (2018), Charupat and Mui (2011), Israeli, Lee and Sridharan (2017), and Malamud (2016), the superior liquidity of these ETFs might attract short-term traders that could introduce noise into the prices for constituent securities.

To test this proposition, Figure 3 allows us to compare the timing of price adjustments between an ETF and its constituent securities. First, we see that the 40-minute event window provides enough time to observe total convergence between the fund's price, represented by the solid red line, and that of its portfolio. In three of the four size classifications, we also observe that the value of the underlying securities begins moving several minutes before the arbitrage opportunity materializes. Meanwhile, the ETF price does not begin adjusting to new information about the constituent securities until $t + 1$. As before, we observe different behavior when the ETF is large relative to the underlying portfolio. Here, the fund's price shifts several minutes before the arbitrage event. Furthermore, the prices of constituent securities appear to follow those of the ETF, at least temporarily, before the fund and portfolio shift back towards their $t - 10$ values. Thus, for the *Small Underlying* and *Large ETF* case, Figure 3 provides some evidence that short-term fluctuations in ETF prices could introduce noise into their underlying portfolios. Yet, for all other size classifications, we observe persistent shocks to the constituent securities with ETF price adjustments lagging by several minutes.

One remaining empirical detail visible in each of the cases described by Figure 3 is the peculiar short-term reversal in ETF midpoint returns between minutes $t-1$ and $t+1$. To better understand this artifact, we also provide the sample-average cumulative bid and ask returns around arbitrage opportunities in Figure 4. The top and bottom of each vertical line represent the average path of the best offer and bid price, respectively, and the horizontal dash in the center represents the average path of the midpoint.

For the underlying securities, the bid and ask prices adjust smoothly and symmetrically to the arrival of new information during the minutes preceding an arbitrage opportunity. Conversely, the adjustment process is less balanced for the ETF's quotes. In Panel A, negative information lowers the prices of constituent securities and, at $t+1$, the ETF's best offer price begins to fall as expected. For bid prices, however, we observe a sharp increase during minute t that temporarily pushes up midpoint prices. Here, the market appears to temporarily offer liquidity to ETF sellers at favorable prices before adjusting quotes downward to reflect new information about the constituent securities. We observe a similar pattern in Panel B, except that ETF bid prices adjust smoothly while offer prices fluctuate. In either case, these asymmetrical shifts in quoted prices could arise from conditioning our graphical analysis on arbitrage opportunities. Had the quotes adjusted smoothly and symmetrically, instead, the arbitrage opportunity would never materialize. Regardless, the overall behavior in Figure 4 is still consistent with hedging on the part of liquidity providers, whereby aggressive price quotes for the ETF might serve to offset liquidity provision in the underlying securities.

B. On-balance volume

Next, we examine the flow of marketable orders in Figure 5 for the ETF and its underlying securities during the same 40-minute event windows depicted in Figure 3. To calculate on-balance volume, we cumulate $VolDiffRaw_t$, defined in Equation (2), after each minute from $t-10$ through

$t + 30$.[26] Then, we take the average event-window-minute on-balance volume across all observations within a particular size category. Once again, we observe a consistent pattern for all three size classifications where the size of the ETF is not large relative to its constituents. Specifically, we find persistent buying or selling pressure in the underlying portfolio during the minutes immediately surrounding the arbitrage event, followed by relatively balanced trading during the remainder of the event window. Therefore, the underlying price shocks highlighted in Figure 3 appear to be impounded by informed directional trading in the constituent securities. Conversely, on-balance volume for the ETF flows in the opposite direction as midpoint prices throughout the event window. Here, marketable ETF buy orders arrive more frequently as midpoint prices fall, whereas the intensity of ETF sell orders increases with rising prices.

For these situations where the underlying portfolio experiences a persistent price shock, marketable order flow can be classified as informed or uninformed depending on whether the direction of trading is consistent with the ETF or portfolio's return. Negative price shocks are accompanied by informed marketable sell orders for the constituents executed below the midpoint. Thus, for liquidity providers, negative price shocks lead to buildup of inventory in the underlying securities. As exposure to the portfolio rises, liquidity providers might choose to offset their position by selling shares of the ETF. From Figure 3 and Figure 5, lowering the offer price for ETF shares encourages uninformed marketable orders and reduces the number of fund shares in the liquidity providers' inventory. Thus, a series of mispricing events associated with negative price shocks would lead to an accumulation of constituent securities held by liquidity providers. Similarly, repeated positive shocks would lead to a buildup in ETF shares. Should their inventories become large enough,

[26] After cumulative $VolDiffRaw_t$ is calculated, the variable is Winsorized each minute at the 1st and 99th percentiles.

liquidity providers could sell their accrual of constituent securities or fund shares in the ETF's primary market.

For the *Small Underlying* and *Large ETF* case, we observe less informed trading in the underlying portfolio and almost no evidence of liquidity provider hedging in the ETF. Here, transient shocks to the price of the ETF coincide, rather weakly, with fluctuations in ETF order flow. Yet, on-balance ETF volume returns to zero by the end of the window. Thus, it is difficult to determine why ETF prices adjust to create arbitrage opportunities in these situations. Furthermore, cumulative order imbalance for the constituents is negative when $Misp_t^{EB>UA}$, but roughly zero when $Misp_t^{UB>EA}$. Therefore, it is difficult to draw conclusions about whether trading in portfolios of small securities is influenced by short-term price fluctuations in larger ETFs.

C. Stochastic Jumps in ETF and Portfolio Returns

While the arbitrage mechanism proposed in prior literature requires some divergence between fund and underlying values, our graphical analysis suggests that these mispricing events are usually preceded by persistent shocks to constituent prices. At the same time, the very existence of arbitrage opportunities may signal unusual market conditions (e.g. limits-to-arbitrage such as a lack of available arbitrage capital) which limit the generalizability of the results. As an alternative, we additionally examine stochastic jumps in the returns of both ETFs and their constituent securities. Based on the jump identification measure developed by Lee and Mykland (2008), we use the prior fifteen days of 1-minute return volatility to test the null of a diffusion process for each fund and portfolio.[27] Across our entire sample we find that 0.27 percent of 1-minute underlying returns contain a jump, compared with 0.39 percent of ETF observations.[28]

[27] Please see Appendix D for a discussion of stochastic jumps and the Lee and Mykland (2008) jump detection measure.
[28] Given the volatility inherent in the opening auction, we exclude the first 15 minutes of each trading day both as a candidate for a jump as well as in the estimation of prior volatility.

With these jump events centered at $t = 0$, average cumulative midpoint returns are depicted in Panel A of Figure 6. During the minutes surrounding underlying portfolio jumps, pictured on the left side of Figure 6, average fund and constituent cumulative returns overlap almost perfectly whether the shock is positive or negative.[29] Conversely, ETF jumps, depicted on the right side, are anticipated by small price changes in the underlying securities on average. While cumulative fund returns are slightly larger in magnitude than those of the portfolio following ETF jumps, price changes persist for the remaining thirty minutes of the event window in all four cases. Thus, fund and constituent price shocks are not, on average, transient deviations from fundamental values.

As with our graphical analysis conditioned on mispricing events, the paths of average on-balance volume depicted in Figure 6 are consistent with liquidity providers hedging their exposure to informed trading in the underlying shares. Similar to our analysis of arbitrage opportunities, our analysis of stochastic jumps provides little evidence that ETF shocks are transmitting noise into their portfolios.[30]

IV. Daily Analysis

Up to this point, our analyses have been predicated on the search for an intraday arbitrage mechanism whereby non-fundamental ETF demand shocks are transmitted to constituent security prices. It is possible, however, that any trading process propagating shocks from funds into their holdings occurs at a lower frequency than analyzed here. Previous studies have primarily examined such transmissions across days or months, not minutes or hours. Expanding our analysis beyond intraday trading also allows us to incorporate other potential sources of non-fundamental disruption, such as ETF share creations and redemptions or end-of-trading discounts and premiums.

---

[29] While stochastic jumps are generally less common in diversified settings, they have been identified even in market returns (Pukthuanthong and Roll (2015)).

[30] We repeat this analysis using "big" returns instead of stochastic jumps. We define a big return as one larger in absolute value than three times the prior 15 days' 1-minute return standard deviation. We find 1.80 (1.55) percent of underlying portfolio (ETF) returns to be big. Our results are qualitatively similar to those reported in Figure 6.

For most traders, arbitrage profits could be realized by entering into, then later reversing, opposing positions through the secondary market for fund and constituent shares. What makes ETFs unique is that a subset of potential arbitrageurs, the APs, can choose to unwind their positions in the primary market by creating or redeeming shares directly with the fund sponsor. To measure this form of primary market activity, we collect total shares outstanding, $ShrOut_d$, from Bloomberg for each fund at the end of every trading day, d, in our sample. We analyze daily percent changes in shares outstanding, $\Delta ShrOut_d$, after performing a symmetric log transformation similar to what is described in Equations (1) and (3).

Regardless of whether arbitrageurs would seek to unwind their positions in primary or secondary markets, some divergence in ETF and constituent prices must occur to incentivize their activity. Figure 2 describes four situations where intraday mispricing between a fund and its holdings might provide opportunities to earn arbitrage profits. The following summation, across all intraday periods t, captures the persistence of these opportunities during the trading day d:

$$
\begin{aligned}
&MispSumRaw_d \\
&\quad = \sum_t \Big[\big(2 \times Misp_t^{EB>UA}\big) + \big(1 \times Misp_t^{UA>EB>UM}\big) \\
&\quad + \big(-1 \times Misp_t^{UM>EA>UB}\big) + \big(-2 \times Misp_t^{UB>EA}\big)\Big].
\end{aligned}
\tag{5}
$$

As before, $MispSum_d$ represents the summation described by (5) following a symmetric log transformation.

Lastly, arbitrageurs can participate in secondary markets by posting orders for ETF and underlying shares on several different stock exchanges and dark pools throughout the day. However, shares can also be bought and sold in centralized closing auctions hosted by each security's primary listing exchange during the conclusion of trading.[31] The constituent prices determined by these

[31] The Closing Auction brings all buyers and sellers together into one common trade that establishes a clearing price for all interest. This centralized liquidity event allows institutional investors to establish sizeable positions without undue complexity.

auctions are often used by fund sponsors to calculate an ETF's reported NAV. The deviation between this value and the fund's own closing price is described by the widely disseminated daily discount or premium. Most previous studies examining ETF and portfolio dynamics rely on these daily reported values to quantify mispricing and infer intraday arbitrage opportunities. Therefore, we also examine, after performing another symmetric log transformation, each fund's daily basis point discount or premium, $DiscPrem_d$, based on ETF closing prices and their reported NAVs collected from CRSP.

A. Panel Vector Autoregression Analysis

Our lower-frequency analysis of ETF and constituent trading dynamics begins by expanding the PVAR framework described in Equation (4) to include the summation of intraday mispricing, the daily change in shares outstanding and reported basis point discounts or premiums. Daily midpoint returns, Ret_d, and order imbalance, $VolDiff_d$, measure market activity within each trading session and exclude overnight price changes and volume. The calculation of these variables is similar to what was described by Equations (1), (2) and (3), except that $VolDiffRaw_d$ is calculated in percentage terms instead of basis points.

Summary statistics for each time series included in $\boldsymbol{Y}_{fd}$ are reported in Table VII. Even though the summation of intraday fund mispricing is slightly positive on average, at 0.137, median $MispSum_d$ is exactly zero. Likewise, we observe no strong tendency in the percentage deviation between an ETF's closing price and its reported NAV. There appears, however, to be some asymmetry in $\Delta ShrOut_d$, with the frequency of share creations exceeding that of redemptions by at least a small margin.

Figure 7 depicts ten-day cumulative IRFs derived from a PVAR specification where $\boldsymbol{Y}_{fd} =$ is a (1×7) vector of dependent variables.[32] Out of 49 possible impulse and response combinations, we tabulate only those relevant to the transmission of shocks from ETFs to their underlying portfolios. The IRFs in Figure 7 describe the ten-day reactions of Ret_{fd}^{Und}, Ret_{fd}^{ETF}, $VolDiff_{fd}^{Und}$ and $VolDiff_{fd}^{ETF}$ after a one-unit shock to Ret_{fd}^{ETF}, $VolDiff_{fd}^{ETF}$, $MispSum_{fd}$, $\Delta ShrOut_{fd}$ or $DiscPrem_{fd}$ when $d = 0$.

Across the top row, a percentage point shock to fund returns is followed by ETF price increases and downward pressure on constituent values. In practice, however, independent fund shocks of this magnitude are improbable given that the correlation between Ret_{fd}^{Und} and Ret_{fd}^{ETF} is 95.5%. In the second row, we observe positive reactions from ETF demand shocks in fund and underlying order flow, but little response in subsequent prices. Likewise, shocks to intraday mispricing, primary market activity and daily discounts or premiums are only correlated with future trading, not returns, in the ETF and portfolio. Thus, any demand shocks associated with these three variables are unlikely to cause non-fundamental disruptions in market prices.

For a more deliberate evaluation of these causal links, we turn to another FEVD analysis of our estimated parameters. Within $\boldsymbol{Y}_{fd}$, the sequencing of each time series is such that shocks to $MispSum_{fd}$, $\Delta ShrOut_{fd}$ and $DiscPrem_{fd}$ are orthogonalized relative to the instantaneous impacts of Ret_{fd}^{Und}, Ret_{fd}^{ETF}, $VolDiff_{fd}^{Und}$ and $VolDiff_{fd}^{ETF}$.[33] The contributions of an impulse variable to the forecast-error variance of each response variable, after ten days, are tabulated in Table VIII.

[32] The Andrews and Lu (2001) model selection criteria suggest that the specification's order p should equal 4, with q set to $p + 3$.

[33] Both $\Delta ShrOut_{fd}$ and $DiscPrem_{fd}$ are determined after the conclusion of trading and, thus, after the realizations of Ret_{fd}^{Und}, Ret_{fd}^{ETF}, $VolDiff_{fd}^{Und}$, $VolDiff_{fd}^{ETF}$ and $MispSum_{fd}$. It is unlikely that shocks to end-of-day variables contribute to the forecast-error variances of time series measured within the same trading day.

Consistent with our intraday FEVD analysis, we find almost no evidence that independent shocks to ETF returns or order imbalances impact future price changes or trading in their underlying securities. Furthermore, as presaged by Figure 7's IRFs, orthogonalized shocks to intraday mispricing, shares outstanding and daily discounts or premiums have no association with future innovations in Ret_d^{Und} or $VolDiff_d^{Und}$. Finally, while our results imply that future ETF prices respond to potentially simultaneous shocks in their portfolios, independent fund shocks do not affect their constituents reciprocally during later periods. Altogether, the five potential sources of non-fundamental disruption considered in Table VIII appear to have little effect on subsequent portfolio returns and trading.

B. ETF Activity and Portfolio Volatility

Ben-David, Franzoni, and Moussawi (2018) observe that ETFs increase the nonfundamental volatility of the securities in their baskets, and Da and Shive (2018) find that ETF ownership is associated with higher realizations of comovement between their underlying holdings. For fixed-income ETFs, Dannhauser (2017) shows that the liquidity of constituent securities falls as ETF ownership and activity rises. Finally, Israeli, Lee, and Sridharan (2017) report that an increase in the level of ETF ownership for a stock is associated with higher trading costs and lower benefits from information acquisition. All of these studies presume that directional order flow in fund shares impound non-fundamental fluctuations into underlying securities through an arbitrage mechanism. As such, our PVAR and graphical analyses have sought to observe, without much success, the transmission of ETF noise into portfolio share prices. Focusing only on this arbitrage mechanism, we have not examined whether or not any of the deleterious effects of fund ownership reported in prior literature might also be visible in our sample.

To this end, we propose the following OLS specification to capture potential associations between portfolio volatility and ETF activity:

$$\begin{aligned}\ln\left(Volatility_{fd}^{Und}\right) \\ &= \phi_1 \ln\left(Volatility_{fd-1}^{Und}\right) + \beta_0 + \beta_1 \left|MispSum_{fd-1}\right| \\ &+ \beta_2 \left|\Delta ShrOut_{fd-1}\right| + \beta_3 \left|DiscPrem_{fd-1}\right| \\ &+ \sum_{k=4}^{K} \beta_k Control_{kfd-1} + \alpha_t + \varepsilon_{fd},\end{aligned} \tag{6}$$

where $Volatility_{fd}^{Und}$ is measured as the standard deviation of one-minute returns for portfolio f during day d. While estimates of raw volatility are strictly non-negative and clustered near zero, the distribution of log-transformed dependent variable, $\ln\left(Volatility_{fd}^{Und}\right)$, described in Panel B of Table VII is more symmetric with mean and median of -3.14 and -3.22, respectively.

Defined previously, $MispSum_{fd}$, $\Delta ShrOut_{fd}$ or $DiscPrem_{fd}$ are intended to capture potential disruptions from intraday mispricing, shares outstanding or end-of-day discount and premium. However, Equation (6), includes the absolute values of these three variables because it is their magnitude, not direction, that should be correlated with future volatility. Furthermore, it is likely that contemporaneous realizations of $MispSum_{fd}$, $\Delta ShrOut_{fd}$ or $DiscPrem_{fd}$ are themselves responses to changes in volatility occurring during previous trading days. To isolate the independent effects of these variables, our OLS specification also includes the underlying's logged volatility from the day before.

The control variables, $Control_{kfd-1}$, appearing in Equation (6), account for a variety of factors that may also impact the portfolio's standard deviation of one-minute returns. First, the absolute value of underlying return, $\left|Ret_{fd}^{Und}\right|$, should control for material information about the constituents arriving within a given trading session. Next, daily logged dollar volume, in millions, for the portfolio, $TotVol_{fd}^{Und}$, and ETF, $TotVol_{fd}^{ETF}$, account for the documented relation between a security's volume and volatility (see Karpoff (1987) for summary). Likewise, $\left|VolDiff_{fd}^{Und}\right|$ and $\left|VolDiff_{fd}^{ETF}\right|$ capture the magnitude of directional order flow for the underlying and fund, respectively, during each trading day. Finally, as suggested in the existing literature, an ETF's liquidity, relative to that of its constituents, may encourage short term traders that introduce noise into prices. Therefore, we also include the bid-ask spread of the portfolio, $Spread_{fd}^{Und}$, and fund, $Spread_{fd}^{ETF}$, as a proxy for liquidity.

For the OLS specification, disturbances estimated from Equation (6) are likely to contain some unfavorable structure. Despite the inclusion of a lagged dependent variable, logged volatility during day d is likely related to volatility realized by the same portfolio at all other points in time due to persistent ETF- or portfolio-specific characteristics. To address this issue, we also repeat the analysis using a dynamic panel estimator (DPE) approach. The inclusion of fund panel effects can correct for any omitted variable bias associated with these characteristics. However, the OLS estimation is still biased and inconsistent because both $\ln\left(Volatility_{fd}^{Und}\right)$ and $\ln\left(Volatility_{fd-1}^{Und}\right)$ would be functions of any right-hand-side fixed effects. The DPE approach addresses this issue. Rewriting Equation (6) in terms of first differences removes these correlated panel effects,[34] along with any other time-independent variables:

$$\begin{aligned}\Delta\ln\left(Volatility_{fd}^{Und}\right)\\ &= \phi_1\Delta\ln\left(Volatility_{fd-1}^{Und}\right) + \beta_0 + \beta_1\Delta\left|MispSum_{fd-1}\right| \\ &+ \beta_2\Delta\left|\Delta ShrOut_{fd-1}\right| + \beta_3\Delta\left|DiscPrem_{fd-1}\right| \\ &+ \sum_{k=4}^{K}\beta_k\Delta Control_{kfd-1} + \alpha_t + \Delta\varepsilon_{fd}.\end{aligned} \tag{7}$$

Notice that the coefficients ϕ_1 and $\beta_1,\ldots,\beta_K$ are completely unchanged by this transformation, and that the number of estimated parameters has declined drastically with the removal of the time-

[34] During day d, $\ln\left(Volatility_{fd}^{Und}\right)$ is determined by a linear function with panel effects, γ_i, on the right-hand-side. Because $\ln\left(Volatility_{fd-1}^{Und}\right)$ also appears in the function generating $\ln\left(Volatility_{fd}^{Und}\right)$, the time-independent fixed effects are correlated with the disturbances (Greene (2008)).

independent panel effects. Instrumental variables estimation is still required, however, because $\ln\left(Volatility_{fd-1}^{Und}\right)$ is used to calculate both $\Delta\ln\left(Volatility_{fd}^{Und}\right)$ and $\Delta\ln(Volatility_{fd-1}^{Und})$ (Anderson and Hsiao (1981)).[35, 36] While an instrumental variables approach adds some complexity to the estimation of parameters, it also allows for the inclusion of endogenous regressors. Specifically, observations of $\left|Ret_{fd}^{Und}\right|$, $\left|VolDiff_{fd}^{Und}\right|$, $\left|VolDiff_{fd}^{ETF}\right|$, $TotVol_{fd}^{Und}$, $TotVol_{fd}^{ETF}$, $Spread_{fd}^{Und}$, $Spread_{fd}^{ETF}$ and $\left|MispSum_{fd}\right|$ occurring during the same trading day are jointly determined with respect to volatility.

OLS parameter estimates from Equation (6) are reported in the first column of Table IX. We partition our sample into three-year windows for the DPE analysis due to higher computational demands of the Arellano and Bover (1995) and Blundell and Bond (1998) estimation procedure. Even though OLS estimation of Equation (6) is potentially biased and inconsistent, coefficients estimated from the full sample provide a benchmark for DPE parameters estimated from partitioned data. Ultimately, the economic inferences are similar across all four columns, regardless of which estimator is used.

The second order test for serial correlation (p-values reported) was suggested by Arellano and Bond (1991) to detect any pattern in the differenced time series residuals of the individual cross sections. Four additional lagged dependent variables, $\ln(Volatility_{fd-1}^{Und})$, $\ln(Volatility_{fd-2}^{Und})$, etc., and one additional lag of the endogenous variables, are included to remove evidence of serial correlation in the first-differenced residuals and validate the moment conditions of the dynamic panel

[35] Building on the work of Arellano and Bond (1991) and Arellano and Bover (1995), Blundell and Bond (1998) propose a system estimator that uses moment conditions in which lagged differences are used as instruments for the level equation, in addition to the moment conditions of lagged levels as instruments for the differenced equation.

[36] Wintoki, Linck and Netter (2012) and Box, Davis, and Fuller (2019) use a similar dynamic panel estimator to mitigate endogeneity in an empirical corporate finance and investments setting, respectively.

estimator. The resulting p-values from the second order tests for serial correlation are below 2 in all specifications, implying that there is no evidence of persistence in the differenced residuals.

Overall, Table IX suggests end-of-day premiums or discounts do not predict future price volatility. In contrast to previous studies documenting a relation between lagged mispricing and volatility, across all of our specifications, the coefficient of $|DiscPrem_{d-1}|$ is not statistically different from zero. Considering our measure of mispricing, $|MispSum_{fd}|$, we find a positive relation between contemporaneous mispricing and volatility. As our graphical analysis in Section III suggests movement in underlying prices proceeds intraday mispricing events, it is not surprising to find increased volatility on days with mispricing. However, as with end-of-day premiums and discounts, this relation does not hold with lagged mispricing. We find no evidence that mispricing, either throughout the day or at the end of the day, transmits into the volatility of the underlying portfolio the following day. We find limited support for the notion that primary market activity flows to the underlying portfolio. Although the results are weak and inconsistent across our specifications and sample periods, the positive coefficient of $|\Delta ShrOut_{d-1}|$ suggests that the creation and redemption of shares on day $d-1$ is associated with mild increases in volatility on day d.

Our results also indicate that underlying volatility moves with volume, both contemporaneous and intertemporal. While we find that greater $TotVol_d^{ETF}$ is associated with higher levels of underlying volatility, lagged ETF volume, $TotVol_{d-1}^{ETF}$, is negatively related to future underlying volatility. Larger bid-ask spreads in the underlying, $Spread_d^{Und}$, are in line with higher contemporaneous levels of volatility. Interestingly, the coefficient of $Spread_{d-1}^{ETF}$ is also positive. We established in Section I that most ETFs are less liquid than their holdings. However, our daily results here suggest that greater liquidity in the ETF leads to lower volatility in the portfolio. This result is in contrast to previous studies that suggest superior liquidity of ETFs induce volatility into the portfolio through the arbitrage mechanism.

Ben-David, Franzoni and Maussawi (2018) examine the volatility of individual S&P 500 stocks conditional on their ETF ownership and mispricing relative to all ETFs in which they are held. It is possible our results in Table IX diverge from theirs because we examine the underlying portfolio of an ETF. We therefore replicate their analysis by also examining the daily volatility of each of the S&P 500 constituent stocks, $Volatility_{d}^{Stock}$, based on the percentage of shares outstanding held by all ETFs at the end of the previous month, $Ownership_{m}^{Stock}$ and the Ben-David, Franzoni and Maussawi (2018) relative mispricing measure, $\left|MispSum_{d-1}^{Stock}\right|$.[37] Our replication of Ben-David, Franzoni and Maussawi (2018) is found in the first column of Table X, and includes similarly defined control variables to those used in their study.[38] However, we only report the coefficients for variables of interest. We include $Volatility_{d-1}^{Stock}$, a lagged value of our dependent variable, and $CRSPRet_{d-1}^{Stock}$, the stock's lagged total return.

Similar to those of Ben-David, Franzoni and Maussawi (2018), the OLS regression results indicate that ETF ownership is positively related to the daily volatility of individual stocks in the portfolio. Additionally, mispricing between the stock and ETFs which hold it is associated with increased volatility the following two days.

This specification is incomplete across two dimensions.[39] First, it omits known determinants of volatility. Volatility is highly persistent, requiring multiple daily lags to fully model, and both total

---

[37] This measurement of ETF ownership is made from all passive ETFs with holdings reported in the CRSP Open Ended Mutual Fund Database,

[38] Controls include logged market capitalization, inverse share price, book-to-market ratio, and gross profitability (Novy-Marx (2013)) from the end of the prior month, as well as mean daily Amihud ratio (Amihud (2002)) and intradaily quoted spread over the prior month. Finally, we include cumulative returns over the prior year.

[39] One additional concern is the leptokurtic distribution of intraday volatility. Therefore, as in our daily PVAR and ETF portfolio level analyses (Table IX and Table X), we use the natural log of volatility as the dependent variable.

and directional volume directly affect volatility (Karpoff (1987)). Additionally, ETF ownership could propagate ETF shocks into underlying securities through either primary or secondary markets. The exclusion of change in ETF ownership ignores the potential impact of primary markets, either through the build up to or consequence of creation/redemption. We therefore include these additional variables.

Second, the inclusion of a lagged dependent variable in a panel setting induces endogeneity concerns as discussed in the prior section. This necessitates a dynamic panel estimation model that includes sufficient lags of the dependent and independent variables to model their endogenous nature. The coefficients resulting from the dynamic panel estimation are reported in the last three columns on Table X. To reduce computational complexities, we again partition our results using three-year sample windows from 2007 through 2015. For reference, we also include coefficients from an OLS version of this specification in the second column of Table X.

Four lagged dependent variables, $\ln\left(Volatility_{d-1}^{Stock}\right), \ldots, \ln\left(Volatility_{d-4}^{Stock}\right)$, are necessary to remove any evidence of serial correlation in the first-differenced residuals and validate the moment conditions of the dynamic panel estimator. In addition to $CRSPRet_{d}^{Stock}$, we also include its absolute value, $\left|CRSPRet_{d}^{Stock}\right|$, to better capture the arrival of new information. We control for $TotVol_{d}^{Stock}$, total daily volume in each stock, and $\left|VolDiff_{d}^{Stock}\right|$, defined as daily dollar volume in millions. Next, $\left|MispSum_{d-1}^{Stock}\right|$ measures the relative mispricing between a stock and its ETF-implied value based on the weighted average NAV of all funds that hold the security. $Ownership_{m}^{Stock}$ measures the effects of prior-month ETF ownership on volatility, while $\left|\Delta Ownership_{d-1}^{Stock}\right|$ measures the change in ETF ownership either through fund rebalancing or through creation/redemption of ETF shares, and is defined as the change in ownership from one day to the next.

Overall, the results in the last three columns of Table X suggest that prior-month ETF ownership is negatively related to individual stock volatility. Whereas Ben-David, Franzoni and Moussawi (2018) find that prior-month ETF ownership directly influences the daily volatility of individual stocks, after accounting for the fact that many of the variables are jointly determined with volatility, the coefficients of $Ownership_m^{Stock}$ across subperiods are negative. Even daily changes in ETF ownership, $\left|\Delta Ownership_{d-1}^{Stock}\right|$, have little effect on daily volatility, although the results for additional lagged values of $\left|\Delta Ownership_d^{Stock}\right|$ are significant, yet less consistent, across subperiods. In total, our results are more consistent with an alternative hypothesis presented in Ben-David, Franzoni and Moussawi (2018): the "liquidity-buffer hypothesis". This hypothesis suggests that ETFs shield the underlying securities from demand shocks in the fund.

V. Conclusion

Recent studies suggest that ETFs may attract short-term traders that introduce noise in the underlying securities prices through the arbitrage mechanism (Broman and Shum (2018), Ben-David, Franzoni and Moussawi (2018), Charupat and Miu (2011), Israeli, Lee and Sridharan (2017) and Malamud (2016)). As traders take opposing positions in the ETF and underlying shares, price pressures resulting from ETF demand may extend to the constituent securities. We directly examine this proposition by examining the minute-by-minute relation between the returns and order imbalances of ETFs and their constituent securities. Intraday impulse response functions and forecast error variance decompositions generated from a panel vector autoregression (PVAR) suggest ETF returns and order imbalance have little to no impact on underlying returns. Conversely, we find ETF returns follow underlying returns.

Identifying intraday arbitrage opportunities between ETFs and the constituents, we find little evidence that arbitrage opportunities precede trading in the underlying. Instead, arbitrage opportunities are initiated by shocks to the underlying and subsequently corrected through updates in the best bid and offer quotes. Thus, while we observe quote adjustments in response to price discrepancies, we find limited evidence of arbitrage trading. Additionally, our results indicate that bid-ask spreads remain steady during arbitrage opportunities. Not only are we unable to document arbitrage in the face of mispricing, we also find no evidence that the convergence of prices removes liquidity from the market.

After considering the primary market activity of ETFs, we find little evidence that trading the ETF propagates into the underlying in the following days. Daily PVAR results show constituent securities returns are unaffected by the prior day's ETF returns, order imbalance, share creation/redemption, or mispricing. Additionally, the volatility of the underlying securities is unaffected by share creation/redemption and only weakly responsive to mispricing. Our results stand in sharp contrast to recent studies suggesting that ETF shares serves as a shock propagation channel that allows temporary demand shocks to leave an enduring impact on constituent security prices. In fact, our results suggest that in the presence of an information event in the underlying, ETF trading may help to shield the portfolio from demand shocks by offsetting liquidity provision in the underlying securities.

References

Abrigo, M. R., & Love, I. (2016). Estimation of Panel Vector Autoregression in Stata. *The Stata Journal 16 (3)*, 778–804.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5 (1)*, 31-56.

Anderson, T. W., & Hsiao, C. (1981). Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association 76 (375)*, 598-606.

Andrews, D. W., & Lu, B. (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics 101 (1)*, 123-164.

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies 58 (2)*, 277-297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68 (1)*, 29-51.

Barndorff-Nielsen, O. E., & Shephard, N. (2006). Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. *Journal of Financial Econometrics 4 (1)*, 1-30.

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

Bessembinder, H. (1999). Trade Execution Costs on NASDAQ and the NYSE: A Post-Reform Comparison. *Journal of Financial and Quantitative Analysis 34 (3)*, 387-407.

Bessembinder, H. (2003a). Issues in assessing trade execution costs. *Journal of Financial Markets 6 (3)*, 233–257.

Bessembinder, H. (2003b). Quote-based competition and trade exectuion costs in NYSE-listed stocks. *Journal of Financial Economics 70 (3)*, 385-422.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87 (1)*, 115-143.

Box, T., Davis, R. L., & Fuller, K. P. (2019). ETF Competition and Market Quality. *Financial Management 48 (3)*, 873-916.

Box, T., Davis, R., & Fuller, K. P. (2020). The Dynamics of ETF Fees. *Financial Analysts Journal 76 (1)*, 11-18.

Brogaard, J., Hendershott, T., & Riordan, R. (2019). Price Discovery without Trading: Evidence from Limit Orders. *Journal of Finance 74 (4)*, 1621-1658.

Broman, M. S. (2016). Liquidity, style investing and excess comovement of exchange-traded fund returns. *Journal of Financial Markets 30*, 27–53.

Broman, M. S., & Shum, P. (2018). Relative Liquidity, Fund Flows and short-Term Demand: Evidence from Exchange-Traded Funds. *Financial Review 53*, 87-115.

Brown, D. C., Davies, S., & Ringgenberg, M. (2019). ETF Arbitrage, Non-Fundamental Demand, and Return Predictability. *Working paper*.

Charupat, N., & Miu, P. (2011). The pricing and performance of leveraged exchange-traded funds. *Journal of Banking & Finance 35 (4)*, 966-977.

Chen, G., & Strother, T. S. (2008). On the Contribution of Index Exchange Traded Funds to Price Discovery in the Presence of Price Limits Without Short Selling. *Working paper*.

Cremers, K. J., & Petajisto, A. (2009). How Active Is Your Fund Manager? A New Measure That Predicts Performance. *The Review of Financial Studies 22 (9)*, 3329-3365.

Da, Z., & Shive, S. (2018). Exchange Traded Funds and Asset Return Correlations. *European Financial Management 24 (1)*, 136-168.

Dannhauser, C. D. (2017). The Impact of Innovation: Evidence from Corporate Bond ETFs. *Journal of Financial Economics 125 (3)*, 537-560.

Dufee, G. R. (2018). Expected Inflation and Other Determinants of Treasury Yields. *Journal of Finance 73 (5)*, 2139-2180.

Fang, Y., & Sanger, G. C. (2012). Index Price Discovery in the Cash Market. *Working paper*.

Greene, W. H. (2008). Minimum Distance and the Generalized Method of Moments. In *Econometric Analysis Sixth Edition* (pp. 428-481). Upper Saddle River, NJ: Pearson Prentice Hall.

Hamilton, J. D. (1985). Uncovering Financial Market Expectations of Inflation. *Journal of Political Economy 93 (6)*, 1224-1241.

Hasbrouck, J. (2003). Intraday Price Formation in U.S. Equity Index Markets. *The Journal of Finance 58 (6)*, 2375-2399.

Hilscher, J., Pollet, J. M., & Wilson, M. (2015). Are Credit Default Swaps a Sideshow? Evidence That Information Flows from Equity to CDS Markets. *Journal of Financial and Quantitative Analysis 50 (3)*, 543-567.

Hollifield, B., Neklyudov, A., & Spatt, C. (2017). Bid-Ask Spreads, Trading Networks, and the Pricing of Securitizations. *The Review of Financial Studies 30 (9)*, 3048–3085.

Holtz-Eakin, D., Newey, W., & Rosen, H. S. (1988). Estimating Vector Autoregressions with Panel Data. *Econometrica 56 (6)*, 1371-1395.

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 (3)*, 313-357.

Israeli, D., Lee, C. M., & Sridharan, S. A. (2017). Is there a Dark Side to Exchange Traded Funds (ETFs)? An Information Perspective. *Review of Accounting Studies 22*, 1048-1083.

Ivanov, S. I., Jones, F. J., & Zaima, J. K. (2013). Analysis of DJIA, S&P 500, S&P 400, NASDAQ 100 and Russell 2000 ETFs and their influence on price discovery. *Global Finance Journal 24 (3)*, 171-187.

Judson, R. A., & Owen, A. L. (1999). Estimating dynamic panel data models: a guide for macroeconomists. *Economics Letters 65 (1)*, 9-15.

Karpoff, J. M. (1987). The Relation Between Price Changes and Trading Volume: A Survey. *The Journal of Financial and Quantitative Analysis 22 (1)*, 109-126.

Koijen, R. S., Lustig, H., & Van Nieuwerburgh, S. (2017). The cross-section and time series of stock and bond returns. *Journal of Monetary Economics 88,* 50-69.

Lee, J., Naranjo, A., & Velioglu, G. (2018). When do CDS spreads lead? Rating events, private entities, and firm-specific information flows. *Journal of Financial Economics 130 (3)*, 556-578.

Lee, S. S., & Mykland, P. A. (2008). Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *The Review of Financial Studies 21 (6)*, 2535–2563.

Love, I., & Zicchino, L. (2006). Financial development and dynamic investment behavior: Evidence from panel VAR. *The Quarterly Review of Economics and Finance 46 (2)*, 190-210.

Madhavan, A., & Sobczyk, A. (2016). Price Dynamics and Liquidity of Exchange-Traded Funds. *Journal of Investment Management 14 (2)*, 1-17.

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

Marshall, B. R., Nguyen, N. H., & Visaltanachoti, N. (2013). ETF arbitrage: Intraday evidence. *Journal of Banking & Finance 37 (9)*, 3486–3498.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108 (1)*, 1-28.

Petajisto, A. (2017). Inefficiencies in the Pricing of Exchange-Traded Funds. *Financial Analysts Journal 73 (1)*, 24-54.

Pukthuanthong, K., & Roll, R. (2015). Internationally Correlated Jumps. *The Review of Asset Pricing Studies 5 (1)*, 92–111.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica 48 (1)*, 1-48.

Wermers, R. (2000). Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses. *The Journal of Finance 55 (4)*, 1655-1695.

Wintoki, M. B., Linck, J. S., & Netter, J. M. (2012). Endogeneity and the dynamics of internal corporate governance. *Journal of Financial Economics 105 (3)*, 581-606.

Yu, L. (2005). Basket Securities, Price Formation, and Informational Efficiency. *Working paper*.

Table I
Summary statistics for midpoint return and order imbalance variables

Ret_t^{Und} represents the midpoint return of the underlying securities during period t defined in Equation (1). Likewise, Ret_t^{ETF}, represents the midpoint return of the ETF. $VolDiff_t^{Und}$ and $VolDiff_t^{ETF}$, defined in Equation (3), measure the difference between buy and sell volume for the underlying securities and ETF shares, respectively. Correlations are estimated across the entire sample.

	Summary Statistics								Correlations		
	Mean	P1	P10	P25	P50	P75	P90	P99	Ret_t^{Und}	Ret_t^{ETF}	$VolDiff_t^{Und}$
Panel A: 1-Minute Window											
Ret_t^{Und}	0.00	-0.17	-0.05	-0.02	0.00	0.02	0.05	0.18			
Ret_t^{ETF}	0.00	-0.18	-0.05	-0.02	0.00	0.02	0.05	0.18	81.6%		
$VolDiff_t^{Und}$	-0.01	-8.66	-7.49	-6.66	0.00	6.66	7.49	8.68	9.5%	8.4%	
$VolDiff_t^{ETF}$	0.24	-11.01	-7.41	0.00	0.00	0.00	8.01	11.28	-1.2%	-2.0%	2.2%
Panel B: 5-Minute Window											
Ret_t^{Und}	0.00	-0.38	-0.13	-0.05	0.00	0.05	0.12	0.38			
Ret_t^{ETF}	0.00	-0.38	-0.13	-0.05	0.00	0.05	0.12	0.38	91.8%		
$VolDiff_t^{Und}$	-0.09	-9.72	-8.57	-7.77	-3.45	7.73	8.55	9.74	19.9%	18.8%	
$VolDiff_t^{ETF}$	0.63	-12.64	-9.89	0.00	0.00	8.20	10.36	12.83	-0.3%	-0.3%	2.0%
Panel C: 10-Minute Window											
Ret_t^{Und}	0.00	-0.49	-0.17	-0.07	0.00	0.07	0.17	0.49			
Ret_t^{ETF}	0.00	-0.50	-0.17	-0.07	0.00	0.07	0.17	0.50	94.3%		
$VolDiff_t^{Und}$	-0.10	-10.19	-9.05	-8.26	-4.04	8.20	9.03	10.22	23.2%	22.2%	
$VolDiff_t^{ETF}$	0.87	-13.25	-10.75	-8.14	0.00	9.56	11.20	13.41	0.2%	0.4%	2.3%

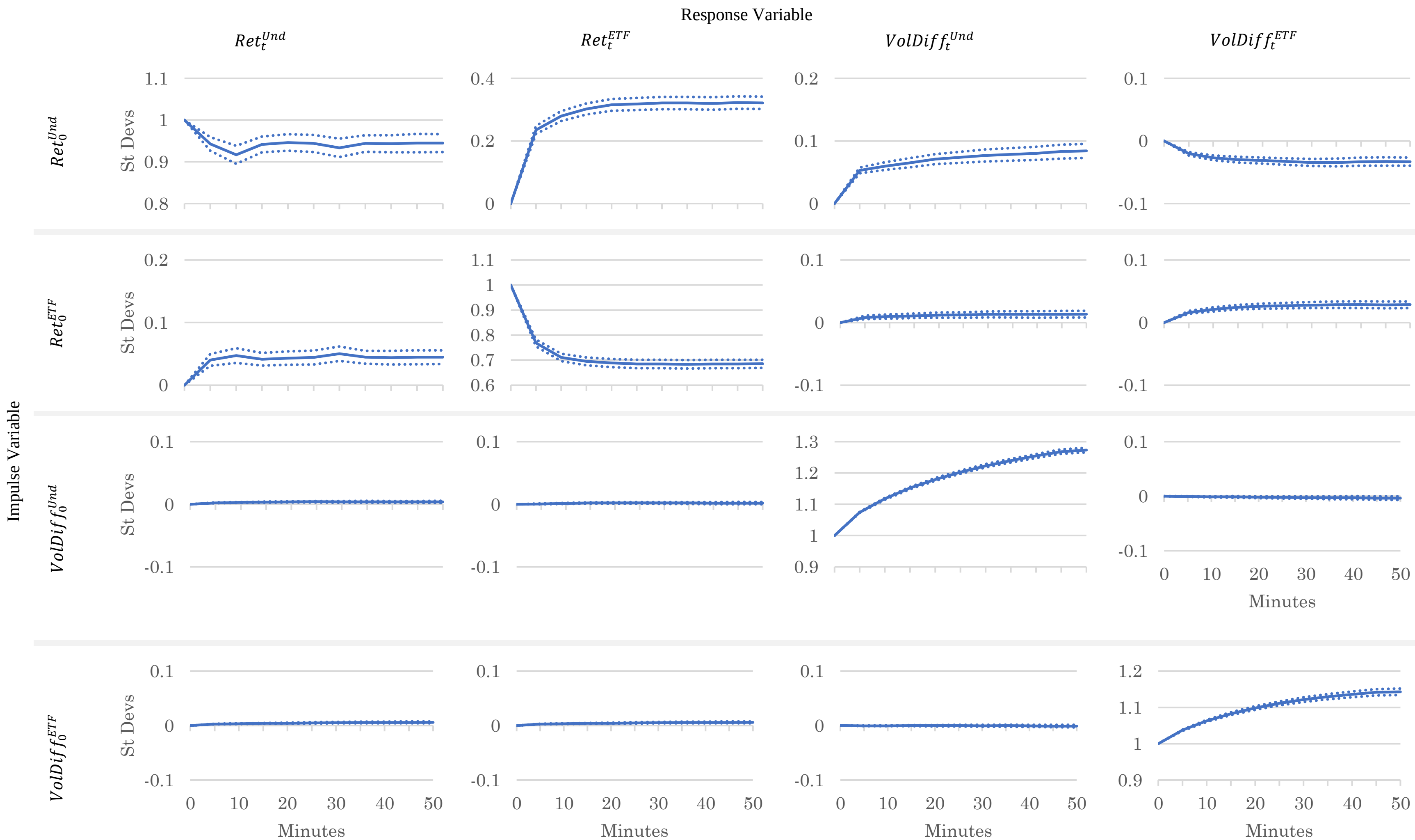


Figure 1. 5-minute cumulative impulse response functions

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 Ret_t^{Und} and Ret_t^{ETF}, as defined in Equation (1), and $VolDiff_t^{Und}$ and $VolDiff_t^{ETF}$, as defined in Equation (3), measured over 5-minute intervals. Confidence intervals, 97.5% and 2.5%, are denoted by dotted lines.

Table II
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 Ret_t^{Und} and Ret_t^{ETF}, as defined in Equation (1), and $VolDiff_t^{Und}$ and $VolDiff_t^{ETF}$, as defined in Equation (3), measured over 1-, 5-, and 10-minute intervals. Shocks are orthogonalized from top to bottom in the order presented.

Panel A: 1-Minute Window

		Response Variable			
		Ret_{10}^{Und}	Ret_{10}^{ETF}	$VolDiff_{10}^{Und}$	$VolDiff_{10}^{ETF}$
Impulse Variable	Ret_0^{Und}	99.73%	18.67%	0.19%	0.00%
	Ret_0^{ETF}	0.26%	81.32%	0.01%	0.05%
	$VolDiff_0^{Und}$	0.01%	0.00%	99.80%	0.00%
	$VolDiff_0^{ETF}$	0.00%	0.01%	0.00%	99.95%
	Observations	87,421,327			

Panel B: 5-Minute Window

		Response Variable			
		Ret_{10}^{Und}	Ret_{10}^{ETF}	$VolDiff_{10}^{Und}$	$VolDiff_{10}^{ETF}$
Impulse Variable	Ret_0^{Und}	99.83%	37.14%	0.41%	0.01%
	Ret_0^{ETF}	0.16%	62.86%	0.01%	0.01%
	$VolDiff_0^{Und}$	0.00%	0.00%	99.58%	0.00%
	$VolDiff_0^{ETF}$	0.00%	0.00%	0.00%	99.99%
	Observations	15,309,936			

Panel C: 10-Minute Window

		Response Variable			
		Ret_{10}^{Und}	Ret_{10}^{ETF}	$VolDiff_{10}^{Und}$	$VolDiff_{10}^{ETF}$
Impulse Variable	Ret_0^{Und}	99.89%	43.20%	0.51%	0.01%
	Ret_0^{ETF}	0.11%	56.80%	0.00%	0.00%
	$VolDiff_0^{Und}$	0.00%	0.00%	99.48%	0.00%
	$VolDiff_0^{ETF}$	0.00%	0.00%	0.00%	99.99%
	Observations	6,821,930			

Table III

Intraday 5-Minute Panel Vector Autoregression Forecast-Error Variance Decomposition Across Size Categories

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 (1), and $VolDiff_t^{Und}$ and $VolDiff_t^{ETF}$, 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 four subsamples divided on the market capitalization of the ETF and underlying securities. Funds and their constituents are considered large if their mean daily market capitalization falls into the top tercile of our sample during a particular month.

		Small underlying				Large underlying			
		Response Variable				Response Variable			
	Impulse Variable	Ret_{10}^{Und}	Ret_{10}^{ETF}	$VolDiff_{10}^{Und}$	$VolDiff_{10}^{ETF}$	Ret_{10}^{Und}	Ret_{10}^{ETF}	$VolDiff_{10}^{Und}$	$VolDiff_{10}^{ETF}$
Small ETF	Ret_0^{Und}	99.93%	39.37%	0.40%	0.00%	99.85%	19.76%	0.54%	0.00%
	Ret_0^{ETF}	0.07%	60.63%	0.00%	0.01%	0.15%	80.24%	0.00%	0.01%
	$VolDiff_0^{Und}$	0.00%	0.00%	99.60%	0.00%	0.00%	0.00%	99.46%	0.00%
	$VolDiff_0^{ETF}$	0.00%	0.00%	0.00%	99.99%	0.00%	0.00%	0.00%	99.99%
Large ETF	Ret_0^{Und}	99.68%	51.93%	0.35%	0.03%	99.75%	46.37%	0.46%	0.01%
	Ret_0^{ETF}	0.32%	48.07%	0.00%	0.00%	0.25%	53.63%	0.03%	0.00%
	$VolDiff_0^{Und}$	0.00%	0.00%	99.64%	0.01%	0.00%	0.00%	99.51%	0.00%
	$VolDiff_0^{ETF}$	0.00%	0.00%	0.00%	99.96%	0.00%	0.00%	0.00%	99.98%

Table IV
Intraday 5-Minute Panel Vector Autoregression Forecast-Error Variance Decomposition Across Spread Difference 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 (1), and $VolDiff_t^{Und}$ and $VolDiff_t^{ETF}$, 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 mean minute-by-minute difference in spread between the fund and its underlying portfolio over the prior 200 days.

		ETF Spread Difference Quintile				
		ETF Spread Low Relative to Underlying ⟵⟶ ETF Spread High Relative to Underlying				
Impulse Variable	Response Variable	1	2	3	4	5
Ret_0^{Und}	Ret_{10}^{Und}	99.90%	99.90%	99.74%	99.95%	99.98%
	Ret_{10}^{ETF}	59.92%	64.46%	48.68%	43.51%	32.50%
	$VolDiff_{10}^{Und}$	0.66%	0.49%	0.48%	0.59%	0.53%
	$VolDiff_{10}^{ETF}$	0.02%	0.00%	0.00%	0.01%	0.01%
Ret_0^{ETF}	Ret_{10}^{Und}	0.10%	0.10%	0.25%	0.05%	0.01%
	Ret_{10}^{ETF}	40.07%	35.53%	51.31%	56.48%	67.50%
	$VolDiff_{10}^{Und}$	0.00%	0.00%	0.01%	0.00%	0.00%
	$VolDiff_{10}^{ETF}$	0.01%	0.01%	0.00%	0.00%	0.00%
$VolDiff_0^{Und}$	Ret_{10}^{Und}	0.00%	0.00%	0.01%	0.00%	0.00%
	Ret_{10}^{ETF}	0.00%	0.00%	0.00%	0.00%	0.00%
	$VolDiff_{10}^{Und}$	99.33%	99.51%	99.51%	99.40%	99.46%
	$VolDiff_{10}^{ETF}$	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%
	$VolDiff_{10}^{ETF}$	99.97%	99.99%	99.99%	99.99%	99.99%

Table V
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 (1), and $VolDiff_t^{Und}$ and $VolDiff_t^{ETF}$, 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%

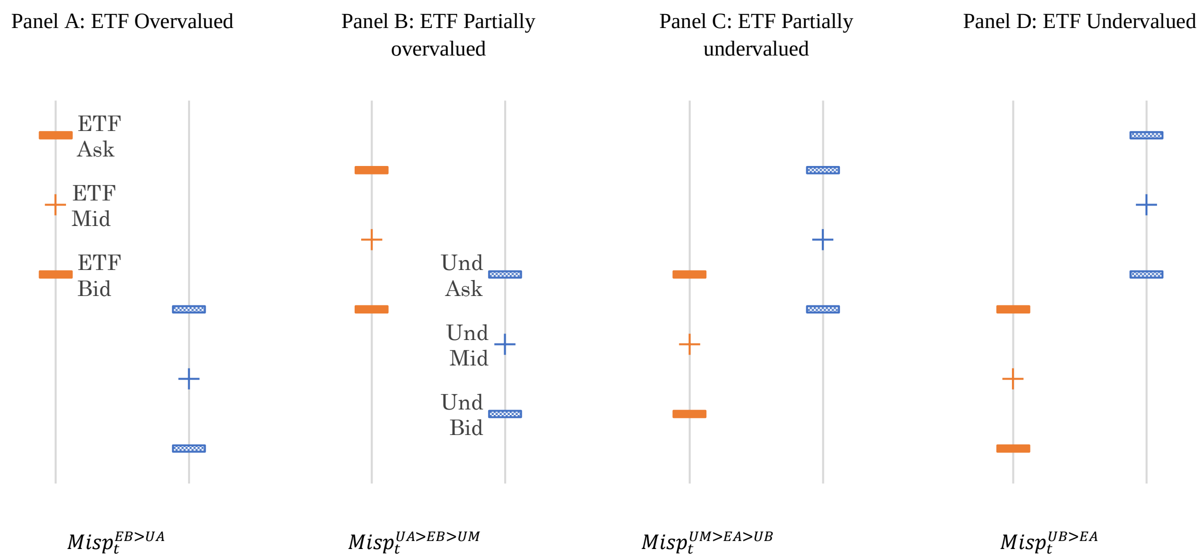


Figure 2. Mispricing events

Table VI
Frequency of mispricing events across size categories

The following table reports the frequency of mispricing events, as well as the average quoted spreads from the end of each minute t, across fund and ETF size categories. All fund-minute observations are designated as $Misp_t^{EB>UA}$, $Misp_t^{UA>EB>UM}$, $Misp_t^{UM>EA>UB}$ or $Misp_t^{UB>EA}$, in accordance with Figure 2, or classified as $NoMisp_t$ when none of these conditions is satisfied.

		Frequency					Average Quoted Spread in Basis Points	
	N	$Arb_t^{EB>UA}$	$Arb_t^{UA>EB>UM}$	$NoArb_t$	$Arb_t^{UM>EA>UB}$	$Arb_t^{UB>EA}$	Und	ETF
All Observations	94,305,900	0.86%	3.69%	91.02%	3.63%	0.79%	12.4	19.3
Small ETF	58,886,490	0.85%	1.70%	95.04%	1.65%	0.76%	13.7	26.0
Large ETF	35,419,410	0.88%	7.00%	84.34%	6.93%	0.85%	10.4	8.2
Small Underlying	58,121,310	0.83%	4.24%	89.95%	4.19%	0.78%	16.3	21.4
Large Underlying	36,184,590	0.91%	2.80%	92.75%	2.73%	0.81%	6.2	16.0
Small Und/Small ETF	41,385,630	0.80%	2.09%	94.33%	2.04%	0.74%	16.8	26.5
Small Und/Large ETF	16,735,680	0.91%	9.56%	79.12%	9.53%	0.89%	15.2	8.8
Large Und/Small ETF	17,500,860	0.97%	0.76%	96.72%	0.73%	0.81%	6.4	24.9
Large Und/Large ETF	18,683,730	0.85%	4.71%	89.02%	4.60%	0.82%	6.0	7.6

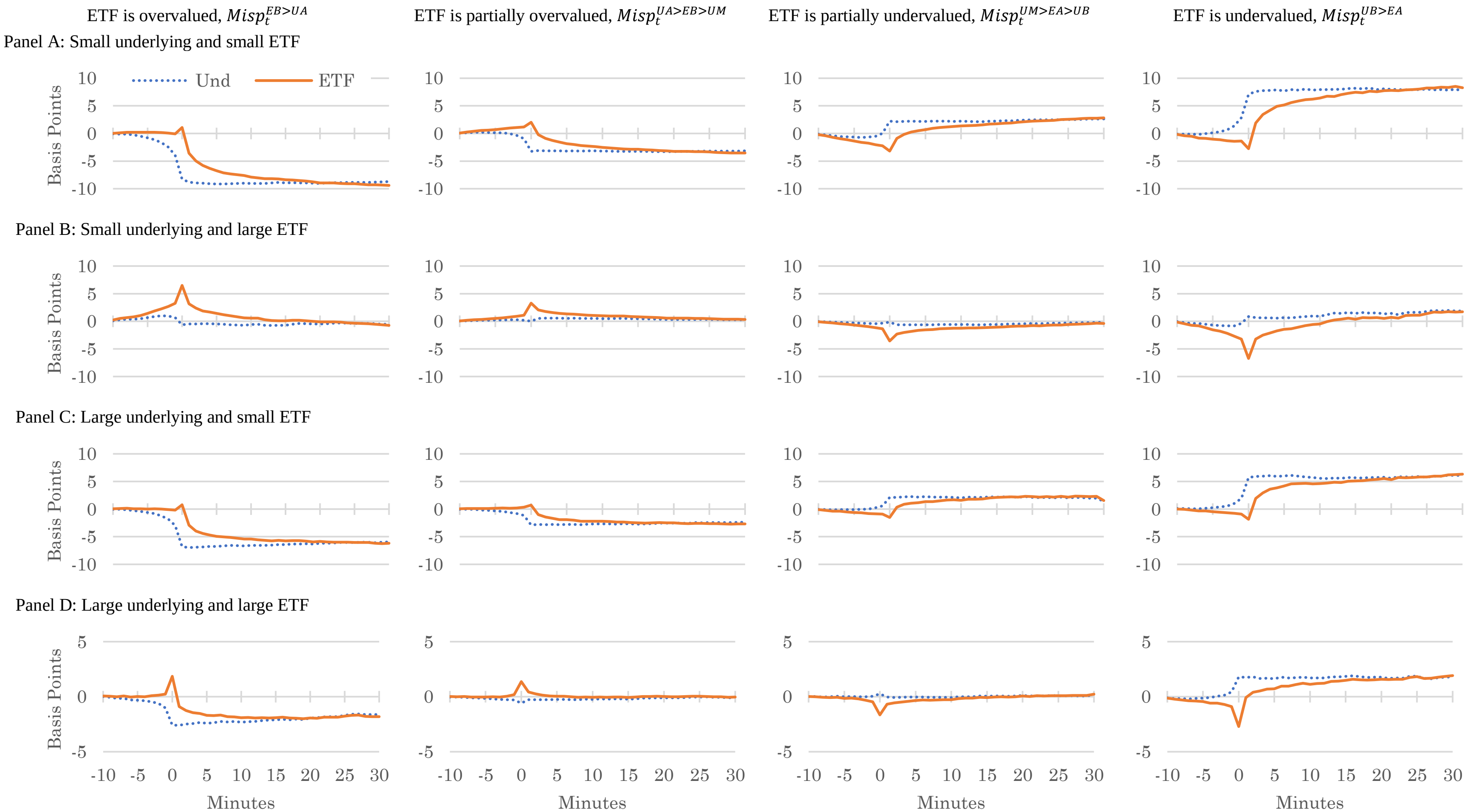


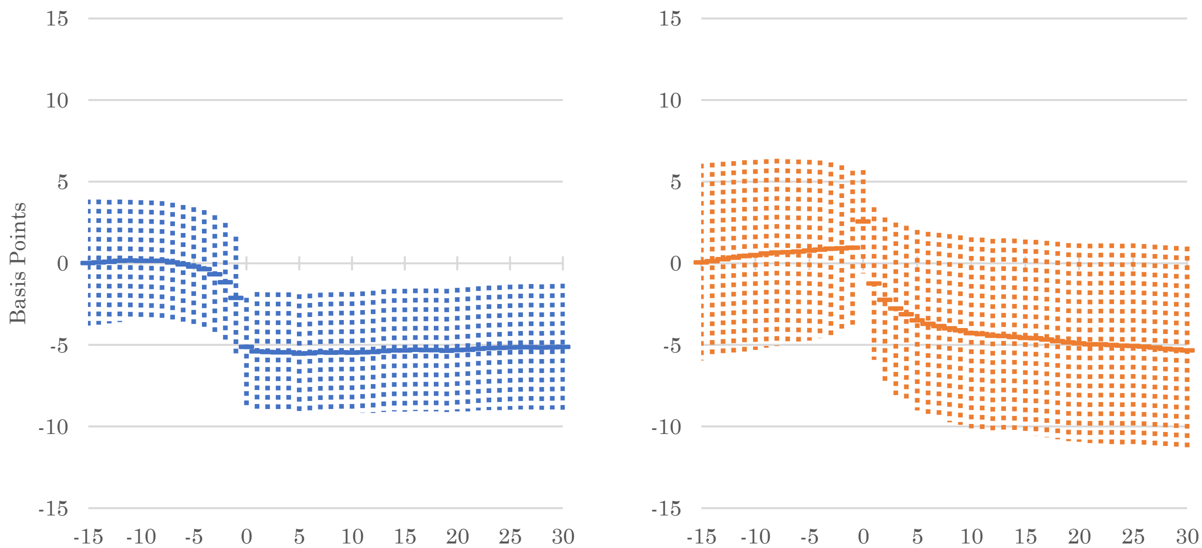
Figure 3. Average midpoint returns around arbitrage opportunities

This figure describes average midpoint returns for ETFs and constituents in a 40-minute window around arbitrage. Each panel presents average midpoint returns based on the size of the ETF and the underlying. The results are partitioned using the four mispricing conditions described in Figure (2).

Underlying

ETF

Panel A: ETF becomes overvalued, $Misp_t^{EB>UA}$



Panel B: Underlying becomes overvalued, $Misp_t^{UB>EA}$

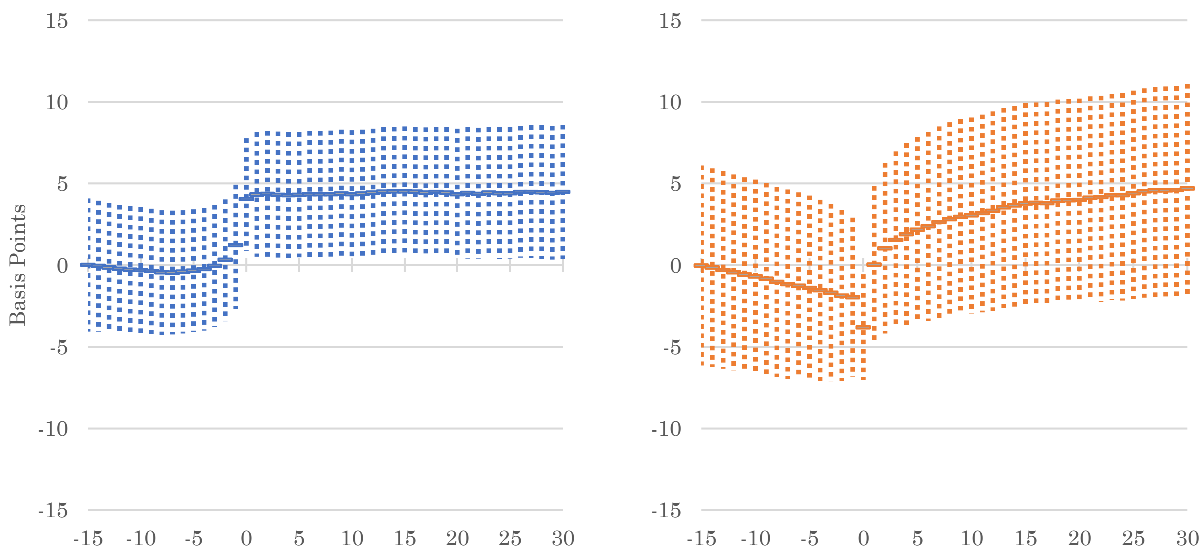


Figure 4. Cumulative bid and ask returns around arbitrage opportunities

This figure presents cumulative bid and ask returns for ETFs and constituents in the 45-minute window around arbitrage, classified using the two clear arbitrage opportunities presented in Figure 2 (i.e. when the ETF bid exceeds the underlying ask or when the underlying bid exceeds the ETF ask).

Figure 5. Average on-balance volume around arbitrage opportunities

This figure describes average on-balance volume for ETFs and constituents in a 40-minute window around arbitrage. Each panel presents average on-balance volume based on the size of the ETF and the underlying. The results are partitioned using the four mispricing conditions described in Figure (2).

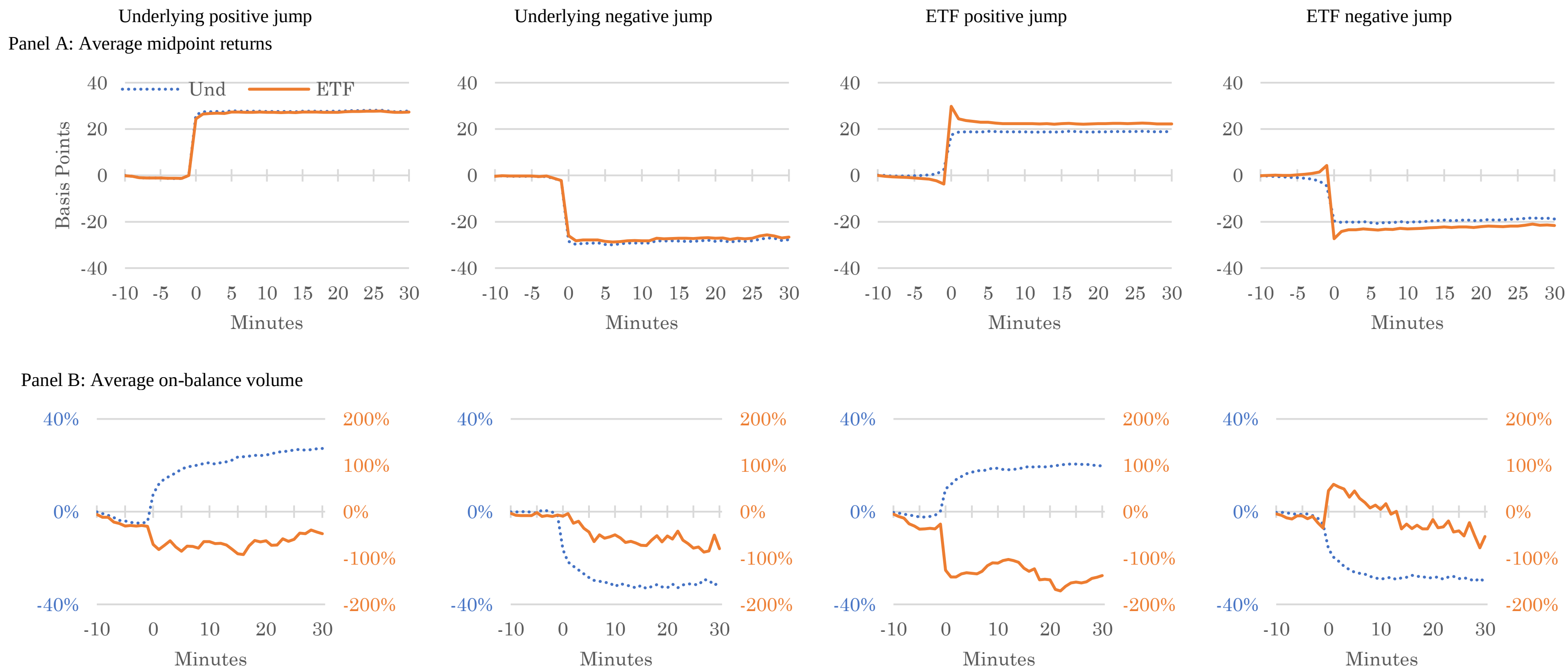


Figure 6. Average midpoint returns and on-balance volume around stochastic jumps in ETF returns

This figure describes average midpoint returns and on-balance volume for ETFs and constituents in a 40-minute window around stochastic jumps in either the fund or the underlying. Panel A shows midpoint returns. Panel B shows on-balance volume. Stochastic jumps are identified for both funds and their underlying portfolios using the Lee and Mykland (2008) jump detection measure using the prior 15 days of minute-by-minute returns. Please see Appendix D for a discussion of the measure and its calculation.

Table VII
Summary statistics for daily analysis

Ret_d^{Und} represents the cumulative midpoint return of the underlying securities over day d. Likewise, Ret_d^{ETF}, represents the cumulative midpoint return of the ETF. $VolDiff_d^{Und}$ and $VolDiff_d^{ETF}$ measure the total daily difference between buy and sell volume for the underlying securities and ETF shares, respectively. $\Delta ShrOut_d$ and $MispSum_d$ are symmetrically logged percentage change in ETF shares outstanding and logged daily sum of mispricing events, respectively. $DiscPrem_d$ is the symmetrically logged percent deviation in ETF price from NAV as reported by CRSP at the close of day d. $Volatility_d^{Und}$ is the standard deviation of underlying midpoint minute returns over day d. $TotVol_d$ is the sum of daily dollar volume in millions, and $Spread_d$ is the average daily percentage minute-by-minute quoted spread in percent, both symmetrically logged. $Volatility_d^{Stock}$ is the standard deviation of stock midpoint returns over day d. $CRSPRet_d^{Stock}$ is the daily total stock return reported in CRSP. $VolDiff_d^{Stock}$ is total daily difference between buy and sell volume of a stock. $MispSum_d^{Stock}$ is the mispricing measure proposed by Ben-David, Franzoni and Moussawi (2018), which is the dollar holdings weighted premium/discount between ETF and NAV of all ETFs which hold the stock at the end of day d. $Ownership_m^{Stock}$ is the fraction of a stock's shares outstanding held by ETFs at the end of the month, and $\Delta Ownership_d^{Stock}$ is the daily percentage change in $Ownership$. ln ($MktCap_m^{Stock}$) is the log of stock market capitalization in millions and $1/Price_m^{Stock}$ is the inverse of the stock price at the end of the month. $Amihud_m^{Stock}$ and $Spread_m^{Stock}$ are the monthly averages of the daily Amihud illiquidity measure (Amihud (2002)) and minute-by-minute quoted spreads, respectively. Panels A and B include all funds included in our intraday sample, as described in Section I. Panel C includes all S&P 500 constituent securities over our sample period.

	N	Mean	P1	P10	P25	P50	P75	P90	P99
				Panel A: PVAR Analysis					
Ret_d^{Und}	241,734	0.035	-1.71	-0.91	-0.43	0.09	0.51	0.87	1.68
Ret_d^{ETF}	241,734	0.008	-1.83	-0.95	-0.46	0.06	0.49	0.86	1.75
$VolDiff_d^{Und}$	241,544	0.020	-2.18	-1.50	-1.00	0.02	1.02	1.55	2.37
$VolDiff_d^{ETF}$	236,717	0.682	-4.61	-3.98	-2.84	2.04	3.70	4.30	4.61
$MispSum_d$	241,734	0.137	-4.45	-2.94	-1.61	0.00	1.95	3.04	4.45
$\Delta ShrOut_d$	219,943	0.033	-1.926	0.000	0.000	0.000	0.000	0.350	2.097
$DiscPrem_d$	241,734	-0.002	-0.283	-0.082	-0.035	0.000	0.030	0.071	0.280
				Panel B: Portfolio Volatility Panel Specifications					
$Volatility_d^{Und}$	241,734	0.053	0.015	0.022	0.028	0.040	0.061	0.098	0.221
ln ($Volatility_d^{Und}$)	241,734	-3.142	-4.20	-3.83	-3.58	-3.22	-2.79	-2.32	-1.51
$TotVol_d^{Und}$	241,734	4.756	1.58	2.64	3.84	4.91	5.90	6.35	7.17
$TotVol_d^{ETF}$	241,734	1.587	0.00	0.06	0.25	0.88	2.35	4.28	7.63
$Spread_d^{Und}$	241,734	-2.093	-3.75	-3.13	-2.71	-2.21	-1.58	-0.93	0.81
$Spread_d^{ETF}$	241,734	-2.025	-4.67	-3.32	-2.66	-2.01	-1.41	-0.85	1.30
				Panel C: Individual Stock Volatility Panel Specifications					
$Volatility_d^{Stock}$	1,196,318	0.085	0.026	0.038	0.049	0.066	0.097	0.147	0.355
ln ($Volatility_d^{Stock}$)	1,196,318	-2.638	-3.64	-3.26	-3.02	-2.71	-2.33	-1.92	-1.03
$CRSPRet_d^{Stock}$	1,196,318	0.000	-0.066	-0.022	-0.009	0.000	0.010	0.022	0.068
$VolDiff_d^{Stock}$	1,195,572	0.004	-0.224	-0.105	-0.052	0.002	0.057	0.115	0.250
$MispSum_d^{Stock}$	1,196,318	0.095	0.001	0.017	0.036	0.062	0.102	0.175	0.664
$\Delta Ownership_d^{Stock}$	1,196,318	0.00000	-0.00099	-0.00004	-0.00001	0.00000	0.00001	0.00005	0.00124
$Ownership_m^{Stock}$	1,196,318	2.630	0.08	0.74	1.67	2.43	3.48	4.51	7.18
ln ($MktCap_m^{Stock}$)	1,196,318	9.469	7.35	8.25	8.79	9.42	10.13	10.88	11.49
$1/Price_m^{Stock}$	1,196,318	0.033	0.005	0.011	0.015	0.023	0.037	0.062	0.194
$Amihud_m^{Stock}$	1,196,318	0.0002	0.0000	0.0000	0.0001	0.0001	0.0002	0.0004	0.0012
$Spread_m^{Stock}$	1,196,318	0.070	0.020	0.026	0.033	0.046	0.072	0.126	0.468

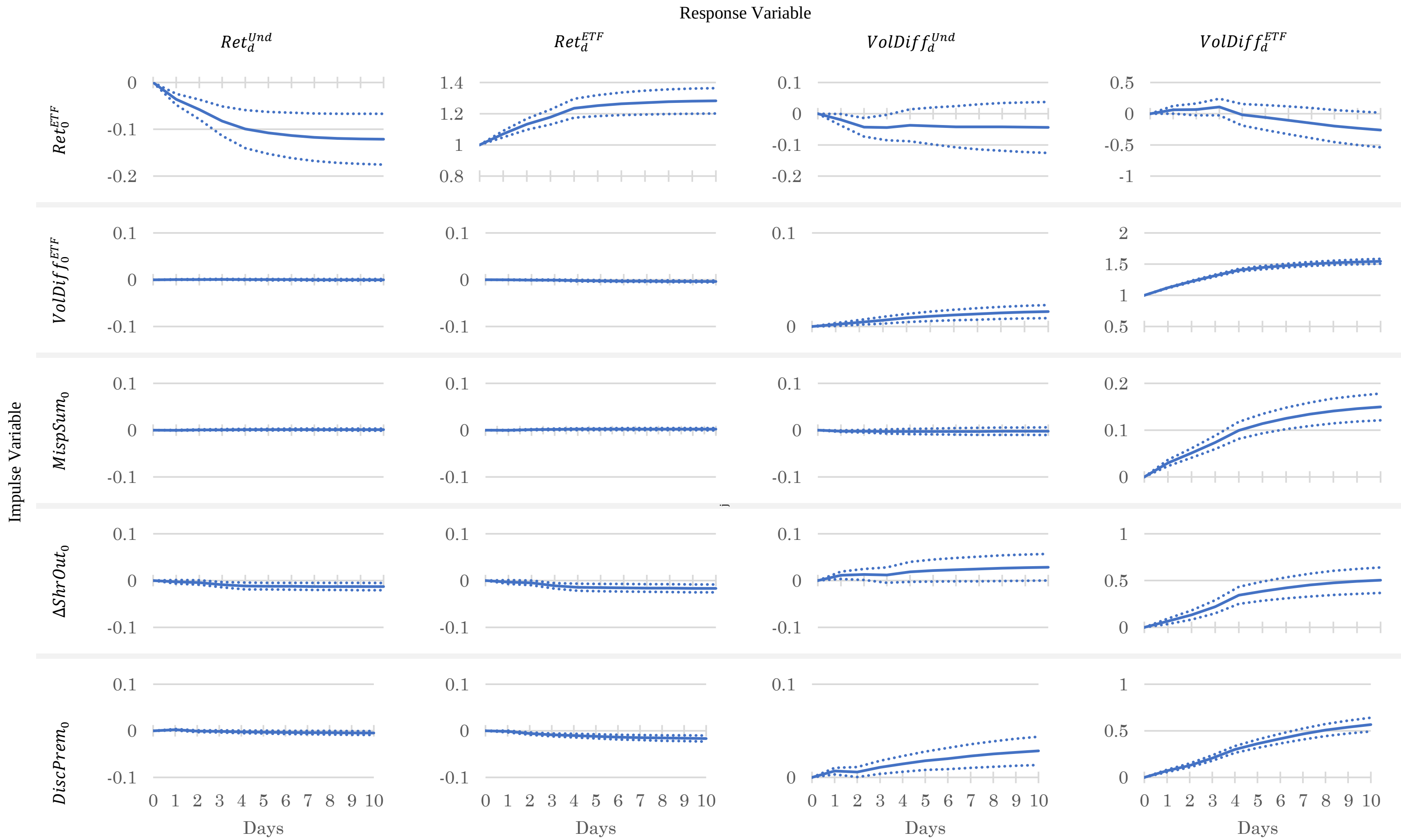


Figure 7. Daily cumulative impulse response functions

This figure depicts the cumulative impulse response functions (IRF) after expanding the PVAR framework introduced in Equation (4) to a daily setting. These IRFs describe each dependent variable's evolution following a one standard deviation shock in the associated impulse variable. The seven variables included are Ret_d^{Und}, Ret_d^{ETF}, $VolDiff_d^{Und}$ and $VolDiff_d^{ETF}$, as in Figure 1, and $MispSum_d$, $\Delta ShrOut_d$, and $DiscPrem_d$, as defined in Table VII. Confidence intervals, 97.5% and 2.5%, are denoted by dotted lines.

Table VIII

Daily 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 days. The seven variables included are Ret_d^{Und}, Ret_d^{ETF}, $VolDiff_d^{Und}$ and $VolDiff_d^{ETF}$, as in Figure 1, and $MispSum_d$, $\Delta ShrOut_d$, and $DiscPrem_d$, as defined in Table VII. Shocks are orthogonalized from top to bottom in the order presented.

		Response Variable						
		Ret_d^{Und}	Ret_d^{ETF}	$VolDiff_d^{Und}$	$VolDiff_d^{ETF}$	$MispSum_d$	$\Delta ShrOut_d$	$DiscPrem$
Impulse Variable	Ret_0^{Und}	99.75%	58.56%	1.63%	0.46%	0.05%	0.27%	0.17%
	Ret_0^{ETF}	0.16%	41.29%	0.03%	0.07%	0.29%	0.01%	0.68%
	$VolDiff_0^{Und}$	0.02%	0.02%	98.22%	0.06%	0.01%	0.02%	0.01%
	$VolDiff_0^{ETF}$	0.00%	0.02%	0.04%	89.19%	0.79%	1.74%	2.30%
	$MispSum_0$	0.00%	0.01%	0.01%	0.22%	98.20%	0.13%	0.16%
	$\Delta ShrOut_0$	0.01%	0.01%	0.01%	0.07%	0.04%	94.07%	0.19%
	$DiscPrem_0$	0.05%	0.09%	0.08%	9.94%	0.63%	3.76%	96.51%

Table IX
Daily Volatility of Underlying Portfolio

Each specification is estimated with Arellano and Bover (1995) and Blundell and Bond (1998) dynamic panel estimation methodology with bias-corrected robust variance-covariance estimates of the model parameters. All of the independent variables are included as endogenous instruments in the dynamic panel estimation. Variables are defined in Table VII. Coefficients marked * and ** are significant at the 5% and 1% level, respectively, and t-statistics are reported in parenthesis.

	Ordinary Least Squares	Dynamic Panel Estimator		
	2007-2015	2007-2009	2010-2012	2013-2015
ln ($Volatility_{d-1}^{Und}$)	0.318** (23.19)	0.242** (10.59)	0.319** (27.84)	0.225** (19.76)
$\lvert Ret_d^{Und}\rvert$		0.0822** (5.766)	0.0530** (9.905)	0.232** (11.11)
$\lvert Ret_{d-1}^{Und}\rvert$	0.0406** (11.27)	0.0505** (11.48)	0.0466** (15.55)	0.0663** (14.58)
$\lvert VolDiff_d^{Und}\rvert$		-0.0111** (-4.570)	-0.0102** (-7.388)	-0.0107** (-5.101)
$\lvert VolDiff_{d-1}^{Und}\rvert$	-0.000603 (-0.499)	0.000253 (0.127)	0.00327* (2.439)	-0.00881** (-4.501)
$\lvert VolDiff_d^{ETF}\rvert$		-0.00618** (-4.408)	-0.00301** (-4.081)	-0.00885** (-5.233)
$\lvert VolDiff_{d-1}^{ETF}\rvert$	-0.000863* (-2.476)	0.000826 (0.772)	-0.000718 (-0.870)	-0.00442** (-3.720)
$TotVol_d^{Und}$		0.0113 (0.999)	0.0226** (4.599)	0.0128 (1.529)
$TotVol_{d-1}^{Und}$	-0.000436 (-0.504)	0.00212 (0.186)	-0.0220** (-4.709)	-0.0125 (-1.556)
$TotVol_d^{ETF}$		0.0157** (5.501)	0.0122** (5.642)	0.0162** (6.537)
$TotVol_{d-1}^{ETF}$	0.000926 (1.883)	-0.00950** (-3.037)	-0.00549* (-2.447)	-0.0103** (-3.638)
$Spread_d^{Und}$		0.0282** (5.724)	0.00450 (1.458)	0.0184** (4.125)
$Spread_{d-1}^{Und}$	-0.00105 (-0.759)	0.00230 (0.454)	-0.00302 (-1.210)	-0.0203** (-4.756)
$Spread_d^{ETF}$		0.0142** (6.403)	0.00625** (3.033)	0.00439 (1.777)
$Spread_{d-1}^{ETF}$	0.00297** (4.362)	0.00725** (3.635)	0.00863** (5.103)	0.00732** (2.907)
$\lvert MispSum_d\rvert$		0.00301* (2.150)	0.00248** (2.918)	0.00272* (2.150)
$\lvert MispSum_{d-1}\rvert$	0.000494 (1.507)	0.00108 (0.985)	0.000204 (0.298)	0.000129 (0.124)
$\lvert \Delta ShrOut_{d-1}\rvert$	0.00315** (3.829)	0.00482 (1.886)	0.00101 (0.562)	0.00839** (2.868)
$\lvert DiscPrem_{d-1}\rvert$	0.00737 (1.345)	0.0141 (1.839)	-0.0144 (-1.267)	0.00536 (0.282)
Lagged Dependent Variables	5	5	5	5
ETF Panel Effects	No	Yes	Yes	Yes
Daily Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.792			
AR(2) Test		0.543	1.829	-0.738
Observations	191,343	50,286	62,129	74,295

Table X
Daily Volatility of Individual Stocks

Each specification is estimated with Arellano and Bover (1995) and Blundell and Bond (1998) dynamic panel estimation methodology with bias-corrected robust variance-covariance estimates of the model parameters. All of the independent variables are included as endogenous instruments in the dynamic panel estimation. Variables are defined in Table VII. Coefficients marked * and ** are significant at the 5% and 1% level, respectively, and t-statistics are reported in parenthesis.

	Ordinary Least Squares		Dynamic Panel Estimator		
		2007-2015	2007-2009	2010-2012	2013-2015
$Volatility_{d-1}^{Stock}$	0.474** (19.88)				
ln ($Volatility_{d-1}^{Stock}$)		0.303*** (30.12)	0.185** (35.60)	0.187** (36.01)	0.134** (24.73)
ln ($Volatility_{d-2}^{Stock}$)		0.159*** (23.81)	0.101** (28.93)	0.107** (27.15)	0.0706** (18.11)
ln ($Volatility_{d-3}^{Stock}$)		0.131*** (22.93)	0.107** (26.77)	0.118** (30.19)	0.0933** (21.02)
ln ($Volatility_{d-4}^{Stock}$)		0.136*** (28.99)	0.0897** (29.60)	0.0950** (32.51)	0.0835** (23.50)
$\|\Delta Ownership_{d-1}^{Stock}\|$		1.635 (1.490)	-2.148 (-1.213)	1.759 (1.582)	0.956 (0.783)
$\|\Delta Ownership_{d-2}^{Stock}\|$		0.248 (0.205)	2.143 (1.153)	1.026 (0.776)	-2.823** (-2.856)
$\|\Delta Ownership_{d-3}^{Stock}\|$		2.357** (2.002)	3.270 (1.832)	0.855 (0.654)	6.564** (2.883)
$Ownership_{m}^{Stock}$	0.0325** (3.902)	0.00625*** (4.367)	-0.0183** (-4.183)	-0.0131** (-4.796)	-0.00841** (-2.708)
$CRSPRet_{d}^{Stock}$			-0.382** (-10.01)	-0.248** (-5.147)	-0.552** (-7.120)
$CRSPRet_{d-1}^{Stock}$	-0.501* (-2.215)	-0.172*** (-5.219)	-0.399** (-10.09)	-0.191** (-4.330)	-0.651** (-9.190)
$CRSPRet_{d-2}^{Stock}$		-0.174*** (-6.535)	-0.266** (-8.913)	-0.304** (-7.478)	-0.443** (-7.391)
$CRSPRet_{d-3}^{Stock}$		-0.139*** (-5.296)	-0.199** (-6.455)	-0.134** (-3.762)	-0.272** (-4.244)
$\|CRSPRet_{d}^{Stock}\|$			0.329** (4.493)	1.308** (10.37)	4.265** (16.61)
$\|CRSPRet_{d-1}^{Stock}\|$		0.976*** (17.50)	1.141** (18.32)	1.570** (18.74)	2.916** (19.44)
$\|CRSPRet_{d-2}^{Stock}\|$		0.281*** (7.256)	0.717** (13.53)	0.722** (10.06)	1.313** (12.47)
$\|CRSPRet_{d-3}^{Stock}\|$		0.197*** (5.371)	0.605** (11.18)	0.459** (7.236)	0.932** (8.860)
$TotVol_{d}^{Stock}$			0.333** (48.62)	0.342** (60.17)	0.306** (35.66)
$TotVol_{d-1}^{Stock}$		0.0474*** (12.63)	-0.0212** (-3.298)	-0.0262** (-5.955)	0.0137** (3.025)
$TotVol_{d-2}^{Stock}$		-0.0311*** (-11.29)	-0.0470** (-7.983)	-0.0523** (-14.42)	-0.0351** (-9.400)
$TotVol_{d-3}^{Stock}$		-0.0244*** (-10.46)	-0.0738** (-16.36)	-0.0678** (-21.77)	-0.0678** (-17.13)
$\|VolDiff_{d}^{Stock}\|$			-0.0487** (-20.23)	-0.0450** (-28.53)	-0.0360** (-16.50)
$\|VolDiff_{d-1}^{Stock}\|$		-0.00295*** (-6.080)	-0.0107** (-4.753)	-0.00572** (-3.959)	-0.0131** (-5.801)
$\|VolDiff_{d-2}^{Stock}\|$		-0.000659 (-1.522)	-0.00889** (-3.933)	-0.00233 (-1.452)	-0.00967** (-4.761)

$\lvert VolDiff_{d-3}^{Stock} \rvert$		0.000236 (0.577)	-0.00872** (-3.809)	0.00145 (0.947)	-0.00617** (-3.140)
$\lvert MispSum_{d-1}^{Stock} \rvert$	0.373** (3.670)	0.0202* (1.912)	-0.00517 (-0.784)	0.0283* (1.985)	0.0404* (2.302)
$\lvert MispSum_{d-2}^{Stock} \rvert$	0.302** (3.945)	0.0179*** (3.003)	0.00585 (1.007)	0.0264 (1.859)	0.0445* (2.476)
$\lvert MispSum_{d-3}^{Stock} \rvert$		0.0162** (2.105)	-0.0175** (-3.420)	0.0361* (2.205)	0.142** (7.171)
Lagged Dependent Variables	1	4	4	4	4
Firm-Specific Controls	Yes	Yes	Yes	Yes	Yes
Firm Panel Effects	No	No	Yes	Yes	Yes
Daily Fixed Effects	Yes	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.777	0.857			
AR(2) Test			-0.650	-1.936	-0.466
Observations	1,121,023	1,118,101	372,553	372,226	368,860

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.[40] 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, $DolVal_{fh}$:

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

$$iTNAV_{fh} = \frac{DolVal_{fh}}{\%Val_{fh}}. \tag{8}$$

If an ETF consists of 300 holdings, equation (8) 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.[41] 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.[42]

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

[41] 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.
[42] 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.[43]

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.

[43] 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, $ShrOut_f$:

$$DNAV_f = \frac{\sum_h Price_h \times Shares_{fh}}{ShrOut_f}, \tag{9}$$

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 (9) adjusts accordingly. The total net asset value, the numerator of Equation (9), 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 (9) by the ratio of ETF shares outstanding and the sum of constituent shares:

$$DNAV_f \times \frac{ShrOut_f}{\sum_h Shares_{fh}} = \sum_h \frac{Price_h \times Shares_{fh}}{\sum_h Shares_{fh}}. \tag{10}$$

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, $\sum_h \left[\left(Price_h \times Shares_{fh} \right) / \sum_h Shares_{fh} \right]$.

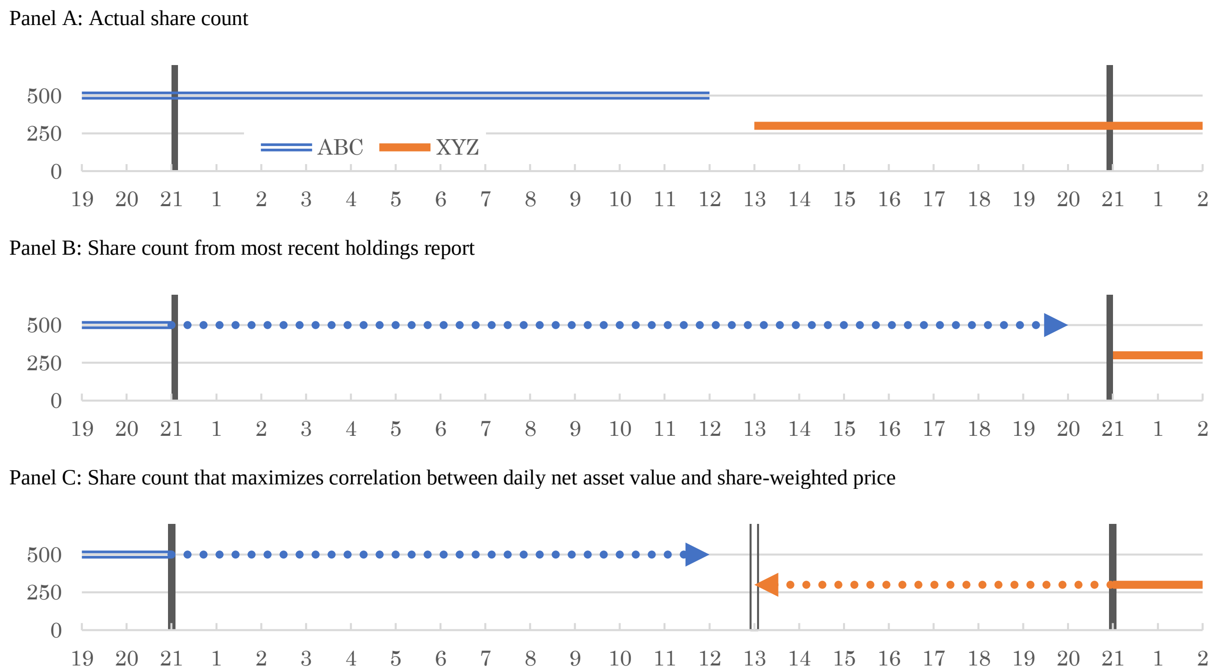


Figure A-1. Share count interpolation example

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 (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 A, 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 (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.

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

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 $MisAsk_t^{Und}$

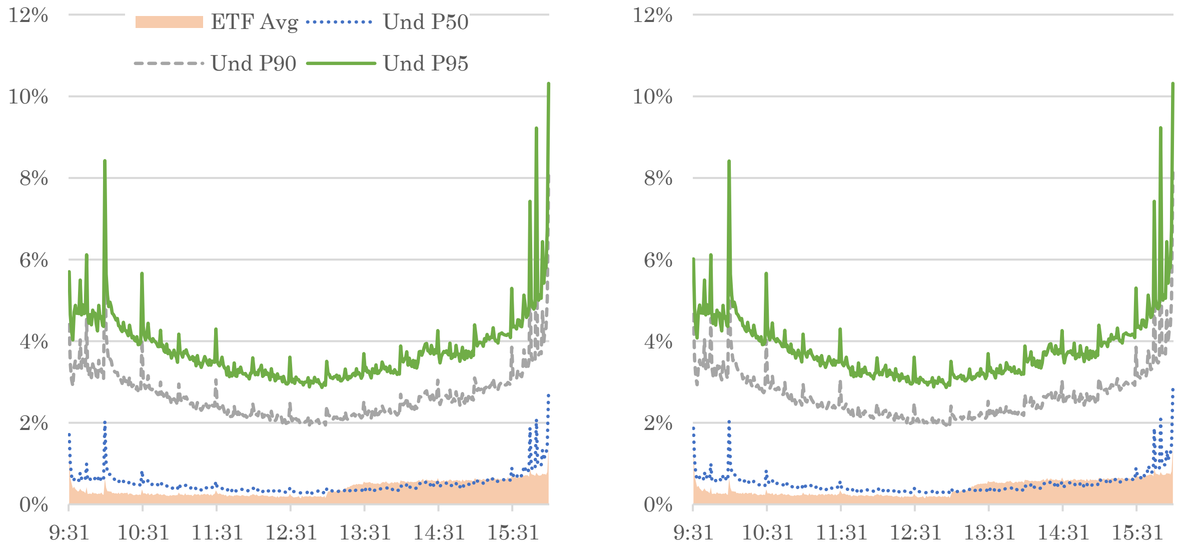


Figure B-1. Percentage of missing NBBO quotes during the trading day

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.

spread is negative or greater than 50% of the quoted midpoint.[44] 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. 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

[44] 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.

$MisAsk_t^{ETF}$, 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 (10). 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:

$$
\begin{aligned}
Und_f^{Bid} \times \frac{ShrOut_f}{\sum_h Shares_{fh}} &= \sum_h \frac{Bid_h \times Shares_{fh}}{\sum_h Shares_{fh}} \\
Und_f^{Ask} \times \frac{ShrOut_f}{\sum_h Shares_{fh}} &= \sum_h \frac{Ask_h \times Shares_{fh}}{\sum_h Shares_{fh}},
\end{aligned}
\tag{11}
$$

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:

$$
\begin{aligned}
&1.\quad ETF_f^{Ask} \geq Adj_{Max} \times \sum_h \frac{Bid_h \times Shares_{fh}}{\sum_h Shares_{fh}} \\
&2.\quad ETF_f^{Bid} \leq Adj_{Min} \times \sum_h \frac{Ask_h \times Shares_{fh}}{\sum_h Shares_{fh}} \\
&3.\quad \frac{ETF_f^{Bid}}{\sum_h \frac{Ask_h \times Shares_{fh}}{\sum_h Shares_{fh}}} = Adj_{Min} \leq \left(\frac{ShrOut_f}{\sum_h Shares_{fh}}\right)^{-1} \leq Adj_{Max} = \frac{ETF_f^{Ask}}{\sum_h \frac{Bid_h \times Shares_{fh}}{\sum_h Shares_{fh}}},
\end{aligned}
\tag{12}
$$

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 (10), gives:

$$\frac{ShrOut_f}{\sum_h Shares_{fh}} = \frac{\sum_h Price_h \times \frac{Shares_{fh}}{\sum_h Shares_{fh}}}{DNAV_f}. \tag{13}$$

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. 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 e.g. Barndorff-Nielsen and Shephard (2006)). More formally, their jump detection measure for a return series is defined as:

$$L_t = \frac{log\left({}^{S_t}/_{S_{t-1}}\right)}{\widehat{\sigma_t}}, \tag{14}$$

where

$$\widehat{\sigma_t}^2 = \left(\frac{1}{k-2}\right) \sum_{i=t-k+2}^{t-1} \left|log\left({}^{S_i}/_{S_{i-1}}\right)\right| \left|log\left({}^{S_{i-1}}/_{S_{i-2}}\right)\right|, \tag{15}$$

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:

$$\theta = \beta^* S_n + C_n, \tag{16}$$

where

$$\beta^* = -\log(-\log(\alpha)), \tag{17}$$

and

$$C_n = \frac{\left(2log(n)\right)^{1/2}}{c} - \frac{log(\pi) + log\left(log(n)\right)}{2c\left(2log(n)\right)^{1/2}}, \tag{18}$$

and

$$S_n = \frac{1}{c\big(2log(n)\big)^{1/2}}, \tag{19}$$

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 observation daily, 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 III 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).