The determinants of ETF liquidity: Theory and evidence from European markets

ANNA CALAMIA*
LAURENT DEVILLE†
FABRICE RIVA‡

January 2014
Preliminary version - Do not circulate

Abstract

Despite the importance ETFs have recently gained, little is known about their liquidity. The conventional view on ETF liquidity is that what really matters is neither the size of the ETF nor its trading volume but the liquidity of its benchmark index. We argue that while creation/redemption effectively creates a tight link between the ETF and the index liquidity, other factors are likely to affect the former. We develop a traditional inventory 2-period model where a risk averse ETF market maker liquidates her ETF inventory. She then bears illiquidity costs when closing out her stock basket position. One first result is that ETFs replicating illiquid indices should trade with higher spreads. Second, with risk-averse market makers, our model also predicts that ETF spreads should be positively correlated with the underlying index volatility. Third, we also show that the maker maker will act so as to minimize her expected inventory at the end of the day. This will be easier on more heavily traded ETFs, that should thus exhibit tighter spreads. We then provide empirical evidence of the determinants of the spreads in the European equity ETF markets from their inception in 2000 to the end of 2012. We find that our theoretical predictions are confirmed by the data as the spread appears to positively depend on variables related to the inventory risk (index volatility, low turnover, funding liquidity cost, currency risk). The stock basket spread does not affect ETF spread in the whole sample, as would be expected, but still does so for the 20% ETFs with the lowest daily trading volumes. Collectively, these findings suggest that ETFs are dealt as stocks by market makers when the trading volume is sufficiently high to manage the inventory with low risk. It is only when trading volume is too low that market makers have to account for the illiquidity of the underlying stock basket in their quoted spread .

*GREDEG; 250 rue Albert Einstein, 06560 Valbonne, France; e-mail: anna.calamia@gmail.com

†CNRS and EDHEC Business School; GREDEG; 250 rue Albert Einstein, 06560 Valbonne, France; Tel.: +33 (0)4 93 95 43 97; Fax: +33 (0)4 93 65 37 98; e-mail: Laurent.Deville@gredeg.cnrs.fr

‡IAE de Lille and LEM, 104 avenue du Peuple Belge, 59043 Lille, France; e-mail: fabrice.riva@iae.univ-lille1.fr

1 Introduction

ETFs are one of the major and most successful financial innovations over the past 20 years. Since the inception of the SPDR on the AMEX in 1993, the market for ETFs has experienced dramatic growth in terms of listed instruments, number of providers, assets under management (AuM), trading volume, investment styles and types of replicated assets among others . ETFs have become increasingly popular alternatives to conventional index mutual funds, introducing a new competing investment vehicle, rather than perfectly substituting for their conventional counterparts (Agapova, 2011; Guedj and Huang, 2009). One feature that has made ETFs strong competitors to mutual funds is their ability to provide efficient intraday exposure to an index at low cost as ETFs need not trade the underlying assets when facing demands for creation and redemption. According to Guedj and Huang (2009), this could explain why, in the case of indices exhibiting high volatility, illiquidity or industry concentration ETFs should on average perform better than mutual funds.

Understanding the liquidity of ETFs is of prime importance to answer various questions raised by the development of this market. ETFs now serve a variety of purposes, including hedging and arbitrage. Liquidity is crucial for the efficiency of these operations (Deville and Riva, 2007; Ben David et al., 2012). Initially small in comparison to the stocks they hold, ETFs have significantly increased in terms of AuM and trading volume. They also cover a broader universe of stock baskets including small, exotic and infrequently-traded assets. As these ETFs represent a larger - and in some cases significant - proportion of the market value and trading volume of the underlying stocks, it may become problematic for market makers to act as liquidity providers in all circumstances.

Despite the importance ETFs have recently gained, little is known about their liquidity. The industry, when promoting ETFs, often argues that liquidity in ETFs is a simple matter that does not deserve much attention: the right measure for the liquidity of an ETF is the liquidity of the basket of stocks constituting the benchmark index. Although the role of other factors such as the trading volume is sometimes acknowledged, in the end, it is the immediate access to the liquidity of the underlying that really matters. A widespread opinion among practitioners is that the liquidity of an ETF is determined by the liquidity of the underlying index. The reason for this presumed dependence comes from the open-ended feature of ETFs and from the structure they traditionally use to achieve their replication objective, i.e. physical replication. In this structure, ETF APs (who may also make the market for the ETF) deliver the constituent stocks of the replicated index to the ETF sponsor in exchange for ETF creation units to be traded on the ETF secondary market. Providing liquidity to a large buy order on the ETF secondary market thus implies the creation of new ETF units, which involves the purchase of the corresponding basket on the stock market. Conversely, any market maker holding a large inventory of ETF shares can redeem them against the stock basket at their Net Asset Value. Arguably, the authorized participants will reflect in their ETF quoted spread the implicit costs borne when trading the basket shares, and illiquidity on the stock market will smear over the ETF liquidity. However, we argue that while creation/redemption effectively creates a tight link between the ETF and the index liquidity, other factors related to the inventory risk management by the market maker are likely to affect the liquidity of the ETF, with the effect of separating it from the liquidity of the index constituent stocks.

To address this issue, we first develop a simple two-period model of physical ETF market making. The model builds on traditional inventory risk models (Stoll, 1978 and Ho and Stoll, 1981) but accounts for the liquidation costs incurred by ETF market makers when they create or redeem ETF shares. To illustrate this feature, consider the following example. At the end of the trading day, a market maker holds a short position on ETFs. To provide the ETF shares to her customers, the market maker must instruct her sponsor to create ETF units. The sponsor will provide the ETF units to the ETF market maker in exchange for the delivery of the ETF underlying basket. Since delivering the basket implies the purchase of its constituent stocks on the secondary market, the ETF market maker will bear the illiquidity cost of these stocks.

Consistent with the general view on ETF liquidity, our model thus predicts that ETFs replicating illiquid indices should trade with higher spreads. Since we consider risk-averse market makers, our model also predicts that ETF spreads should be positively correlated with the underlying index volatility, a feature which generally does not deserve much attention in the ETF industry. Finally, our model includes the possibility of active inventory management by ETF market makers. Active management allows market makers to reduce their exposure to adverse price movements and liquidation costs by keeping their inventory low. Since active management is made easier on actively-traded assets, our model also predicts that large ETFs should trade with lower spreads.

We test these predictions on a comprehensive sample of European equity ETFs from their inception in 2000 to the end of 2012. Although the European markets share most of the characteristics of US markets, we think that these markets deserve specific analysis due to some notable differences, in legal structure and fragmentation, in particular. In Europe, the development of more exotic underlying indices combined with a more liberal regulation soon called for synthetic replication based on total return swaps. In such synthetic ETFs, the fund does not hold the underlying securities of the benchmark index. It enters into a total return swap by which the counterparty commits to delivering the exact index return in exchange for the effective performance of the ETF holdings and a fee. The development of these new replication structures has been accompanied by the possibility of creation and redemption in cash: instead of the underlying stock basket, APs can ask for the creation of shares against cash or even futures. For synthetic ETFs we may thus find a weaker link between the liquidity of the ETF and that of its benchmark index.

We first relate the size of ETF spreads to the liquidity of the benchmark index, as measured by the spread of the underlying stock basket. Actually, the alleged capacity of ETFs to build on the liquidity of their benchmark indices relies on their specific open-ended structure. This structure aims at permitting direct arbitrage between the ETFs and the underlying stock basket, which in turn ensures efficient pricing of ETFs. It can also be used by market makers facing orders exceeding the normal capacity of the ETF market: they can always provide liquidity by trading the underlying basket and, as authorized participants (APs), by creating or redeeming shares at their Net Asset Value. The industry thus claims that, the more liquid the benchmark index, the more efficient the liquidity provision by market makers, even for small and new ETFs or when there is little trading volume on the ETF secondary market.

To our knowledge, this argument, though theoretically sound, has not been empirically tested on a large sample of ETFs. It is true that blue chip ETFs generally present tighter spreads compared to those of ETFs replicating small-cap indices. However, several empirical observations cast doubt on this argument. Hougan (2008) computes average spreads of NYSE-listed ETFs for the first months of 2008. He finds that ETFs tracking the same index often display very different levels of spreads. On Euronext Paris, the ETFs competing for the replication of the CAC 40 index trading on Euronext Paris initially displayed spreads that could vary threefold. Competition itself may also impact the size of spreads. For the CAC 40 ETFs, the introduction of two competitors (EasyETF CAC 40 and Amundi ETF CAC 40) in March 2005 after 4 years of monopoly for Lyxor CAC 40 ETFs resulted in a significant decrease in the spreads of the incumbent.

Actually, we find that ETF closing spreads are only weakly associated with the liquidity of the underlying index, and that this relationship is not very powerful. Accounting for conventional determinants of the spreads in equity markets, we are able to significantly improve the model fit. Contrary to the conventional argument, it appears that bid-ask spreads in ETF markets are driven by the variables associated with inventory risk: the underlying index volatility, the funding cost and the ETFs' own characteristics (size and trading volume). We interpret these results as evidence that market making in ETFs is not limited to the ability of creating and redeeming shares. Rather, ETFs are dealt as stocks by market makers as long as the trading volume is sufficiently high to manage the inventory with low risk. We confirm these results on

ETFs exhibiting the lowest and highest trading volumes. While any dependency on the underlying stock basket spread disappears for the more heavily traded ETFs, the stock basket spread affects ETF spread for the 20% ETFs with the lowest daily trading volumes. It is only when trading volume is too low that market makers have to account in their spread the illiquidity of the underlying stock basket.

Related studies focus on how the advent of ETFs impacts the trading and market quality of index component stocks. Building on the theoretical work addressing the impact of the introduction of index futures (Subrahmanyam (1991), Fremault (1991)), Hedge and McDermott (2004) compare measures of information asymmetry on the underlying stocks before and after the ETF starts trading and also investigate which of the ETF or its underlying stock basket displays the highest liquidity. Overall, they find that, in the case of the Diamonds and Cubes (two well-known ETFs tracking the DJIA and the NASDAQ 100 index, respectively), the ETF attracts very little informed trading and is more liquid than the individual stock basket. Although very insightful about the possible impact of the introduction of ETFs, these studies address the question for a very limited number of ETFs trading on US exchanges, and focus on a very limited time period around their introduction. Hamm (2010) investigates the effect of the introduction of ETFs on the liquidity of individual stocks with a large dataset of US equity ETFs. She finds a positive relationship between the availability of ETFs and the adverse selection component of underlying stocks' bid-ask spreads. However, this liquidity effect is in turn transferred to the ETFs, thus reducing the benefit of switching to ETFs for liquidity traders.

The papers of Ben-David et al. (2012) and Petajisto (2011) are also related to liquidity in ETFs, though in a more distant way. Ben-David et al. (2012) use data on ETFs traded on the major US exchanges to assess the role of arbitrage in the propagation of liquidity shocks between ETFs and the underlying stocks. They provide evidence of mispricing in the ETF market and relate it to measures of limits to arbitrage, among which spreads appear specifically significant. Furthermore, they show that arbitrage activity takes place between ETFs and their underlying securities and that it is tightly connected to ETF mispricing. As a result of this arbitrage activity, liquidity shocks can spill over to the underlying assets, increasing volatility and correlations. Petajisto (2011) also highlights the importance of the underlying (and funding) liquidity in the arbitrage activity to exploit the pricing inefficiencies in US ETFs.

Using an empirical framework closely related to ours, Chiu, Chung, Ho and Wang (2012) use ETFs to explore the relation between funding liquidity and equity liquidity during the subprime crisis period for a sample of index and financial ETFs trading in the US. They investigate a very simple regression model relating the liquidity of ETF to conventional determinants and a funding liquidity variable. They find that an increase in funding liquidity can improve equity liquidity, with a stronger effect for the financial ETFs than for the index ETFs. However, they do not account for the underlying index liquidity in their analysis, implicitly ignoring the creation/redemption process and treating ETFs as if simple stocks. The improvement in liquidity observed in Chiu and al. (2012) may in fact channel through an improved liquidity of the underlying stock basket rather than simply through a direct decrease in funding constraints faced by liquidity providers.

Our empirical analysis is based on closing bid-ask spread data as the measure of liquidity. In general, "liquidity is an elusive concept" (Amihud, 2002). Its measurement is difficult due to the different dimensions it encompasses. This is particularly true for ETF markets where (i) shares can be created or redeemed to absorb liquidity shocks, and (ii) the available data is of relatively poor quality. Measuring liquidity in European ETFs is even more challenging as the order flow for an ETF is often fragmented among different trading venues and OTC trades are generally not reported . We restrict our attention to a simple but effective measure of liquidity, the closing spread. We consider that it represents a robust, though noisy, indicator of the liquidity provided by market-makers. Although investors may trade off-exchange, spreads are a visible way through which ETFs can demonstrate liquidity. In addition, despite the alleged capacity of any ETF to raise liquidity, investors still prefer more heavily traded ETFs. As liquidity begets liquidity, spreads are of particular importance to issuers in competition with other issuers and to markets that compete for the order flow.

The rest of the paper proceeds as follows. Section 2 presents a simple two-period model of ETF market making. Its testable implications are developed in Section 3. Section 4 describes our sample of ETFs and the variables we use in the regressions. Section 5 shows that the liquidity of an ETF not only depends on its underlying benchmark but also on its own characteristics among which the turnover plays a major role. Section 6 concludes.

2 A model of ETF market making

This section presents a simple two-period model of ETF market making. The model builds on traditional inventory risk models (Stoll, 1978; Ho and Stoll; 1981, 1983). We extend this framework to the ETF case by accounting for the liquidation costs incurred by ETF market makers when they create or redeem ETF shares.

2.1 Assumptions

We consider a 4-date model. Time is indexed by t, with $t = 0, \ldots, 3$. The market comprises $n+1$ assets where assets $i = 1, \ldots, n$ are ordinary stocks and asset $n+1$ is an ETF whose underlying basket consists of a price-weighted average[1] of stocks $i = 1, \ldots, n$. Denoting $p_{i,t}$ the price of asset i at date t, the Net Asset Value (hereafter NAV) of the ETF at date t is computed as:

$$NAV_t = \sum_{i=1}^{n} p_{i,t} \tag{1}$$

Trades on the ETF market occur at dates 0 and 1. Date 2 is the date at which creation or redemption takes place. At that date, the ETF market marker liquidates her inventory according

[1] Though most benchmark indexes are value weighted, we use a price-weighting average to alleviate the notational burden. Our model can easily be extended to any weighting scheme.

to the following process:

- If the market maker holds a short position in ETFs, she instructs her sponsor to create additional ETF units in exchange for shares. Denoting I the market maker's inventory on ETFs, the sequence of events at date 2 unfolds as follows: (i) The sponsor creates I ETF units that she transfers to the market maker. (ii) The ETF market maker buys I units of the underlying basket, i.e. I units of each constituent stock i, which she transfers to the sponsor in exchange for the I ETF units. (iii) Settlement takes place, after which the ETF market maker holds a flat position (both in ETFs and in stocks).

- If the market maker holds a long position the sequence of events at date 2 unfolds as follows: (i) the I ETF units are redeemed to the sponsor. (ii) Upon redemption, the market maker receives I units of the underlying basket. (iii) The market maker sells the basket shares on the market for constituent stocks and ends up with a flat position (both in ETFs and in stocks).

While a market maker who is short on ETFs has no choice but to buy the basket of constituent stocks in exchange for the corresponding number of ETF units, a market maker who is long might be willing to hold her position rather than redeeming the ETFs. To make the problem symmetric, we will consider that both short and long positions in ETFs and stocks must be closed at the creation/redemption time.

Trades on the underlying stocks occur at dates 0 to 2. Date 3 is introduced only as a terminal condition for valuing constituent stocks as of date 2. The fundamental value $\tilde{\mu}_{i,t}$ of constituent stock i evolves according to the following process:

$$\tilde{\mu}_{i,t} = \bar{\mu}_i + \sum_{\tau=1}^{t} \tilde{\varepsilon}_{i,\tau}, t = 1, 2, 3 \tag{2}$$

where $\tilde{\varepsilon}_{i,\tau}$ is normally distributed with mean zero. We also make the following assumptions: (i) $\text{cov}(\tilde{\varepsilon}_{i,\tau}, \tilde{\varepsilon}_{j,\tau+k}) = 0 \; \forall k \neq 0$ and $\forall (i,j)$, including $i = j$. (ii) $\text{cov}(\tilde{\varepsilon}_{i,\tau}, \tilde{\varepsilon}_{j,\tau}) = \sigma_{i,j} \; \forall (i,j)$ and

$\forall \tau$. We denote Ψ the (cross-sectional) variance-covariance matrix of the ε terms.

2.2 ETF pricing

We assume that prior to date 0, the ETF market maker has an initial cash of c and is endowed with I ETF units. We denote q_t the number of ETF units traded at date t with $q_t > 0$ denoting a purchase by the market maker. Denoting $p_{E,t}$ the trading price of the ETF at date t, the cash position c_1 and the inventory I_1 of the market maker *after* the date 1 trade are equal to:

$$c_1 = c - q_0 p_{E,0} - q_1 p_{E,1} \quad (3)$$

$$I_1 = I + q_0 + q_1 \quad (4)$$

2.2.1 Creation / Redemption

At date 2, the ETF market maker must liquidate the position I_1 she accumulated on the ETF. Based on the discussion in the previous section, this boils down to trading I_1 units of the underlying basket on the market for constituent stocks. We assume that constituent stocks are imperfectly liquid and we denote λ_i the (constant) illiquidity parameter of stock i[2]. As the market maker must liquidate I_1 units of the basket, the cash flow associated with the liquidation trade is equal to $I_1 \sum_{i=1}^{n}(\mu_{i,2} - \lambda_i I_1)$. Note that on a perfectly liquid market, liquidation would occur at the fundamental value so that the associated cash flow would be equal to $I_1 \sum_{i=1}^{n} \mu_{i,2}$. Thus, the cost of illiquidity borne by the ETF market maker is equal to $I_1^2 \sum_{i=1}^{n} \lambda_i$.

[2] Illiquidity on the constituent stock market can arise from the inventory risk borne by liquidity suppliers and/or adverse selection. Illiquidity still manifests at date 2 since the liquidation of constituent stocks occurs at date 3.

2.2.2 The program of the market maker

We are now able to derive the program that the market maker must solve for at date 0. Using previous notations, the expression of the random wealth $\tilde{W}_2$ of the market maker at date 2 is:

$$\tilde{W}_2 = c - q_0 p_{E,0} - \tilde{q}_1 \tilde{p}_1 + (I + q_0 + \tilde{q}_1) \sum_{i=1}^{n} [\tilde{\mu}_{i,2} - \lambda_i (I + q_0 + \tilde{q}_1)] \tag{5}$$

We assume that the market maker has a CARA utility function with an absolute risk aversion coefficient equal to A. Her optimization program thus can be written:

$$\max_{q_0, \tilde{q}_1} \mathrm{E}(-\exp\{-A\tilde{W}_2\}) \tag{6}$$

Program (6) will be solved for by backward induction. To solve the programm on date 1, we first derive the expression for the wealth of the market maker at date 2. Denoting $c_0 = c - q_0 p_{E,0}$ and $I_0 = I + q_0$, it is given by:

$$\tilde{W}_2 = c_0 - q_1 p_{E,1} + (I_0 + q_1) \sum_{i=1}^{n} [\tilde{\mu}_{i,2} - \lambda_i (I_0 + q_1)] \tag{7}$$

The market maker solves for the following program:

$$\max_{q_1} \; c_0 - q_1 p_1 + (I_0 + q_1) \mathrm{E}_1(\tilde{\mu}_{B,2}) - (I_0 + q_1)^2 \left[\lambda_B + \frac{A}{2} \sigma_B^2 \right] \tag{8}$$

where $\tilde{\mu}_{B,2} = \sum_{i=1}^{n} \tilde{\mu}_{i,2}$, $\lambda_B = \sum_{i=1}^{n} \lambda_i$, $\sigma_B^2 = \mathbb{1}'\Psi\mathbb{1}$ and $\mathrm{E}_1(.)$ denotes the conditional expectation based on the information available at date 1. From (8), the expression for the optimal quantity at date 1 is:

$$q_1 = \frac{\mathrm{E}_1(\tilde{\mu}_{B,2}) - p_{E,1}}{\Lambda_1} - I_0 \tag{9}$$

where $\Lambda_1 = 2\lambda_B + A\sigma_B^2$.

Market clearing imposes that the quantity traded by the market maker equals the order flow from ETF traders. We hypothesize that the order flow x_1 at date 1 is price sensitive and takes the form:

$$x_1 = \beta[\mathrm{E}_1(\tilde{\mu}_{B,2}) - p_{E,1}] \tag{10}$$

where $\beta > 0$ captures the interest of customers for the traded ETF. Intuitively, the higher the β, the more traders are willing to exploit the price discrepancy between the ETF price and its fundamental value. Note that high β values allow the market maker to elicit a demand that moves her inventory level toward its optimal value even if the price concession $\mathrm{E}_1(\tilde{\mu}_{B,2}) - p_{E,1}$ is small.

Setting $q_1 = -x_1$, the equilibrium price at date 1 is found to be:

$$p^{\star}_{E,1} = \mathrm{E}_1(\tilde{\mu}_{B,2}) - \frac{\Lambda_1}{1+\beta\Lambda_1} I_0 \tag{11}$$

and the corresponding quantity traded by the market maker is:

$$q^{\star}_1 = -\frac{\beta\Lambda_1}{1+\beta\Lambda_1} I_0 \tag{12}$$

Replacing $\tilde{q}_1$ and $\tilde{p}_1$ in (7) by their expressions in (11) and (12), yields the following program the market maker must solve for at date 0:

$$\max_{q_0} \; c - q_0 p_{E,0} + (I + q_0)\bar{\mu}_B - (I + q_0)^2 \frac{\beta\Lambda_1^2 + \lambda_B - \frac{A}{2}\sigma_B^2(\beta^2\Lambda_1^2 + 2)}{(1+\beta\Lambda_1)^2} \tag{13}$$

where $\bar{\mu}_B = \sum_{i=1}^{n} \bar{\mu}_i$. Differentiating (13) with respect to q_0, the expression for the price at date 0 is:

$$p_{E,0} = \bar{\mu}_B - (I + q_0)\frac{2\beta\Lambda_1^2 + 2\lambda_B + A\sigma_B^2(\beta^2\Lambda_1^2 + 2)}{(1+\beta\Lambda_1)^2} \tag{14}$$

which can be rewritten:

$$p_{E,0} = \bar{\mu}_B - \Lambda_0(I + q_0) \tag{15}$$

where $\Lambda_0 = \frac{2\beta\Lambda_1^2+2\lambda_B+A\sigma_B^2(\beta^2\Lambda_1^2+2)}{(1+\beta\Lambda_1)^2}$.

Assuming that the market maker post quotes for the same quantity at the bid and at the ask, the size of the ETF bid-ask spread at date 0 is given by:

$$s_0 = 2\Lambda_0|q_0| \tag{16}$$

3 Determinants of the spread and testable implications

The spread quoted by the market maker on date 0 for a quantity q_0 of ETFs depends on her risk aversion, on the volatility and liquidity cost of the stock basket and on the interest of customers for the traded ETF. Differentiating (16) with respect to the corresponding variable, we get the following proposition:

Proposition 1. Properties of the ETF spread

i *The spread is increasing in* A

ii *The spread is increasing in* σ_B^2

iii *The spread is increasing in* λ_B

iv *The spread is decreasing in* β

Properties (i) and (ii) are a direct consequence of the ETF market maker risk aversion. Since σ_B^2 measures the variance of the price of the underlying basket, property (ii) implies that ETFs which replicate riskier indices should trade with higher spreads. Apart from general market conditions and underlying stock characteristics, there exist three reasons why some indices may exhibit higher variance. The first one is diversification: indices based on a few constituent stocks or indices focused on a particular industry are likely to be riskier since a greater fraction of idiosyncratic risk is not diversified away. Sector ETFs and ETFs relying on a few constituent stocks should thus exhibit higher spreads. The second one is currency risk, since this risks adds to fundamental risk. Currency risk arises when the ETF's benchmark is a broad indices encompassing stocks that trade in different monetary zones or for ETFs whose trading currency differs from that of their underlying stocks. The third one is the horizon over which the ETF market makers carry index risk. Though this horizon is not accounted for in our model (σ_B^2 is the variance of prices per unit of time), longer horizons translate into higher overall risk. This implies that ETFs trading while the market of their constituent stocks is closed should exhibit higher spreads. Some indices serve as underlying assets for futures contracts. Though we do not account for hedging decisions, a direct implication of our model is that the existence of a futures contract should result in a lower value for the actual σ_B^2 borne (through hedging) by market makers. ETFs replicating a basket for which a futures contract exists should thus exhibit lower spreads.

Property (iii) implies that ETFs replicating baskets that contain illiquid stocks should trade with higher spreads.[3] Blue-chips ETFs should thus exhibit lower spreads compared to ETFs whose basket includes small-cap stocks. Likewise, ETFs replicating baskets on developed country stocks should be cheaper than ETFs replicating emerging markets indices. This implication is consistent with the widespread opininion in the ETF industry, according to which the liquidity of an ETF is determined by the liquidity of its underlying basket. Yet, as our model shows, this is not the whole story.

Property (iv) implies that ETFs that do not attract interest from traders (low β ETFs) exhibit higher spreads. In our model, this arises from the fact that low β values reduce the scope for maker makers to elicit trades that help reducing their inventory at date 1. Since market makers expect to hold a larger inventory at the creation/redemption date, both expected liquidation costs

[3] A potential limitation of our model is that we consider the liquidity of constituent stocks as an exogenous parameter. A broader model should account for the fact that λ_i is actually endogenous. In inventory models, λ_i reflects the risk exposure of market makers. In asymmetric information models, λ_i reflects the losses incurred by market makers from trading with informed investors. In both cases, λ_i is thus partly determined by the variance of stock i prices. This implies that σ_B^2 should not be considered as independent from λ_i.

and risk exposure increase, which results in larger spreads at date 0. The direct implication of property (iv) is that infrequently traded ETFs should exhibit larger spreads.

Taken together, properties (iii) and (iv) may explain why some ETFs use physical replication while others rely on synthetic replication. Though our model is primarily suited for physical replication, the joint implication of properties (iii) and (iv) is that creation/redemption costs should be low for heavily traded ETFs replicating liquid baskets. For such ETFs, the choice between physical and synthetic replication[4] should be irrelevant and both are likely to exhibit tight spreads. On the contrary, creation/redemption costs should be prohibitive for infrequently-traded ETFs operating on illiquid baskets. In this case, synthetic replication seems inevitable.

4 Data description and summary statistics

We use BlackRock's ETF Landscape Global Handbook (2012) to identify ETFs traded on European exchanges. This document is a directory of Exchange Traded Products (ETPs) listed around the world, as at the end of October 2012. For each equity ETF trading in European exchanges, we gather the following information: product name and Bloomberg ticker, benchmark index, type of asset class exposure (broad, region, country and/or sector), total expense ratio (TER), issuer, primary market with listing date and listing date on other exchanges for cross-listed ETFs.

Our primary interest lies in the links existing between the ETF and the index through its constituent stocks. We thus select the ETFs tracking indifferently the price index, the total return index or the equally-weighted index, as long as they all share the same component stocks. We extract from Bloomberg the daily time series of the composition of these indices (from 2 January 2002 to 31 December 2012), which allows us to build the list of all the component stocks over the period.

[4]While immune to the liquidity of the underlying basket, synthetic ETFs are subject to other costs such as collateral swap funding.

We work with a sample of 369 ETFs (1,121 listings) listed on 22 different European exchanges, replicating the 177 different indices for which we have the time-series of the composition. A few large issuers (iShares, Lyxor, DBX) dominate the market, in terms of number of ETFs and AuM. Issuers are highly specialized in terms of replication style but most offer both types of ETFs, which is probably due to specific advantages offered by each technique. Overall, synthetic ETFs are preferred on emerging indices while physical ETFs dominate national and regional indices.

Table 1 reports the description of our ETF sample by replication style (physical or synthetic), as of the end of 2012. We divide the sample according to the type of asset class exposure (broad, sector, regional or single country) and to the geographical area (global, developed, emerging). Synthetic ETFs account for 63% of the total number of ETFs but only for 42% of the total AuM. Almost all types of indices have both synthetic and physical ETFs. Although physical replication can be used to replicate emerging market or sector indices, synthetic ETFs dominate these index types both in terms of number of ETFs and AuM. The proportion of physical ETFs is higher only for developed non-sector indices. For developed markets indices, synthetic ETFs dominate in terms of number of ETFs but physical ETFs dominate in terms of AuM. Cross-listing is the rule in European equity ETFs: most ETFs trade on multiple exchanges either for physical ETFs (70%) or synthetic ETFs (61%).

To construct our liquidity measures for ETFs and their benchmark indices and the other explanatory variables, we extract the following variables from Datastream: for both stocks and ETFs we take high, low, opening and closing prices, closing bid-ask spreads and turnover, and for ETFs we complement these measures with the fund net asset value (NAV) and the number of shares outstanding. The dependent variable in our regressions is the ETF daily relative bid-ask spread measured at the close. [5]

[5]These closing data that smooth out any intraday changes and may not be representative of the market conditions faced by investors trading before the close or OTC. Another possibility is to build daily measures by averaging intraday spreads as in Hassine and Roncalli (2013). However, relying on closing spreads is the easiest way of capturing liquidity even on markets where the quality of data is unreliable. This is of particular importance here as

Other ETF liquidity variables are the turnover and the AuM (computed as the shares outstanding multiplied by NAV). For cross-listed ETFs, we distinguish the trading volume observed on the primary listing exchange from the total volume traded over all listings. We define the primary market as the market on which most of the trading takes place on a given date. Corresponding index variables are the equally weighted daily closing spread for indices (average closing spread for the stocks constituting the index), the value weighted daily closing spread (for the indices for which we have the weights) and the daily index volatility (standard deviation).

We include in our regressions variables related to the ETF market environment: the number of markets where the ETF is listed (cross-listings), the percentage that trades on the primary market, the normalized Herfindahl index of market fragmentation (fragmentation of the order flow over the different listings of a given ETF on a given day). We also compute a measure of the degree of competition on a given index benchmark as the normalized herfindahl index of the trading volumes of all ETFs sharing this benchmark. Following Chiu, Chung, Ho and Wang (2012), we include a funding liquidity variable, measured as the spread between the 3-month Euribor rate and the overnight index swap from Thomson Datastream. We also include dummy variables indicating whether a futures contract on the same index exists or if the ETF and the index do not trade in the same currency (currency risk).

To account for differences in market structure, we include market indicator dummies as control variables. Finally, we include variables related to ETF characteristics: a dummy variable for the replication style indicates whether the ETF uses physical or synthetic replication techniques and takes on value 1 for synthetic and 0 for physical ETFs; as a proxy for the type of asset class, we use a dummy that takes on value 1 for ETFs that replicate a sector index and 0 for non-sector indices (broad, national or regional). To improve the quality of data and avoid extremely noisy observations, we discard in the regressions the first month of data as well as

we work with ETFs trading on 22 different exchanges replicating indexes with 13,817 different constituent stocks trading on 93 exchanges over the world. Moreover, as found with closing data, McInish and Wood (1992) demonstrate that the behavior of equity spreads measured intraday can also be explained by variables measuring activity, competition, risk, and information.

the ETFs that have less than one year of data available.

Table 2 reports summary statistics on the trading characteristics of ETFs and their underlying baskets. ETF liquidity variables are characterized by large variances, with a few very large spreads, trading volumes and AuM pushing up the average values of these variables. Averages are always higher than medians suggesting that a significant proportion of our sample is composed of thinly traded ETFs characterized by large closing spreads. Though the market is not very competitive with an average of 4.8 ETFs replicating the same benchmark, and an average competition index of 0.72, some indices attract more than 20 competing ETFs. The order flow faces some additional fragmentation coming from the different listings of the same ETF. On average, an ETF is listed on 3 different exchanges, and ETFs on popular indices can count up to 7 listings. Fragmentation of the trading volume indicates that in most cases, with an average of 0.87, and with 90% of the trading volume traded on the primary market, one listing is dominant. [6]

5 Empirical determinants of the spread

In this section, we investigate the links existing between the liquidity of the ETFs and the underlying index, as well as the impact of other variables related to ETF characteristics, to the structure of the market, as well as to the benchmark index and its constituent stocks. To explore the determinants of the spread of ETFs and its link with the benchmark spreads in a multivariate setting, we run OLS regressions with standard errors clustered by index benchmark and year-month (Petersen, 2009).

Table 3 reports the determinants of the liquidity of ETFs for the full sample. In model 1, we regress the relative quoted spread of the ETF against the equally-weighted quoted spread of the

[6] The trading volume concentration and competition indices are computed from daily normalized Herfindahl indices on trading volume, over the different listings of ETFs and over the ETFs competing on a given index, respectively. Their value is comprised between 0 (All ETFs have equal market shares) and 1 (there is only one ETF or it concentrates the whole trading volume on that day).

index. We only find a slightly significant (positive) relationship between the two variables and the spread of the basket only explains 8% of the variance in ETF spreads. This is at odds with the claim that the ETF liquidity is that of its underlying index.

Model 2 includes additional variables that are known to be classic determinants of the spread in equity markets (Stoll, 2000). As in equity markets, the size of the spread decreases with the ETF turnover and increases with the benchmark index volatility. This is consistent with the predictions of our model, as stated in properties *iv* and *ii*: Heavily traded ETFs that attract interest from traders exhibit lower spreads and ETFs replicating riskier indices trade with higher spreads. The liquidity of the underlying basket remains a slightly significant determinant of the spread. ETFs spreads also exhibit persistence: positive (negative) spreads at day t-1 are followed by positive (negative) spreads at day t. Following Chiu *et al.* (2012), we add a funding liquidity variable (measuring the degree of funding cost) in our regression. Consistent with their findings that an increase in funding liquidity can improve equity liquidity, we find a positive relation between this variable and the spread. In terms of our model, an increase in funding cost implies an increase in the illiquidity costs borne by the market maker.

To account for the heterogeneity of our sample, we include in model 3 the type of replication and the investment style of the index. Predictions *ii*, *iii* and *iv* imply that infrequently traded ETFs or ETFs which replicate riskier indices or baskets that contain illiquid stocks should trade with higher spreads. Consistent with these predictions, we find that ETFs that replicate sector indices or emerging markets indices have spreads significantly higher than others, while the replication style does not appear to significantly affect the spreads. We investigate more extensively the effects of the replication style and of the benchmark type in Table 4. To account for other risk factors we also include risk-related indicators. We find that the existence of futures contract on the same index has the effect of tightening the spreads (prediction *ii*), while there is no evidence of significant differences in spreads whether the ETF and the index do not trade in the same currency. In all our regressions we have also included market indicators to control for the impact of market-related characteristics: the market on which the ETF is listed explains an important proportion of the variance.

In model 4, we introduce two other variables related to the issues of the competition among ETFs and market places. Fragmentation of the order flow, both across different listings (fragmentation) and across ETFs competing on the same benchmark (competition) is associated with lower spreads. However, while competition among different ETFs (on a given benchmark) has a significant (positive) impact on ETF liquidity, the impact of the fragmentation of the trading volume of a given ETF among market places is not significant. Such results are at odds with previous US results (Boehmer and Boehmer, 2003 and Tse and Erenburg, 2003) where cross-listing is followed by a significant decrease in spreads and may suggest that European ETF markets are segmented.[7] Relative to model 3, the quality of fit improves only slightly.

Overall, in the European equity ETF market, we evidence multiple determinants of closing spreads, among which the liquidity of the benchmark index only accounts for a small, slightly significant part. As predicted by our model, a major contribution is given by variables related to the ETF market environment, such as ETF turnover, competition among ETFs on the same benchmark and funding liquidity, as well as to the difficulties in the replication of the index, such as the volatility of the index or the fact that the ETF replicates a sector or emerging market index.

5.1 Determinants by replication technique and investment style

Given the relevance of the distinction between synthetic and physical ETFs and between sector and non-sector ETFs, we further explore ETF liquidity relationships by dividing our sample according to the replication style and benchmark type. Table 4 compares the regression coefficients across these sub-samples, **for European ETFs**. As for the full sample, the basket

[7] It must be noted that we do not study here the impact of a change in the cross-listing status (cross-listed vs not cross-listed) as in Boehmer and Boehmer (2003) and Tse and Erenburg (2003) but rather the degree of fragmentation which may have a very different impact.

closing spread does not significantly affect the ETF spread, except for non-sector ETF. Consistent with our model, lagged ETF spreads, ETF turnover, index volatility and funding liquidity still explain most of the ETF spread in all sub-samples. According to their benchmark type, ETFs seem differently affected by the replication technique: while non-sector ETFs spreads are not significantly affected, among sector ETF, synthetic ETFs exhibit lower spreads than physical ETFs (prediction *iii* and *iv* implies that since creation/redemption costs are too high for infrequently-traded ETFs operating on less liquid baskets, synthetic replication is preferred). Sector ETFs do not appear to be affected by the dummy indicating whether a futures contract on the same index exists, which may be explained by the fact that more than 95% of these ETFs replicate an index on which a futures exists. The variable emerging market is not included in the investment style regressions because, in our sample, there are no ETFs replicating emerging markets sector indices. **(MSCI)** The impact of competition among different ETFs (on a given benchmark) is more significant for physical than for synthetic ETFs and for sector than for non-sector ETFs.

5.2 Determinants by turnover level

The results we have obtained so far evidence that the spread of ETFs is but weakly related to the liquidity of their underlying index constituent stocks. Based on the predictions of our model, this suggests that market makers can alleviate the illiquidity costs incurred at the time of the creation/redemption process by keeping their ETF inventory low. To further test whether our model provides a fair description of the functioning of the market, we re-estimate our specification of ETF spread determinants by turnover level. Every month, we sort ETFs into five different groups based on their median daily turnover over that month. We define high (low) turnover ETFs as those pertaining to the group that exhibits the highest (lowest) 20% turnovers. The results are reported into table 5. Consistent with our model, we find that low turnover ETFs exhibit spreads that are positively and significantly related to the spread of their underlying basket whereas we find no evidence of such relationship for high turnover ETFs. The two groups also differ regarding the impact of ETF turnover, AuM, index volatility and futures availability. While high turnover ETF spreads strongly depend on these variables, they do not appear as significant determinants of low turnover ETF spreads. Collectively, these results suggest that high turnover ETFs support the market making story that underlies our model and that market conditions and market characteristics that facilitate market makers' management of their inventory translate into lower spreads for these ETFs. For low turnover ETFs, the scope of active management by market makers is reduced, ETF spreads are highly persistent and mostly depend on the liquidity of the underlying basket.

6 Conclusion

We develop a model of ETF market making with inventory risk that accounts for the illiquidity cost incurred by market makers when they create or redeem shares for closing out their position. Our model yields several testable implications. First, the spread of an ETF should be positively related to the spread of its underlying basket. Second, market maker's risk aversion implies that ETF spreads should increase with the volatility of the underlying index. Active management by market makers should mitigate the dependence of ETF spreads on index spreads by keeping inventory low. Since active management is made easier on actively-traded assets, our model also predicts that the underlying basket liquidity should be less critical for the spread of large ETFs.

We test these predictions on European equity ETF markets from their inception in 2000 to the end of 2012. We find that our model provides a fair description of the functioning of the market as spreads appear to positively depend on variables related to inventory risk (index volatility, low turnover, funding liquidity cost, currency risk). The stock basket spread does not affect ETF spread in the whole sample, as would be expected, but still does so for the 20%

ETFs with the lowest daily trading volumes. All these findings suggest that ETFs are dealt as stocks by market makers when the trading volume is sufficiently high to manage their inventory and maintain low levels of risk and costs. It is only when trading volume is too low that market makers have to account for the illiquidity of the underlying stock basket in their quoted spread.

Table 1: ETF sample description by replication type

This table reports the summary characteristics of our sample of ETFs in terms of number of ETFs (# ETFs), number of listings (# Listings) and AuM (in million euros). Asset under Management are as of end of 2012. Cross-listed reports the percentage of ETFs that are cross-listed. Statistics are computed by type of asset class exposure (Broad, Sector, Regional and Single country) and by geographical area (Developed, Emerging and Global).

	# ETFs		# Listings		Cross-listed (%)		AuM	
	Stocks	**Swaps**	**Stocks**	**Swaps**	**Stocks**	**Swaps**	**Stocks**	**Swaps**
Developed Markets	128	194	385	584	69.53 %	62.37 %	59,944.80	41,031.49
Asia-Pacific	2	13	7	41	100 %	76.92 %	175.21	10,767.86
Europe	110	149	315	440	69.09 %	59.73 %	48,660.80	26,205.30
Country	51	40	95	102	43.14 %	60 %	30,335.00	12,653.96
Regional	36	30	157	103	86.11 %	60 %	16,288.42	9,440.28
Sector	23	79	63	235	100 %	59.49 %	2,037.38	4,111.06
North America	16	32	63	103	68.75 %	68.75 %	11,108.79	4,058.32
Emerging Markets	6	28	23	81	50 %	53.57 %	980.25	3,602.21
Global	4	9	17	31	100 %	66.67 %	182.10	529.48
Broad	1	1	2	3	100 %	100 %	86.43	42.56
Sector	3	8	15	28	100 %	62.5 %	95.67	486.92
All	138	231	425	696	69.57 %	61.47 %	61,107.15	45,163.18

Table 2: Summary statistics - Whole sample

This table reports summary statistics on the trading characteristics of ETFs and their underlying baskets. Variables labeled Primary market report statistics on the trading characteristics of the ETFs on their primary market. Variables labeled All report statistics on the trading characteristics of ETFs based on data consolidated across the various trading venues of cross-listed ETFs. The trading volume concentration and competition indices are computed from daily normalized Herfindahl indices on trading volume, over the different listings of ETFs and over the ETFs competing on a given index, respectively. their value is comprised between 0 (All ETFs have equal market shares) and 1 (there is only one ETF or it concentrates the whole trading volume on that day).

	N. obs.	**Mean**	**S.D.**	**Quantiles**				
				1%	**10%**	**50%**	**90%**	**99%**
ETF closing spread (%)	373,987	0.7108%	3.4082%	0.03%	0.09%	0.31%	1.13%	7.17%
ETF daily trading volume (EUR mn.) Primary	373,987	3.26	13.41	0.00	0.00	0.25	5.64	57.96
ETF daily trading volume (EUR mn.) All	373,987	3.81	14.96	0.00	0.00	0.29	6.81	66.09
ETF AuM (EUR mn.)	322,456	308.73	898.93	2.49	11.72	63.23	567.44	4353.07
Number of ETFs competing on the same benchmark	373,987	4.30	4.81	1	1	3	8	24
Competition index	373,987	0.72	0.26	0.18	0.34	0.77	1.00	1.00
Number of listings	373,987	3.03	1.83	1	1	3	6	7
Trading volume concentration index	373,987	0.87	0.21	0.29	0.50	1.00	1.00	1.00
Proportion on trading volume on primary (%)	373,987	90.03%	17.02%	37.55%	59.52%	100.00%	100.00%	100.00%
Basket closing spread (equally weighted) (%)	373,987	0.306%	0.672%	0.039%	0.082%	0.214%	0.566%	1.704%
Basket closing spread (value weighted) (%)	119,983	0.230%	0.809%	0.024%	0.055%	0.164%	0.453%	1.019%
Daily index volatility over the previous month (%)	371,497	1.371%	0.858%	0.410%	0.626%	1.146%	2.377%	4.624%
Daily ETF volatility over the previous month (%)	372,969	1.487%	1.227%	0.445%	0.666%	1.200%	2.594%	5.358%

Table 3: Determinants of ETF spreads

This table reports OLS regression results of ETF spread determinants. Basket closing spread is the daily equally-weighted spread of the underlying index stocks; Lagged ETF daily closing spread is the ETF spread at t-1; ETF turnover is the turnover (in euros) of the ETF across all listings; AuM is the Asset under Management (in euros) computed as the shares outstanding multiplied by NAV; Index volatility is the daily standard deviation of the underlying index; Funding liquidity is measured as the spread between the 3-month Euribor rate and the overnight index swap; Replication is a dummy variable that takes on the value 1 for synthetic ETFs and 0 for physical ETFs; Sector ETF is a dummy variable that takes on the value 1 for ETFs offering exposure to a particular subset of a broader index and 0 otherwise; Currency risk is a dummy variable that takes on the value 1 if the ETF trades in a currency which is different from that of its underlying index; Futures is a dummy variable that takes on the value 1 if the ETF replicates a benchmark that serves as an underlying asset for a futures contract; Emerging Market is a dummy variable that takes on the value 1 for ETFs replicating emerging market indexes; Fragmentation is the Herfindahl index for the concentration of the trading volume of the primary ETF: its value is comprised between 0 (the ETF trading volume is equally split across all the markets where the ETF is listed) and 1 (one trading venue dominates the others or the ETF is not cross-listed); Competition is computed as 1 minus the Herfindahl index of the trading volumes of all ETFs sharing the same index benchmark. Competition ranges from 0 (one ETF attracts all turnover) and 1 (all ETFs have the same turnover share). Market dummies are dummy variables aimed at controlling for differences in the trading characteristics of the markets ETFs trade on. Year dummies aim at controlling for time trends in overall liquidity. We report t-statistics based on clustered standard errors using a double clustering on both year-motnhs and index benchmark. ***, ** and * denote significance at the 10%, 5% and 1% level, respectively.

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Intercept	0.0041 ***	0.0068 ***	0.0082 ***	0.0074 ***
t-stat	3.26	5.64	7.43	6.46
Basket closing spread (EW)	0.1041 *	0.0644 *	0.0192	0.0171
t-stat	1.82	1.68	1.23	1.15
Lagged ETFs closing spread		0.0366 ***	0.0323 ***	0.0322 ***
t-stat		2.81	2.81	2.81
ETF turnover		-0.0004 ***	-0.0001 ***	-0.0001 ***
t-stat		-10.22	-5.77	-5.47
AuM			-0.0005 ***	-0.0005 ***
t-stat			-7.84	-7.14
Index volatility		0.0763***	0.0440 ***	0.0453 ***
t-stat		6.11	4.91	5.02
Funding liquidity		0.0012 ***	0.0017 ***	0.0017 ***
t-stat		3.54	4.96	4.92
Replication			-0.0003	-0.0002
t-stat			-1.32	-1.02
Sector ETF			0.0013 ***	0.0013 ***
t-stat			4.37	4.43

< Continued on next page >

< Continued from previous page >

Variable	Model 1	Model 2	Model 3	Model 4
Currency risk			0.0005	0.0005
t-stat			1.58	1.63
Futures			-0.0019 ***	-0.0018 ***
t-stat			-4.58	-4.33
Emerging Market			0.0037 ***	0.0036 ***
t-stat			4.97	4.83
Fragmentation				0.0003
t-stat				1.21
Competition				-0.0009 ***
t-stat				-3.14
Market dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
N. obs.	351,853	316,938	316, 938	316,938
Adj. R-squared	8.19%	17.50%	24.90%	25.05%

Table 4: Determinants of ETFs spreads by style

This table reports OLS regression results of ETF spread determinants by replication style (physical vs synthetic) and benchmark type (ETFs replicating sectors vs non-sector indices). Basket closing spread is the daily equally-weighted spread of the underlying index stocks; Lagged ETF daily closing spread is the ETF spread at t-1; ETF turnover is the turnover (in euros) of the ETF across all listings; AuM is the Asset under Management (in euros) computed as the shares outstanding multiplied by NAV; Index volatility is the daily standard deviation of the underlying index; Funding liquidity is measured as the spread between the 3-month Euribor rate and the overnight index swap; Replication is a dummy variable that takes on the value 1 for synthetic ETFs and 0 for physical ETFs; Sector ETF is a dummy variable that takes on the value 1 for ETFs offering exposure to a particular subset of a broader index and 0 otherwise; Currency risk is a dummy variable that takes on the value 1 if the ETF trades in a currency which is different from that of its underlying index; Futures is a dummy variable that takes on the value 1 if the ETF replicates a benchmark that serves as an underlying asset for a futures contract; Emerging Market is a dummy variable that takes on the value 1 for ETFs replicating emerging market indexes; Fragmentation is the Herfindahl index for the concentration of the trading volume of the primary ETF: its value is comprised between 0 (the ETF trading volume is equally split across all the markets where the ETF is listed) and 1 (one trading venue dominates the others or the ETF is not cross-listed); Competition is computed as 1 minus the Herfindahl index of the trading volumes of all ETFs sharing the same index benchmark. Competition ranges from 0 (one ETF attracts all turnover) and 1 (all ETFs have the same turnover share). Market dummies are dummy variables aimed at controlling for differences in the trading characteristics of the markets ETFs trade on. Year dummies aim at controlling for time trends in overall liquidity. We report t-statistics based on clustered standard errors using a double clustering on both year-months and index benchmark. ***, ** and * denote significance at the 10%, 5% and 1% level, respectively.

	Replication style		**Benchmark type**	
Variable	**Stocks**	**Swaps**	**Sector**	**Non-Sector**
Intercept	0.0066***	0.0101***	0.0133***	0.0067***
t-stat	6.04	7.21	8.72	6.11
Basket closing spread (EW)	-0.0123	0.0192	0.0455	0.0379*
t-stat	-0.43	1.17	1.53	1.80
Lagged ETFs closing spread	0.0229**	0.0351***	0.0252***	0.0255**
t-stat	2.13	3.06	2.92	2.08
ETF turnover	-0.0001***	-0.0001***	-0.0001***	-0.0001***
t-stat	-3.87	-3.92	-3.65	-2.99
AuM	-0.0004***	-0.0005***	-0.0004***	-0.0003***
t-stat	-4.81	-4.74	-4.49	-4.25
Index volatility	0.0637***	0.0366***	0.0574***	0.0515***
t-stat	6.66	2.86	5.85	4.65
Funding liquidity	0.0009***	0.0024***	0.0033***	0.0005**
t-stat	2.74	5.03	5.54	2.50
Replication			-0.0017***	0.0000
t-stat			-6.14	-0.16

< Continued on next page >

< Continued from previous page >

	Replication style		Benchmark type	
Variable	**Stocks**	**Swaps**	**Sector**	**Non-Sector**
Sector ETF	0.0014***	0.0011***		
t-stat	2.96	3.26		
Currency risk	0.0001	0.0006	0.0016	0.0010
t-stat	0.23	1.46	1.83	1.93
Futures	-0.0016***	-0.0022***	-0.0003	-0.0018***
t-stat	-3.55	-2.96	-0.51	-3.77
Emerging Market	0.0017***	0.0033***		
t-stat	3.74	3.79		
Fragmentation	-0.0001	0.0003	0.0004	0.0002
t-stat	-0.40	0.85	1.01	0.44
competition	-0.0012***	-0.0008*	-0.0008**	-0.0007*
t-stat	-3.61	-1.96	-2.48	-1.89
Market dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
N. obs.	147,623	169,315	116,475	127,383
Adj. R-squared	34.70%	22.65%	27.41%	18.11%

Table 5: Determinants of ETFs spreads by turnover level

This table reports OLS regression results of ETF spread determinants by turnover level (high turnover vs low turnover). Every month, we assign ETFs to 5 portfolios based on their median turnover over that month. Low turnover corresponds to ETFs exhibiting the lowest 20% median turnovers. High turnover corresponds to the ETFs exhibiting the highest 20% median turnover. Basket closing spread is the daily equally-weighted spread of the underlying index stocks; Lagged ETF daily closing spread is the ETF spread at t-1; ETF turnover is the turnover (in euros) of the ETF across all listings; AuM is the Asset under Management (in euros) computed as the shares outstanding multiplied by NAV; Index volatility is the daily standard deviation of the underlying index; Funding liquidity is measured as the spread between the 3-month Euribor rate and the overnight index swap; Replication is a dummy variable that takes on the value 1 for synthetic ETFs and 0 for physical ETFs; Sector ETF is a dummy variable that takes on the value 1 for ETFs offering exposure to a particular subset of a broader index and 0 otherwise; Currency risk is a dummy variable that takes on the value 1 if the ETF trades in a currency which is different from that of its underlying index; Futures is a dummy variable that takes on the value 1 if the ETF replicates a benchmark that serves as an underlying asset for a futures contract; Emerging Market is a dummy variable that takes on the value 1 for ETFs replicating emerging market indexes; Fragmentation is the Herfindahl index for the concentration of the trading volume of the primary ETF: its value is comprised between 0 (the ETF trading volume is equally split across all the markets where the ETF is listed) and 1 (one trading venue dominates the other or the ETF is not cross-listed); Competition is computed as 1 minus the Herfindahl index of the trading volumes of all ETFs sharing the same index benchmark. Competition ranges from 0 (one ETF attracts all turnover) and 1 (all ETFs have the same turnover share). Market dummies are dummy variables aimed at controlling for differences in the trading characteristics of the markets ETFs trade on. Year dummies aim at controlling for time trends in overall liquidity. We report t-statistics based on clustered standard errors using a double clustering on both year-months and index benchmark. ***, ** and * denote significance at the 10%, 5% and 1% level, respectively.

Variable	**High turnover**	**Low turnover**
Intercept	0.0073***	0.0027
t-stat	4.39	1.44
Basket closing spread (EW)	0.0200	0.1946***
t-stat	1.60	3.75
Lagged ETFs closing spread	0.0261	0.0235***
t-stat	1.31	2.83
ETF turnover	-0.0001**	-0.0001
t-stat	-2.28	-1.64
AuM	-0.0003***	-0.0002
t-stat	-3.72	-0.88
Index volatility	0.0685***	0.0295
t-stat	5.93	1.38
Funding liquidity	0.0007***	0.0024***
t-stat	3.27	3.81
Replication	-0.0001	-0.0002
t-stat	-0.60	-0.29
Sector ETF	0.0011***	0.0019***
t-stat	4.14	2.66

< Continued on next page >

< Continued from previous page >

Variable	High turnover	Low turnover
Currency risk	-0.0001	0.0001
t-stat	-0.44	0.13
Futures	-0.0014***	-0.0015*
t-stat	-3.21	-1.86
Emerging Market	0.0015**	0.0072***
t-stat	2.37	6.27
Fragmentation	-0.0004	0.0005
t-stat	-1.14	0.73
Competition	-0.0007***	-0.0003
t-stat	-2.77	-0.50
Market dummies	Yes	Yes
Year dummies	Yes	Yes
N. obs.	76,352	48,773
Adj. R-squared	19.44%	29.46%

References

[1] AGAPOVA, A. Conventional mutual index funds versus exchange traded funds. *Journal of Financial Markets 14*, 2 (2011), 323–343.

[2] AMIHUD, Y. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5* (2002), 31–56.

[3] BEN-DAVID, I., FRANZONI, F., AND MOUSSAWI, R. Etfs, arbitrage and shock propagation. Fisher College of Business Working Paper. Available at SSRN: http://ssrn.com/abstract=1967599, 2012.

[4] CHIU, J., CHUNG, H., HO, K.-Y., AND WANG, G. Funding liquidity and equity liquidity in the subprime crisis period: Evidence from the etf market. *Journal of Banking and Finance 36* (2012), 2660–2671.

[5] DEVILLE, L., AND RIVA, F. Liquidity and arbitrage in options markets: A survival analysis approach. *Review of Finance 11*, 3 (2007), 497–525.

[6] FREMAULT, A. Stock index futures and index arbitrage in a rational expectations model. *Journal of Business 64* (1991), 523–547.

[7] GUEDJ, I., AND HUANG, J. Are etfs replacing index mutual funds? Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1108728, 2009.

[8] HAMM, S. J. The effect of etfs on stock liquidity. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1687914, 2010.

[9] HASSINE, M., AND RONCALLI, T. Measuring performance of exchange traded funds. Working paper, Available at SSRN:http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2212596, 2013.

[10] HEDGE, S., AND MC DERMOTT, J. The market liquidity of diamonds, q?s and their underlying stocks. *Journal of Banking and Finance 28* (2004), 1043–1067.

[11] HO, T. S. Y., AND STOLL, H. R. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics 9* (1981), 47–73.

[12] HO, T. S. Y., AND STOLL, H. R. The dynamics of dealer markets under competition. *Journal of Finance 38* (1983), 1053–1074.

[13] HOUGAN, M. The impact of spreads on etfs. *Institutional Investors Guide to Exchange Traded Funds* (2008), 137–141.

[14] MCINISH, T., AND WOOD, R. An analysis of intraday patterns in bid/ask spreads for nyse stocks. *Journal of Finance 47*, 2 (1992), 753–764.

[15] PETAJISTO, A. Inefficiencies in the pricing of exchange-traded funds. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2000336, 2011.

[16] PETERSEN, M. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies 22*, 1 (2009), 435–480.

[17] STOLL, H. Friction. *Journal of Finance 55*, 4 (2000), 1479–1514.

[18] STOLL, H. R. The supply of dealer services in securities markets. *Journal of Finance 33* (1978), 11335–1151.

[19] SUBRAHMANYAM, A. A theory of trading in stock index futures. *Review of Financial Studies 4* (1991), 17–51.