G Model
QUAECO-1079; No. of Pages 15
ARTICLE IN PRESS

The Quarterly Review of Economics and Finance xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

The Quarterly Review of Economics and Finance

journal homepage: www.elsevier.com/locate/qref

Equivalent volume and comovement

Arsenio Staer[a,*], Pedro Sottile[b]

[a] Department of Finance, California State University at Fullerton, Fullerton, SGMH 5159, Fullerton, CA 92831, United States
[b] Department of Accounting and Finance, University of Wisconsin, Eau Claire, SSS 300C, Eau Claire, WI 54701, United States

ARTICLE INFO

Article history:
Received 30 November 2016
Received in revised form 21 October 2017
Accepted 4 November 2017
Available online xxx

JEL classifications:
G12
G14
G23

Keywords:
ETF
Liquidity
Arbitrage
Comovement
Trading volume
Asset pricing

ABSTRACT

We introduce a new indicator of relative liquidity, equivalent volume (EV), based on the amount of a stock traded indirectly through its inclusion in ETFs. We hypothesize that the EV of an ETF component stock is related to its comovement with other component stocks through the relative liquidity channel under trading caused by arbitrage. Using daily ETF holdings and several comovement estimators, we find that a one-unit increase in daily equivalent volume is associated with increase in comovement ranging from 1.1% to 27.6%. Our findings contribute to the literature on trading volume, liquidity and comovement by relating arbitrage-induced trading pressure to the underlying stock comovement.

1. Introduction

In this paper, we investigate the relation among arbitrage-induced correlated price impact, relative liquidity and underlying asset comovement. The exponential increase in ETF trading volume, number of funds, and their assets under management, provides a novel opportunity to study those effects. For instance, on August 6, 2012, *Bloomberg* reported that for the first time the dollar volume of S&P 500 tracking Exchange Traded Funds (ETFs) had reached a 12-month average of $28 billion a day. This average trading volume represented 98% of the trading in the underlying stocks.[1] In other words, trading in S&P 500 ETFs essentially achieved parity with the trading of the S&P 500 stocks. More generally, one third of the volume on the U.S. stock exchanges in the period from 2011[2] to 2016[3] was due to trading of ETFs, including sector and bond funds.

ETFs have a distinguishing feature that the ETF price and NAV are kept in line by the intraday arbitrage and the daily creation-redemption mechanism at the market close. Usually, NAV values are updated intraday and can be compared with an ETF's share price almost instantly through their Intraday Indicative Value (IIV), also called intraday value of the ETF assets (iNAV).[4] Sufficiently large deviations from the NAV would prompt arbitrageurs to intervene by buying (selling) the underlying asset basket and selling (buying) the ETF shares, hence reaping arbitrage profits until the spread between the NAV and the ETF price is reduced below transaction costs.[5] Furthermore, the NAV-price arbitrage is strengthened by the creation-redemption mechanism that allows Authorized Participants (APs), usually large broker-dealers and institutions, to exchange the underlying assets with the fund for the equivalent ETF shares at the market close — an operation also called an "in-

* Corresponding author.
E-mail addresses: astaer@fullerton.edu (A. Staer), sottilpd@uwec.edu (P. Sottile).

[1] See "ETFs Poised to Exceed Trade in S&P 500 as Spiders Beat Apple", Bloomberg, August 6, 2012.

[2] See "ETF trading volumes surge in market turmoil", FT.com, April 10, 2011.

[3] See "NYSE ARCA ETF Report Q4 2017 ", NYSE.com, 2017.

[4] Intraday Indicative Value (IIV) or intraday value of the ETF assets (iNAV) is published every 15 s by the exchange and is an indicator of approximate value of the ETF assets using most recent market prices. IIV is similar in nature to the net asset value, however the former is calculated every 15 s by the exchange and is a non-binding indicator, while the latter is calculated around 4:00PM by the ETF and is binding in the fund's creation-redemption activity with APs.

[5] Petajisto (2013) studies ETF premiums and finds that the magnitude of the spread between ETF price and NAV is relatively small especially for the broad-sector ETFs, but the volatility of that spread entails significant arbitrage opportunities most likely exploited by arbitrageurs on the intraday basis.

https://doi.org/10.1016/j.qref.2017.11.001
1062-9769/

Please cite this article in press as: Staer, A., & Sottile, P. Equivalent volume and comovement. *The Quarterly Review of Economics and Finance* (2017), https://doi.org/10.1016/j.qref.2017.11.001

kind" transaction. Hence, APs need not to wait until prices converge but can transact the second leg of the arbitrage trade directly with the ETF, thus reducing potential inventory holding risks.

These small and frequent waves of arbitrage-caused *en masse* buying and selling of the ETF and the underlying stocks occur during the day, exerting price pressure on the underlying prices — not unlike rocking a boat — and making them covary. This covariation is unrelated to the commonalities of stock fundamentals, and thus represents excess comovement. This excess comovement is temporary given the nature of the originating shock and quickly decays if there are no subsequent arbitrager-induced shocks. However, evidence suggests (Petajisto, 2013) that arbitrage opportunities are very frequent and are exploited on a continuous basis and, hence, the excess comovement generated by these trades represents a non-trivial amount. One of the corollaries of this excess comovement is the reduction of the diversification benefits for investors and market participants.

Furthermore, the basic intuition behind the ETF arbitrage implies that the ETF price will converge with the NAV and the effect on the NAV will be minimal. However, this is not always the case — in the situations when the liquidity of the underlying assets is lower relative to the liquidity of the ETF, the arbitrage-related trading will exert more price pressure on the less liquid instrument (Kyle, 1985; Amihud & Mendelson, 1986), in this case the underlying stocks. For large ETFs, the turnover and the trading volume, which is often associated with liquidity proxies (Amihud, 2002; Holden, Jacobsen, & Subrahmanyam, 2014), are generally much higher than those of the underlying stocks, suggesting that the ETF liquidity is usually higher than that of the underlying stocks. Hence, a measure of relative liquidity of the ETF and the underlying should be associated with the differential price impact of the arbitrage trade and, consequently, the comovement of the underlying stocks.

In order to study the effects of liquidity and arbitrage-related trading on excess comovement, we develop a stock-level indicator called "equivalent volume" (EV) that attempts to capture the relative liquidity of the individual stock vs. the liquidity of the (same) stock as part of the ETF in the arbitrage framework. This indicator is a normalized ratio of weighted trading volumes of the ETF and the underlying stock.

This indicator of relative liquidity, equivalent trading volume, is economically significant, ranging from 5% to 70% of the daily volume of an average stock in our sample. Put another way, the average stock's weight in the trading volume of the ETF that it belongs to, comprises up to 70% of the stock's own trading volume. Although already large, we underestimate equivalent volume because a majority of stocks on U.S. exchanges belong to several ETFs. For instance, Google is a component in two ETFs in our sample and also in other smaller 109 domestic equity ETFs[6] not included in the sample. Hence, the aggregated equivalent volume for the stocks in our sample is most likely larger than the 5% to 75% range mentioned earlier.

We hypothesize that an increase in the weighted liquidity of the ETF relative to the liquidity of the underlying stock, represented by the equivalent trading volume, is related to the increase in the comovement of the underlying stock with the rest of the stocks in the ETF. We do not model arbitrage trading directly, but rather measure the impact of the arbitrage on the comovement of the underlying conditional on the relative liquidity.[7] We conjecture that the more frequently the stock is traded through the ETF, which is reflected as an increase in its equivalent volume, the more frequently its return, on average, comoves with the other component returns in the ETF basket.

This study contributes to the literature in several ways. Although there are concurrently initiated studies investigating the ETF-related activity and comovement (Da & Shive, 2013), to our knowledge, this is a first look at the ETF arbitrage-induced trading and daily comovement of the underlying assets linked via a new differential liquidity measure based on trading volumes and daily ETF holdings. Furthermore, the empirical findings in this paper provide a new understanding of price shocks contagion (Antón & Polk, 2014) during daily ETF arbitrage-driven trading. We also contribute to the burgeoning ETF literature (Da & Shive, 2012; Ben-David, Franzoni, & Moussawi, 2014) by exploring unexpected consequences of ETF related daily trading activities on the pricing of the underlying assets. Interestingly, industry experts also suspect that ETF activity may pose unexplored risks (Johnson & Newlands, 2016; McNulty, 2017). Moreover, our findings extend the arbitrage literature by documenting a channel through which arbitrage, a presumably benign activity, can produce unexpected consequences via shock propagation to the underlying assets in the vein of Greenwood and Thesmar (2009) and Hong, Kubik, and Fishman (2012). Finally, we furnish the liquidity literature with a new measure of relative liquidity based on a ratio of trading volumes similar to O/S Roll, Schwartz, and Subrahmanyam (2010).

To address our research objectives, we use two approaches to estimate daily co-movement. The first approach employs dynamic conditional correlations (DCC) from the multivariate volatility model family (Engle 2002). Conditional correlations, and in particular DCC, offer sufficient flexibility to parameterize a change in the correlation as a function of the weighted average of the past and most recent return shocks. Dynamic conditional correlations also allow for the ability to specify additional independent variables in the GARCH mean equation such as market return, providing a natural way to control for fundamentals. Although comovement can be estimated more parsimoniously using intraday data, DCC has the advantage in markets where high-quality intraday data is hard to obtain or simply not available as is the case for many markets outside of the U.S.[8]

The second approach uses a shorter intraday sample to estimate short-horizon co- movement based on the Pearson correlation of intraday returns calculated using 5-min prices. Tests using intraday data should have more power in testing the comovement hypothesis given that the arbitrage-induced trading and, hence, the comovement shocks occur during the day.

To preview our results, we find that there is a strong, positive association between the liquidity of the ETF relative to the underlying stocks, proxied by equivalent volume and comovement of a stock's returns with those of the other ETF components using dynamic conditional correlations and intraday Pearson correlation estimators. In economic terms, using daily data from January 2002 to September 2011 for 12 ETFs and over 800 stocks, a daily change in the equivalent volume from 10% to 20% is associated with an increase in the dynamic conditional correlation of 0.01 or 1%. Using log transformed equivalent volume and dynamic conditional correlations, a 1% increase in equivalent volume is associated with a

[6] ETFs With Google, Inc. (GOOG) Exposure", August 12, 2015, http://etfdb.com/stock/GOOG/.

[7] Extracting intraday arbitrage activity is a daunting task. Some researchers use the end of day ETF price-NAV spreads but these are poor indicators of the intraday spreads which ultimately drive arbitrage trading. Furthermore, historical intraday spreads themselves are hard to obtain and the spread dynamics may point to arbitrage opportunities but not necessarily to arbitrageurs exploiting these opportunities. One potential variable correlated with arbitrage volume could be the separate ETF creation and redemption flows. However, the only public information available is the net creation and redemption, or the net change in shares outstanding. Hence, even when the creation and redemption flows are large but close in magnitude signaling large arbitrage volume by the APs, the net change in shares outstanding is close to zero.

[8] We thank an anonymous referee for drawing our attention to this point.

roughly 0.02% increase in daily correlation. Given that the pooled standard deviation for equivalent volume in 2010 is of the same magnitude as the mean and that the equivalent volume distribution is positively skewed with high kurtosis, daily movements of twice the average are not uncommon, translating into a 4% increase in correlation. Using an intraday sample from 2010 to 2011 and a comovement estimator based on the Pearson correlations of the intraday returns, we find that the magnitude of the association is 13 times stronger, from 0.02% to 0.26% basis points.

Summarizing, the arbitrage-induced trading is positively associated with comovement of the underlying assets when the relative liquidity of the underlying asset is low compared to the liquidity of that asset as part of the ETF, proxied by equivalent volume, leading to reduced diversification benefits for investors and market participants. This relationship between arbitrage and equivalent volume is robust to a variety of controlling variables and estimation procedures. Given the high growth rate of the ETF industry in terms of assets and trading volume, it is logical to expect the magnitude of the association to grow with time.

2. Related literature

2.1. Trading volume, liquidity and price impact

The premise of this paper lies in the intersection of the trading-related price impact, liquidity and trading volume. This section briefly summarizes the relevant research in these areas.

The empirical price impact literature usually centers on the effect of the order flow shocks and returns. For instance, Shleifer (1986) and Kaul, Mehrotra, and Morck (2000) look at downward sloping demand curves, while Lakonishok, Shleifer, and Vishny (1992) and Chan and Lakonishok (1993) look at the transactions of money management firms and their effect on returns on the intraday basis. Warther (1995), Wermers (1999), and Goetzmann and Massa (2003) look at mutual fund flows and returns while Cai and Zheng (2004) focus on institutional trading and Sun (2008) examines the association of investor clienteles, order flow and comovement.

The part of the trading volume literature relevant to the premise in this paper looks at the relationship between volumes of different assets and the role volume plays in liquidity studies. Lo and Wang (2009) provide a survey of trading volume research and build a volume-based asset pricing model. Furthermore, recently a stream of literature has emerged that is based on analyzing information contained in relative volumes of different assets. Roll, Schwartz, and Subrahmanyam (2012) analyze the informational impact of trading volumes of contingent claims on the S&P500. Chakravarty, Gulen, & Mayhew (2004) investigate contribution of the options volume to price discovery and Roll et al. (2010) study information contained in the option/stock volume ratio called O/S. Johnson and So (2012) further extend this line of research and link O/S and future returns. Ge, Lin, and Pearson (2016) use signed O/S and explore the effect of embedded leverage on future returns. On an accounting side, Rai and Tartaroglu (2015) focus on the link between O/S and market response to earnings surprises.

The most recent survey of empirical liquidity research by Holden et al. (2014) lists a variety of general proxies for liquidity classified along cost, quantity and time dimensions. The cost dimension relates to explicit transaction costs and comprises a variety of bid-ask spread-based measures. The quantity dimension is composed of several depth measures and the slope of a price function based on Kyle (1985)'s λ. The time dimension is based on measures of execution speed and fill rates. Liquidity proxies for equity are discussed separately with a particular emphasis on the liquidity measures from daily data, which are computationally more efficient than the intraday ones. Among those mentioned are Goyenko, Holden, and Trzcinka (2009)'s extended Amihud and Mendelson (1986) measures in form of ratios constructed using existing cost-based liquidity proxies such as Hasbrouck (2009)'s Gibbs-based measure in the numerator and average trading volume in the denominator.

The positive relation between volume and liquidity arises from the intuitive result that it is easier to transact in more active markets (Johnson 2008). From the theoretical standpoint, there are several types of models that seem to provide a more rigorous guidance. In particular, Kyle (1985) introduced a model with two types of the λ, whose inverse is the proxy for market depth, and which decreases with the higher trading volume. The concept of λ was further built on by Amihud (2002) who introduced the Amihud illiquidity measure (ILLIQ) based on the ratio of average absolute return to the daily volume, or daily price impact of the order flow. The implementation of trading volume as a liquidity proxy stems from the Amihud and Mendelson (1986) model which states that, in equilibrium, trading volume is correlated with liquidity.

Additionally, empirical research by Darolles and Le Fol (2003) and Darolles, Le Fol, and Mero (2015) have contributed to the literature by exploring liquidity based volume decompositions. On the other hand, Brennan, Chordia, and Subrahmanyam (1998) found that the trading volume is statistically significant as the liquidity proxy in their analysis of the expected stock returns. Datar, Naik, and Radcliffe (1998) also employed trading volume scaled by the shares outstanding as a liquidity proxy to explain cross-sectional variation in stock returns. Chordia, Subrahmanyam, and Anshuman (2001) continue this line of research and look at the second moment of liquidity factor expressed by the variability of trading volume and find a negative relation between returns and variability of liquidity.

2.2. Exchange-traded funds

Exchange-traded funds have recently become more prominent in academic research with the tracking error and pricing literature probably forming the largest group. Engle and Sarkar (2002), Shin and Soydemir (2010), and Petajisto (2013) study ETF pricing vis-à-vis the underlying assets while Doran, Boney, and Peterson (2006), Svetina and Wahal (2008), and Huang and Guedj (2009) investigate tracking error differences between ETFs and mutual funds. Box, Davis, and Fuller (2016) look at the effects of competition among ETFs and find a decrease in market quality for the incumbent fund when a new fund is introduced. In the global markets, fund premia for international vs. domestic ETFs are studied by Delcoure and Zhong (2007) and Milonas and Rompotis (2010) with an extension into the arbitrage and liquidity context. Furthermore, tracking error in leveraged funds is explored by Cheng and Madhavan (2009), Lu, Wang, and Zhang (2009), and Charupat and Miu (2011), while Madhavan and Sobczyk (2014) develop a more general model for ETF pricing and apply it to the bond funds. A new paper by Cheng, Massa, and Zhang (2015) looks at European markets and finds a surprising return predictive power in stocks in which ETF weights diverge from those of the index.

Price discovery via ETFs is investigated in U.S. markets by Hasbrouck (2003), Yu (2005), Tse, Bandyopadhyay, and Shen (2006), and Hseu, Chung, and Sun (2007), while in overseas markets Chen and Strother (2008) and Schlusche (2009) find that price innovation is usually faster in the ETF and futures markets than in the underlying cash index. Fang and Sanger (2011) and Wallace, Kalev, and Lian (2014) confirm the role of the ETFs as preeminent channels for price discovery.

Following the mutual fund literature, several papers look at the effects and drivers of the ETF flows. Staer (2017) studies the effect of ETF flows on the underlying returns and find evidence of a transitory price impact, while Clifford, Fulkerson, and Jordan (2014) investigate the factors driving ETF flows. Fulkerson, Jordan, and

Travis (2014) find return chasing and "smart money" evidence in bond ETF flows.

Another strand of research has emerged building on previous studies to link ETFs to price impact, liquidity, volatility and comovement. Krause, Ehsani, and Lien (2013) and Ben-David et al. (2014) study the propagation of shocks through ETFs to volatility of the underlying assets. Hegde and McDermott (2004), Richie and Madura (2007), Ackert and Tian (2008), and Hamm (2010) look at the link between ETFs and component liquidity. Broman and Shum (2015) study short-term trading and liquidity clienteles, and Broman (2015) link liquidity and comovement across ETFs. Malamud (2015) develop a dynamic model describing how liquidity shocks propagate through creation-redemption mechanism. A paper most similar to this one is Da and Shive (2013) that links ETF arbitrage and comovement. Although there are substantial differences in methodology, scope and data frequency, the authors reach similar conclusions to this paper. On a more general note, Wurgler (2010) and DeLisle, French, and Schutte (2015) point at a widespread reduction in information discovery due to an increase of usage of index investing and asset baskets, such as ETFs. Recently, a new stream of research emerged investigating the particularities of leveraged and inverse ETFs which usually hold derivatives and aim to replicate a multiple of an index. Cheng and Madhavan (2009), Guedj, Li, and McCann (2010), Tang and Xu (2011), Bai, Bond, and Hatch (2012), and Tuzun (2014) look at the pricing and leverage-related anomalies that arise with derivatives-based ETFs.

2.3. Asset comovement

A variety of recent studies challenge the traditional view based on frictionless economies and rational investors, where asset returns comove due to covariation in fundamental factors, and show that returns often comove in excess of fundamentals.

The index comovement literature is based on the evidence that the inclusion of a stock in the index is exogenous to the firm's fundamentals and should not affect a stock's comovement with the other index constituents. Early papers from Shleifer (1986) and Vijh (1994) study the effect of inclusion of stocks in the S&P 500 index on returns and comovement. A seminal paper by Barberis, Shleifer, and Wurgler (2005) also uses the S&P 500 index inclusions and defines several types of asset return comovement: category based on style investing (Barberis and Shleifer, 2003), habitat based on investor constraints, and information diffusion in part due to differential information access speeds. Greenwood and Sosner (2007) study the singular event of a replacement of 30 stocks in the Nikkei 225 index, which results in increased comovement with the index after the inclusion. Furthermore, Boyer (2011) find that a mechanical procedure of stock assignment by S&P/Barra from Value to Growth indexes produces excess comovement. Claessens and Yafeh (2011) expand the index inclusion-based comovement to international markets.

Recent work by Lou and Polk (2013) builds on Barberis et al. (2005) and relates momentum strategy arbitrage to the increase in comovement of the stocks through correlated price impacts, which is similar to the premise for the arbitrage-related comovement argument in this paper. Domowitz, Hansch, and Wang (2005) not only relate order flow to the return comovement but also relate order types to commonality in liquidity.

The relation of excess comovement with other corporate events has also been studied in the literature. Green and Hwang (2009) find that the stocks after splits comove more with the lower-priced stocks than with the higher-priced stocks. Furthermore, Kumar, Page, and Spalt (2013) extend their Kumar and Lee (2006) sentiment and comovement paper and explore the link between retail investor trading and the excess comovement using stock splits, headquarter changes and other trading-related measures.

In the international markets, Hardouvelis, Porta, and Wizman (1994) discover that U.S. closed-end funds that hold only foreign assets comove more with the U.S. market than they should. Froot and Dabora (1999) study twin stocks, e.g., Royal Dutch and Shell, and find that these stocks comove with the exchanges they are traded on instead of moving in lockstep as implied by the fundamentals. Greenwood (2008) explore the differences on comovement of equally-weighted Nikkei 225 index vs. the value-weighted TOPIX index and find that overweighted stocks in Nikkei 225 have higher betas and comove more with the market.

A different strand of the comovement literature targets longitudinal patterns in comovement between asset classes. Their methodology usually includes a time-varying correlation estimator, as in Tang and Xiong (2012) who compute rolling correlations and perform regressions on the year and factor dummies. Büyüksahin, Haigh, and Robe (2010) and Chong and Miffre (2009) investigate time-varying comovement between assets time series using the dynamic conditional correlations (DCC).

3. Variable construction and estimation

3.1. Equivalent volume construction

This section discusses the construction of equivalent volume as a relative liquidity indicator. The arbitrageur engaging in ETF arbitrage usually seeks to reduce tracking error and holds a basket of assets constructed to replicate the ETF as closely as possible.

To execute the arbitrage trade, the arbitrageur will buy (sell) $w_{i,t}P_{etf}N_{etf}$ of the underlying stock i for each sell (buy) of $P_{etf}N_{etf}$ of ETF where N_{etf} is the number of ETF shares, P_{etf} is the ETF price per share and $w_{i,t}$ is the weight of the stock in the ETF. The price impact on the arbitraged assets, however, will depend on their relative liquidity. The same trade will disproportionately affect more the less liquid asset on one leg of the trade than the relatively more liquid asset on the other leg of the trade. The higher the stock's liquidity as part of the ETF as proxied by its weight of the stock in the ETF trading volume, and the lower the liquidity of the stock as proxied by the stock's own trading volume, the higher the impact of the ETF arbitrage trade on the underlying stock's price. In order to capture that differential liquidity, we construct equivalent volume, a new indicator of relative liquidity between the ETF and the stock it holds. Equivalent volume is formalized in two equations. In Eq. (1), $w_{i,t}$ is the weight of the stock in the ETF, $volume_{etf,t}$ is the ETF trading volume in shares, $volume_{i,t}$ is the trading volume for the stock i held by the ETF and $w_{i,t}$ denotes the weight of the stock i in the ETF at time t.

$$Equivalent\ Volume_{i,t} = \frac{w_{i,t} \times volume_{etf,t}}{volume_{i,t}} \tag{1}$$

The equivalent volume from Eq. (1) theoretically varies along a $[0, \infty]$ interval. We normalize it to a $[0, 1]$ interval by keeping the same numerator, $w_{i,t} \times volume_{etf,t}$, and dividing by the total volume of the stock composed out of the stock's trading volume and the stock's share in the ETF trading volume, $volume_{i,t} + w_{i,t} \times volume_{etf,t}$, as in Eq. (2).[9]

$$Normalized\ Equivalent\ Volume_{i,t} = \frac{w_{i,t} \times volume_{etf,t}}{volume_{i,t} + w_{i,t} \times volume_{etf,t}} \tag{2}$$

where $w_{i,t}$ is the weight of the stock i in the ETF at time t, $volume_{i,t}$ is the trading volume of the stock at time t, and $volume_{etf,t}$ is the ETF trading volume at time t. The LHS in Eq. (2) is essentially a proportion of the trading volume of the stock as part of the ETF to

[9] Our results are qualitative similar for the non-normalized variant of the equivalent volume in Eq. (1) as well.

Please cite this article in press as: Staer, A., & Sottile, P. Equivalent volume and comovement. *The Quarterly Review of Economics and Finance* (2017), https://doi.org/10.1016/j.qref.2017.11.001

the total volume of the stock composed from the stock's trading volume and the trading volume of the stock as part of the ETF. We drop the "normalized" term from the name of the variable on the LHS in Eq. (2) and call it equivalent volume for brevity from now on.

The rationale behind the numerator, the weight of the stock in the ETF multiplied by the ETF trading volume, is that in order to exploit the price disparity, the arbitrageurs have to replicate the ETF holdings where every ETF share is equivalent to a weighted sum of underlying stock shares. By multiplying the ETF share with the weights of the stocks, we can find the equivalent partial shares of the underlying stocks that the arbitrageur would have to buy to replicate the ETF. Extending this intuition, we can extract the share of the ETF trading volume equivalent to that stock by multiplying the ETF trading volume by the weight of the stock. Consequently, when the trading volume of the ETF is high, the equivalent volume and, hence liquidity (Amihud, 2002), of the stock as part of the ETF is high too[10] and the arbitrage-driven trading does not move the ETF price as much as the price of the underlying stock. On the contrary, when the denominator, i.e., the total stock volume including stock volume and the equivalent volume of the stock as part of the ETF, is high, then the total stock liquidity is high, hence the same magnitude of the arbitrage trade should have less effect on the price of the underlying stock and more effect on the price of the ETF, which has lower stock-level equivalent liquidity.

We illustrate the equivalent volume calculation on Google, stock that is held by S&P 500 ETF (SPY), on June 10–11, 2010. Panel A of Table 1, shows that, on June 11, 2010, the S&P 500 ETF trading volume was 214 million shares, while Google's trading volume was 1.9 million shares.

On June 10, 2010, we calculate Google's share in the S&P 500 ETF trading volume, $w_{i,t} \times volume_{etf,t}$, where the weight of Google shares in the S&P 500 ETF, 0.27%, is multiplied by that ETF trading volume, 317.9 million shares, which results in a value of 0.9 million shares. Equivalent volume then is the ratio of Google's share in the S&P 500 ETF trading volume calculated earlier, 0.9 million shares, to the total volume comprised of Google's own trading volume and Google's share in the S&P 500 ETF trading volume, 0.9 + 2.8 = 3.7 million shares, resulting in a value of 23.29%. Therefore, out of Google shares traded on the market and those traded indirectly through the S&P 500 ETF, 23.29% of the total were traded indirectly as the part of the ETF.

3.2. Comovement estimation

We employ two approaches to measure daily return comovement. For both approaches, the comovement is measured as a correlation between the stock's return and the ETF NAV return *excluding* the return of that stock (Barberis et al. 2005), but the frequency for returns and methodologies for correlations differ.[11] The first approach uses the residuals from the regression of daily returns of the stock and the ETF NAV return on the market to calculate excess daily dynamic conditional correlations (DCC), while the second approach uses 5-min intraday returns of the stock and the ETF NAV to calculate the intraday correlations.

In the first approach, we employ the well-known dynamic conditional correlations (DCC) introduced by Engle (2002) as the proxy for comovement between the stock and the ETF NAV. Estimation of the comovement via DCC has the distinct advantage of applicability in the markets where intraday data is hard to obtain while the daily data is readily available. We condition the stock return, R_s, and ETF NAV return, R_{etf}, on the value-weighted market return from CRSP, R_{VWRETD}, with the residuals as the vector of ε at each t:

$$\begin{aligned} R_{s,t} &= \alpha_s + \beta_{VWRETD,s} R_{VWRETD,t} + \varepsilon_{s,t} \\ R_{etf,t} &= \alpha_{etf} + \beta_{VWRETD,etf} R_{VWRETD,t} + \varepsilon_{etf,t} \end{aligned} \tag{3}$$

where the vector $\varepsilon_t \sim N(0, H_t)$ with the conditional variance, H_t, following:

$$H_t = D_t C_t D_t \tag{4}$$

D_t $k \times k$ diagonal matrix of the time varying standard deviations obtained from univariate GARCH with $\sqrt{h_{it}}$ on the ith diagonal and C_t is the time varying correlation matrix,

$$C_t = Q_t^{*-1} Q_t Q_t^{*-1} \tag{5}$$

where,

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha(\tilde{\varepsilon}_{t-1}\tilde{\varepsilon}'_{t-1}) + \beta Q_{t-1} \tag{6}$$

where Q_t^* is a diagonal matrix composed of the square root of the diagonal elements of Q_t,$\tilde{\varepsilon_t}$ is an $m \times 1$ vector of standardized residuals, ${D_t}^{-1/2}\varepsilon_t$. Hence, a typical element of R_t and the time varying correlation estimate is,

$$\rho_{s,etf,t} = \frac{q_{s,etf,t}}{\sqrt{q_{s,etf,t} q_{s,etf,t}}} \tag{7}$$

where $\rho_{s,etf,t}$ is the time-varying excess correlation between the stock s and the ETF basket that the stock belongs to, at time t.

Furthermore, the DCC by its nature is a next period dynamic conditional correlation forecast. In other words, Eq. (6) is based on the information up to and including $t-1$ period when estimated for each t. To align the DCC forecasts with the other variables measured on a daily basis, we lag the other variables one day, so that a DCC forecast for t corresponds to the 1-day lag of independent variables. Following the suggestions of Asai and McAleer (2009) and Hafner and Reznikova (2010), we estimate the k pairwise correlations of k stocks with the ETF NAV return instead of estimating $k \times k$ correlation matrix for each ETF. Specifically, at each t and for k stocks, we estimate k DCC_i coefficients where DCC_i for each stock i represents the dynamic conditional correlation between the returns of a ith stock and the returns of the portfolio of remaining ETF components. One caveat of the DCC estimation is non-convergence, especially for time series with higher volatility. In the sample, out of 1453 series corresponding to stock-fund combinations about 124 series did not converge. The absence of the non-converged series from the sample should not bias our estimates.

4. Data

The sample period starts in 2002 and ends in 2011, spanning nine years. The U.S. equity ETF sample used in this study covers 12 equity ETFs from two large fund families, SPDRs issued by State Street Global Advisors (SSgA) and HOLDRs issued by Merrill Lynch.[12] HOLDRs historical daily holdings were obtained from *Bloomberg*. Sector SPDRs ETFs and S&P 500 ETF (SPY) historical daily holdings were obtained from the fund distributor.[13],[14]

[10] Here we assume the weight of the stock in the ETF is constant for simplicity of exposition.

[11] We also test the results in the paper using comovement between the stock returns and the ETF NAV return including the stock with similar results.

[12] Sector SPDRs and S&P 500 SPDR share the same issuer, State Street Global Advisors (SSgA), however, within the issuer's internal structure they are separated in different fund subfamilies. HOLDRs were managed by Merrill Lynch, but in December 2011 (after the end of the sample) several of them were sold to another fund family Van Eck and the rest were terminated. The reason for fund sale was unrelated to the fund performance according to Merrill Lynch.

[13] We are grateful to Alps Distributors for the data on the Select SPDR ETFs.

[14] The component weights or the composition of the asset basket for some ETFs are usually available on contemporaneous basis on the funds webpages and financial

Table 1
Creation and Redemption vs. Trading Volume: S&P 500 ETF and Google.

Panel A: ETF						
Fund	Date	ETF Volume	Price	Shares	Net Asset Value	Creation/ Redemption
S&P500 SPDR	6/10/2010	317.9	109.15	651	109.16	
	6/11/2010	214.1	109.68	717	109.67	66

Panel B: Stock					
Stock	Date	Stock Volume	Weight	Stock's share of ETF Volume	Equivalent Volume
Google	6/10/2010	2.8	0.27%	0.9	23.29%
	6/11/2010	1.9	0.22%	0.5	19.55%

First row corresponds to the S&P500 ETF (SPY) and the 2nd row to the component of that ETF, Google Inc. (GOOG). *Price* and *Net Asset Value* are in U.S. Dollars. *Stock Volume*, *ETF Volume*, *Shares* and *Creation/Redemption* are in millions of shares. *Weight* is Google's weight in S&P500 SPDR ETF (SPY) on that date. *Stock's share of ETF volume* is calculated following Eq. (1) and expressed in millions of shares and *Equivalent Volume* is calculated following Eq. (2).

Table 2
Sample ETF Descriptive Statistics.

ETF	S&P500 Index SPDR	Select Sector SPDRs	HOLDRs	Total Sample
Issuer	SSGA	SSGA	Merrill Lynch	
Number of funds	1	9	2	12
Inception Date	1/22/1993	12/16/1998	2000–2001	1993–2001
Sample Start	6/20/2003	6/20/2003	1/2/2002	2002–2003
Ticker	*SPY*	*XLB – XLY*	*SMH – OIH*	
Index Ticker	*SPX*	*IXB – IXY*	*XSH – OXH*	
AUM	93,327	38,944	2288	134,559
Number of Holding	501	501	39	706
Avg Spread	0.0001	0.0003	0.00035	0.00025
Avg Daily Volume	47,098	4,993	907	52,997
Avg Daily Share Volume	414	189	16	618
Avg Turnover (days)	1.98	7.80	2.36	2.54

This table presents descriptive statistics for the ETF sample. Funds for the SPDRs fund family are separated into 2 subfamilies due to substantial differences in fund characteristics. *# of Funds* is the number of the funds from that issuer/fund family in the sample. *AUM* is a cross-section of assets under management summed over funds for each issuer/fund family. *# of Holdings* for the fund level is the number of stocks in each fund and for the total sample level is the total amount of stocks that have been held by the funds at any point in time since the sample start. *Avg spread* is calculated as the difference between fund intraday price and fund NAV normalized by the fund NAV. *Avg daily volume* and *Avg daily share volume* are the daily trading volume in $ millions and shares respectively averaged over the previous 30 days. The sample start and the fund inception differ between funds due to the data constraints, which also means that the sample does not include the period right after the fund inception. For comparison purposes, total assets under management (AUM) for the sample are $134 billion which is about 28% of the total assets under management for the unleveraged U.S. equity ETFs of $490 billion (Investment Company Institute 2011). Data as of September 30, 2011.

The first family, Sector SPDRs and S&P 500 SPDR by State Street Global Advisors, tracks different subsets of S&P 500 index. Specifically, the index is partitioned in 9 subindices, each one covering a sector based on the general industry classification. The industries with the respective ETF tickers in parentheses are: Consumer Discretionary (XLY), Consumer Staples (XLP), Energy (XLE), Financial (XLF), Health Care (XLV), Industrial (XLI), Materials (XLB), Technology (XLK), Utilities (XLU). The partitioning is done in such a way that the relative stock weights in S&P 500 are preserved and only the absolute weights change.[15] Summarizing, in our sample, we have 10 funds tracking the S&P 500 Index and 9 Select Sector subindices.

The second family, HOLDRs by Merrill Lynch, is one of the very few fund families that have some of their historical weights series available on *Bloomberg*, making data collection relatively easier. We chose two of their funds due to data availability and their use in the index arbitrage research (Avellaneda & Lee 2010).

Daily returns and trading volume are from Center for Research in Security Prices (CRSP). The market return is the value-weighted NYSE/AMEX/NASDAQ index, also known as the *CRSP Value-Weighted Index*. The ETF asset basket weights for the HOLDRs fund family and intraday prices are from *Bloomberg*. The intraday data in *Bloomberg* is 1-min frequency and only the last 6 months are available. Other by studies, for instance by Hegde and McDermott (2004), Alexander and Barbosa (2005), and Svetina and Wahal (2008), have used *Bloomberg* as a source of ETF data, such as assets under management, fund family characteristics and intraday quotes. As described in Anderson and Dyl (2008) and Roll et al. (2010), NASDAQ volume is subject to double dealer reporting and is thus overstated by up to 50% as compared to the NYSE; consequently, we control for that by including an exchange dummy in the regressions in Section 5.2.

Table 2 describes the ETFs in our sample. Total assets under management (AUM) for the sample are $134 billion which is about 28% of the total assets under management for the unleveraged U.S. equity ETFs of $490 billion (Investment Company Institute 2011). All funds in the sample exhibit high turnover especially when compared to the average turnover of 192 days for U.S. equities in 2011.[16] Turnover across funds varies in a tight range from 1.98 to 7.80 days even though their assets under management differ by orders of magnitude.

Table 3 shows the descriptive statistics for cross-sectional distribution of equivalent volume each year. There is substantial evidence of positive kurtosis and skewness across all years. The distribution of the ETF assets under management (AUM) is positively skewed as a few funds command a disproportionate amount of assets which allows us to use a smaller but still representative sam-

portals. The limitation however is that only the most recent cross-section of holdings is shown. Our data sample has historical weights for all the components in the ETF asset basket.

[15] For instance, on September 14, 2010, Google comprised 1.36% of S&P 500 index and 5.80% of Technology Select Sector Index. Hence, Technology Select Sector Index constitutes 23.44% of the S&P 500 index.

[16] "Annual Report", World Bank, 2012.

Please cite this article in press as: Staer, A., & Sottile, P. Equivalent volume and comovement. *The Quarterly Review of Economics and Finance* (2017), https://doi.org/10.1016/j.qref.2017.11.001

Table 3
Descriptive Statistics for Equivalent Volume.

Year	P1	P25	Median	Mean	P75	P99	Kurtosis	Skewness
2002	0.01	0.02	0.04	0.04	0.06	0.12	1.63	1.03
2003	0	0	0.01	0.02	0.03	0.12	10.98	2.3
2004	0	0	0.02	0.03	0.04	0.14	9.34	2.36
2005	0	0.01	0.02	0.04	0.06	0.19	5.51	1.97
2006	0	0.01	0.03	0.04	0.06	0.21	5.89	2.02
2007	0	0.01	0.04	0.06	0.09	0.27	7.3	2
2008	0	0.02	0.05	0.08	0.11	0.41	7.85	2.25
2009	0	0.02	0.05	0.07	0.1	0.37	9.28	2.39
2010	0	0.02	0.05	0.08	0.11	0.38	8.08	2.25
2011	0	0.03	0.06	0.09	0.12	0.42	7.11	2.18

This table presents percentiles, the mean and the median for the pooled equivalent volume for each year in the sample from 2002 to 2011. Equivalent volume measure follows Eq. (2). *P*1, *P*25, *P*75 and *P* 99 denote the 1st, 25th, 75th and 99th percentiles of the pooled equivalent volume distribution respectively. *Mean*, *Median*, *Kurtosis*, *Skewness* corresponds to the mean, median, excess kurtosis and skewness of pooled equivalent volume, respectively. Values are rounded to two decimal places. Due to rounding some small values may appear as zeros.

ple. In Section 5.1, we discuss the way to approach these deviations from normality via differencing and logarithmic transformations.

Fig. 1 shows the average and maximum daily equivalent volume for our sample from 2002 to 2011. The magnitude of the equivalent volume is considerable for a majority of the ETF components. For example, for some stocks in 2010, the equivalent volume is close to 80% on some dates.

5. Comovement and equivalent volume regressions

5.1. Estimation: methodology

We apply first differencing and log transformations to the variables of interest in the daily and intraday regressions to adjust for serial autocorrelation and to facilitate the economic interpretation. For equivalent volume, the first difference is $\Delta Equivalent\ Volume_{i,t} = Equivalent\ Volume_{i,t} - Equivalent\ Volume_{i,t-1}$ And for the DCC, the first difference is $\Delta DCC_{i,t} = DCC_{i,t} - DCC_{i,t-1}$.

We apply log transformation to the comovement proxies and equivalent volume to mitigate high kurtosis and positive skewness of the latter to obtain a log–log regression setup with an economic interpretation based on elasticities.

Given that equivalent volume is a ratio of non-negative random variables, log transformation for equivalent volume is straightforward, but the correlations can assume zero or negative values; hence, log transformations should be used with caution. In our sample, DCC values are negative only in 0.01% of cases. To maintain interpretability, we code zero and negative values as missing.

Table 4 presents descriptive statistics for the various transformations of equivalent volume and DCC. As expected, the log transformation reduces kurtosis for the equivalent volume and slightly increases it for the DCC. On the other hand, the differencing transformation decreases kurtosis and skewness for equivalent volume and increases kurtosis for the DCC.

5.2. Regressions: daily comovement

We employ a panel estimation with cross-sectionally AR(1) adjusted heteroscedasticity robust errors (Wooldridge, 2001) that corrects for autocorrelation between stocks within the same ETF and across ETFs using a flexibly specified error correlation matrix. Results for more parsimonious specifications such as no autocorrelation or IID errors generally provide similar estimates, but with higher *t*-stats. Additionally, we perform more tests, such as fixed and random effects panel models with cross-correlation heteroscedasticity adjusted errors, with similar results. Following Petersen (2009), we also implement two dimensional clustering using varying time clustering frequencies (week, month, quarter and year) that provide similar coefficient magnitudes and inferences.

The panel unit is a stock-fund combination totaling 1453 stock-fund panels in the sample that corresponds to 706 unique stocks from 12 ETFs. After removing the stocks that failed to converge during the DCC estimation, 1329 panels remain, which correspond to 684 unique stocks held by the same 12 ETFs. The stock-fund combination also implies that a stock present in more than one fund is considered a different observation unit. Under the equivalent volume-comovement hypothesis a stock present in several ETFs covaries with the ETF basket the stock belongs to. That covariation should be proportional to that stock's equivalent volume that is ETF-specific by construction. We test the robustness of this assumption in fund-panel regressions that are run separately for each fund with the stock as an observation unit.

There are several other considerations. First, according to Gallant, Rossi, and Tauchen (1992) and Chordia and Swaminathan (2000), lagged returns and volume should be included in the daily return-volume regressions to reduce autocorrelation. Following Lo and Wang (2009), we select the equivalent volume-related independent variables: two leading, one contemporaneous and two lags, thus, covering 5 trading days. This specification adjusts for autocorrelation in residuals and adds two leads as controls to test whether leading equivalent volume is associated with contemporaneous correlation.

If the equivalent volume–comovement hypothesis were true, then the coefficients on the leading equivalent volume should be close to zero and coefficients on the lagging equivalent volume should be smaller in magnitude than for the concurrent equivalent volume coefficient. If the leading coefficients had comparable estimates to the contemporaneous coefficients, it would imply either persistence in equivalent volume or an unobserved confounding factor. Furthermore, according to the institutional and MF trading literature (Edelen & Warner 2001; Ben-David et al. 2014), price pressure effects should last 1-2 days; hence, under the alternative hypothesis, the lagging coefficients should be smaller in magnitude than the concurrent coefficient and the magnitude should decrease with the lag distance. Under the null hypothesis of no comovement, the equivalent volume coefficients should not be significantly different from zero. Additionally, if leading or lagging equivalent volume coefficients were larger than the contemporaneous coefficient, then the arbitrage-induced trading would be less likely to be related to the return comovement via the contemporaneously correlated price pressure.

The general panel regression specification is presented in Eq. (8). Model 1 contains the fully controlled specification with a lagged dependent variable as shown in Eq. (8), while Model 2 excludes the lagged dependent variable and Model 3 is the least controlled specification with the equivalent volume variables only.

$$
\begin{aligned}
DCC^{s}_{i,t} = \alpha + & \underbrace{\beta_{DCC^{s},t-1} DCC^{s}_{i,t-1}}_{\text{Lagged Dependent Varas Control}} + \underbrace{\sum_{t-3}^{t+1} \beta_{EqVol,t} EqVol^{s}_{i,t} +}_{\text{Equivalent Volume and Lagsas Controls}} \\
& \underbrace{\sum_{t-2}^{t-1} \beta_{stock\_ret,t} R_{i,t} + \sum_{t-2}^{t-1} \beta_{eft\_ret,t} R_{eft,t} + \sum_{t-2}^{t-1} \beta_{VWRETD\_ret,t} R_{VWRETD,t} +}_{\text{Return Controls}} \\
& \underbrace{\sum_{i=1}^{j} \beta_{control,i} C_{i,t-1}}_{\text{Various Controls}} + \varepsilon_{i,t}
\end{aligned}
\tag{8}
$$

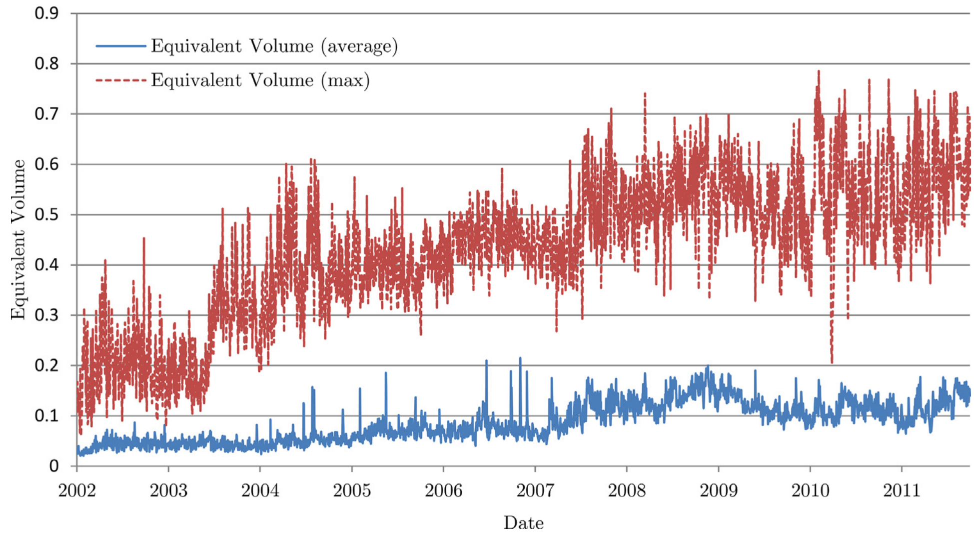

Fig. 1. Sample Equivalent Volume.
The figure depicts two daily time series: the *average* and the *maximum* cross-sectional equivalent volume (EV) for the daily sample of 12 ETFs with descriptive statistics in Table 2. Equivalent volume is defined in Eq. (2). Sample period is from 2002 to 2011.

Table 4
Descriptive Statistics for Equivalent Volume and DCC Transformations: Daily Sample.

Variable	Mean	SD	Skewness	Kurtosis	N
Panel A: Descriptive Statistics for Equivalent Volume					
Equivalent Volume	0.06	0.07	2.67	11.65	1,981,608
ΔEquivalent Volume	0	0.03	−0.14	10.47	1,980,271
Equivalent Volume Log	−3.66	1.56	−0.94	0.87	1,981,608
Panel B: Descriptive Statistics for Dynamic Conditional Correlations					
DCC	0.6	0.15	−0.46	3.22	1,981,608
ΔDCC	0	0.02	−1.82	199.8	1,977,649
DCC Log	−0.56	0.31	−2.02	12.2	1,981,344

This table presents descriptive statistics for equivalent volume and DCC and their difference and log transformations for the daily sample. *Mean*, *Median*, *Kurtosis*, *Skewness* correspond to the mean, median, excess kurtosis and skewness of the pooled sample for the appropriate variable respectively. There are 1329 stock-fund groups with approximately 1.98 million daily observations. Values are rounded to two decimal places. Due to rounding some small values may appear as zeros.

where $DCC^s_{i,t}$ is the transformation of DCC correlation, either log transformation $log\left(DCC_{i,t}\right)$ or innovations $\Delta DCC_{i,t}$. The same notation applies to the equivalent volume $EqVol_{i,t}$. Furthermore, $R_{i,t}$ is the return of the stock i, $R_{etf,t}$ is the ETF return, $R_{VWRETD,t}$ is the return on the value-weighted CRSP market index, $C_{i,t-1}$ are the control variables which for this model are the stock exchange dummies for every stock–fund combination.

Table 5 presents the panel estimation results adjusted for cross-panel autocorrelation (AR1) and heteroscedasticity across stock-fund panels. Panel A shows that the contemporaneous equivalent volume innovation, $EqVol_{t-1}$, has a strong, positive effect on the contemporaneous DCC, with the innovations for higher lags presenting a smaller and still positive effect. The first and second lags of the equivalent volume show that the effect decays quickly with the second lag being only one-third of the zero lag coefficient's magnitude (from 0.079 for the zero lag to 0.022 for the second lag). Fig. 2 plots equivalent volume coefficient estimates for each specification. The downward slope in coefficients of equivalent volume lags is consistent with the time decay in the association with comovement and is generally present in all specifications and models. On the other hand, leading equivalent volume exhibits a different behavior with an immediate drop from 0.078 to 0.008 at the first lead, while the second lead is insignificant across all but one innovation specifications, which is consistent with the arbitrage-induced trading pressure hypothesis.

Table 5 Panel B presents the same analysis performed on the log transformations of dependent and independent variables. Taking

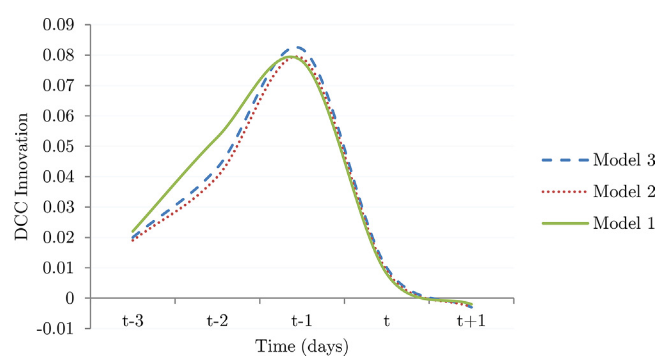

Fig. 2. Equivalent Volume Innovation Coefficients from Daily Regressions.
The graph depicts equivalent volume innovation coefficient estimates using daily sample under Models 1–3 from Table 5. Equivalent volume innovation is specified as first difference of the equivalent volume in Eq. (2). The regression model estimated is from Eq. (8), where equivalent volume innovations coefficients are specified as two leading coefficients ($t+1$ to t), one contemporaneous coefficient ($t1$) and two lagging coefficients (t-2 to t-3). Comovement at t is estimated via dynamic conditional correlations (DCC) and is contemporaneous with our variable of interest, equivalent volume at $t-1$. Please note that this Fig. 2 is different from Fig. 4 where comovement at t is estimated via intraday correlations and is contemporaneous with equivalent volume at t. Model 1 is the fully controlled specification, Model 2 excludes the lagged dependent variable (DCC^s_{t-1}) and Model 3 is the equivalent volume only specification. All three specifications show that the relation between comovement and equivalent volume as the measure of relative liquidity is positive and highly significant at $t-1$ and decays quickly afterwards at $t-2$ and $t-3$.

Table 5
Panel Regression of DCC on Equivalent Volume.

Variables	Panel A			Panel B		
	3 ΔDCC	2 ΔDCC	1 ΔDCC	3 $\log(DCC_t)$	2 $\log(DCC_t)$	1 $\log(DCC_t)$
DCC^s_{t-1}			−0.21***			0.97***
			(−43.89)			(1630.93)
$EqVol^s_{t+1}$	0.00*	0.00	0.00	0.00***	0.00***	0.00**
	(−1.73)	(−1.51)	(−1.18)	(16.00)	(16.20)	(−2.41)
$EqVol^s_t$	0.01***	0.01***	0.01***	0.01***	0.01***	0.00***
	(4.93)	(4.70)	(3.78)	(28.46)	(27.96)	(12.82)
$EqVol^s_{t-1}$	0.08***	0.08***	0.08***	0.02***	0.02***	0.01***
	(40.48)	(40.29)	(36.26)	(82.20)	(79.61)	(56.98)
$EqVol^s_{t-2}$	0.04***	0.04***	0.05***	0.01***	0.01***	−0.01***
	(21.53)	(21.06)	(25.90)	(53.79)	(52.69)	(−31.29)
$EqVol^s_{t-3}$	0.02***	0.02***	0.02***	0.01***	0.01***	−0.01***
	(11.55)	(11.37)	(13.24)	(36.29)	(36.16)	(−26.64)
$R_{stock,t-1}$		−0.03***	−0.03***		−0.04***	−0.05***
		(−17.43)	(−17.35)		(−13.64)	(−15.77)
$R_{stock,t-2}$		0.00	−0.01***		−0.02***	−0.01***
		(−0.03)	(−4.19)		(−8.62)	(−3.53)
$R_{eft,t-1}$		0.03***	0.03***		0.04***	0.06***
		(4.82)	(4.95)		(4.78)	(5.74)
$R_{eft,t-2}$		0.00	0.01		0.02**	0.01
		(−0.51)	(0.92)		(2.51)	(0.96)
$VWRETD_{t-1}$		−0.09***	−0.09***		−0.07***	−0.14***
		(−8.40)	(−8.36)		(−4.62)	(−7.78)
$VWRETD_{t-2}$		−0.01	−0.03***		−0.01	−0.03*
		(−1.33)	(−2.61)		(−0.61)	(−1.81)
Exchange=NYSE		0.00	0.00		0.03**	0.00*
		(−0.95)	(−0.94)		(2.52)	(−1.90)
Exchange=NASDAQ		0.00	0.00		0.06***	−0.000
		(−0.98)	(−0.96)		(4.68)	(−0.10)
Constant	0.00	0.00	0.00	−0.42***	−0.46***	−0.01***
	(1.39)	(1.62)	(1.58)	(−108.86)	(−37.02)	(−7.02)
Obs (in '000)	1872	1871	1872	1871	1871	1871
N	1329	1329	1329	1329	1329	1329
R^2	0.008	0.011	0.055	0.545	0.529	0.969

The row with lagged equivalent volume, $EqVol^s_{t-1}$, corresponds to the contemporaneous dependent variable DCC_t. Both dependent and independent variables of interest are specified as innovations in Panel A and as logs in Panel B, where superscript ^{s} in $EqVol^s_t$ denotes the appropriate *innovation* or *log* specification, so that the specifications of dependent and independent variables match. Control variables in the regression are contemporaneous and lagged stock and ETF returns. Exchange dummy variable codes three exchanges (NYSE, NASDAQ and AMEX) with AMEX coded with a value of zero. Regression fitted with an intercept using cross-sectionally AR(1) corrected, heteroscedasticity robust errors. Panel units are the stock-fund combinations. The column after each coefficient estimate contains the respective *z*-stats. *Obs* is a number of observations (in thousands). *N* denotes the number of the stock-fund combinations. The asterisks denote significance at *p*-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

logs normalizes high kurtosis and positive skewness in equivalent volume distribution, which results in higher R^2 when compared to the innovation specification. The economic intuition is that a shock in the daily equivalent volume from 10% to 60% of the daily total volume would imply a change of that stock's correlation with the rest of the ETF asset basket from 0.5 to 0.55. The abnormally high R^2 in Panel B Model 1 is due to high autocorrelation for the logs of DCC as confirmed by Wooldridge panel autocorrelation tests with p-values < 0.01. Interestingly, in the case when most of the variation in $log\,(DCC)$ is picked up by its lag, the $EqVol$ lags from $t-2$ to $t-3$ become significantly negative indicating a reversal in comovement.

Fund heterogeneity described in Table 2 may affect the extent of the relationship between equivalent volume and comovement. Table 6 shows the result of fund-by-fund panel regressions. Only the coefficients on the variables of interest are reported. The estimates for the contemporaneous $EqVol^s_{t-1}$ are remarkably similar to Table 5. The coefficients on the equivalent volume innovation, $\Delta EqVol_{t-1}$, range from 0.069 for the Semiconductors ETF (SMH) to 0.342 for the Technology ETF (XLK), with S&P500 ETF (SPY) closer to the average mean of the estimates with 0.200.

Forbes and Rigobon (2002) and Bekaert and Harvey (2003) mention that in some market states comovement and market activity can be highly correlated. To address the concern of the panel regression results driven by subperiods, we run yearly regressions and report our findings in Table 7.[17] The estimates on the contemporaneous equivalent volume innovation, $\Delta EqVol_{t-1}$, remain stable and significantly positive across the subperiods. The increase in the estimates from 0.079 to 0.125 for the year 2003 is most likely due to the start of the SPDR ETFs data. In the following years, the coefficient tapers off to a table range between 0.064 and 0.083.

Given high kurtosis of the equivalent volume in Table 4, it is possible that the association between equivalent volume and comovement may vary over the distribution of the equivalent volume.[18] We investigate the non-linearity of the equivalent volume—comovement relation via a piece-wise regression using splines with knots at quintiles. The coefficients are estimated using a log transformation of equivalent volume under 2-way clustering (Petersen, 2009) and panel-specific robust SE with AR(1) used throughout the paper. The EV–comovement relation is consistently positive and highly significant across all EV quintiles with coefficients ranging from 0.009 to 0.014. As an additional test, we reestimate these regressions using a Fisher Z transformation for the raw correlation coefficient and get the same consistently positive and significant results.[19] These tests suggest that the relation

[17] We also include yearly fixed effects in the panel regressions in Table 5 without any material changes in our findings.
[18] We thank an anonymous referee for pointing this issue out.
[19] Detailed results available upon request.

Table 6
Fund-by-fund Panel Regression of DCC on Equivalent Volume.

ETFs	$EqVol^s_{t+1}$		$EqVol^s_t$		$EqVol^s_{t-1}$		$EqVol^s_{t-2}$		$EqVol^s_{t-3}$		N
XLK	−0.013	(−0.9)	0.013	(0.7)	0.189***	(9.7)	0.144***	(7.7)	0.090***	(5.9)	122
XLY	−0.007	(−0.7)	0.006	(0.5)	0.166***	(14.4)	0.104***	(9.3)	0.048***	(4.9)	118
XLV	−0.037**	(−2.4)	−0.023	(−1.2)	0.159***	(8.0)	0.123***	(6.4)	0.053***	(3.4)	65
XLI	−0.012*	(−2.0)	0.017**	(2.1)	0.112***	(13.1)	0.083***	(10.2)	0.032***	(5.1)	76
XLE	−0.006	(−1.2)	0.012*	(1.9)	0.110***	(17.1)	0.075***	(12.2)	0.034***	(6.8)	52
SPY	0.002	(0.9)	0.014***	(5.3)	0.082***	(29.7)	0.054***	(20.6)	0.019***	(9.2)	620
XLP	−0.006	(−0.7)	−0.001	(−0.1)	0.082***	(7.7)	0.061***	(6.0)	0.034***	(3.8)	46
XLB	−0.002	(−0.6)	0.001	(0.1)	0.074***	(14.0)	0.042***	(8.2)	0.028***	(6.3)	40
OIH	−0.002	(−0.6)	0.002	(0.5)	0.059***	(15.3)	0.044***	(11.7)	0.022***	(6.6)	20
XLF	−0.002	(−0.9)	0.001	(0.3)	0.057***	(17.2)	0.033***	(10.4)	0.018***	(6.4)	113
XLU	−0.010**	(−2.2)	−0.004	(−0.9)	0.051***	(9.9)	0.032***	(6.4)	0.019***	(4.3)	42
SMH	−0.008**	(−2.3)	−0.005	(−1.3)	0.050***	(12.9)	0.037***	(9.7)	0.019***	(5.5)	15

Dependent variable is the innovation in dynamic conditional correlation (DCC), $\Delta DCCt$, as the comovement estimator. Independent variable of interest is the innovation in the lagged equivalent volume, $\Delta EqVol_{t-1}$, represented as $EqVol^s_{t-1}$, contemporaneous with the dependent variable $\Delta DCCt$. Both dependent and independent variables of interest are specified as innovations. Control variables in the regression but omitted from the table are the contemporaneous and lagged stock, ETF and market returns as in Table 5. All regressions are fitted with an intercept using cross-sectionally AR(1) corrected, heteroscedasticity robust errors. N is the number of the stock-fund combinations (panels). The column after each coefficient estimate contains the respective t-stats. The asterisks denote significance at p values smaller than 0.01, 0.05, and 0.1. The funds are sorted by the magnitude of the innovation in the lagged equivalent volume coefficient $\beta_{EqVol^s_{t-1}}$.

Table 7
Yearly Panel Regressions of DCC on Equivalent Volume.

	Time Period									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Variables	ΔDCC_t	ΔDCC_t	ΔDCC_t	ΔDCC_t	ΔDCC_t	ΔDCC_t	ΔDCC_t	ΔDCC_t	ΔDCC_t	ΔDCC_t
ΔDCC_{t-1}	−0.020	−0.239***	−0.226***	−0.223***	−0.208***	−0.214***	−0.242***	−0.175***	−0.173***	−0.199***
	(−0.839)	(−8.844)	(−16.476)	(−17.385)	(−17.623)	(−14.214)	(−13.556)	(−9.443)	(−11.858)	(−11.530)
$EqVol^s_{t+1}$	0.004	0.017	0.006	0.001	−0.006	−0.003	0.001	−0.004	−0.010**	0.004
	(0.283)	(1.594)	(0.996)	(0.305)	(−1.419)	(−0.491)	(0.154)	(−0.969)	(−2.271)	(0.928)
$EqVol^s_t$	0.015	0.052***	0.034***	0.017***	0.007	0.004	0.003	0.001	−0.004	0.016***
	(1.049)	(3.589)	(4.254)	(3.343)	(1.358)	(0.557)	(0.497)	(0.132)	(−0.755)	(2.871)
$EqVol^s_{t-1}$	0.079***	0.125***	0.119***	0.083***	0.081***	0.069***	0.064***	0.078***	0.069***	0.079***
	(5.194)	(8.158)	(14.071)	(16.254)	(14.499)	(10.365)	(11.031)	(15.974)	(13.608)	(13.985)
$EqVol^s_{t-2}$	0.061***	0.098***	0.082***	0.062***	0.064***	0.046***	0.037***	0.048***	0.046***	0.051***
	(4.217)	(7.028)	(10.353)	(12.517)	(11.759)	(7.350)	(6.674)	(10.127)	(9.468)	(9.572)
$EqVol^s_{t-3}$	0.056***	0.058***	0.032***	0.021***	0.028***	0.018***	0.019***	0.023***	0.013***	0.023***
	(4.479)	(5.639)	(5.142)	(5.411)	(6.590)	(3.403)	(4.107)	(5.704)	(3.215)	(5.177)
$R_{stock,t-1}$	−0.019***	−0.032***	−0.040***	−0.059***	−0.019***	−0.057***	−0.021***	−0.014***	−0.004	−0.041***
	(−2.989)	(−2.599)	(−6.635)	(−9.459)	(−3.534)	(−9.943)	(−4.712)	(−4.040)	(−0.845)	(−5.963)
$R_{stock,t-2}$	−0.003	−0.014	−0.009	−0.009	−0.002	−0.016***	−0.004	−0.000	−0.011**	−0.009
	(−0.401)	(−1.184)	(−1.513)	(−1.446)	(−0.377)	(−2.795)	(−0.782)	(−0.046)	(−2.082)	(−1.376)
$R_{etf,t-1}$	0.023*	0.058**	0.007	−0.043**	−0.065***	−0.018	0.062***	0.041***	0.037	−0.040
	(1.701)	(2.512)	(0.305)	(−2.160)	(−3.348)	(−0.635)	(4.598)	(4.059)	(1.537)	(−1.368)
$R_{etf,t-2}$	0.005	−0.011	0.023	−0.005	0.005	−0.003	0.010	−0.003	0.043*	−0.021
	(0.362)	(−0.487)	(1.034)	(−0.253)	(0.232)	(−0.106)	(0.759)	(−0.297)	(1.791)	(−0.715)
$VWRETD_{t-1}$	0.023	0.037	−0.022	−0.005	0.079**	−0.146***	−0.105***	−0.080***	−0.117***	−0.126***
	(0.965)	(0.674)	(−0.512)	(0.134)	(1.973)	(−3.142)	(−4.596)	(−4.040)	(−3.121)	(−3.168)
$VWRETD_{t-2}$	0.002	−0.010	−0.030	−0.050	−0.027	−0.020	−0.023	−0.026	−0.095**	−0.012
	(0.066)	(−0.177)	(−0.702)	(−1.347)	(−0.672)	(−0.430)	(−0.953)	(−1.312)	(−2.516)	(−0.289)
Constant	0.001**	0.000	0.000	0.000	0.000	0.000	0.000	−0.000	0.000	0.000
	(2.563)	(1.406)	(0.605)	(0.795)	(0.347)	(0.895)	(0.187)	(−0.445)	(0.337)	(0.935)
Obs	5945	78,528	219,361	226,404	225,682	238,470	223,913	236,580	233,197	183,995
R^2	0.028	0.061	0.056	0.056	0.050	0.065	0.066	0.044	0.047	0.067
N	30	968	997	1017	1036	1053	1048	1027	1015	993

Both dependent and independent variables of interest are specified as innovations. The column with lagged Equivalent Volume, $\Delta EqVol_{t-1}$, represented as $EqVol^s_{t-1}$, is contemporaneous with dependent variable $\Delta DCCt$. Control variables in the regression are contemporaneous and lagged stock, ETF and market returns as in Table 5. All regressions are fitted with an intercept using cross-sectionally AR(1) corrected, heteroscedasticity robust errors. Panel units are the stock-fund combinations. The t-stats are in the parenthesis below each coefficient estimate. The asterisks denote significance at p- values: *** $p < 0.01$, ** $p < .05$, * $p < 0.1$.

between EV and comovement is positive and robust across the whole distribution of EV and is not driven by the top quintile of the observations.

5.3. Regressions: intraday comovement

In this section, we use six months of intraday data to test the relation between comovement and equivalent volume with an intraday comovement estimator. The comovement estimator is the simple intraday return correlation between the stock and the ETF return. The intraday sample is limited by the data availability to one ETF and 130 trading days (approximately 80 observations of 5-min price points per stock per day) for each stock in that ETF. No particular differences between this fund and the ETFs in the daily sample were noticed.

Comparing the two samples, the relation between comovement and equivalent volume is assessed on a daily level in both samples. The only difference between the daily and intraday sample analysis is the comovement estimator. In the daily sample, comovement is measured with dynamic conditional correlation (DCC), while in the intraday sample it is measured with intraday correlation.

We manually verify the sample observations, control for quote artifacts, and exclude data after 3:55 p.m. We calculate intraday correlations between the underlying stocks and the ETF returns using 80 observations per day representing 5-min intervals following Bollen and Inder (2002). Although 1-min interval data are also available, higher frequency sampling would introduce microstructure artifacts, including price discreteness, infrequent trading and bid-ask bounce effects (Andersen, Bollerslev, Diebold, & Ebens, 2001).

There are no particular restrictions on the timing of intraday arbitrage-induced trades. The arbitrageurs trade when they observe the spread widening such that the difference is significant after transaction costs, which is not constrained to a specific time period within the day. Under equivalent volume-comovement hypothesis, the arbitrage-induced order flows exert price impact on the underlying assets when the underlying assets have lower liquidity than the ETF, which produces return covariation among the underlying assets. If correlated order flows produce comovement with a decay period longer than 5 min, then it will be identified in the 5 min interval price data even in small samples. Furthermore, we do not calculate correlations conditional on the market return like we did with the DCC in Section 3.2 due to the lack of data on intraday market returns, but instead control for market return via control variables in the comovement on equivalent volume regressions.

Fig. 3 shows a plot of daily correlation and equivalent volume innovations for Analog Devices, Inc. (ADI), a component in the Semiconductors ETF (SMH). Correlation between innovations in intraday correlation and equivalent volume is 0.29 and significant at the 1% level.

We employ a similar regression specification to the one from the daily regressions,

$$\rho^s_{i,t} = \alpha + \underbrace{\beta_{\rho^s,t-1}\rho^s_{i,t-1}}_{\text{Lagged DepVar}} + \underbrace{\sum_{t-3}^{t}\beta_{EqVol^s,t}\text{EqVol}^s_{i,t}}_{\text{Equivalent volume and Lags}} + \underbrace{\sum_{t-2}^{t-1}\beta_{stock\_ret,t}R_{i,t} + \sum_{t-2}^{t-1}\beta_{etf\_ret,t}\ R_{etf,t} + \sum_{t-2}^{t-1}\beta_{VWRETD\_ret,t}R_{VWRETD,t}}_{\text{Return Controls}} + \varepsilon_{i,t} \tag{9}$$

where $\rho^s_{i,t}$ is the transformation of the intraday correlation, using either logs $log\left(\rho_{ij,t}\right)$ or innovations $\Delta\rho_{ij,t}$. The same notation applies to the equivalent volume measure, $EqVol^s$. Furthermore, $R_{i,t}$ is the return of the stock i, $R_{etf,t}$ is the ETF $R_{VWRETD,t}$ is the return on the value-weighted CRSP market index. There are no exchange dummied in this model as all of the stocks trade on the same exchange.

Given that the posited relation between arbitrage-induced trading and comovement occurs on the intraday basis, intraday correlations should capture that effect better than the daily comovement estimator from the Section 5.2 should. Under the null hypothesis of no comovement, the coefficient estimates for the equivalent volume variables should be insignificantly different from zero or negative. Although we do not explicitly specify the decay period for the effect of the shock in equivalent volume, the resulting comovement shock is unlikely to last more than one or two trading sessions (Ben-David et al. 2014). The relation between comovement and equivalent volume is assessed on the daily level in both samples. The only difference between the daily and intraday sample analysis is the comovement estimator. In the daily sample, the comovement is measured with dynamic conditional correlation (DCC), while in the intraday sample it is measured with intraday correlation. Hence, if the leads and lags of equivalent volume had similar magnitudes as the contemporaneous variable, it would be harder to explain them with an arbitrage- induced comovement story. Downward slopes in leads and lags, forming a peak at the contemporaneous equivalent volume, are consistent with the transitory comovement shock hypothesis, while the flatter slopes lend credence to the null hypothesis of no comovement.

Since the size of the cross-section for the intraday sample is small, we estimate Eq. (9) individually for each stock in Table 8 and in panel settings in Table 9. For the individual regressions, we employ Newey-West standard error adjustments (Newey & West, 1987) with 5 lags to account for serial correlation of up to 5 trading days. The coefficients for the contemporaneous equivalent volume, $EqVol_t$ are positive and significant. Almost all 18 stocks have significant coefficients at the 10% level, while 12 stocks (66% of the intraday sample stocks) have coefficients significant at the 1% level. Only one stock, Amkor Technology, Inc. (AMKR), has a coefficient insignificantly different from zero on all of the equivalent volume variables, with a relatively low fit ($R^2 = 0.16$). The lowest and highest equivalent volume coefficients are 0.179 for Amkor Technology, Inc. (AMKR) and 0.455 for Applied Materials, Inc. (AMAT), which are 10–25 times larger than the coefficient of 0.018 in the full sample daily panel regressions in Table 5. In the individual regressions, equivalent volume leads and lags are, on average, smaller in magnitude with far fewer significant estimates. Indeed, the contemporaneous equivalent volume has the largest coefficients out of all 5 equivalent volume variables. Thus, as with our earlier findings in the daily sample, results from individual fund regressions provide strong support for the equivalent volume–comovement hypothesis.

As a full-scale test, we implement panel regressions using Eq. (9) as shown in Table 9.[20] The overall model is similar to the daily sample reported in Section 5, except for the comovement estimator calculation.

Fig. 4 plots equivalent volume coefficient estimates from Table 9 for each specification. The peak at the contemporaneous relation between equivalent volume and comovement is evident especially when compared to the daily sample. Leads and lags are insignificant across most of the specifications, only Model 2 has the first lead coefficient with a value of −0.325 and a t-stat of −1.78 that becomes insignificant when we include the lagged dependent variable in Model 1. Furthermore, Model 1 exhibits a flatter slope across the two lags, suggesting a slower decay in the comovement when compared to Models 2 and 3. Interestingly, the coefficient for the lagged equivalent volume in Model 1 becomes negative and significant indicating reversal of comovement. This pattern is similar to the panel regressions using DCC in Table 9.

Coefficient magnitudes for equivalent volume in the regressions using the intraday comovement estimator stand in stark contrast to the coefficients in the regressions using the DCC comovement estimator. Under innovations specification, the coefficient for the concurrent equivalent volume for the most controlled model (Model 1) is 15 times larger than the same coefficient in the daily sample. Moreover, in the log-transformed specification, the difference for the most controlled model is 25 times larger, increasing from 0.011 in the daily sample to 0.276 in the intraday sample.

Given the much stronger estimates of the equivalent volume and comovement relation using the intraday sample, we conjecture that DCC lacks statistical power to fully identify comovement based on the intraday returns. Nonetheless, it is hard to draw inferences from samples with different number of panels and durations. We speculate, however, that the true estimate of the relation between equivalent volume and comovement lies above the lower bound defined by the estimates in the DCC regressions and is better

[20] We also perform robustness tests (untabulated) using two-way clustered standard errors as suggested by Petersen (2009) with varying time unit clustering and find similar results.

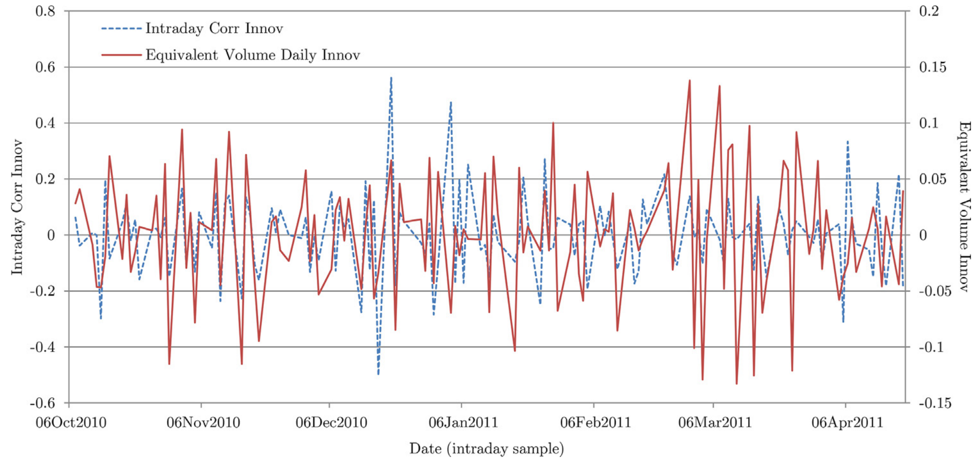

Fig. 3. Intraday Correlation and Equivalent Volume Innovations.
The time series represent intraday correlation and equivalent volume innovations for Analog Devices, Inc. (ADI), a component in Semiconductors ETF (SMH). Frequency is daily and the sample period is from October 2010 to April 2011 — total 6 months equivalent to 130 observations. Intraday correlation between the stock and the ETF returns is calculated using 5 min bar intraday data.

Table 8
Individual Regressions of Intraday Correlations on Equivalent Volume.

Stock	ρ^s_{t-1}		$EqVol^s_{t+2}$		$EqVol^s_{t+1}$		$EqVol^s_t$		$EqVol^s_{t-1}$		$EqVol^s_{t-2}$		Obs	R^2
ADI	0.108	(0.79)	0.100	(1.22)	−0.026	(−0.51)	0.192***	(3.44)	−0.034	(−0.86)	−0.026	(−0.51)	132	0.378
ALTR	0.236*	(1.68)	−0.003	(−0.06)	−0.069	(−1.46)	0.271***	(2.77)	−0.064	(−0.72)	−0.003	(−0.04)	130	0.414
AMAT	0.251	(1.39)	0.177	(1.47)	−0.124	(−1.30)	0.455***	(2.73)	−0.164*	(−1.81)	−0.057	(−0.51)	132	0.407
AMD	0.242**	(2.36)	−0.134*	(−1.87)	0.007	(0.09)	0.356***	(3.94)	0.079	(0.71)	−0.089	(−0.90)	130	0.382
AMKR	0.150**	(2.18)	0.089	(0.77)	−0.079	(−0.97)	0.179	(1.23)	−0.157	(−0.87)	0.422	(1.22)	130	0.296
ATML	0.363***	(3.00)	−0.015	(−0.31)	0.008	(0.18)	0.217***	(4.77)	−0.132**	(−2.14)	0.069*	(1.89)	132	0.404
BRCM	0.345***	(3.01)	0.080	(1.58)	−0.224**	(−2.62)	0.451***	(4.46)	−0.152***	(−2.67)	0.039	(0.83)	132	0.534
INTC	0.523***	(5.90)	−0.058	(−0.65)	0.111	(0.80)	0.200*	(1.91)	−0.223*	(−1.72)	0.093*	(1.75)	132	0.351
KLAC	0.460***	(3.14)	−0.009	(−0.11)	0.042	(0.32)	0.386***	(2.90)	−0.280*	(−1.78)	0.064	(1.52)	132	0.218
LLTC	0.037	(0.33)	0.154***	(3.06)	−0.048	(−0.73)	0.355***	(3.93)	0.001	(0.01)	−0.144**	(−2.45)	132	0.282
LSI	0.341**	(2.42)	−0.010	(−0.10)	0.059	(0.64)	0.242**	(2.29)	−0.060	(−0.45)	0.085	(0.85)	128	0.238
MU	0.195	(1.46)	0.074	(0.93)	0.037	(0.46)	0.233*	(1.72)	0.000	(0.00)	−0.039	(−0.37)	130	0.303
NSM	0.310*	(1.68)	0.085	(0.91)	−0.042	(−0.36)	0.385*	(1.92)	−0.091	(−1.10)	0.175	(1.14)	132	0.356
NVLS	0.327**	(2.32)	0.063	(1.02)	−0.032	-(0.62)	0.226***	(3.03)	−0.080	(−1.50)	−0.017	(−0.29)	132	0.211
SNDK	0.096	(0.91)	0.197***	(3.15)	−0.152**	(−2.35)	0.229***	(3.02)	−0.081	(−0.97)	0.053	(0.77)	132	0.376
TER	0.287**	(2.24)	−0.021	(−0.47)	0.013	(0.28)	0.332***	(6.37)	−0.180**	(−2.14)	0.100*	(1.81)	132	0383
TXM	0.323**	(2.17)	0.084	(1.67)	−0.039	(−0.63)	0.250*	(1.95)	−0.063	(−0.87)	−0.027	(−0.44)	132	0.283
XLNX	0.438**	(2.59)	−0.013	(−0.40)	−0.064*	(−1.70)	0.308***	(3.67)	−0.166*	(−1.84)	0.065	(1.25)	132	0.166

Both dependent and independent variables of interest are specified as logs. The column with the coefficient estimate of the transformation of contemporaneous equivalent volume $EqVol^s_t$ corresponds to the transformation of the contemporaneous intraday correlation ρ^s_{t-1}. Control variables present in the regression but omitted from the table are contemporaneous and lagged ETF returns $R_{etf,t}$ market returns $VWRETD_t$ and stock returns $R_{i,t}$. The estimation is done using Newey-West standard errors under 5 lags specification. The t-stats are in the parenthesis in the adjacent column to each coefficient estimate. The asterisks denote significance at p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

approximated by the values in the intraday correlations regressions. Regardless of the true value of the estimator, the relation between equivalent volume and comovement is positive and significant in both daily and intraday regressions, supporting the equivalent volume–comovement hypothesis.

Additionally, given high kurtosis of the EV in Table 4 we extend the piece-wise regression with splines based on quintiles to the intraday sample and obtain similar results.[21] The coefficients are positive and significant across the EV quintiles (ranging from 0.225 to 0.358) indicating consistency of the EV–comovement relation across the EV distribution.[22]

6. Conclusions

By 2016, the ETFs have amassed $2.5 trillion in assets (ICI, 2017) and reached 29% of the total equity trading volume (NYSE, 2017) since their introduction in 1993. An essential feature of the ETF structure is the arbitrage condition between the ETF shares and the underlying stocks to ensure an exact pricing relationship between the two. This arbitrage is conducted via trading of the underlying stocks and ETF shares. The order flows associated with arbitrage affect the ETF and the underlying assets differently. If the underlying assets have lower liquidity relative to their proportion in the ETF, then the order flows will have a larger impact on the underlying assets resulting in correlated price changes, or comovement.

To measure this relative liquidity we introduce an indicator based on the ratio of weighted trading volumes to reproduce the differential liquidity impact on the ETF and the underlying assets in the context of the arbitrage trades. We call this new indicator, equivalent volume, and use it to estimate the effect of arbitrage-induced price impact on the underlying asset comovement. We use two different estimators and samples to measure asset comovement: dynamic conditional correlations for the daily sample and 5-min intraday return correlations for the intraday sample.

We find strong evidence supporting the hypothesis that equivalent volume and co-movement are related using both daily and

21 We thank an anonymous referee for pointing this issue out.
22 Detailed results available upon request.

Table 9
Panel Regression of Intraday Correlation on Equivalent Volume.

	Panel A: Innovations			Panel B: Logs		
Variables	3 $\Delta\rho_t$	2 $\Delta\rho_t$	1 $\Delta\rho_t$	3 $\log(\rho_t)$	2 $\log(\rho_t)$	1 $\log(\rho_t)$
ρ^s_{t-1}			−0.344***			0.379***
			(−10.208)			(10.345)
$EqVol^s_{t+2}$	−0.031	−0.012	0.025	0.002	0.001	−0.000
	(−0.162)	(−0.064)	(0.138)	(0.061)	(0.036)	(−0.004)
$EqVol^s_{t+1}$	−0.223	−0.325*	−0.271	0.040	−0.052*	−0.051
	(−1.151)	(−1.786)	(−1.388)	(−1.239)	(−1.654)	(−1.431)
$EqVol^s_{t}$	1.429***	1.218***	1.166***	0.279***	0.256***	0.276***
	(7.104)	(6.432)	(5.867)	(8.746)	(8.181)	(7.820)
$EqVol^s_{t-1}$	−0.267	−0.186	0.226	−0.048	−0.037	−0.143***
	(−1.385)	(−1.017)	(1.128)	(−1.493)	(−1.170)	(−3.925)
$EqVol^s_{t-2}$	−0.247	−0.142	−0.169	−0.002	0.020	0.032
	(−1.277)	(−0.771)	(−0.927)	(−0.071)	(0.631)	(1.018)
$R_{stock,t}$		−0.391***	−0.379***		−0.212	−0.258
		(−2.982)	(−2.864)		(−0.445)	(−0.559)
$R_{stock,t-1}$		0.123	−0.041		−0.406	−0.395
		(0.945)	(−0.309)		(−0.857)	(−0.862)
$R_{etf,t}$		−1.098	−0.952		−2.324	−2.059
		(−1.209)	(−1.019)		(−1.038)	(−0.923)
$R_{etf,t-1}$		1.228	0.815		0.118	2.127
		(1.349)	(0.865)		(0.052)	(0.947)
$VWRETD_t$		−3.308**	−3.427***		−3.227	−3.929
		(−2.441)	(−2.475)		(−0.992)	(−1.207)
$VWRETD_{t-1}$		0.717	−0.571		−2.429	−2.835
		(0.528)	(−0.410)		(−0.747)	(−0.871)
Constant	−0.001	0.003	0.005	0.106***	0.109***	0.063**
	(−0.098)	(0.497)	(0.694)	(2.798)	(2.784)	(2.330)
Obs	2358	2356	2356	2371	2369	2364
R^2	0.080	0.157	0.258	0.206	0.225	0.419
N	18	18	18	18	18	18

Intraday sample corresponds to Semiconductors ETF, (SMH) fund with 18 holdings during the sample period from October 2010 to April 2011. Both dependent and independent variables of interest are specified as innovations in Panel A and as logs Panel B. The $EqVol^s$ is either log transformation or first-difference of the $EqVol$ as specified in Eq. (9). Control variables in the regression are stock, ETF and market returns with the corresponding lags as in Table 5. Regression is fitted with an intercept using cross-sectionally AR(1) corrected, heteroscedasticity robust errors. N denotes the number of the stock-fund combinations. The column after each coefficient estimate contains the respective z-stats. The asterisks denote significance at p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

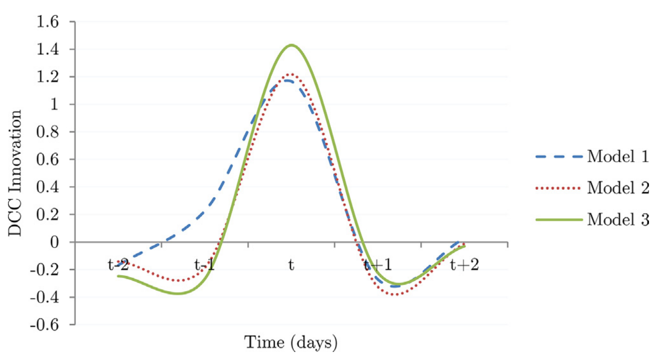

Fig. 4. Equivalent Volume Innovation Coefficients from Intraday Regressions.
The graph depicts equivalent volume innovation coefficient estimates using intraday sample under Models 1–3 from Table 9 Panel A. Equivalent volume innovation is specified as first difference of the equivalent volume in Eq. (2). The regression model estimated is from Eq. (9), where equivalent volume innovations coefficients are specified as two leading coefficients ($t+2$ to $t+1$), one contemporaneous coefficient (t) and two lagging coefficients ($t-1$ to $t-2$). Comovement at t is estimated via 5-min return intraday correlations and is contemporaneous with our variable of interest, equivalent volume at t. Please note that this Fig. 4 is different from Fig. 2 where comovement at t is estimated via dynamic conditional correlations and is contemporaneous with our variable of interest, equivalent volume at $t-1$. Model 1 is the fully controlled specification, Model 2 is the fully controlled specification without the lagged dependent variable (DCC^s_{t-1}) and Model 3 is the equivalent volume only specification. In this Figure comovement at t corresponds to the equivalent volume at t, which is different from Fig. 2 where comovement at t corresponds to an equivalent volume at t-1. All three specifications show that the relation between comovement and equivalent volume is positive and highly significant at t then drops switching signs at $t-1$ indicating slight overshooting and then finally levels off slightly below zero at $t-2$.

intraday data. In economic terms, for the period 2002–2011, a daily equivalent volume shock of 100% is associated with a 1.1% change in the daily comovement as estimated by dynamic conditional correlations (DCC). Using intraday data, we estimate daily correlations for the period from October 2010 to April 2011 and find a 25 times larger and equally significant positive relationship between equivalent volume and stock comovement. Economically, a daily equivalent volume shock of 100% corresponds to the increase in daily comovement of 27.6%, which reduces diversification benefits for investors and market participants. Given that the intraday comovement estimator offers a more powerful test, it is likely that the coefficient estimate from the daily sample using DCC understates the true value of the relation between equivalent volume and comovement.

On average, moderate changes in equivalent volume are associated with statistically significant but relatively small changes in daily comovement. Since large changes in equivalent volume have a higher probability of occurring due to high positive skewness and kurtosis, the magnitude of the associated effects on comovement is largely understated.

The results seem to be in line with the general idea of Wurgler (2010) that argues that index-linked investing might distort fundamental-based asset pricing of equities. Furthermore, our results support the growing literature on the unexpected consequences of arbitrage (Greenwood & Thesmar, 2009; Hong et al. 2012). We also find evidence of the relation between relative liquidity expressed as a ratio of weighted trading volumes (Amihud, 2002; Roll et al. 2010) and the asset comovement contributing to the study of the relative liquidity as the trading-induced contagion propagation channel. Moreover, our findings seem to confirm recent conclusions by Ben-David et al. (2014) and Da and Shive (2016) that the ETF-related activities serve as shock propagation channels from the ETFs to the underlying assets.

The growth of trading volume in ETFs will most likely not abate unless the regulatory landscape changes and, hence, the arbitrage-induced comovement shocks will most likely stay. Given that the relation between equivalent volume and comovement was much stronger using intraday data and that other ETF-related studies use monthly data, we speculate that studying the intraday ETF trading activity (Petajisto, 2013) relative to the underlying stocks will contribute further insights into the pricing of the arbitrage-linked securities and the potential implications for market participants.

Acknowledgements

We thank Chris Schwarz, Kerry Vandell, Lu Zheng, Zheng Sun, Phil Bromiley, Aurelio Vasquez, Alexander Barinov, Maria Schutte, Travis Box, Ana Paula Serra, Ethan Watson and participants at the brown bag presentation and seminars at UC Irvine, ITAM, CSU Fullerton and at the annual meetings of the Southwestern and Southern Finance Association, Eastern Finance Association, Midwest Finance Association, FMA in Europe and USA and for helpful comments and discussions. We are grateful to Alps Distributors/Sector SPDR ETFs for the data provided. All remaining errors are ours.

References

Ackert, L. F., & Tian, Y. S. (2008). Arbitrage, liquidity, and the valuation of exchange traded funds. *Financial Markets, Institutions & Instruments*, *17*(5), 331–362.
Alexander, C., & Barbosa, A. (2005). *Is minimum variance hedging necessary for equity indices? A study of hedging and cross-hedging exchange traded funds*. SSRN eLibrary.
Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, *17*(2), 223–249.
Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*(1), 43–76.
Anderson, A. M., & Dyl, E. A. (2008). *Trading volume: NASDAQ and the NYSE*.
Antón, M., & Polk, C. (2014). Connected Stocks. *Journal of Finance*, *69*(3), 1099–1127.
Asai, M., & McAleer, M. (2009). The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics*, *150*(2), 182–192.
Avellaneda, M., & Lee, J. H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, *10*(7), 761–782.
Büyüksahin, B., Haigh, M. S., & Robe, M. A. (2010). Commodities and equities: ever a 'Market of one'? *The Journal of Alternative Investments*, *12*(3), 76–95 [6]
Bai, Q., Bond, S. A., & Hatch, B. C. (2012). The impact of leveraged and inverse ETFs on underlying stock returns. *46th annual AREUEA conference paper*.
Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, *68*(2), 161–199.
Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, *75*(2), 283–317.
Bekaert, G., & Harvey, C. R. (2003). *Market integration and contagion, working paper 9510*. National Bureau of Economic Research.
Ben-David, I., Franzoni, Fr. A., & Moussawi, R. (2014). *Do ETFs increase volatility? SSRN scholarly paper ID 1967599*. Rochester, NY: Social Science Research Network.
Bollen, B., & Inder, B. (2002). Estimating daily volatility in financial markets utilizing intraday data. *Journal of Empirical Finance*, *9*(5), 551–562.
Box, T., Davis, R. L., & Fuller, K. P. (2016). *ETF competition and market quality, SSRN scholarly paper ID 2721813*. Rochester NY: Social Science Research Network.
Boyer, B. (2011). Style-related comovement: Fundamentals or labels? *The Journal of Finance*, *66*(1), 307–332.
Brennan, M. J., Chordia, T., & Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, *49*(3), 345–373.
Broman, M. S., & Shum, P. M. (2015). *Short-term trading and liquidity clienteles: Evidence from exchange-traded funds, SSRN scholarly paper ID 2361514*. Rochester NY: Social Science Research Network.
Broman, M. S. (2015). *Liquidity, style investing and excess comovement of exchange-traded fund returns, SSRN scholarly paper ID 2307229*. Rochester NY: Social Science Research Network.
Cai, F., & Zheng, L. (2004). Institutional trading and stock returns. *Finance Research Letters*, *1*(3), 178–189.
Chakravarty, S., Gulen, H., & Mayhew, S. (2004). Informed trading in stock and option markets. *The Journal of Finance*, *59*(3), 1235–1257.
Chan, L. K. C., & Lakonishok, J. (1993). Institutional trades and intraday stock price behavior. *Journal of Financial Economics*, *33*(2), 173–199.
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*. SSRN eLibrary.
Cheng, M., & Madhavan, A. (2009). *Dynamics of leveraged and inverse ETFs*. Barclays Global Investors [October].
Cheng, S., Massa, M., & Zhang, H. (2015). *The multiple facets of ETF investing: A world-wide analysis, SSRN scholarly paper ID 2224424*. Rochester, NY: Social Science Research Network.
Chong, J., & Miffre, J. (2009). *Conditional correlation and volatility in commodity futures and traditional asset markets*. SSRN eLibrary.
Chordia, T., & Swaminathan, B. (2000). Trading volume and cross-autocorrelations in stock returns. *The Journal of Finance*, *55*(2), 913–935.
Chordia, T., Subrahmanyam, A., & Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, *59*(1), 3–32.
Claessens, S., & Yafeh, Y. (2011). Additions to market indices and the comovement of stock returns around the world. In *Revised version of a paper formerly circulated as CEPR Discussion Paper*.
Clifford, C. P., Fulkerson, J. A., & Jordan, B. D. (2014). What drives ETF flows? *Financial Review*, *49*(3), 619–642.
Da, Z., & Shive, S. (2012). *Exchange-traded funds and asset return correlations, Tech. Rep., working paper*. University of Notre Dame.
Da, Z., & Shive, S. (2013). *When the bellwether dances to noise: Evidence from exchange-Traded funds, SSRN scholarly paper ID 2158361*. Rochester, NY: Social Science Research Network.
Da, Z., & Shive, S. (2016). *Exchange-traded funds and asset return correlations, working paper*. University of Notre Dame.
Darolles, S., & Le Fol, G. (2003). *Trading volume and arbitrage*.
Darolles, S., Le Fol, G., & Mero, G. (2015). Measuring the liquidity part of volume. *Journal of Banking & Finance*, *50*, 92–105.
Datar, V. T., Naik, N. Y., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, *1*(2), 203–219.
DeLisle, J., French, D. W., & Schutte, M. G. (2015). *Comovement, passive investing and price informativeness, SSRN scholarly paper ID 2166118*. Rochester NY: Social Science Research Network.
Delcoure, N., & Zhong, M. (2007). On the premiums of iShares. *Journal of Empirical Finance*, *14*(2), 168–195.
Domowitz, I., Hansch, O., & Wang, X. (2005). Liquidity commonality and return co-movement. *Journal of Financial Markets*, *8*(4), 351–376.
Doran, J. S., Boney, V., & Peterson, D. R. (2006). *The effect of the spider exchange traded fund on the cash flow of funds of S&P index mutual funds*. SSRN eLibrary.
Edelen, R. M., & Warner, J. B. (2001). Aggregate price effects of institutional trading: A study of mutual fund flow and market returns. *Journal of Financial Economics*, *59*(2), 195–220.
Engle, R. F., & Sarkar, D. (2002). *Pricing exchange traded funds*. SSRN eLibrary.
Engle, R. (2002). Dynamic conditional correlation. *Journal of Business and Economic Statistics*, *20*(3), 339–350.
Fang, Y., & Sanger, G. C. (2011). *Index price discovery in the cash market, SSRN scholarly paper ID 1926287*. Rochester, NY: Social Science Research Network Forbes.
Forbes, K. J., & Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. *The Journal of Finance*, *57*(5), 2223–2261.
Froot, K. A., & Dabora, E. M. (1999). How are stock prices affected by the location of trade? *Journal of Financial Economics*, *53*(2), 189–216.
Fulkerson, J. A., Jordan, S. D., & Travis, D. H. (2014). *Are bond ETF investors smart?, SSRN scholarly paper ID 2560515*. Rochester, NY: Social Science Research Network.
Gallant, A. R., Rossi, P. E., & Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies*, *5*(2), 199–242.
Ge, L., Lin, T.-C., & Pearson, N. D. (2016). Why does the option to stock volume ratio predict stock returns? *Journal of Financial Economics*.
Goetzmann, W. N., & Massa, M. (2003). Index funds and stock market growth. *The Journal of Business*, *76*(1), 1–28.
Goyenko, R. Y., Holden, C. W., & Trzcinka, C. A. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, *92*(2), 153–181.
Green, T. C., & Hwang, B.-H. (2009). Price-based return comovement. *Journal of Financial Economics*, *93*(1), 37–50.
Greenwood, R. M., & Sosner, N. (2007). Trading patterns and excess comovement of stock returns. *Financial Analysts Journal*, *63*(5), 69–81.
Greenwood, R., & Thesmar, D. (2009). *Stock price fragility*.
Greenwood, R. (2008). Excess comovement of stock returns: Evidence from cross-sectional variation in nikkei 225 weights. *Review of Financial Studies*, *21*(3), 1153–1186.
Guedj, I., Li, G., & McCann, C. (2010). Leveraged and inverse ETFs, holding periods, and investment shortfalls. *The Journal of Index Investing*, *1*(3), 45–57.
Hafner, C. M., & Reznikova, O. (2010). On the estimation of dynamic conditional correlation models. *Computational Statistics & Data Analysis* [in press, Corrected Proof].
Hamm, S. (2010). *The effect of ETFs on stock liquidity*. SSRN eLibrary.
Hardouvelis, G., Porta, R. L., & Wizman, T. A. (1994). What moves the discount on country equity funds? In *The Internationalization of equity markets*.

Hasbrouck, J. (2003). Intraday price formation in U.S. equity index markets. *The Journal of Finance*, *58*(6), 2375–2399.
Hasbrouck, J. (2009). Trading costs and returns for U.S. equities: Estimating effective costs from daily data. *The Journal of Finance*, *64*(3), 1445–1477.
Hegde, S. P., & McDermott, J. B. (2004). The market liquidity of DIAMONDS, Q's, and their underlying stocks. *Journal of Banking & Finance*, *28*(5), 1043–1067.
Holden, C. W., Jacobsen, S. E., & Subrahmanyam, A. (2014). *The empirical analysis of liquidity, SSRN scholarly paper ID 2402215*. Rochester NY: Social Science Research Network.
Hong, H., Kubik, J. D., & Fishman, T. (2012). Do arbitrageurs amplify economic shocks? *Journal of Financial Economics*, *103*(2), 454–470.
Hseu, M.-M., Chung, H., & Sun, E.-Y. (2007). Price discovery across the stock index futures and the ETF markets: Intra-day evidence from the S&P 500, Nasdaq-100 and DJIA indices. *Review of Pacific Basin Financial Markets & Policies*, *10*(2), 215–236.
Huang, J. C., & Guedj, I. (2009). *Are ETFs replacing index mutual funds?* SSRN eLibrary.
Investment Company Institute. (2011). *2011 investment company fact book*.
Investment Company Institute. (2017). *2017 investment company fact book*.
Johnson, M., & Newlands, C. (2016). *ETFs: everywhere but not yet a systemic risk*. Financial Times., url: https://www.ft.com/content/59739d3a-bd60-11e6-8b45-b8b81dd5d080. [visited on 10/16/2017]
Johnson, T. L., & So, E. C. (2012). The option to stock volume ratio and future returns. *Journal of Financial Economics*, *106*(2), 262–286.
Johnson, T. (2008). Volume, liquidity, and liquidity risk*. *Journal of Financial Economics*, *87*(2), 388–417.
Kaul, A., Mehrotra, V., & Morck, R. (2000). Demand curves for stocks do slope down: New evidence from an index weights adjustment. *The Journal of Finance*, *55*(2), 893–912.
Krause, T. A., Ehsani, S., & Lien, D. D. (2013). *Exchange traded funds, liquidity, and market volatility, SSRN scholarly paper ID 2153903*. Rochester NY: Social Science Research Network.
Kumar, A., & Lee, C. M. (2006). Retail investor sentiment and return comovements. *The Journal of Finance*, *61*(5), 2451–2486.
Kumar, A., Page, J. K., & Spalt, O. G. (2013). Investor sentiment and return comovements: Evidence from stock splits and headquarters changes. *Review of Finance*, *17*(3), 921–953.
Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1335.
Lakonishok, J., Shleifer, A., & Vishny, R. W. (1992). The impact of institutional trading on stock prices. *Journal of Financial Economics*, *32*(1), 23–43.
Lo, A. W., & Wang, J. (2009). Stock market trading volume. In *Handbook of financial econometrics: Applications*.
Lou, D., & Polk, C. (2013). *Comomentum: Inferring arbitrage activity from return correlations*.
Lu, L., Wang, J., & Zhang, G. (2009). *Long term performance of leveraged ETFs*. SSRN eLibrary.
Madhavan, A., & Sobczyk, A. (2014). *Price dynamics and liquidity of exchange-traded funds, SSRN scholarly paper ID 2429509*. Rochester, NY: Social Science Research Network.
Malamud, S. (2015). *A dynamic equilibrium model of ETFs, SSRN scholarly paper ID 2662433*. Rochester, NY: Social Science Research Network.
McNulty, L. (2017). *El-Erian warns of ETF contagion risks*. Financial News., url: https://www.fnlondon.com/articles/el-erian-warns-of-etf-contagion-risks-20170928. [visited on 10/16/2017]
Milonas, N. T., & Rompotis, G. G. (2010). Dual offerings of ETFs on the same stock index: US vs. Swiss ETFs. *The Journal of Alternative Investments*, *12*(4), 97–113.
NYSE. (2017). *Nyse arca etf report Q4 2017*.. NYSE.com, url: https://www.nyse.com/publicdocs/nyse/products/etp-funds/ETP_Quarterly_Report_Q42017.pdf. [visited on 10/20/2017]
Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.
Petajisto, A. (2013). Inefficiencies in the pricing of exchange-traded funds. In *Available at SSRN 2000336*.
Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Rev. Financ. Stud*, *22*(1), 435–480.
Rai, A., & Tartaroglu, S. (2015). The relative option to stock volume and market response to earnings surprises. *Journal of Accounting, Auditing & Finance*, *30*(4), 431–460.
Richie, N., & Madura, J. (2007). Impact of the QQQ on liquidity and risk of the underlying stocks. *The Quarterly Review of Economics and Finance*, *47*(3), 411–421.
Roll, R., Schwartz, E., & Subrahmanyam, A. (2010). O/S: The relative trading activity in options and stock. *Journal of Financial Economics*, *96*(1), 1–17.
Roll, R., Schwartz, E., & Subrahmanyam, A. (2012). *Trading activity in the equity market and its contingent claims: An empirical investigation*. SSRN eLibrary.
Schlusche, B. (2009). *Price formation in spot and futures markets: Exchange traded funds vs. index futures*. SSRN eLibrary.
Shin, S., & Soydemir, G. (2010). Exchange-traded funds, persistence in tracking errors and information dissemination. *Journal of Multinational Financial Management*, *20*(4–5), 214–234.
Shleifer, A. (1986). Do demand curves for stocks slope down? *The Journal of Finance*, *41*(3), 579–590.
Staer, A. (2017). Fund flows and underlying returns: The case of ETFs. *International Journal of Business* [Forthcoming].
Sun, Z. (2008). *Clustered institutional holdings and stock comovement*. SSRN eLibrary.
Svetina, M., & Wahal, S. (2008). Exchange traded funds: Performance and competition. In *Arizona state university, working paper*.
Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, *68*(6), 54–74.
Tang, H., & Xu, X. E. (2011). *Solving the return deviation conundrum of leveraged exchange-traded funds, SSRN scholarly paper ID 1972009*. Rochester, NY: Social Science Research Network.
Tse, Y., Bandyopadhyay, P., & Shen, Y.-P. (2006). Intraday price discovery in the DJIA index markets. *Journal of Business Finance & Accounting*, *33*(9–10), 1572–1585.
Tuzun, T. (2014). Are leveraged and inverse ETFs the new portfolio insurers? In *Paris december 2014 meeting EUROFIDAI-AFFI paper*.
Vijh, A. M. (1994). S&P 500 trading strategies and stock betas. *Review of Financial Studies*, *7*(1), 215–251.
Wallace, D. G., Kalev, P. S., & Lian, G. (2014). *The evolution of price discovery in US equity and derivatives markets, SSRN scholarly paper ID 2485022*. Rochester, NY: Social Science Research Network.
Warther, V. A. (1995). Aggregate mutual fund flows and security returns. *Journal of Financial Economics*, *39*(2–3), 209–235.
Wermers, R. (1999). Mutual fund herding and the impact on stock prices. *The Journal of Finance*, *54*(2), 581–622.
Wooldridge, J. M. (2001). *Econometric analysis of cross section and panel data* (st ed.). The MIT Press.
Wurgler, J. (2010). *On the economic consequences of index-linked investing*. SSRN eLibrary.
Yu, L. (2005). *Basket securities, price formation, and informational efficiency*. SSRN eLibrary.