A Work Project, presented as part of the requirements for the Award of a Master's

Degree in Finance from the NOVA – School of Business and Economics.

THE EFFECTS OF ETF CREATION

ON THE PRICE EFFICIENCY OF UNDERLYING STOCKS

BRUNA HELENA BELCHIOR MACHADO DA SILVA BUENO, 336

A Project carried out in the Master in Finance under the supervision of

Professor Melissa Prado

JUNE 4, 2012

THE EFFECTS OF ETF CREATION

ON THE PRICE EFFICIENCY OF UNDERLYING STOCKS

In this assignment I build an intuitive panel regression model, in order to achieve a clear isolation of the impact of the inception of the first Exchange Traded Fund created on the FTSE100 index on the price efficiency of its underlying stocks. The main finding of this analysis is that price efficiency at the individual stock market decreases after ETF introduction. Thus, the adverse selection hypothesis highlights the shift of liquidity traders to the basket security, leaving informed traders exposed in the individual market. This decrease is evident and significant for different time range samples employed, as well as for the several measures of price efficiency used.

Purpose of Project

The main purpose of this assignment is to investigate the impact of the creation of Equity Exchange Traded Funds (ETFs) on the pricing efficiency of their underlying stocks.

The market of ETFs shows vivid increase since its creation and mainly in the most recent years, due to its unique advantageous properties, in such dimensions that its possible effects on the financial market cannot be left unexplored. An interesting task is to investigate how the inception of ETFs impacts the price dynamics of the underlying component stocks. Is the speed at which prices incorporate information altered after ETF creation? Or do ETFs have insignificant impact on the price formation process of its component stocks?

In order to include the returns of a given index in their portfolio, an investor can either buy all the shares belonging to the index, or buy it from a mutual fund which already assembles the index composition in a share, or invest in an ETF. Buying the ETF brings transaction cost, tax benefits and trading flexibility benefits in comparison to buying the index from the open end fund, which are the major reasons why this recent financial product has become so successful. The first ETF in the form as one sees them today, the Standard & Poor's Depositary Receipts (SPDR), is created in the USA in 1993 and it replicates the S&P 500 Index.
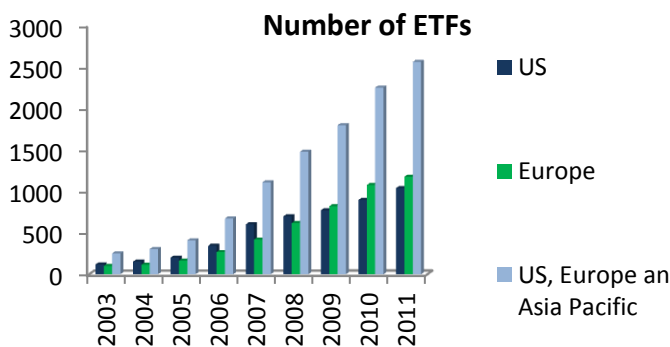
Graph 1- Evolution of ETFs – Number of ETFs, Source: Deutsche Bank
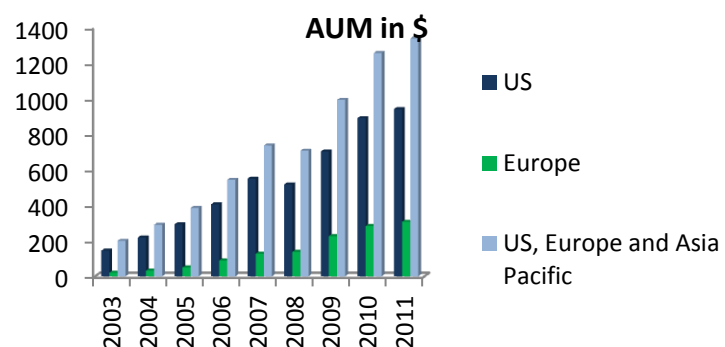
Graph 2 – Evolution of ETFs – Assets under Management in US dollars, Source: Deutsche

Following that, the growth of both the number of ETFs and the value of assets under management (AuM) reflects the popularity of the ETF industry.

The various ETFs are distributed among the world, mainly in the U.S. and Europe, as shown in the graphs above. Although Europe has a higher number of ETFs available, the U.S. is still a long step ahead in terms of AuM. Furthermore, ETFs are not constrained to replicating an equity index, but they range from equity, fixed income, currency and commodity, to inverse, leveraged, physical or synthetic.

Therefore, as the ETF can assume large magnitude and complexity, whose effects remain unclear and not extensively explored, I find it indispensable to analyze it. Aroused by the consequences of ETF creation, I discuss the issue of their effect on price efficiency of the underlying component stocks.

It would be unreasonable to take the financial market as comprising a completely efficient price mechanism which provides a complete incorporation of all available information instantaneously. The existence and degree of market efficiency is a keystone in all financial theory. By one of the most important academic developments, the efficient market hypothesis of Eugene Fama (1970), we can declare markets eventually produce the price embedding all available information. Still, it is inevitable to realize that anomalies occur and arbitrage opportunities exist, even if only for a given time period after which they are eliminated. This same time of reaction to new information, in which mispricing is eliminated, is variable (Merton, 1970). Efficiency in price is strongly connected to what and how investment decisions are made, thus, one is left to explore the extent to which markets are efficient and what drivers affect that same efficiency.

Literature Review

The effect of ETF creation on the price efficiency of the underlying stocks is still not fully and surely defined and explained academically, so puzzles still remain unsolved.

There is vast evidence supporting a significant and positive contribution of ETF creation on the price discovery and efficiency of the respective index and underlying component stocks. Chu et al (1999) propose 3 hypotheses that follow the logic of the informed investor taking the leading role in the process of bringing prices to equilibrium. The first idea implies that informed investors prefer to choose leveraged markets; secondly, the trading cost hypothesis reasonably indicates that they avoid high trading costs; the third hypothesis concerns trading restrictions and similarly to the previous idea, suggests that informed investors will avoid restrictions to trading. Thus, all these premises are in accordance with the ETF market. Therefore, if informed investors allocate to the basket of securities, its informational power is higher. The main point is that in the case of the shift of the informed investor to the ETF market, this index linked security market brings the opportunity to arbitrage exiting mispricing (Fremault (1991), Kumar and Seppi (1994) and Holden (1995)), the so called *arbitrage hypothesis*.

The following literature specifically relates exchange traded funds to the price efficiency of their underlying market, and relies on models of asymmetric information among investors, where their objective is to discover the relative informational role of various parallel markets, as futures and ETFs.

Hasbrouck (2003) is the first who investigates the relationship between ETFs in particular and the price discovery process in their respective index. This study is not only conducted in the ETF environment, but it also analyzes E-mini futures contracts and the regular index futures contracts, in which it tries to quantify each market's

informational contribution to the configuration of the index price. The motivation behind this research lies not only on the high success but also on the curiosity about how these relatively new securities (ETFs and E-Mini futures) could modify the short-run dynamics of price formation, specifically price discovery and leadership. The main findings are that the E-mini market plays the major role in contributing to price discovery of both the S&P 500 and the Nasdaq 100 indices, only then followed by the ETF market.

Tse, Bandyopadhyay and Shen (2006) build on the previous study by similarly analyzing the dynamics of price discovery between the Dow Jones Industrial Average Index (DJIA) and its correspondent ETF (DIAMOND), E-mini futures and regular futures. They add value to previous studies by working with a different index, but apparently the most important twist influencing their findings is their use of electronically traded ETFs, in addition to the floor traded ETFs used by Hasbrouck. They claim that using just floor traded ETFs may not be consistent with the use of E-minis which are electronically traded. With this approach we can exclude the possibility of accounting for investors who are choosing E-Minis and not ETFs simply because they would rather trade on electronic platforms. The fact is that their findings enhance the importance of ETFs in the process of price discovery when compared to the previous findings.

Bernd Schlusche (2009) examines price leadership from the ETF and futures markets as previously introduced by Hasbrouck, although this study extends to analyzing the issue whether liquidity and volatility contribute to the fact that the futures market keeps on being the major contributor. The findings report that when volatility is high, the ETF market plays a stronger role in the process of price discovery. The research is conducted

on the German Stock Index, Deutscher Aktien Index (DAX30), which to the best of my knowledge, is one of very few studies on this specific topic on a European index.

Furthermore, Lei Yu (2003) shifts the focus from the respective index to the individual component securities, in their analysis about the impact of ETFs on the price formation of the underlying assets. They attempt to investigate the informational function that ETFs perform on the price formation process, the informational efficiency and market quality of the underlying index stocks. These questions are driven by the fact that alterations in the market and price efficiency represent important practical implications for investment decisions of the firms representing the component stocks. Once again, they find that the Standard & Poor's Depositary Receipts (SPDR) ETFs play a significant role in the price discovery process, this time at the individual stock level, and also conclude that they may have a permanent impact on the stock's price.

Finally, Chen and Strother (2008), already working at the individual stock level, uncover the price dynamics impacts for the Shanghai Stock Exchange 50 Index (SSE50) ETF. Besides the different index and geography explored, they extend the study to the case where stock price limits are implemented by law. If a stock price is regulated not to fall below or rise past a certain threshold per day, then a clear case of price inefficiency is present. They analyze the effect the ETF, which continues to trade normally even if a certain stock hits its limit price, has on the price discovery process, and conclude that they indeed play an important role in the process.

Besides all this investigation, there is also the point of view that investors shifting to the basket security market when it is created, rather than being informed investors, they are liquidity traders who are rational but less informed than privately informed traders. In this case, assuming that liquidity traders do not bring fundamental value knowledge to the more liquid market, the imbalances are not removed. The major idea is that the shift

of uninformed investors to the basket security market, leaves the adverse selection cost to increase in the underlying security market and thus decreases price efficiency. Yet these studies, as Subrahmanyam (1991), Jegadeesh and Subrahmanyam (1993), Gammill and Perold (1989) and Gorton and Pennacchi (1991) focus on the stock index futures market, and not in the ETF market in particular.

I build on the previous research by shifting my analysis to the European market, which remains to be quite unexplored concerning this particular research topic, which relates ETFs and price efficiency. In particular, I study the impact of the first ETF built on the FTSE100 index, the iShares FTSE100 in April 2000.

Methodology

For the purpose of this assignment I examine the impact of ETF creation on the stock price efficiency, using a fixed effects regression analysis with panel data, so as to isolate and quantify the impact of the ETF creation. Thus, the dependent variable of the regression is price efficiency and the independent variables are chosen considering they affect stock price efficiency.

Measures of price efficiency

Firstly, I determine how to measure price efficiency. Therefore I use four different measures of price efficiency: The Cross-autocorrelation (Bris, Goetzmann and Zhu (2007)), the Variance Ratio (Lo and MacKinlay (1988)) and two Delay measures (Hou and Moskowitz (2005)).

The first measure is the *Cross-autocorrelation*, used by Bris, Goetzmann and Zhu (2007) with the purpose of computing the speed of stock price adjustment to market movement, in their study of the impact of short sale constraints on market efficiency. This measure is the correlation between the stock return at time period t and the lagged local market return at time period t-1, represented as:

$$\rho_{i,t}^{cross} = corr(r_{i,t}, r_{m,t-1})$$

However, correlations are restricted to assume values between -1 and 1 only. In order to solve this problem I apply the transformation $log\left[\frac{(1+\rho)}{(1-\rho)}\right]$ and define it as the Cross-autocorrelation as in Bris, Goetzmann and Zhu's (2007) work, so the measure now assumes values ranging from $-\infty$ to $+\infty$. It suggests that the higher is the correlation between the stock return and the market return of the previous period, the lower is the price efficiency. In other words, the closer the stock moves along with the market, the more it appears to be dependent on the market information. The fact that a stock is less impacted by the news of the overall market from the previous period, in this case one week, it means that it enjoys from more efficiency and moves randomly, similarly to the subsequently presented measures. Nevertheless, a drawback of this measure is if correlation is negative, an increase does not translate into more correlation with the market, just less negative correlation.

The second measure of price efficiency is the absolute value of the Variance Ratio (*VR*) presented in the study of Lo and MacKinlay (1988), in which their evidence rejects prices following a random walk. In their work, this measure is used to test whether stock prices follow a random walk for various portfolios, in which their null hypothesis is that *VR* is equal to 1. This means that the closer the *VR* values resulting from the present data are to 1, the more efficient the stock prices hint to be. For the purpose of this assignment I subtract 1 to *VR* so that the Null becomes *VR* is equal to 0. Subsequently I show how the VR of each year is computed for each stock using monthly and weekly returns:

$$VR = abs\left[\frac{Var_{i,t,m}}{4 \times Var_{i,t,w}} - 1\right]$$

In which $Var_{i,t,m}$ is the variance of monthly stock returns for stock i and year t, and $Var_{i,t,w}$ is the variance of weekly stock returns for stock i and year t. The reasoning

behind this measure is, if stock returns behave randomly, then they are independent and identically distributed (iid). If returns are indeed iid, then it is proven by Lo and MacKinlay (1988) that: $Var(r_{t,t+k}) = Var(r_t + r_{t+1} + \cdots + r_{t+k}) = kVar(r_{t+1})$, in which $Var(r_{t,t+k})$ is the variance of returns at the higher frequency and $kVar(r_{t+1})$ is the variance of returns at lower frequency multiplied by the number of shorter time periods there are in the longer time period. This measure has the advantage of providing an absolute measure which will clearly define whether efficiency is closer or further from a random walk.

At last, the effect of the lagged local market returns on the return of the stock today is introduced by the "delay" measures, as in the work of Hou and Moskowitz (2005), complementing the previously presented measures. They define the delay realized by the stock price to reflect information as the "severity of market frictions affecting a stock". Thus the delay can be interpreted as a measure of price inefficiency. These will be calculated by evaluating the importance the lagged local market returns have on today's stock returns, which is not taken into account in the other measures of price efficiency. Therefore, the following regressions are estimated for each stock I for each year:

$$r_{i,t} = \alpha_i + \beta_i \times r_{m,t} + \gamma_i r_{M,t} + \varepsilon_{i,t}$$

$$r_{i,t} = \alpha_i + \beta_i \times r_{m,t} + \gamma_i r_{M,t} + \sum_{n=1}^{4} \delta_{i,(-n)} \times r_{m,(t-n)} + \varepsilon_{i,t}$$

Where $r_{i,t}$ is the weekly stock return, $r_{m,t}$ is the local market weekly return and $r_{M,t}$ stands for the world market weekly return, all in week t. Thus, the lagged local market weekly returns from the previous week (t-1) until the fourth previous week return (t-4) are represented by: $r_{m,(t-n)}$. Obtained the R-squared of both regressions (R_a^2 is defined

as the R2 of the first equation and R_b^2 of the second equation), the computation of the first delay measure ($D1$) is:

$$D1_{i,t} = 1 - \frac{R_a^2}{R_b^2}$$

$D1$ captures the relative explanatory value that the four previous market returns have on the stock return. The higher the importance of that value, the higher delay will be, for it means the stock takes more time to assimilate the market information and reflect it on the price. Finally, the second delay measure is calculated using the estimated coefficients associated with both the contemporaneous and the lagged weekly market returns as:

$$D2_{i,t} = \frac{\sum_{n=1}^{4} n \times \left| \delta_{i,(-n)} \right|}{\left| \beta_i \right| + \sum_{n=1}^{4} \left| \delta_{i,(-n)} \right|}$$

$D2$ is a complementary delay measure, which already accounts for the magnitude of the impact the lagged returns have. The more pronounced the weight of coefficients δ relatively to their sum with β, the higher the delay is.

Variables influencing Price Efficiency – Independent Variables

Having covered measures of price efficiency, the dependent variable of the regression analysis, the next step is to find which variables influence this variable.

Hou and Moskowitz (2005) find that "most delayed stocks are small, volatile, less visible, and neglected by many market participants." Indeed, size, liquidity, volatility, and visibility are variables consistent with the literature relating frictions and price efficiency. Arbel, Carvell and Strebel (1983) state that small neglected firms enjoy a return premium associated to pricing inefficiencies eventually caused by lack of available information and analyst coverage. Also, according to Verrecchia (1979), market efficiency is positively influenced by the number of active traders, "as *if* traders all knew the 'true' distribution of returns on the security". Furthermore, bringing up

Merton's (1970) investor recognition hypothesis, stocks become more liquid after enjoying of additional exposure to investors. Bearing in mind that investors would only be active towards a stock on which they possess information (Merton (1970)), liquidity is strongly related to the previously mentioned investor recognition related measures.

Moreover, the volatility of the stock returns also constitutes one of the input variables in the main regression, to separate the impact of periods of high uncertainty and turbulence on price efficiency. It is verified that periods of high volatility would be thought as linked to lower price efficiency due to possible irrational and extreme investments. Therefore, not only it is reasonable to deduce that volatility should have an impact in the price efficiency, but also, as mentioned in the literature review section of this study, Bernd Schlusche's paper (2009) supports this argument.

All in all, the variables included in this analysis are: The log of the yearly market value (*Mkt*) computed as the total number of outstanding shares per year multiplied by the yearly average of the weekly closing price, in order to control for size; the weekly average bid-ask spread divided by closing price (*BA*) for each year, the log of the number of trades per year (*NT*) and the share turnover (*TS*) which is calculated as the number of shares traded in a year divided by the number of shares outstanding, as measures of liquidity; the weekly average standard deviation of returns per year (*Vol*) and three dummies (*Ind, ETF*, and *IndETF*), which are subsequently described.

The first dummy variable, *Ind,* differentiates the events of being in the FTSE 100 index (1) or not (0). It is of high importance to isolate this "index effect" since the fact that a stock is part of a widely known and the highest capitalized stocks index of the UK may have influence on the speed the stock assimilates new information due to increased visibility and recognition. This index membership dummy can be seen as a proxy for the variable category of institutional ownership mentioned and used in Hou and Moskowitz

(2005) work, as being an indicator of investor recognition. The constitution of the index is checked for the beginning of each year and name changes are taken into account in order to create this dummy variable. Since 2001, data on constituents' composition is retrieved from Bloomberg, although before that date, it is taken from the archived documents provided by FTSE. Finally, with the purpose of reaching the main variable of the regression, there is the need to include a second dummy variable, *ETF*, assumes the value 1 for all observations after 2000, when the ETF is already active, and 0 otherwise. This variable provides the differential of the ETF creation alone, regardless of the stocks in question. Then the last variable, *IndETF*, responds to the main purpose of the assignment, which is to evaluate the effect of the ETF creation on the index underlying stocks. This is the intersection variable between both previous dummies, assuming 1 in the situation where the ETF is already active and if the stock is part of the FTSE 100 index, and is 0 if one of the conditions is false.

Main Regression

Four regressions are estimated for each one of the four measures of price efficiency as a dependent variable: *Cross-Correlation, VR, D1* and *D2*. Being the independent variables Log Market Cap (*Mkt*), Bid Ask spread relative to Price (*BA*), Log Number of Trades (*NT*), Shares Turnover (*TS*), Volatility (*Vol*), and the dummies Index (*Ind*), ETF (*ETF*) and the Index with ETF intersection (*IndETF*). The main regression is described as follows:

$$PEff_{i,t} = \alpha + \beta Mkt_{i,t} + \gamma BA_{i,t} + \delta NT_{i,t} + \tau TS_{i,t} + \theta Vol_{i,t} + \vartheta Ind_{i,t} + \omega ETF_{i,t} + \mu IndETF_{i,t}$$

In which $PEff$ assumes one of the price efficiency measures, at a yearly frequency for each stock i and year t. There is the need to consider two dimensions in the analysis, as the purpose is to see what influences a change through time (time series) but also across different stocks (cross section), meaning that there is the presence of a repeated cross sectional time series - panel data. Also, there is the possibility to control for individual

unobserved heterogeneity, potential omitted variables which are specific to one stock. The approach followed in this study as to get the most accurate results is the Fixed Effects model (FE). The objective is to get unbiased estimators in which stock specific characteristics that are fixed in time are separated from the estimation. The point is, if these static variables which are intrinsic to one stock alone are considered, they could be influencing the estimators, so the issue needs to be addressed and this effect is removed. In practice, there are two error terms, the idiosyncratic error and the stock specific error (related to the invariant characteristics), being the latter fixed over time. This is done due to the possible correlation between the stock specific error and the estimators. At this stage the model would be:

$$PEff_{i,t} = \beta Mkt_{i,t} + \gamma BA_{i,t} + \delta NT_{i,t} + \tau TS_{i,t} + \theta Vol_{i,t} + \vartheta Ind_{i,t} + \omega ETF_{i,t} + \mu IndETF_{i,t} + \varphi_i + \varepsilon_{i,t}$$

By averaging the equation above and subtracting the averaged equation to the previous, one obtains the fixed effects regression model:

$$PEff_{i,t}^* = \beta Mkt_{i,t}^* + \gamma BA_{i,t}^* + \delta NT_{i,t}^* + \tau TS_{i,t}^* + \theta Vol_{i,t}^* + \vartheta Ind_{i,t}^* + \omega ETF_{i,t}^* + \mu IndETF_{i,t}^* + \varepsilon_{i,t}^*$$

Where $x_{i,t}^*$ (x standing for the variables in the previous regression model), stands for: $(x_{i,t} - \bar{x}_i)$. Notice that the stock specific time invariant error term has disappeared because it is constant over time, leaving the correlation between errors solved, as the model now provides only the "net effect" of the estimators. This final equation grants the econometric model used to perform the main regression analysis.

Data Description

The first ETF created to track the FTSE100 index is the iShares FTSE100 launched in April 2000. It is originally provided by Barclays Global Investors but it is Blackrock who supplies them today after a deal realized between both companies. The FTSE100 index comprises the 100 largest capitalization companies in the UK. This ETF has the point of granting the investor with the same return of the FTSE100 index, and in order to do so, it holds the same composition of securities as the tracking index, such that it is

a physical ETF. There is the need to include in this analysis a sample of stocks not belonging to the index, so I gather data of the stocks of the FTSE250 index today, in order to create a control sample. This index represents the most capitalized stocks in the UK right after the FTSE100 stocks. Evidently, some of the stocks in the FTSE100 have previously been part of the FTSE250 instead, and the inverse also applies, which is controlled by the dummy variable *Ind* . Both stock groups represent highly capitalized and liquid companies, with the difference that the FTSE250 stocks are naturally smaller. This has to be taken into account in the first general analysis of the data, but it will be mostly controlled for in the main regression analysis with the introduction of the size variable *Mkt*.

Finally, this leaves this analysis with a total of 355 stocks, from 1991 until 2011.

I use weekly data of stock returns to compute yearly measures of price efficiency. According to Hou and Moskowitz (2005), using monthly returns to compute price efficiency measures is not the best procedure, because most stocks take less than a month to respond to information. On the other hand, using daily data would bring additional estimation error due to bid ask bounce and non-synchronous trading. In order to match the independent variables according to the frequency of the dependent variables but to include as more information as possible, the yearly averages are computed from weekly data.

Descriptive Statistics

Before proceeding to the more thorough analysis provided by the main regressions estimations, in an attempt to understand the most outstanding properties of the data used in this assignment, the following lines provide a brief analysis of some statistical properties of the variables as well as of the possible impact of the first ETF creation on the price efficiency. Table1 provides some descriptive statistics computed for the

sample of stocks which underlie the ETF (Sample1), for the control sample which are the stocks which at the observation time are not part of the FTSE100 index ETF (Sample2), and for the entire sample which includes the previous two. For each of the three groups, one sample includes the whole time frame in analysis – 1991-2011, another sample goes from 1991 until 1999, and the last from 2000, when the ETF is created, until 2011, making a total of 9 samples.

Starting with some remarks regarding the independent variables, *Mkt* is on average lower for Sample2 than for Sample1, 6.079 compared to 8.840, corresponding to a market value of £436 million vs. £6,903 million. This is already expected since the main criterion for the inclusion of a stock in the FTSE100 is exactly its market capitalization, so it is natural that the control sample is smaller in terms of size. The previous is also reflected on the range between the minimum and maximum *Mkt* values, -0.871 to 10.481 vs 5.447 to 11.990, in which Sample2 reaches negative values of *Mkt* and Sample1 does not. The market value is on average higher after the ETF creation for both groups of stocks and for the entire sample, being that the smaller stocks enjoy from a higher increase relatively to the other group, of 13.03% against 4.36%. This shows that the mid cap index shows higher growth relative to the stocks which are always considered the largest in the country.

Descriptive Statistics		Entire Sample				Sample1				Sample2			
		Mean	St Dev	Max	Min	Mean	St Dev	Max	Min	Mean	St Dev	Max	Min
1991-2011 (All time range)	Mkt	6,751	1,650	11,990	-0,871	8,840	1,062	11,990	5,447	6,079	1,175	10,481	-0,871
	BA	0,013	0,018	0,446	-0,051	0,004	0,004	0,073	0,000	0,016	0,020	0,446	-0,051
	NT	10,130	2,221	15,217	0,000	12,069	1,542	15,217	6,353	9,403	1,988	14,981	0,000
	TS	1,007	1,031	29,077	0,000	1,256	0,853	8,424	0,004	0,914	1,076	29,077	0,000
	Vol	0,045	0,025	0,273	0,000	0,044	0,023	0,273	0,014	0,045	0,025	0,263	0,000
	Corr	0,013	0,275	2,389	-16,256	-0,044	0,138	0,371	-0,462	0,031	0,304	2,389	-16,256
	VR	0,505	1,174	35,822	0,000	0,328	0,257	3,327	0,000	0,565	1,343	35,822	0,000
	D1	0,311	0,263	0,999	0,000	0,202	0,186	0,966	0,000	0,346	0,274	0,999	0,002
	D2	1,261	0,585	4,000	0,000	1,096	0,519	3,234	0,055	1,313	0,595	4,000	0,000
1991-1999 (Before)	Mkt	6,262	1,760	11,574	-0,871	8,582	0,938	11,574	5,447	5,607	1,332	9,288	-0,871

	BA	0,017	0,018	0,246	0,000	0,007	0,003	0,028	0,002	0,020	0,019	0,246	0,000
	NT	8,563	1,675	13,158	0,000	10,311	0,974	13,158	6,353	7,852	1,345	11,422	0,000
	TS	0,627	0,544	7,674	0,000	0,710	0,412	2,736	0,004	0,594	0,585	7,674	0,000
	Vol	0,039	0,021	0,263	0,000	0,039	0,013	0,106	0,018	0,039	0,023	0,263	0,000
	Corr	0,046	0,408	1,631	-16,256	-0,025	0,128	0,294	-0,408	0,066	0,455	1,631	-16,256
	VR	0,583	1,340	35,822	0,000	0,321	0,237	1,242	0,000	0,663	1,515	35,822	0,000
	D1	0,360	0,272	0,999	0,000	0,216	0,177	0,966	0,000	0,400	0,281	0,999	0,002
	D2	1,348	0,581	4,000	0,000	1,131	0,477	3,234	0,055	1,408	0,593	4,000	0,000
2000-2011 (After ETF)	Mkt	7,007	1,529	11,990	2,181	8,956	1,095	11,990	5,971	6,338	0,988	10,481	2,181
	BA	0,011	0,018	0,446	-0,051	0,003	0,004	0,073	0,000	0,013	0,020	0,446	-0,051
	NT	10,775	2,093	15,217	2,485	12,855	1,007	15,217	9,645	10,021	1,862	14,981	2,485
	TS	1,160	1,136	29,077	0,000	1,486	0,886	8,424	0,079	1,039	1,193	29,077	0,000
	Vol	0,048	0,026	0,273	0,002	0,046	0,027	0,273	0,014	0,048	0,026	0,263	0,002
	Corr	-0,005	0,164	2,389	-2,346	-0,053	0,141	0,371	-0,462	0,012	0,168	2,389	-2,346
	VR	0,466	1,078	35,232	0,001	0,330	0,265	3,327	0,004	0,513	1,241	35,232	0,001
	D1	0,286	0,254	0,997	0,001	0,196	0,189	0,947	0,001	0,317	0,266	0,997	0,002
	D2	1,216	0,582	4,000	0,081	1,080	0,537	3,163	0,083	1,261	0,589	4,000	0,081

Table 1 – Descriptive Statistics including average, standard deviation, maximum and minimum values for all variables for the 6 different samples described above.

The average bid-ask spread (relative to price) (*BA*) is larger for the smaller stocks, 1.6% against 0.4%, and declines for the three samples after the introduction of the ETF. A larger spread translates into lower liquidity, so that this result is again proof of the former propositions that not only small stocks are less liquid, but are also in accordance to the hypothesis that after the ETF, liquidity increases due to the arbitrage hypothesis. The exact same indications come from both the number of trades (*NT*) and the share turnover (*TS*). *NT* is higher for the larger firms and it increases after 2000 evenly for all samples. In the case of the third measure of liquidity, *TS* is on average lower for the control group, which is around 91.4% compared to 125.6% for Sample1. Also, it increases after the creation of ETF for all samples. Even if the three measures of liquidity are higher after 2000, after ETF creation, it is reasonable to suppose the increased liquidity does not only originate from ETF creation. The development of the financial system technologies and overall higher spread of information of every kind around the globe make it cheaper and easier to participate in the stock market. As

previously said, investment decisions are only realized if the agent holds information about the security on which it is trading. This information is most probably faster spread and easier to obtain if they have access to the necessary tools of both financial analysis and of information transmission.

The average return volatilities of all samples are very similar to each other, they all increase after 2000 and also tend to get more disperse.

Finally, conclusions are very homogeneous among price efficiency measures, in which all indicate price efficiency is lower for the control group of stocks and that it increases after 2000 for all samples. *D1* translates the weight of the R2 of the regression with 4 lagged weekly market returns relative to the R2 of the regression without lags. The average *D1* of Sample1 decreases from 21.6% to 19.6% and from 40% to 31.7% for the control sample. *D2* yields very similar results, with a decrease of 4.46% in its value for Sample1 and of 10.43% for Sample2. The speed of convergence to price is lower for the sample with smaller stocks, where *D1* is 34.6% vs. 20.2% for Sample1, and *D2* is 1.313 vs. 1.083. This matches the previous indication that the most delayed firms are the smallest.

Moreover, the *VR* also respects the supposition that efficiency increases after the ETF inception, and also that it is lower for smaller stocks. In fact, the average absolute *VR* is higher for Sample2, with a mean of 0.565 vs. 0.328 of Sample1 for the whole time frame, which indicates that the smallest stocks are further from a random walk, implying they are less efficient.

The *Cross-Correlation* tends to be negative for the blue chip companies (Sample1) and positive for the smallest. This denotes that the group of solid big stocks which are present in the main index seem to move against the rest of the market, while the others are more correlated with the lagged market returns. It is on average even more negative

after 2000 for Sample1, decreases from 6.6% to 1.2% for Sample2, and goes from being positive to negative for the entire sample of stocks, due to Sample1.

For the independent variables, the differences between means are always statistically significant using a t-test, which supports the previous results. However, the following table shows the significance of the differences of means concerning the price efficiency measures using a t-test, because there are a few insignificant differences to notice. Namely, for the stocks underlying the ETF (Sample1), the difference in the mean of *VR* between the observations before and after 2000 is not statistically significant at any level (1%-10%). For the same group of stocks and the same comparison, but regarding *D1* and *D2*, their differences are only significant at a 10% significance level. Hence, it is rather premature to draw conclusions solely from the descriptive statistics.

It is also interesting to study what variables are more correlated to price efficiency, by computing and analyzing the correlations between the independent and dependent variables, which are later used for the deeper and more thorough multivariable regression analysis. The signs of the correlations are the expected, in which the efficiency increases with market value and liquidity. *Vol* seems to be slightly correlated with price efficiency, presenting small values of correlations. When measured by *cross-correlation* or *VR*, volatility seems to increase efficiency, although when measured by *D1* or *D2* it seems to decrease it, being the two latter measures more complete by construction.

It is clear that *Mkt*, *BA* and *NT* are the variables which present the most significant correlations with the price efficiency variables. Thus, these liquidity measures seem to be very important in the following regression analysis, besides being consistent with the reasons why these variables are chosen to be included in the first place, as well as size.

The delay measures are the variables which most seem to be grasped by the independent variables, still the others present significant values of correlation.

Correlations	Mkt	BA	NT	TS	Vol
Corr	-0.1625	0.1219	-0.3858	-0.0904	-0.025
VR	-0.1003	0.0059	-0.1035	-0.0586	-0.0958
D1	-0.3787	0.3193	-0.3996	-0.0779	0.0305
D2	-0.2649	0.2443	-0.2757	-0.0178	0.0266

Table 3: Correlations between independent variables and price efficiency measures (dependent variables).

Results

The main regressions are run in Stata using the Fixed Effects technique with panel data. Three separate datasets are used, representing different time frames around the creation of the first ETF (iShares) on the FTSE100 index: 1991-2011, 1995-2005 and 1999-2001. The independent variables are checked for multicollinearity being the results negligible with all variance inflation factors (VIF) below 3.4. Also, besides the variables included in the regression specification previously presented in this text, a control for year is also introduced in order to control for any specific effects of the time evolution or any specific events in a given year. The results are presented in table 4 in the end, and the following comments are relative to the variables which show statistical significance.

The increase in *Mkt* seems to be associated with an increase of the speed at which prices incorporate new information of the market, measured by both delays in the three time ranges. Being the coefficients associated with *Mkt* -0.038(*D1*) and -0.096(*D2*) for the wider time period, -0.049(*D1*) and -0.107(*D2*) for the medium and -0.121(*D1*) and -0.206(*D2*) for the shortest time frame. This shows that when market value increases 1%, then delays decrease by the coefficient divided by 100 which results in a very small value for all above. As the delay measures the dependence of the stock return on the total market return, this results point to the fact that the bigger the company in terms of

value and shares, the lower the delay, and so the higher the price efficiency. *Mkt* also seems to have significant negative impact on the *cross-correlation*, for first and last time periods, strengthening the previous results. Also, the absolute *VR* declines with *Mkt* with a coefficient of -0.072 for 1995-2005, thus a stock sees its price reach a random walk closer as its market value increases. All of the above support the initial suppositions of this assignment that small stocks are less efficient and so size is a variable of significant impact in the regression analysis.

The greater the gap between the bid and ask prices relatively to the closing price, meaning lower liquidity, the higher should be the delay as it is indeed the fact, although only for *D1* in the longer time sample. *NT* shows to impact the delays negatively, thus meaning that more liquidity yields higher efficiency, with coefficients of -0.045 (*D1*) and -0.044(*D2*) in the largest sample and -0.019 for the medium, as it is also taken from the significance of the *cross-correlation* variable, which has a negative coefficient too, -0.019. Regarding *TS*, it does not produce statistically significant impact on any of the dependant variables. Thus the results for the liquidity measures *BA* and *NT* are consistent with each other.

Concerning *Vol*, the results are indefinite, showing that *D2* increases with volatility, thus resulting in less efficiency measured by a 1.857 impact. However, it also shows that returns move more randomly, consequently there is more price efficiency, by looking at the coefficients associated with *VR* for 1991-2001(-2.057) and 1995-2005 (-3.106).

The dummy variable Ind, serving as a measure of investor recognition, proves to explain the cross-correlation (-0.042) and *D1*(-0.035) for the longer time sample. The fact that a stock is part of the FTSE100 index appears to have a negative impact on both

measures, meaning a positive impact on efficiency. This is again consistent with the initial idea that higher investor recognition should translate into higher price efficiency.

The dummy variable *ETF*, which provides the differential effect of the introduction of the ETF regardless of whether the stock is in the index or not, is mostly positively related to our measures, hence negatively related to price efficiency. This happens for *D1*(0.309) and *D2*(0.474) in the first sample and for cross-correlation in the last two samples (0.156 and 0.100). However, inconsistently with the other coefficients, the etf creation in the first sample provides a negative impact on *cross-correlation* of -0.0866, although it is to mention that it is the smallest impact concerning this independent variable. It means that the creation of an ETF on FTSE100 has a negative impact on all stocks' efficiency, which might be due to the shift in liquidity of investors from the individual security market to the ETF market. A possible explanation is that the new ETF product might attract investors who invest in the ETF underlying stocks, but also from other stocks in a similar market, which would be the mid cap section of FTSE.

The variable representing the main question of this study, the dummy *IndETF* appears to significantly explain *D1*(0.073) and *D2*(0.087) for the longer time range and *D1*(0.0969) for 1999-2001. In fact, according to the result, stocks which are part of the index see their price efficiency decrease with the inception of the ETF. This variable represents the differential of being part of the index at a time when the ETF already exists. This is consistent with the *VR* associated coefficient of 0.209 which shows that these stocks suffer an increase in the distance from a random walk. Also, after 2000 a constituent stock of the index exposes a higher dependence with the rest of the market in the previous week, proven by an increase in *Cross-Correlation* by 0.0364 considering the 1991-2011 sample. So it seems that the creation of the FTSE100 index-linked ETF results in a deterioration of the price efficiency in the underlying stock market.

For the shortest time frame, less variables seem to be statistically significant. One of the reasons which may stand behind these results is that this range of observations hits exactly the period of the Dot-com bubble collapse in 2000-2001. The extremely unstable and perhaps more than usually irrational of a bubble burst has certainly effects in the results. As this time period choice is exactly meant to analyze the closest period possible to the ETF creation and yearly data is being used, it leaves no option to isolate this effect.

Concluding Remarks

In a world with asymmetric information, there is the distinction between the informed investor who beholds private information and trades in his favor, against the liquidity trader who is less informed and therefore faces a disadvantaged trading scenario. If a basket of securities is introduced, in this case the ETF, then the adverse selection, derived from the private information of each underlying stock, is reduced because the specific security risk is diversified away. This naturally creates an investor allocation restructuring. Firstly, the liquidity trader has a big incentive to invest in the ETF because it is a cheap way of getting rid of the adverse selection risk. Even though the basket may also be interesting for the informed investor's allocations because this market provides of course other type of cost advantages, my results evidence that their proportion must be smaller. Indeed, if very well informed investors optimize their portfolios using individual securities, investing in a composite security would not allow mixed positions (long and short) on different securities as a response to liquidity traders, thus it would be a disadvantage for them to transfer to the index market (Gorton and Pennacchi (1993)). This means that the singular stock market is left with a predominance of informed investors (Jegadeesh and Subrahmanyam (1993)). Given that investors are risk averse, this market sees a decrease in market liquidity because the

higher number of more informed investors creates a higher expected profit even though more informed investors enter the market and there is higher competition among them, proven by Subrahmanyan (1991). This is because with the exit of liquidity traders, the adverse selection cost in this market of the underlying security is even higher so that prices take longer to respond to new information. In the figure below (graph 3) it is clear that during the three months after ETF inception, the bid-ask spread of the ETF underlying securities increase. Furthermore, corroborating this theory, Gammill and Perold (1989) "envision macro markets gaining in liquidity at the expense of deteriorating liquidity in the micro market", in which they claim the creation of a basket security leans towards the formation of market-factor rather than firm-specific information.

Still, Subrahmanyam (1991) finds that price efficiency tends to increase with the precision of information held by investors. With the introduction of the first ETF, when the properties and consequences of the brand new investment product is still not thoroughly analyzed and observed, the liquidity shift may lie on a point where the proportion of informed investors in the underlying security market is higher than before, so that price efficiency declines, yet not as large so that the weight of extremely informed investors is enough to boost price efficiency again. Subsequently, I graphically show (graph 4) that price efficiency tends to increase with the introduction of a more considerable number of ETFs on the FTSE100, which means an even higher devotion of traders to the ETF market, leaving the proportion of precisely informed investors to increase and surpass the "turning point" at which price efficiency starts to increase. Obviously it is precipitate to argue the accurateness of the previous statement, and as such, a profound analysis on this matter would be of value as a continuation of this work.
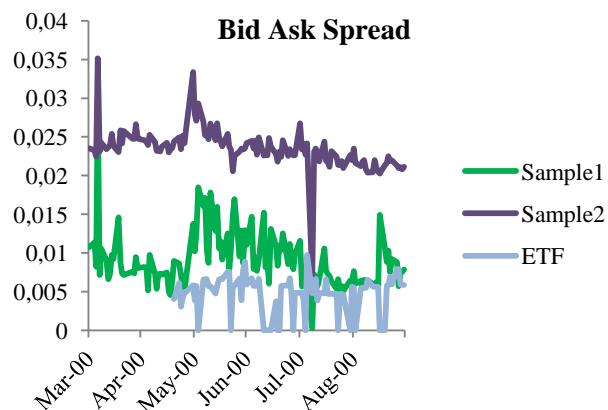
Graph 3- Bid-Ask Spread divided by price for Sample1, Sample2, and for the iShares FTSE100 ETF from March 2000 until September 2000, Source: Datastream
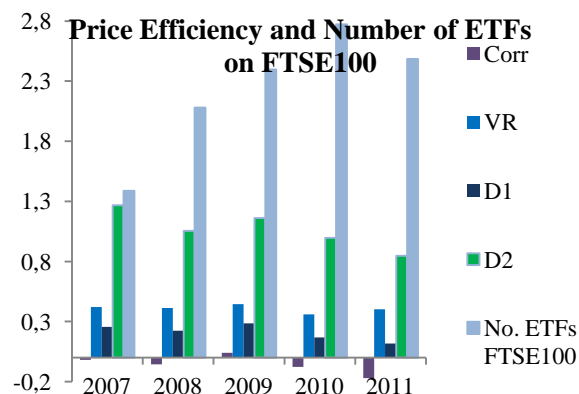
Graph 4- Price Efficiency measures for the entire sample and Number of ETFs created on FTSE100 index from 2007 until 2011.Source: Datastream

The hypothesis that the index security creates the opportunity to arbitrage away any price inefficiencies among markets, despite being reasonable, it does not prove to work at least in this case. A plausible explanation is that, given that there is an increase in liquidity in both markets, the increase in liquidity in the index linked market should be higher. Also, this increase should be due to the migration of less informed investors, which do not bring fundamental value (Lei Yu (2003)), who do not provide the appropriate price correction as in the presence of precisely informed investors. The arbitraging effect created by the introduction of the ETF would probably be more pronounced for smaller, less visible and more delayed stocks.

It is already certain that the ETF market faces high adherence of investors and growing popularity since its early years. Nonetheless, the interest lies in the origin of this liquidity, and as such, how liquidity in the underlying securities' market alters. As such, an analysis of the sources of liquidity, thus a distinction between arbitrage and adverse selection effect would be another step into a thorough knowledge on this issue.

References

Arbel, A., Carvell, S., Strebel, P. 1983. "Giraffes, Institutions, and Neglected Firms."
Financial Analysts Journal, 39

Bris, Arturo, Goetzmann, Willian N. and Zhu, Ning. 2007. "Efficiency and the Bear:
Short Sales and Markets Around the World." *The Journal of Finance*, Volume 62 No 3.

Brüderl, Josef. 2005. "Panel Data Analysis." *Tutorial*, University of Mannheim,
Germany.

Chelley-Steeley, P. and Park, Keebong. 2010. "The adverse selection component of
exchange traded funds." *International Review of Financial Analysis*, 19, 65-76.

Chen, Guang and Strother, T. Shawn. 2008. "On the Contribution of Index Traded
Funds to Price Discovery in the Presence of Price Limits without Short Selling." *Master
Thesis*, University of Otago, Dunedin, New Zealand.

Costandinides, Christos and Arnold, Daniel J. 2011. "In the ETF labyrinth, where
does the thread begin?" *ETF Research – Industry Perspective*, Deutsche Bank

Fama, Eugene F. 1979. "Efficient Capital Markets: A Review of Theory and Empirical
Work." *Journal of Finance*, Volume 25 Issue 2

Gorton, Gary B. and Pennacchi, George G. 1993. "Security Baskets and Index-
Linked Securities." *The Journal of Business*, Vol.66, No1, 1-27.

Hasbrouck, Joel. 2003. "Intraday Price Formation in U.S. Equity Index Markets."
Journal of Finance, Volume 58 No.6

Hegde, S. P. and McDermott, J. B. 2003. "The market liquidity of DIAMONDS, Q's,
and their underlying stocks." *Journal of Banking & Finance*, 28, 1043-1067.

Hou, Kewei and Moskowitz, Tobias J. 2005. "Market Frictions, Price Delay, and the
Cross-Section of Expected Returns." *The Review of Financial Studies*, Volume 18 No 3

James F. Gammill, Jr., and Andre F. Perold. 1989. "The changing character of stock market liquidity." *The Journal of Portfolio Management*, 15, 3, 13-18.

Jegadeesh, Narasimhan and Subrahmanyam, Avanidhar. 1993. "Liquidity Effects of the Introduction of the S&P 500 Index Futures Contract on the Underlying Stocks." *The Journal of Business*, 66, 2, 171-197

Lo, Andrew W., MacKinlay, A. Craig. 1988. "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test." *The Review of Financial Studies*, Volume 1 No 1.

Merton, Robert C. 1987. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *The Journal of Finance*, Volume 42 No.3

Saffi, Pedro A. C. and Sigurdsson, Kari. 2008. "Price Efficiency and Short Selling." *Research Paper*, IESE Business School.

Schlusche, Bernd. 2009. "Price Formation in Spot and Futures Markets: Exchange Traded Funds vs. Index Futures." *Journal of Derivatives*, Volume 17 No 2

Subrahmanyam, Avanidhar. 1991. "Risk Aversion, Market Liquidity, and Price Efficiency." *The Review of Financial Studies*, Vol 4, No3.

Tse, Yiuman, Bandyopadhyay, Paramita and Shen, Yang-Pin. 2006. "Intraday Price Discovery in the DJIA Index Markets." *Journal of Business Finance & Accounting*, 33(9) & (10), 1572-1585

Verrecchia, Robert E. 1979. "On the theory of market information efficiency." *Journal of Accounting and Economics*, 1, 77-90.

Yu, Lei. 2003. "Basket Securities, Price Formation, and Informational Efficiency." *Job Market Paper,* New York University.

Main Regressions	1991-2011				1995-2005				1999-2001			
	ρ^{cross}	VR	D1	D1	ρ^{cross}	VR	D1	D1	ρ^{cross}	VR	D1	D1
Mkt	-0.00809*	-0.0373	-0.038***	-0.096***	-0.0108	-0.0724*	-0.0490***	-0.107***	-0.068***	-0.045	-0.12***	-0.206*
	(-1.721)	(-1.169)	(-4.197)	(-4.981)	(-1.601)	(-1.742)	(-3.212)	(-3.115)	(-3.047)	(-0.62)	(-2.650)	(-1.95)
BA	-0.0486	-1.293	0.353*	0.757	0.162	-0.223	0.318	0.828	-0.186	1.222	-1.356	-2.474
	(-0.310)	(-0.971)	(1.809)	(1.429)	(0.920)	(-0.161)	(1.335)	(1.399)	(-0.211)	(0.484)	(-0.746)	(-0.70)
NT	-0.019***	0.000594	-0.045***	-0.044***	-0.0146***	0.0495	-0.0186**	-0.0147	0.00599	-0.0144	-0.0253	-0.0948
	(-6.085)	(0.0161)	(-7.131)	(-3.513)	(-3.001)	(1.371)	(-2.094)	(-0.803)	(0.461)	(-0.26)	(-0.812)	(-1.48)
TS	0.000907	0.0109	0.00774	-0.00140	-0.00201	0.0198	0.0105	0.00703	-0.0203	0.0803	-0.0322	-0.115
	(0.353)	(1.097)	(1.438)	(-0.102)	(-0.480)	(1.525)	(1.246)	(0.291)	(-1.171)	(0.968)	(-0.974)	(-1.46)
Vol	-0.0609	-2.057**	0.122	0.200	-0.305	-3.106*	0.261	1.857**	-0.0816	-3.302	0.653	2.635
	(-0.490)	(-2.219)	(0.487)	(0.369)	(-1.437)	(-1.872)	(0.649)	(2.044)	(-0.172)	(-1.44)	(0.743)	(1.375)
Ind	-0.042***	-0.112	-0.0353*	-0.0428	-0.0140	-0.0188	-0.0248	-0.0167	0.0295	-0.0173	0.00197	0.0800
	(-3.937)	(-1.600)	(-1.900)	(-0.896)	(-0.866)	(-0.301)	(-0.782)	(-0.206)	(0.741)	(-0.12)	(0.0252)	(0.373)
ETF	-0.087***	-0.151	0.309***	0.474***	0.156***	-0.186	-0.00121	0.101	0.100***	-0.176	-0.0423	0.00310
	(-5.541)	(-0.627)	(11.91)	(7.537)	(8.471)	(-0.874)	(-0.0395)	(1.370)	(7.306)	(-0.94)	(-1.488)	(0.053)
IndETF	0.0364***	0.209***	0.073***	0.0871**	-0.00141	0.0632	0.0301	-0.0180	0.00247	0.123	0.0969**	0.0272
	(3.783)	(2.697)	(4.687)	(1.994)	(-0.116)	(0.913)	(1.347)	(-0.292)	(0.112)	(0.883)	(2.380)	(0.264)
Constant	0.335***	0.88***	0.820***	2.172***	0.215***	0.78***	0.843***	2.106***	0.432***	1.24**	1.503***	3.63***
	(11.74)	(3.924)	(16.10)	(16.26)	(4.816)	(3.231)	(9.350)	(9.153)	(2.777)	(2.147)	(4.028)	(4.222)
Observations	4,699	4,536	4,680	4,680	2,358	2,247	2,347	2,347	622	590	619	619
R-squared	0.332	0.023	0.277	0.157	0.155	0.013	0.099	0.055	0.187	0.009	0.076	0.109
Number of id	338	337	337	337	275	266	274	274	224	213	223	223

Table 4 – Results of the Main Regressions using Stata. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1. The variables which are statistically significant are colored in blue. The independent variables – Mkt: log of the market value, BA: bid ask spread divided by price, NT: log of the number of trades, TS: share turnover = number of shares traded divided by the total number of shares outstanding in a year, Vol: volatility of returns, Ind: dummy variable assuming 1 if stock is part of the FTSE100 index and 0 otherwise, ETF: dummy variable assuming 1 if ETF is already created and 0 otherwise, IndETF: dummy variable assuming 1 if stock is part of FTSE100 index and if the ETF is already created and 0 if one of the conditions is not applicable; The dependent variables - ρ^{cross}: Cross-autocorrelation, VR: aboslute value of the variance ratio, D1: delay measure 1, D2: delay measure 2.