N O T A S

C O N Ó M I C A S 48

Imprensa da Universidade de Coimbra
Coimbra University Press

Faculdade de Economia da
Universidade de Coimbra

NOTAS

CONÓMICAS 48

Propriedade
Faculdade de Economia
Universidade de Coimbra
Av. Dias da Silva, 165
3004-512 Coimbra

Apoios
Esta revista é financiada por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito do projeto PEst-OE/EGE/UI0031/2014

Número avulso
Portugal 7,5 €
Outros países 15 €

Condições de assinatura bienal
Portugal: 20 €
Outros países: 40 €

ISSN 0872-4733

ISSN Digital 2183-203X

DOI https://doi.org/10.14195/2183-203X_48

Depósito Legal 64932/93

Sumário

Artigos

António Afonso
ISEG
Universidade de Lisboa, REM, UECE
aafonso@iseg.ulisboa.pt
Pedro Cardoso
ISEG
Universidade de Lisboa
pedro.mmartins.cardoso@gmail.com
DOI: https://doi.org/10.14195/2183-203X_48_1

Exchange-traded funds as an alternative investment option

Exchange-traded funds como uma opção alternativa de investimento

António Afonso
Pedro Cardoso

Received for publication: December 20, 2018
Revision accepted for publication: April 10, 2019

ABSTRACT
We conduct an analysis of Exchange-traded Funds (ETFs), Index and Equity mutual funds and their respective benchmark during the 2010-2015 period for the Portuguese fund industry. For the period 2010-2017, we test ETFs for price inefficiency (existence of deviations between prices and the Net Asset Value) and persistence. We find that the studied ETF does not always outperform index funds in replicating the variations of the PSI 20 index, despite exhibiting better tracking ability when facing downside deviations of the benchmark and a better capacity of smoothing tracking deviations. Regarding ETFs price efficiency and its persistence, the study reveals that the examined ETF is priced at a low average discount with evidence of deviations persistence of at least two days. The investment schemes with the highest ability to track the PSI 20 Index were PSI20 (ETF), BBVA PPA Índice PSI20, and the equity mutual fund BPI Portugal.
Keywords: Exchange-traded fund; mutual fund; performance evaluation; tracking error; price efficiency.

JEL Classification: G11; G12; G14

RESUMO
Foi conduzida uma análise a Exchange-traded Funds (ETFs), Fundos de Investimento mobiliário de Índice, Fundos de Investimento mobiliário de ações e respetivo índice de referência (benchmark) no período 2010-2015 para a indústria de fundos portuguesa. Para o período 2010-2017 foi testada a ineficiência de preço para os ETFs (existência de desvios entre os preços de negociação e o valor intrínseco da unidade de participação) e a sua persistência. Concluiu-se que o ETF analisado nem sempre supera os fundos de investimento mobiliários de índice na replicação das variações do Índice PSI 20, não obstante exibir uma melhor capacidade de replicação das variações negativas do benchmark e uma melhor capacidade de alisar os desvios da replicação. Em relação à eficiência de preços dos ETFs e à sua persistência, o estudo revela que os preços do ETF examinado apresentam um valor médio

inferior (embora baixo) face seu valor intrínseco, com evidência de persistência de desvios de pelo menos dois dias. Os organismos de investimento analisados com maior capacidade de acompanhar o Índice PSI 20 foram o PSI20 (ETF), o BBVA PPA Índice PSI20 e o fundo de investimento mobiliário de acções BPI Portugal.
Palavras-chave: Exchange-traded fund; fundos de investimento; replicação de benchmark; ineficiência de preço.

Acknowledgements: We thank an anonymous referee for very useful comments. The opinions expressed herein are those of the authors and do not necessarily reflect those of their employers. Any remaining errors are the authors' sole responsibility.

1. Introduction

Since its initial appearance in 1993, Exchange-Traded Funds (ETFs) have exhibited a steady asset growth as a result of their popularity worldwide and have become a relevant investment alternative for investors. With primary roots in the United States (U.S.), the overcome of the European market happened rapidly in the beginning of the 21st Century.

However, the Portuguese Investors' access to this type of funds traded in the local exchange (Euronext Lisbon) with the Portuguese Stock Index (PSI 20) as the underlying index (i.e. Benchmark) has only happened late in 2010, but, since then, the acceptance in the Portuguese market has been verified.[1] This paper evaluates ETFs as a comparative relevant investment option for Portuguese investors by conducting a descriptive statistics, regression and index tracking comparative examination of returns with other types of investment products normally considered substitutes [index mutual funds and mutual equity funds (sharing the same benchmark)]. It also aims to examine the price efficiency and deviation persistence of the ETF prices.

Our main result shows that ETF do not always outperform index funds in replicating the variations of the PSI 20 index, despite exhibiting better tracking ability when facing downside deviations of the benchmark and a better capacity of smoothing tracking deviations. Regarding ETF price efficiency and its persistence, we find that the evaluated ETF is priced at a low average discount with evidence of deviations persistence of at least two days. In addition, the lowest results of tracking error measures (i.e. the investment schemes with the highest ability to track the PSI 20 Index) were from PSI20 (ETF), BBVA PPA Índice PSI20 (Index Fund) and from the equity mutual fund BPI Portugal.

The remainder of the paper is organized as follows. Section 2 discusses ETFs origins, main characteristics and comparative advantages. Section 3 reviews the related literature. Data and methodologies will be described in Section 4. Section 5 presents the results of the empirical analysis. Section 6 is the conclusion.

2. Exchange-Traded Funds

Origins

Exchange-Traded Funds in the U.S. were first introduced in 1993 by the American Stock Exchange. This first ETF "Standard & Poor's Depositary Receipts" (SPRDS)[2] traded in the U.S. was developed as simple unit trust that invested in the 500 shares of the underlying index[3] (Standard & Poor's 500 Index) and was the consequence of the consistent progress in the financial industry aiming to reduce the costs and to increase the accessibility of this type

[1] In the end of August 2017, the total amount of Assets under Management (AuM) of the Comstage PSI20 (only non-leverage ETF traded in Euronext Lisbon with the PSI20 Index as Benchmark) is close to 68 million euros, a value similar to the average AuM in the previous month of each Undertakings for the Collective Investment of Transferable Securities (UCITS) in Portugal (74 million euros). Sources: Morningstar (2017) and CMVM (2017).

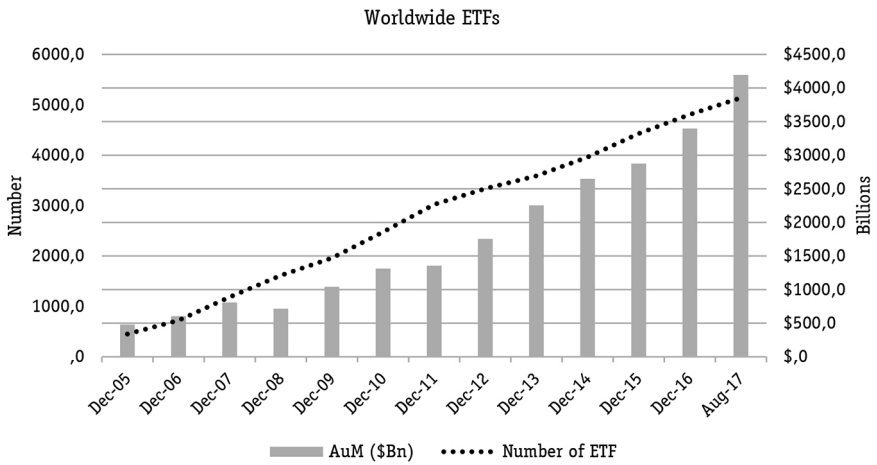
[2] Pronounced "spiders".

[3] The term underlying index and benchmark will be used to refer to the index that is tracked by the ETF.

of financial products to retail investors. The full history behind the precedents which lead to the creation of the first ETFs [which include Index Participation Shares (IPS) and Toronto Stock Exchange Index Participations (TIPs)] was examined in detail by Gastineau (2001).

The innovation was introduced in the Asian continent in 1999 and reached Europe in April 2000 with the launch of the EURO STOXX 50 and STOXX Europe 50 traded in the Frankfurt Stock Exchange. Since then, as showed in Figure 1, the ETF industry has grown worldwide, both in number and Assets under Management (AuM), with more indices as underlying.

Figure 1: Total amount of assets under management (AuM) and number of exchange-traded funds



Source: ETFGI (2017).

Main characteristics

An Exchange-Traded Fund is a form of collective investment scheme whose units or shares[4] are traded in an Exchange market. For the purpose of this investigation, the scope of ETF will be limited to the ones that aim to replicate specific indices as close as possible. It's comparability with mutual funds, in particular with index funds, is understandable since the main portfolio characteristics and fund features are present. Moreover, ETFs combine the attributes of mutual funds with the characteristics of common stock, making it possible to trade each share on an exchange market which leads to the intraday possibility of sell each position instead of having to wait, like in mutual funds, for the process of redemption from the fund (which occurs at the end of the day Net Asset Value (NAV) per share that

[4] For simplification, in the paper the term Units will be omitted and share will represent both realities in discussion.

is calculated with the close of the market prices). Since shares are traded in an exchange market, each ETF has two different prices, the value in which the security (fund's shares) is being traded and also the intrinsic value of the fund assets that results from the net asset value of the ETF divided by the total number of existing shares. As understandable, each deviation between both values leaves space for arbitrage opportunities. These differences have made ETFs a very relevant investment option for investors that demand short-term liquidity and buy in large lots while mutual funds remain a relevant option for an investor looking for high trading of small purchases or sales and for those who do not demand short-term liquidity (Poterba and Shoven, 2002).

ETFs can be divided into multiple subsets by their most relevant specific features. Firstly, they can be divided by the type of management associated, active or passive. An actively managed ETF is an investment fund in which the intervention of the management company is determinant for the portfolio capability of index tracking, while in passively managed ETFs (most common) the involvement of the management company is kept as low as possible which is one of the main reasons for the comparative low annual expense ratio. Secondly, by the type of exposition: a) Physical – in which the ETF holds the shares that are the constituents of the underlying index; or b) Synthetic, in which ETF replicates the underlying index performance through the use of derivative instruments (e.g. replication through the use of swaps or futures contracts). Another distinctive factor among ETFs is the degree of replication of the underlying index, meaning the degree of leverage of the underlying index performance (e.g. an ETF that has the aim of duplicate the effect of the underlying index has a replication degree of 2).

Comparative advantages

In addition to the main characteristic of ETFs that can be considered a comparative advantage (liquidity access), other key aspects are typically presented as advantages to investors. The first advantage to arise is the process associated with creation and redemption of ETF shares (known as in-kind creation/redemption). Specifically, in addition to the possibility of trading shares on an exchange market, some types of investors (known as Authorized Participants) have the possibility of create/redeem shares as in the traditional mutual funds (subscriptions and redemptions) making it possible to these shares to be resold in exchange markets for profits or kept in the investors' portfolio. This process of creation/redemptions is mainly motivated by the arbitrage opportunities mentioned before and by the market pressure on the shares. For example, if investors are buying the shares of an ETF from the market, it generates pressure on the Authorized Participants to create new shares to supply the market demand pressure. As the Authorized Participants, in the case of a non-synthetic replication ETF, have to buy shares of the constituents of the index for the creation process of the new ETF shares (in-kind creation), it's likely to raise the price of the index it tracks, ensuring market prices close to the intrinsic Net Asset Value (Petajisto, 2017; Shin and Soydemir, 2010; Xu and Yin, 2017). Additionally, the in-kind process (for redemption) enhances tax efficiency as it delays capital gains up to the end to pay for redemptions (Gastineau, 2001).

Other advantage that is normally pointed out to this type of financial instruments is the low total expense ratio (which includes the management fee) associated to the passive managed, but efficient, portfolio structure. Also, as mentioned by Gastineau (2001), the low expense ratio comes from the elimination of the transfer agent function (i.e. the elimination of shareholders accounting) at the fund level.

Furthermore, as stated in Rompotis (2011a), another comparative advantage of ETF is associated with the fact that shares can be purchased on margin, traded using limits and stop orders as well as short-sold.

Lastly, the possibility for small quantity transactions allows retail investors to participate in the market, in contrast with equivalent future products that are relatively large in notional size with expensive variation margin requirements for small investors (Kearney *et al.*, 2014).

3. Literature

Since the appearance of the first ETF, the relative performance of its underlying index (i.e. if the return of the ETF outperforms or underperforms the underlying index) has always been examined as the key factor for comparison purposes. However, the first studies on ETFs aimed to examine price efficiency (comparison of market prices vs. the intrinsic NAV per share of the ETF) which is an alternative way of looking the relative performance issue since if it is assumed that NAV perfectly replicates the index, only prices lead to inefficiency (existence of premium/outperformance and discount/underperformance). Using the SPRDS data Ackert and Tian (2000) concluded no economically significant mispricing in the S&P500 SPDRs market and Elton *et al.* (2002) observed an average discount of 1.8 basis points per year to its NAV and that almost all the differences (prices inefficiencies) disappeared within one day. This last investigation also concluded that the amount of income that is lost by the holding dividends received in cash was the main cause of the underperformance of the SPDR. Likewise, Poterba and Shoven (2002) corroborate the previous underperformance conclusions in their study of SPDR for 1993-2001. Additionally, Charteris (2013) conducted a price efficiency analysis for the South African ETFs and found that funds were reasonably efficiently priced (low premium and discounts) to mainly all ETF and justify this conclusion by the efficient execution of arbitrage. Additionally, respectively to the Dow Jones Istanbul 20 Fund and the Taiwan Top 50 Tracker Fund, Kayali (2007) and Lin *et al.* (2005) found that these ETFs were trading at a small discount (₺0.008, which is 0.11% of the average close price) and at a small premium [$0.018, which is 0.041% of the average close price (although not statistically significant)], respectively.

Regarding the comparison between index funds and ETFs, Kotosvestky (2003) found through a multi-period model that the differences between the returns of both types of investment schemes come mainly from transaction and management fees, taxation efficiency and qualitative difference (i.e. convenience and ability to buy on margin and sell short). Within a European geographical focus, Blitz *et al.* (2012) examined the relative performance of ETFs and European Index Funds to their benchmarks and found that both types of funds exhibit an underperformance between 50 to 150 basis points per annum, being the dividend withholding taxes on par with fund expenses the determinants for underperformance.

In complement, but still in the same comparable scope (ETFs and Index Funds), Agapova (2011) concluded that conventional index funds and ETFs are substitutes investment products, but not perfect ones, meaning that ETFs have not replaced the conventional index funds but have introduced a new alternative investment vehicle. This study was conducted through the analysis of ETFs and Index funds flows.

In addition to this type of investigation which aim exclusively to the ETF and Mutual Funds performance, Rompotis (2011a) conducted a cross-section examination of performance on Greek ETFs, Index Mutual Funds and Equity Mutual Funds concluding that classic mutual funds, despite having high expense ratios, performed better and are less volatile for the period under examination. In terms of the tracking error of ETFs, it was found that they were reasonably lower than the tracking error of the actively managed funds but greater than the tracking error of the index fund.

Regarding Risk-adjusted measures of ETF, the examination conducted by Rompotis (2011b) for the 2002-2007 period for 50 iShares ETFs found that the high majority of ETFs outperformed the S&P 500 annually and in aggregate values. This finding was obtained through the calculation of indicators like Sharpe and Sortino Ratios. Moreover, Wong and Shum (2010) found that ETFs perform differently when facing bearish and bullish markets from 1999 to 2007. In their 15 ETFs examination, it was concluded through the Sharpe ratio test that ETF provides relatively higher returns in a bullish market than in a bearish market. Pinheiro and Varela (2018) didn't find evidence that ETFs tracking the PSI20 outperform the market using the risk-return model and analysing Jensen's alpha for the period of December 2012 to June 2017.

In terms of market type comparative analysis of ETFs and the tracking activity, Blitz and Huij (2012) concluded that global emerging markets ETFs exhibit higher levels of tracking error than developed markets ETFs, which the authors relate to the cross-sectional dispersion in stock returns being structurally larger in emerging markets. Lastly, to summarize the different results among the relevant literature regarding ETFs outperformance/underperformance and price premium/discount, Figure 2 is presented.

Figure 2: Literature evidence of ETFs outperformance/underperformance and price premium/discount. The repeated references on both sides (e.g. outperformance and underperformance) are the result of both conclusions among different ETFs in that particular study

OUTPERFORMANCE	**UNDERPERFORMANCE**
BLITZ AND HUIJ, 2012 BUETOW AND HENDERSON, 2012 KEARNEY et al., 2014 ROMPOTIS, 2011a ROMPOTIS, 2011b	BLITZ AND HUIJ, 2012 BLITZ et al., 2012 BUETOW AND HENDERSON, 2012 FRINO AND GALLAGHER, 2001 POTERBA AND SHOVEN, 2002 SHIN AND SOYDEMIR, 2010
PREMIUM	**DISCOUNT**
CHARTERIS, 2013 KEARNEY et al., 2014 PETAJISTO, 2017	CHARTERIS, 2013 ELTON et al., 2002 KAYALI, 2007 PETAJISTO, 2017

4. Methodology

Following the procedures taken by the literature, different measures will be applied to the ETFs, index mutual funds and equity mutual funds in the sample to conduct a complete test of relative performance. Previously to the calculation of tracking error (which is the most used measure of relative performance in the literature [see Buetow and Henderson (2012), Frino and Gallagher (2001), Rompotis (2011a), Shin and Soydemir (2010), Wong and Shum (2010)]), some descriptive statistics and regression related to the binominal return/risk as performance measures will be calculated in line with Blitz *et al.* (2012) and Gastineau (2004). In the end, with the aim of better understanding of the ETF price efficiency, a relation regression, a deviation and persistence analysis between the exchange price and the intrinsic NAV per share will be conducted as in Charteris (2013), Kayali (2007) and Shin and Soydemir (2010).

4.1. Descriptive statistics

The first analysis that will be conducted is related to the average return and risk (obtained by the standard deviation of returns) from the examined investment schemes and indices.

The daily returns of ETF, indices and equity mutual funds are expressed by the following equations:

$$RE_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}} \quad (1) \quad \text{and} \quad RM_{i,t} = \frac{NAV_{i,t} - NAV_{i,t-1}}{NAV_{i,t-1}} \quad (2)$$

where $RE_{i,t}$ = Return of ETF or of the Index in day t; $RM_{i,t}$ = Return of index mutual funds and equity mutual funds in day t; $P_{i,t}$ = End of the day (last) Price of the ETF or value of the Index in day t; $NAV_{i,t}$ = Net Asset Value per share of the Index Mutual Funds and Equity Mutual Funds in day t. The use of the end of the day (last) Price of the ETF instead of Bid or Ask prices is related to the low Bid/Ask percentage average spread (0.42%).[5]

Regarding the risk, it will be expressed by the standard deviation of returns in the following way:

$$\sigma_{pi} = \sqrt{\frac{\Sigma_{t=1}^{n}\left(RE_{i,t} - \overline{RE}_i\right)}{n-1}} \quad (3) \quad \text{and} \quad \sigma_{NAVt} = \sqrt{\frac{\Sigma_{t=1}^{n}\left(RM_{i,t} - \overline{RM}_i\right)}{n-1}} \quad (4)$$

where $\overline{RE}_t$ is the average return of ETF or Index i and $\overline{RM}_t$ is the average return of the index mutual fund and equity mutual fund i; n is the number of observations.

In addition, it is also going to be computed the minimum, maximum, median, Skewness and Kurtosis values to obtain a clear understanding of each distribution of returns with the aim of avoiding a biased analysis. Also, with the same goal, a normality test (Jarque-Bera test) will be applied to the sample.

4.2. Regression statistics

In accordance with the literature mentioned in Section 3, to examine the performance of ETF, index mutual fund and equity mutual fund in comparison with the respective benchmarks a model regressing the return of this investment schemes on the return of the benchmark will be conducted. For that purpose, a Jensen's model [Jensen (1968)] is employed (through ordinary least squares estimation) to each ETF, index mutual fund and equity mutual fund as:

$$R_{i,t} - R_f = \alpha + \beta\,(RI_{i,t} - R_f) + \varepsilon_t \quad (5)$$

In (5) $R_{i,t}$ is the return of the ETF, index mutual fund and equity mutual fund [for simplification RM and RE (for ETFs) were merged in R]; R_f is the risk-free rate proxy and will be the result of the daily one-month interbank (Euribor) rate; $RI_{i,t}$ is the return of the Index; α is the measure of the performance (return part explained by other factors than the replication of the index); β describes the slope of the regression, being the relation of risk adjusted returns of the investment schemes and their benchmarks; ε_t is the residual. The aim is to exam the statistical significance of β with the purpose of examine the benchmark linkage to the ETF returns.

[5] Using all available information. Additionally, we also conducted all calculations using ask, bid and end of the day prices for the ETF, which lead to similar results and conclusions.

4.3. Performance and tracking error

The first examinations that are going to be conducted regarding relative performance among the investment schemes and the benchmark will aim to identify under or outperformance in relation to the benchmark for the sampling period. The identification of the number of days and the average underperformance and outperformance return will also be computed. Additionally, the ex-post Sharpe Ratio (Sharpe, 1966) and Sortino Ratio (Sortino and Price, 1994), will be computed as in (6) and (7) to evaluate comparatively all the funds:

$$SH = \frac{\overline{R}i - R_f}{\sigma_i} \quad (6) \text{ and } \qquad SO = \frac{\overline{R}i - R_f}{\sigma_n} \qquad (7)$$

where σ_i is the standard deviation of returns and σ_n is the standard deviation of negative returns (downside deviations).[6]

Regarding each investment scheme capability of benchmark replication, the tracking error deviations between their performance and the performance of the benchmark will be measured. Although the idea behind the tracking error seems simple, relevant studies applied different approaches to its calculation, specifically regarding ETFs. In the present paper we will compute four tracking error methodologies:

Average of absolute differences between the returns on investment schemes and their benchmark (TE_1) as:

$$TE_1 - \frac{\Sigma_{t=1}^{n} \mid R_{i,t} - RI_{i,t} \mid}{n} \qquad (8)$$

Standard deviation of daily relative negative returns (TE_2). This measure applies the same idea as in Sortino Ratio (downside deviations) (Sortino and Price, 1994). In detail, since investors will not dislike positive or equal to zero tracking errors (outperformance) only the daily negative relative returns will be taken into account in the calculation of the tracking error. This measure, computed as follows, will be helpful for the verification of out or underperformance of the previous calculation of tracking errors (TE_1):

$$TE_2 = \sqrt{\frac{\Sigma_{t=1}^{n} \left(R_{i,t} - RI_{i,t}\right)^2}{n-1}} \qquad (9)$$

in which $R_{i,t} - RI_{i,t} < 0$.

[6] $\sigma_i = \sqrt{\frac{\Sigma_{t=1}^{n} \left(R_{i,t} - \overline{R}_{i,t}\right)^2}{n-1}}$ and $\sigma_n = \sqrt{\frac{\Sigma_{t=1}^{n} \left(R_{n,t} - \overline{R}_{n,t}\right)^2}{n-1}}$, where $R_{n,t} - \overline{R}_{n,t} < 0$.

Standard deviation of return differences between the return of funds and the index (, measured as the following:

$$TE_3 = \sqrt{\frac{\sum_{t=1}^{n}\left(RD_{i,t} - RD_{i,t}\right)^2}{n-1}} \tag{10}$$

where $RD_{i,t}$ is the difference between the return on the investment scheme and its benchmark in day t; $\overline{RD}_i$ is the average difference between the return on the investment scheme and its benchmark.

Single index model (TE_4), which is a simplified version of the regression (5). In the following regression (calculated through an ordinary least squares estimation), the focus will be on the residual that will be the proxy for the standard deviation as if, for example, an ETF perfectly replicates its benchmark it is expected to have a tracking error of zero and an residual also equal to zero.

$$R_{i,t} = \alpha + \beta\left(RI_{i,t}\right) + \varepsilon_t \tag{11}$$

4.4. Etfs price efficiency and persistence

ETFs, due to the fact they're exchange-traded, may experience some price inefficiencies which may be one cause of higher than normal tracking errors. In order to examine this possibility, the link between the ETF Price and the NAV is going to be regressed (12) through an ordinary least squares estimation without any constant term since it is theoretically assumed that with a zero intrinsic NAV value nondifferent from zero-priced transactions will not occur. Deviation calculation (13) and descriptive statistics of its results are going to be computed for all the period available (from 2010 to 2017) since the comparison principle does not need to be verified. Also, with the purpose of examining the persistence of price inefficiency the regressions (14) and (15) will be computed with the same estimation methodology and the expected results for price efficiency are insignificant values of and meaning that the premium and discount do not persist within 1 or 2 trading days.

$$P_{t,i} = \beta\left(NAV_{t,i}\right) + \varepsilon_t \tag{12}$$
$$D_{i,t} = P_{t,i} - NAV_{t,i} \tag{13}$$
$$D_{i,t} = \gamma_0 + \gamma_1 D_{t-1} + \varepsilon_t \tag{14}$$
$$D_{i,t} = \gamma_0 + \gamma_1 D_{t-1} + \gamma_2 D_{t-2} + \varepsilon_t \tag{15}$$

5. Empirical analysis

5.1. Data

The sample used in this paper is focused on the Portuguese fund industry, including one ETF listed on the Euronext Lisbon and seven mutual funds, one of which is an index mutual fund. The index mutual fund and two other equity mutual funds have the Portuguese Stock Index (PSI 20) as clear benchmark, with its identification on their prospectus while the other four are equity mutual funds with an investment strategy limited to Portuguese stocks many of which (in particular the most liquids) are constituents of the PSI 20 Index. The author assumption of including these four equity mutual funds aims to avoid the risk of non-inclusion of funds with closet index behavior.[7]

The detailed information of each fund is presented in the Table 1. All the data used was from Thomson Reuters Datastream (accessed on September 2017) (Thomson Reuters, 2017) and was confronted (sample testing) with data available at CMVM (2017) and Euronext (2017) for validation purposes. The information will cover the period between the 30th of September 2010 and the 31st of August 2017, except for two funds that were liquidated previously (see Table 1). For each comparative examination, the period that is going to be used is from the 30th of September 2010 to the 29th of October 2015, which is the period for each all information for all funds is available.

For the ETF specific analysis, NAV was obtained from the management company website (Comstage, 2017). Lastly, it is important to mention that some daily prices of the ETF arise from valuation prices calculated by Euronext (Euronext, 2017).[8] As an example, in days in which no trade occurs the final price results from the average of the last best bid and ask prices (valuation price).

[7] Actively managed funds that use a portfolio strategy to achieve similar returns to a benchmark, without clearly mentioning this strategy and charging a relatively high management fee.

[8] The use of all observations in the study is justified by the impossibility to distinguish the source of the last price (market vs Euronext valuation) from the data source. However, since no transaction occurred in 22.5% of the observations, it is possible to conclude that this valuation occurred at least on these observations. The studied was also conducted excluding these days leading to the same conclusions.

Table 1: List of investment schemes

Type	Name	Acronym	ISIN	Benchmark	Total Expense Ratio (%NAV)
Equity Mutual Fund	Banif Acções Portugal[a]	BAN	PTYBNKLM0003	PSI 20	2.06
Index Mutual Fund	Bbva Ppa Indice Psi20[b]	BBV	PTYBBGLM0003		0.54
Equity Mutual Fund	Bpi Portugal	BPI	PTYPIGLM0000	NA	1.29
Equity Mutual Fund	Caixagest Acções Portugal	CAI	PTYCXNLP0004		2.03
Exchange-Traded Fund	Comstage Psi20	COM	LU0444605215	PSI 20	0.35
Equity Mutual Fund	Imga Ações Portugal	IMG	PTAFIALM0006		2.30
Equity Mutual Fund	Nb Portugal Ações	NBP	PTYESYLM0009	NA	2.31
Equity Mutual Fund	Santander Acções Portugal	SAN	PTYSAFLM0006		2.03

Notes: (a) Liquidated in January 2017; (b) Liquidated in October 2015.
Source: Thomson Reuters (2017).

5.2. Results

5.2.1. *Descriptive statistics*

The descriptive analysis of the returns was conducted to the subset of data for which all funds could be directly compared and the results are present in Table 2. For the total 1,301 observations, it was found that all the investment schemes in the analysis had a daily return mean higher than the Benchmark (outperformance). In detail, five of the eight funds exhibit a positive daily return mean for the period and only three had a negative daily return mean, as occurred for the benchmark.

The series that presented the results closer to the benchmark performance (-0.0158%) were BBVA PPA Índice PSI20 (-0.0085%) and Comstage PSI20 ETF (-0.006%). Regarding standard deviation, the results were mainly the same, having Comstage PSI20 ETF and BBVA PPA Índice PSI20 values of 1.3306 and 1.3264, respectively, which are very similar to the benchmark standard deviation (1.3217). Regarding minimal and maximal values, it is important to mention that only Comstage PSI20 ETF had a lower maximum daily return than the benchmark. In terms of the normality of all the series it is observed small distributions skewed to the left and a leptokurtic behavior (skewness negative but less than |0.5| and Kurtosis positive and moderately higher than 3). When applying the Jarque-Bera test for normality, the normal distribution hypothesis is rejected (at a significance level of 1%) for all returns under analysis. Its result demonstrates that the used sample is made of statistically different means and median values.

Table 2: Descriptive statistics of daily returns

Name (Acro Nym)	Mean	Median	Maxi mum	Mini mum	Std. Dev.	Skewness	Kurtosis	Obser Vations	Jarque-Bera	P-Value of Jarque-Bera Test
Period: 01-Oct-2010 29-Oct-2015										
BAN	0.0086	0.0000	5.3151	-5.8854	1.1905	-0.2482	4.8142	1,301	191.785	0.00
BBV	-0.0085	0.0000	4.9266	-6.5749	1.3264	-0.2774	4.6720	1,301	168.245	0.00
BPI	0.0115	0.0223	4.9546	-5.8460	1.1833	-0.2544	4.8383	1,301	197.237	0.00
CAI	-0.0010	0.0000	5.5550	-5.5994	1.1533	-0.2080	4.8441	1,301	193.729	0.00
COM	-0.0062	0.0696	4.6914	-5.6161	1.3306	-0.3302	3.9837	1,301	76.0985	0.00
IMG	0.0082	0.0000	5.3395	-5.5248	1.2308	-0.2278	4.8652	1,301	199.847	0.00
NBP	0.0038	0.0000	5.2779	-5.2359	1.1667	-0.2542	4.8872	1,301	207.092	0.00
SAN	0.0077	0.0000	6.6368	-5.7410	1.2467	-0.1191	5.4407	1,301	325.997	0.00
PSI20	-0.0158	0.0061	4.7119	-5.8028	1.3217	-0.2833	4.0207	1,301	73.884	0.00

The overall behavior was also observed biennially (for the period from 2010 to 2017) being the result in accordance with the ones presented for the sample [for detail, see Table A1].

5.2.2. *Regression analysis*

The results of the Jensen's model regression are expressed in Table 3. This examination showed that for seven of eight mutual funds β values were higher than 0.75 with a R-squared of at least 0.71. Moreover, for all funds the result of β was significant at 1% level. Concerning the higher β values, they were observed in the ETF (Comstage PSI20 ETF) and the index fund (BBVA PPA Índice PSI20) with a value of 0.97 and 0.93 for an R-squared of 0.93 and 0.86, respectively. Regarding equity mutual funds that don't have the PSI 20 index as benchmark, diverge results were observed. On one hand, BPI Portugal has a relatively high β value of 0.825 with a considerable data fitness to the regression line (R-squared of 0.85), but, on other hand, NB Portugal Ações reveals the lowest β value among the sample (0.61) with a R-squared of 0.48 making it questionable its linkage to the benchmark performance.

Additionally, it is relevant to mention that BPI Portugal was the only investment fund to exhibit a statistical significant α value (significant at a 10% level) meaning that a significant part of the return is positively explained by other factors than the replication of the index (e.g. active management). For comparison purposes, the results of the BPI Portugal have to be understood in line with the relatively low total expense ratio (TER) among the other equity mutual funds.

Table 3: Regression results

VARIABLES	BAN	BBV	BPI	CAI	COM	IMG	NBP	SAN
α	0.00018	0.00005	0.00022**	0.00007	0.00009	0.00018	0.00008	0.00018
t-Statistic#α	1.16112	0.37599	1.73549	0.46276	0.87709	0.96377	0.35305	0.96134
β	0.79244*	0.92908*	0.82511*	0.75386*	0.96928*	0.78551*	0.61452*	0.80022*
t-Statistic#β	66.59637	88.2812	85.42982	61.7454	128.4362	56.55615	34.92727	57.71576
R^2	0.77346	0.85714	0.84891	0.74587	0.92700	0.71118	0.48430	0.71945
Observations	1,301	1,301	1301	1301	1301	1301	1301	1301

Note: * (**) denotes significance at 1% (10%).

When increasing the sample date for the maximum of observations [which makes it necessary to exclude two investment funds (BBVA PPA Índice PSI20 and Banif Acções Portugal) for comparison purposes] the regression results (summarized in Table A2) show that the indications found in the sample were also observed and reinforced with a higher data range. In detail, the increase of data led to with a relevant increase of the β value and the R-squared for the regressions for Caixagest Acções Portugal, IMGA Ações Portugal and Novo Banco Acções Portugal. For example, Novo Banco Acções Portugal, which was the investment fund with the lowest values in the sample, increased both it's β from 0.61 to 0.67 and the R-squared from 0.48 to 0.56.

5.2.3. *Performance and tracking error*

5.2.3.1. General performance statistics and ratios

The idea behind the computation and study of tracking errors is mainly originated by the need to understand the capability of the fund to underperform or outperform its benchmark. For that purpose, before the examination of tracking errors, a summary of the absolute performance and relevant ratios values for the comparable period is presented in Table 4. During this period all funds outperformed (had a higher return) the PSI 20 Index (which faced a negative variation of 27.38%). The index fund (BBVA PPA Índice PSI20) and the ETF (Comstage PSI20) were the ones presenting closer to benchmark performances (-20.17% and -17.78%, respectively). Also, during the sample period some actively managed mutual funds faced positive returns [e.g. BPI Acções Portugal (6.05%), Banif Acções Portugal (1.98%) and IMGA Ações Portugal (0.85%)]. Although the results seem to demonstrate that outperformance is present for all funds, a decomposition of the analysis in an annual frequency for all the period available shows that for all funds (with except BBVA PPA Índice PSI20) underperformance is found at least once (see Table A3). Regarding the Sharpe and Sortino Ratio results, the evidence shows that BPI Acções Portugal is the investment fund with the best risk-adjusted return. Regarding the other results, they are in line with expected since the worst Sharpe and Sortino ratios come from the two funds (ETF and Index Funds) with the lowest performances in the sample.

Table 4: Absolute performance summary and Sharpe/Sortino ratios results

MEASURES	BAN	BBV	BPI	CAI	COM	IMG	NBP	SAN	PSI20 INDEX
Period: 01-Oct-2010 29-Oct-2015									
Return	1.98%	-20.17%	6.05%	-9.54%	-17.78%	0.85%	-3.80%	-0.07%	-27.38%
Sharpe Ratio	-0.07	-0.27	-0.03	-0.21	-0.24	-0.07	-0.14	-0.08	
Sortino Ratio	-0.10	-0.38	-0.04	-0.28	-0.34	-0.10	-0.19	-0.11	

Besides, regarding relative performance, as seen in Table 5 and suspected from the distributions presented in the previous results, it is observed a marginally higher percentage of outperformance observation in equity mutual funds. Moreover, it is observed that the outperformance and underperformance values are well distributed among the ETF and the equity mutual funds and that their average outperformance and underperformance values are mainly the same (in absolutes terms).

The index fund (BBVA PPA Índice PSI20) is the fund that had opposite results, having a higher percentage of underperformance observations (65.64%). It is also important to highlight that both BBVA PPA Índice PSI20 and the Comstage PSI20 reveal similar values

of average outperformance but different values of average underperformance. This difference will be scrutinized in the TE_2

Table 5: Relative performance summary

MEASURES	BAN	BBV	BPI	CAI	COM	IMG	NBP	SAN
Period: 01-Oct-2010 29-Oct-2015								
Average Return	0.00024	0.00007	0.00027	0.00015	0.00010	0.00024	0.00020	0.00024
Average underperformance	-0.00401	-0.00110	-0.00344	-0.00434	-0.00220	-0.00424	-0.00621	-0.00442
N.º of obs. (underperformance)	598	854	614	624	647	625	621	623
% of underperformance	45.96%	65.64%	47.19%	47.96%	49.73%	48.04%	47.73%	47.89%
Average Outperformance	0.00386	0.00232	0.00359	0.00428	0.00237	0.00438	0.00605	0.00451
N.º of obs. (outperformance)	703	447	687	677	654	676	680	678
% of outperformance	54.04%	34.36%	52.81%	52.04%	50.27%	51.96%	52.27%	52.11%
Total obs.	1,301	1,301	1,301	1,301	1,301	1,301	1,301	1,301

5.2.3.2. Tracking Errors

TE_1

The results of the average of absolute differences between the returns of the investment schemes and their benchmark are observable in Table 6. The first result to be emphasized comes from the lowest value of TE_1 from BBVA PPA Índice PSI20 [0.00152 percentage points (pp)], followed by Comstage PSI20 (0.00229 pp). BPI Portugal (0.0035 pp) and Banif Acções Portugal (0.0039 pp) also reveal a relatively low tracking error values among the actively managed equity mutual funds. Moreover, the result of the equity fund BPI Portugal has to be highlighted since this particular fund does not have the PSI 20 index as a clear benchmark on its prospectus information.

Table 6: Tracking error results

	BAN	BBV	BPI	CAI	COM	IMG	NBP	SAN
Period: 01-Oct-2010 29-Oct-2015								
TE_1	0.00393	0.00152	0.00352	0.0043	0.0022	0.0043	0.00613	0.0044
OBS.	1,301	1,301	1,301	1,301	1,301	1,301	1,301	1,301
TE_2	0.00652	0.00446	0.00501	0.0067	0.0035	0.0074	0.01029	0.0071
OBS.	598	854	614	624	647	625	621	623
TE_3	0.00493	0.00478	0.00377	0.0050	0.0028	0.0057	0.00767	0.0055
OBS.	1,301	1,301	1,301	1,301	1,301	1,301	1,301	1,301

Despite the results presented for the period, the biennial defragmentation also conducted (see Table A4) shows that overall results are not always verified biennially. As examples, in the first two years of the sample (2010-2012), Comstage PSI20 had a lower tracking error than BBVA PPA Índice PSI20 and Caixagest Acções Portugal displayed a more moderate tracking error than BPI Portugal for the last data interval (2016-2017). Finally, it is also important to mention that the average tracking error for Comstage PSI20 is less than the average tracking error of European ETFs computed by Shin and Soydemir (2010) for the 2004-2007 period.

TE_2

The outcome of the tracking error methodology applied to downside deviations is exhibited in Table 6. The results show a different perspective from TE_1 since Comstage PSI20 has a lower value than BBVA PPA Índice PSI20. This measure also shows that BPI Portugal is the equity mutual fund with the lowest tracking error regarding negative deviations, in contrast with NB Portugal Ações which has the highest value. In sum, concerning equity funds, the verified results are mainly in line with TE_1 values. Nevertheless, in the biennial analysis also conducted (for detail see Table A5), seasonality seems to have an effect on the results, having the index mutual fund and the ETF similar results in the biennial 2014-2015, although Comstage PSI20 still has the lowest values of tracking error.

TE_3

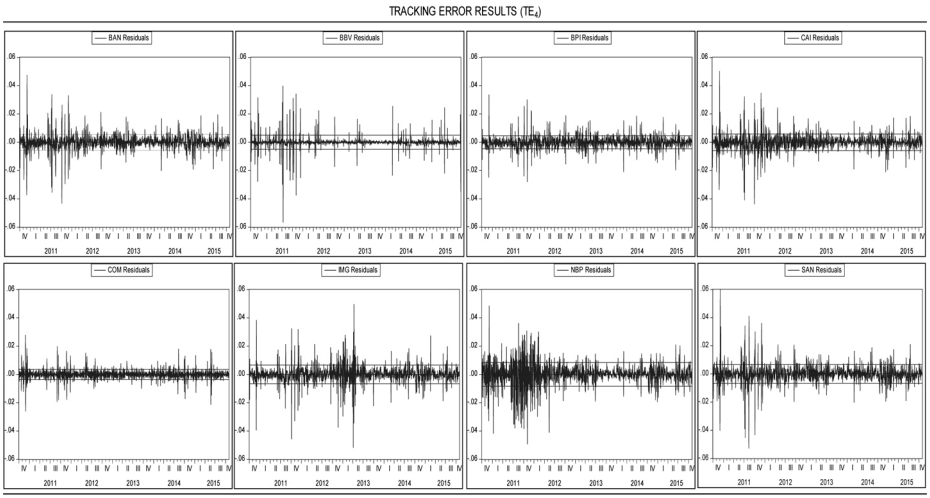
Table 6 also summarizes the results of the third measure of tracking error which is the result of the standard deviation of the return differences between the investment schemes and their benchmark. Within this measure, the fund with lower tracking error is Comstage PSI20 (0.0028). The relative surprise is the fact that BBVA PPA Índice PSI20 comes in third with (0.00478) after BPI Portugal (0.00377). In the biennial decomposition for all years available of this tracking error measure (accessible in Table A6), it is found that the values

of these two funds were only inverted for the biennial (2012-2013). The highest values of TE_3 were verified for NB Portugal Ações both in the period in analysis and in the biennial decomposition. Regarding Santander Acções Portugal, Caixagest Acções Portugal and IMGA Ações Portugal all these three funds show similar tracking error results (0.00554, 0.00509 and 0.00577) despite the fact that only IMGA Ações Portugal has the PSI20 Index as a clear benchmark.

TE_4

The single model was regressed for all the dependent variables in discussion and the residual of each regression is presented in Figure 3. The results show the residual variation among each investment scheme for the comparable period and from its breakdown is obvious the relatively low volatility and mean of the residual from BBVA PPA Índice PSI20 and Comstage PSI20 [for individual residual descriptive statistics detail see Table A7]. Regarding the equity mutual funds, BPI Portugal is the investment scheme that has the relatively lowest volatile residual in its regression. It is also important to mention that the high levels of residuals in actively managed funds are mainly seen in the year of 2011, which was a negative year for the PSI 20 index with a negative global variation of 27.60%. This result may be associated with the ability of investment fund to adjust their portfolio to the material loss faced by the financial sector related to the sovereign debt crisis. For the last years of the comparative period, even investment funds like NB Portugal Ações (which has been the fund with the highest performance differences with the benchmark) it was seen a decrease in volatility of the residual of its regression.

Figure 3: Residual graph for all the regressions



5.2.4. *ETF – Price vs. Net Asset Value and Deviation Persistence*

As expected, results presented in Table VII show that the linkage between the Price and the NAV is quite significant and close to one. Furthermore, in this regression, the high R-squared value (0.99), the significance of β at 1% level and its close to but less than one value suggests that Comstage PSI20 trades at a discount from its NAV. However, the result does not give a clear idea of the discount value, being then necessary to conduct a deviation analysis.

Table 7: Regression results

Variables	β	t-Statistic#β	R^2	Observations
$P_{t,i}$	0.999999*	15010.3958	0.99962	1673

Note: * denotes significance at 1%.

With that goal in mind, Price to NAV deviations were computed and the results are displayed in Table 8. From the results presented in the referred Table, it is possible to conclude that for all the sampling period Comstage PSI20 exhibits a slightly higher number of observations in which it is priced at a discount (854 = 51.04%) than at a premium (807 = 48.24%) though the absolute average premium (€0.01096; 0.167% of the average price) is moderately higher than the absolute average value of the discount (€0.01049; 0.174% of

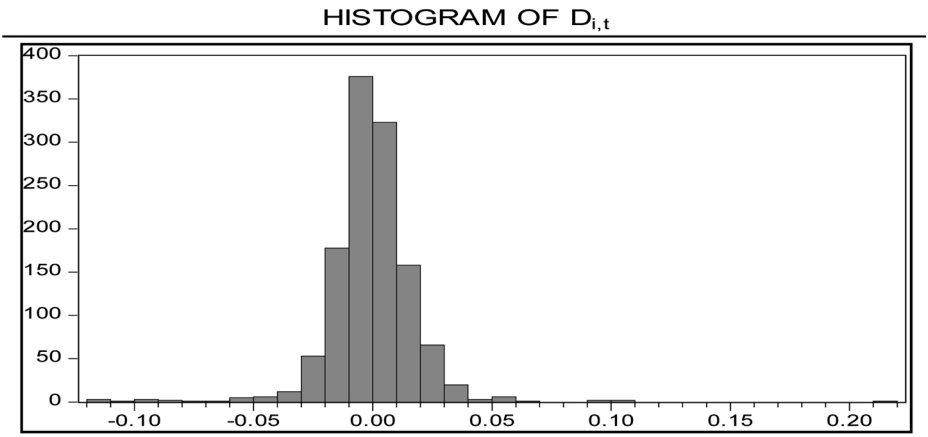
the average price). Notwithstanding, the finding presented have always to be understood in a context in which some of the prices used are originated from the price valuation conducted by Euronext. It is also important to mention the positive impact that the inclusion of more observation has in the series mean and standard deviation, suggesting that as the data range increases the average premium or discount would tend to decrease [For detail see Table A8]. This behavior may be justified by the increased popularity of the ETF (and consequent more trades and bid and ask orders) as an alternative investment schemes option for Portuguese investors within the period range.

Table 8: Performance and descriptive statistics summary

Mean	Median	Maximum	Minimum	Std. Dev.	Obs.	N.º Obs at Premium	Average Premium	N.º Obs at Discount	Average Discount
Period: 01-Oct-2010 31-Aug-2017									
-0.00007	-0.0003	0.2175	-0.1180	0.01729	1673	807	0.01096	854	-0.01049

Regarding the distribution of the difference between the Price of the ETF and its NAV, it can be seen in Figure 4 that is mainly settled around the mean having a few statistical outliers. However, the existence of deviations sets the need to study its persistence.

Figure 4: Histogram of the difference values between the Prices and the NAV



For that purpose, two regressions with the aim of evaluating the persistence of deviation analysis were conducted being the results presented in Table 9. As seen in the referred table, although both the dependent variables in regressions 1 (one lag period) and 2 (two lag period) are significant at 1% with positive values, the R-squared of both regression shows that the explanatory power of both regressions is small and close to zero. Nevertheless, the persistence of small deviations is a verified phenomenon in Comstage PSI20 for at least two days lag. Again, the results may be justified by the low turnover associated to the ETF and the relevant importance of price valuation of Euronext. However, it is important to mention that the average discount is just 0.00007 euros (0.001% of the average price) which, for example, can be favorably compared with the average discount of $0.014[9] (0.018% of average price difference) for SPDRs from Elton et al. (2002) and with Kayali (2007) discount of £0.008[10] (0.11% of average price).

Table 9: Regression results

	REGRESSION 1	REGRESSION 2
γ_0	-0.00008	0.00003
t-Statistic#γ_0	-0.18232	0.06053
γ_1	0.28344*	0.25116*
t-Statistic#γ_1	11.73379	9.53841
γ_2	NA	0.09646*
t-Statistic#γ_2	NA	3.69158
R^2	0.08086	0.08526
Observations	1,567	1,474

Note: * denotes significance at 1%.

6. Conclusion

We applied different approaches to examine the empirical validation of Comstage PSI20 ETF performance as an alternative investment option to Portuguese Investors that aim to be exposed to the fluctuation of constituents of the Portuguese Stock Index (PSI 20 index). The results suggest that an investor that started an investment in the 30th of September 2010 and closed all positions on the 29th of October 2015 would achieve a closer to index performance if he had chosen to invest in the Index Fund (BBVA PPA Índice PSI20) rather than in the ETF or any other mutual fund. Moreover, for the same period, all the investment schemes outperformed the PSI 20. However, within the sample period, investment in these different securities would be associated with different average rates of under and outperformance, being the number of days of outperformance slightly higher than the underperformance

[9] Which corresponds to €0.0133504 [Exchange Rate for the 31/12/2002 from the ECB (2017)].

[10] Which corresponds to €0.0046592 [Exchange Rate for the 31/12/2007 from the ECB (2017)].

ones, with except of the BBVA PPA Índice PSI20, in which in 65% of the trading days the investor would face a lower return than the PSI 20 index variation.

Also, the results from the Jensen's model have shown a high relation (over 0.75) between mainly all the investment schemes performance and the PSI 20 index performance, making the equity mutual funds also a competitive opponent for ETFs and Index funds in terms of PSI 20 index exposure. These values corroborate the importance of including the four equity mutual funds in the study despite all of them not having the PSI20 as a clear benchmark in their prospectus. Comstage PSI 20 is the investment scheme with the highest relation between the benchmark and its returns (0.97), meaning that a daily return of 1% in the PSI 20 index increases in 0.97% the return of this ETF in the 2010-2015 studied period.

Since investors that choose an ETF as an investment option have the aim of tracking the benchmark return, several tracking error measures were calculated in order to have a clear investigation of this relevant ability. As expected, the lowest results of tracking error measures (i.e. the investment schemes with the highest ability to track the PSI 20 Index) were from Comstage PSI20 (ETF), BBVA PPA Índice PSI20 (Index Fund) and, as not so expected, from the equity mutual fund BPI Portugal since it has not the PSI 20 Index as a benchmark. The results suggest that BBVA PPA Índice PSI20 tracks better both positive and negative variations of the benchmark. However, regarding just downside deviations, Comstage PSI20 is found to be the best tracking investment scheme option for the examined period. Additionally, Comstage PSI20 is the best tracking investment scheme option if an investor intends to have a smooth tracking of the index (lowest value of the standard deviation of the difference between the return of the investment scheme and the PSI 20 index) and its results can be positively compared with the Shin and Soydemir (2010) study, meaning that a lower than average tracking error was verified. Within equity mutual funds, only the BPI Acções Portugal exhibited competitive tracking error values that could be comparable to the ones presented by the ETF and index mutual fund.

Solely for the ETF, the price efficiency and deviation persistence results showed that Comstage PSI20 exhibits a slightly higher number of days being traded at a discount (51.04%) than at a premium (48.24%), been verified only 12 trading day in which the price was the same as the NAV (price efficiency). Also, it was concluded that on average Comstage PSI20 faces a discount of €0.00007 between the Price and NAV, with minimum (discount) and maximum (premium) values of −€0.118 (-1.52% difference between the price and NAV) and €0.2175 (2.85% difference between the price and NAV), respectively. The results can be positively compared (since a lower discount was found) with the analysis of Elton et al. (2002) and Kayali (2007).

Persistence of deviations (premium and discounts) was also examined, being concluded the existence of this phenomenon for at least two trading days. However, the results need to be relativized by the low average value of deviations and by the low explanatory power of the examined regressions.

Our paper also highlighted some topics that may be relevant for further research. Future research may focus on the study of seasonality of the tracking errors for the investigated funds and the computing price efficiency of the ETF solely for traded values, avoiding the Euronext estimations that were not perceived in the current study. Also, regarding the ETF traded in the local exchange (Euronext Lisbon), a comparable examination including leverage ETFs may be conducted.

Referências

Ackert, L. F.; Tian, S. (2000) Arbitrage and valuation in the market for Standard and Poor's Depositary Receipts. *Financial Management*, 29(3): 71-87.

Agapova, A. (2011) Conventional mutual index funds versus Exchange-Traded Funds. *Journal of Financial Markets* 14(2): 323-343.

Blitz, D.; Huij, J. (2012) Evaluating the performance of global emerging markets equity Exchange-Traded Funds. *Emerging Markets Review*, 13(2): 149-158.

Blitz, D., Huij, J.; Swinkels, L. (2012) The performance of European Index Funds and Exchange-Traded Funds. *European Financial Management*, 18(4): 649-662.

Buetow, G. W.; Henderson, B. J. (2012) An empirical analysis of Exchange-Traded Funds. *Journal of Portfolio Management*, 38(4): 112-127.

Charteris, A. (2013) The price efficiency of South African Exchange Traded Funds. *Investment Analysts Journal*, (78): 1-11.

CMVM (2017) Individual NAV information from the Portuguese Securities Market Commission website [online]. Available from: http://web3.cmvm.pt/english/sdi /fundos/app/index.cfm.

Comstage (2017) Individual NAV information from the Comstage PSI20 [online]. Available from: https://www.comstage.pt/Products/ProductSearchPage.aspx? pac=7&c=220009.

ECB (2017) Historical Euro foreign exchange reference rates [online]. Available from: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_referen ce_exchange_rates/ html/index.en.html.

Elton, E. J.; Gruber, M. J.; Comer, G.; Li, K. (2002) Spiders: Where are the bugs? *Journal of Business*, 75(3): 453-472.

ETFGI (2017) Exchange-Traded Ffunds statistics from ETFGI LLP [online]. Available from: http://etfgi.com/index/home.

Euronext (2017) Individual Price information of the Comstage PSI20 from the Euronext website [online]. Available from: https://www.euronext.com/nl/ products/etfs/LU0444605215-XLIS/quotes.

Frino, A.; Gallagher, D. (2001) Tracking S&P 500 Index Funds. *Journal of Portfolio Management*, 28: 44-55.

Gastineau, G. L. (2001) Exchange-Traded Funds: An introduction–And further likely evolution. *Journal of Portfolio Management*, 27(3): 88-96.

Gastineau, G. L. (2004) The benchmark index ETF performance problem: A simple solution. *Journal of Portfolio Management*, 30(2): 96-103.

Jensen, M. C. (1968) The performance of Mutual Funds in the period 1945-1964. *Journal of Finance*, 23(2): 389-416.

Kayali, M. M. (2017) Pricing efficiency of Exchange Traded Funds in Turkey: Early evidence from the Dow Jones Istanbul 20. *International Research Journal of Finance and Economics*, 10: July, 14-23.

Kearney, F.; Cummins, M.; Murphy, F. (2014) Outperformance in Exchange-Traded Fund pricing deviations: Generalized control of data snooping bias. *Journal of Financial Markets*, 19: 86-109.

Kostovetsky, L. (2003) Index mutual funds and Exchange-Traded Funds: A comparison of two methods of passive investment. *Journal of Portfolio Management*, 29(4): 80-92.

Lin, C.; Chan, S. J.; Hsu, H. J. (2006) Pricing efficiency of Exchange Traded Funds in Taiwan. *Journal of Asset Management*, 7: 60-68.

Morningstar (2017). NAV information of the Comstage PSI20 [online]. Available from: http://www.morningstar.co.uk/uk/etf/snapshot/snapshot.aspx?id=0P0000S AMD.

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

Pinheiro, C. M.; Varela, H. H. (2018) Do Exchange Traded Funds (ETFs) outperform the market? Evidence from the Portuguese Stock Index. GEE Papers No. 0109, Gabinete de Estratégia e Estudos, Ministério da Economia, Lisboa.

Poterba, J. M.; Shoven, J. B. (2002) Exchange-Traded Funds: A new investment option for taxable investors. *American Economic Review*, 92(2): 422-427.

Rompotis, G. G. (2011a). ETFs vs. Mutual Funds: Evidence from the Greek market. *South-Eastern Europe Journal of Economics*, 9(1): 27-43.

Rompotis, G. G. (2011b). Predictable patterns in ETFs' return and tracking error. *Studies in Economics and Finance*, 28(1): 14-35.

Sharpe, W. F. (1966). Mutual Fund performance. *Journal of Business*, 39(1): 119-138.

Shin, S.; Soydemir, G. (2010) Exchange-Traded Funds, persistence in tracking errors and information dissemination. *Journal of Multinational Financial Management*, 20(4): 214-234.

Sortino, F.; Price, L. (1994) Performance measurement in a downside risk framework. *Journal of Investing*, 3: 59-65.

Thomson Reuters (2017) Thomson Reuters Datastream [Online]. Available at: Subscription Service.

Wong, K. H. Y.; Shum, W. C. (2010) Exchange-Traded Funds in bullish and bearish markets. *Applied Economics Letters*, 17(16): 1615-1624.

Xu, L.; Yin, X. K. (2017) Does ETF trading affect the efficiency of the underlying index? *International Review of Financial Analysis*, 51: 82-101.

ANNEX

Table A1: Descriptive statistics

NAME (ACRONYM)	MEAN	MEDIAN	MAXIMUM	MINIMUM	STD. DEV.	SKEWNESS	KURTOSIS	OBSERVATIONS
Period: 01-Oct-2010 31-Dec-2011								
BAN	-0.1005	0.0000	5.3151	-4.0629	1.2220	-0.0341	4.7771	323
BBV	-0.0780	0.0000	4.9266	-6.5749	1.3978	-0.3328	5.5585	323
BPI	-0.0916	0.0000	4.9546	-4.0880	1.2242	-0.1038	4.5550	323
CAI	-0.0927	0.0000	5.5550	-4.3040	1.2792	0.1091	4.6358	323
COM	-0.0813	0.0000	3.4276	-5.3398	1.3523	-0.3122	3.8432	323
IMG	-0.0961	0.0000	5.3395	-4.5411	1.2807	-0.0794	4.8285	323
NBP	-0.1103	0.0000	5.2779	-5.1322	1.2397	-0.1391	5.0037	323
SAN	-0.0910	0.0000	6.6368	-4.0682	1.3902	0.3227	5.9163	323
PSI20	-0.0873	-0.0573	3.2216	-5.2156	1.3608	-0.3034	3.8162	323
Period: 01-Jan-2012 31-Dec-2013								
BAN	0.0754	0.0927	4.0956	-5.2184	1.0775	-0.2542	4.7789	511
BBV	0.0558	0.0181	4.3516	-5.3151	1.1585	-0.1512	4.2811	511
BPI	0.0758	0.0848	4.6332	-5.8460	1.1109	-0.3039	5.4405	511
CAI	0.0514	0.0394	3.1223	-3.5430	0.9796	-0.2945	4.2949	511
COM	0.0557	0.0818	4.6914	-5.3366	1.1750	-0.2350	4.2512	511
IMG	0.0678	0.0530	4.6697	-5.5248	1.1328	-0.2395	5.2814	511
NBP	0.0732	0.0294	4.0229	-5.2359	1.0625	-0.2839	4.9422	511
SAN	0.0772	0.0428	4.5093	-5.7410	1.0963	-0.2893	5.4152	511
PSI20	0.0416	0.0266	4.3531	-5.3148	1.1762	-0.1610	4.1902	511

Period: 01-Jan-2014 31-Dec-2015								
BAN	0.0154	0.0462	4.8380	-5.8854	1.2682	-0.3557	4.6915	511
BBV	-0.0308	0.0000	4.7953	-5.9289	1.4424	-0.2615	3.9377	467
BPI	0.0138	0.0426	4.3842	-5.4034	1.2194	-0.3206	4.5864	511
CAI	0.0088	0.0322	4.4897	-5.5994	1.2271	-0.3693	4.8044	511
COM	-0.0219	0.0750	4.4405	-5.6161	1.4483	-0.3641	3.7116	511
IMG	0.0145	0.0036	4.5407	-5.4821	1.2827	-0.3224	4.5524	511
NBP	0.0053	0.0271	4.2717	-5.1266	1.2040	-0.2953	4.6520	511
SAN	-0.0006	0.0176	4.4492	-5.5288	1.2812	-0.3284	4.6866	511
PSI20	-0.0311	-0.0065	4.7119	-5.8028	1.4184	-0.3184	3.8634	511
Period: 01-Jan-2016 31-Aug-2017								
BAN	-0.0217	0.0187	3.1503	-6.1327	1.2233	-0.8620	6.1780	264
BBV	NA	NA	NA	NA	NA	NA	NA	0
BPI	0.0040	0.0000	3.1953	-6.0542	1.0697	-0.7736	7.1950	428
CAI	0.0132	0.0101	3.1647	-5.9462	1.0493	-0.7983	7.1863	428
COM	0.0000	0.0000	3.3270	-7.0642	1.1335	-0.8322	7.4440	428
IMG	0.0117	0.0021	3.3627	-6.4032	1.0779	-0.8455	7.4666	428
NBP	0.0023	0.0160	3.6662	-6.9296	1.1311	-0.8204	7.6159	428
SAN	-0.0008	0.0229	3.2344	-6.9905	1.1096	-0.8136	7.9180	428
PSI20	0.0211	0.0000	3.5680	-6.5879	1.0645	-0.7287	7.4911	428

Table A2: Regression results (01-Oct-2010 to 31-Aug-2017)

VARIABLES	BAN	BBV	BPI	CAI	COM	IMG	NBP	SAN
α	-0.00027	0.00005	0.00020	0.00013	-0.00008	0.00019	0.00010	0.00021
t-Statistic#α	-1.78685**	0.37599	1.92113**	1.00863	-1.06441	1.30202	0.57073	1.41956
β	0.97461	0.92908	0.83948	0.78264	0.95507	0.81036	0.67909	0.81799
t-Statistic#β	77.7967*	88.2812*	100.7108*	77.9738*	154.1708*	71.6205*	47.2244*	72.1111*
R^2	0.79019	0.85714	0.85135	0.77442	0.93066	0.74335	0.55738	0.74595
Observations	1,609	1,301	1773	1773	1773	1773	1773	1773

Note: * (**) denotes significance at 1% (10%).

Table A3: Absolute performance summary

	BAN	BBV	BPI	CAI	COM	IMG	NBP	SAN	PSI20 INDEX
Period: 01-Oct-2010 31-Dec-2010	-2.27%	1.33%	-0.80%	0.25%	1.60%	-0.83%	-2.77%	-1.09%	1.08%
Period: 01-Jan-2011 31-Dec-2011	-27.81%	-25.68%	-26.81%	-28.00%	-26.51%	-28.02%	-29.76%	-26.95%	-27.60%
Period: 01-Jan-2012 31-Dec-2012	14.40%	7.93%	14.01%	2.61%	7.88%	12.39%	15.16%	9.04%	2.93%
Period: 01-Jan-2013 31-Dec-2013	24.72%	19.09%	25.19%	23.67%	18.92%	21.73%	22.60%	31.94%	15.98%
Period: 1-Jan-2014 31-Dec-2014	-13.19%	-26.06%	-11.76%	-13.63%	-25.68%	-13.21%	-9.26%	-11.63%	-26.83%
Period: 01-Jan-2015 31-Dec-2015	19.58%	NA	17.05%	16.53%	13.99%	18.96%	9.11%	8.20%	10.71%
Period: 01-Jan-2016 31-Dec-2016	-6.05%	NA	-11.91%	-6.83%	-10.87%	-11.07%	-13.70%	-7.23%	-11.93%
Period: 01-Jan-2017 31-Aug-2017	NA	NA	12.69%	10.91%	9.15%	15.29%	13.83%	15.13%	10.20%

Table A4: Tracking error results (TE_1)

Tracking Error	BAN	BBV	BPI	CAI	COM	IMG	NBP	SAN
Period: 01-Oct-2010 31-Dec-2011								
TE_1	0.00544	0.00329	0.00371	0.00565	0.00285	0.00410	0.01208	0.00607
OBS.	323	323	323	323	323	323	323	323
Period: 01-Jan-2012 31-Dec-2013								
TE_1	0.00326	0.00070	0.00333	0.00420	0.00199	0.00498	0.00440	0.00380
OBS.	511	511	511	511	511	511	511	511
Period: 01-Jan-2014 31-Dec-2015								
TE_1	0.00354	0.00120	0.00349	0.00343	0.00216	0.00362	0.00385	0.00396
OBS.	511	467	511	511	511	511	511	511
Period: 01-Jan-2016 31-Aug-2017								
TE_1	0.00312	NA	0.00275	0.00238	0.00162	0.00274	0.00300	0.00310
OBS.	264	NA	428	428	428	428	428	428
Period: 01-Oct-2010 31-Aug-2017								
TE_1	0.00376	0.00152	0.00330	0.00381	0.00211	0.00389	0.00530	0.00409
OBS.	1,609	1,301	1,773	1,773	1,773	1,773	1,773	1,773

Table A5: Tracking error results (TE_2)

Tracking Error	BAN	BBV	BPI	CAI	COM	IMG	NBP	SAN
Period: 01-Oct-2010 31-Dec-2011								
TE_2	0.01025	0.00843	0.00593	0.00992	0.00465	0.00822	0.01764	0.01133
OBS.	150	168	167	160	164	167	156	154
Period: 01-Jan-2012 31-Dec-2013								
TE_2	0.00431	0.00177	0.00427	0.00550	0.00252	0.00846	0.00694	0.00486
OBS.	237	381	240	248	249	250	242	244
Period: 01-Jan-2014 31-Dec-2015								
TE_2	0.00501	0.00357	0.00500	0.00487	0.00334	0.00515	0.00512	0.00526
OBS.	226	305	221	232	256	226	242	247
Period: 01-Jan-2016 31-Aug-2017								
TE_2	0.00437	NA	0.00398	0.00342	0.00318	0.00397	0.00436	0.00417
OBS.	130	NA	217	216	211	205	204	204
Period: 01-Oct-2010 31-Aug-2017								
TE_2	0.00616	0.00446	0.00476	0.00603	0.00338	0.00672	0.00911	0.00649
OBS.	743	854	845	856	880	848	844	849

Table A6: Tracking error results (TE_3)

Tracking Error	BAN	BBV	BPI	CAI	COM	IMG	NBP	SAN
Period: 01-Oct-2010 31-Dec-2011								
TE_3	0.00779	0.00781	0.00500	0.00792	0.00384	0.00668	0.01120	0.00893
OBS.	323	323	323	323	323	323	323	323
Period: 01-Jan-2012 31-Dec-2013								
TE_3	0.00318	0.00254	0.00306	0.00369	0.00190	0.00650	0.00529	0.00351
OBS.	511	511	511	511	511	511	511	511
Period: 01-Jan-2014 31-Dec-2015								
TE_3	0.00554	0.00760	0.00380	0.00568	0.00314	0.00491	0.00824	0.00628
OBS.	511	467	511	511	511	511	511	511
Period: 01-Jan-2016 31-Aug-2017								
TE_3	0.00345	NA	0.00309	0.00288	0.00221	0.00298	0.00311	0.00304
OBS.	264	NA	428	428	428	428	428	428
Period: 01-Oct-2010 31-Aug-2017								
TE_3	0.00469	0.00487	0.00363	0.00469	0.00266	0.00523	0.00690	0.00504
OBS.	1,609	1,301	1,773	1,773	1,773	1,773	1,773	1,773

Table A7: Descriptive statistics of residuals

REGRESSION (IDENTIFIED BY THE DEPENDET VARIABLE)	MEAN	MEDIAN	MAXIMUM	MINIMUM	STD. DEV.	SKEWNESS	KURTOSIS	OBSERVATIONS
Period: 01-Oct-2010 29-Oct-2015								
BAN	4.97E-20	2.03E-05	0.0474	-0.0431	0.0057	-0.2024	16.80	1,301
BBV	2.35E-19	-9.33E-05	0.0397	-0.0565	0.0050	-0.6849	33.93	1,301
BPI	-5.60E-20	-5.40E-05	0.0336	-0.0279	0.0046	0.4239	11.50	1,301
CAI	-2.13E-19	8.37E-05	0.0502	-0.0436	0.0058	-0.0399	16.84	1,301
COM	-7.72E-20	-7.09E-05	0.0277	-0.0260	0.0036	0.1730	14.10	1,301
IMG	4.06E-19	-7.64E-05	0.0494	-0.0516	0.0066	-0.3741	15.83	1,301
NBP	-1.91E-19	2.16E-04	0.0484	-0.0492	0.0084	-0.6486	8.79	1,301
SAN	-1.35E-20	2.87E-05	0.0606	-0.0527	0.0066	-0.0184	19.51	1,301

Table A8: Performance and descriptive statistics summary

MEAN	MEDIAN	MAXIMUM	MINIMUM	STD. DEV.	OBS.		N.º Obs at Premium	Average Premium	N.º Obs at Discount	Average Discount
Period: 01-Oct-2010 29-Oct-2015										
-0.00030	-0.00070	0.21750	-0.11800	0.01889	1223		577	0.01229	641	-0.01165

Catarina Pinto
School of Economics and Management
Universidade do Porto
catarinap1994@gmail.com
Miguel Sousa
School of Economics and Management
Universidade do Porto
msousa@fep.up.pt

DOI: https://doi.org/10.14195/2183-203X_48_2

Impact of double taxation treaties on cross-border acquisitions

Impacto dos tratados de dupla tributação nas aquisições transfronteiriças

Catarina Pinto
Miguel Sousa

Received for publication: December 15, 2018
Revision accepted for publication: May 10, 2019

ABSTRACT
In order to evaluate the impact of Double Taxation Treaties (DTTs) on the Foreign Direct Investment (FDI), we analysed the impact of a DTT implementation on both the number of cross-border acquisitions and the average value of M&A deals between companies from the countries that signed the DTT. Moreover, the impact of DTTs on the takeover bid premiums is analysed in order to access if companies are willing to pay higher premiums after the DTT is implemented and whether the impact on the premium is immediate or gradual. Overall, our findings lead us to conclude that DTTs effectively promote FDI.
Keywords: Double taxation treaties; cross-border acquisitions; takeover bid premiums; foreign direct investment.

JEL Classification: F21; F23; F38; G34; H25; H26; H87

RESUMO
Para avaliar o impacto dos Tratados de Dupla Tributação (TDT) sobre o Investimento Direto Estrangeiro (IDE), analisamos o impacto da assinatura de um TDT no número de aquisições transfronteiriças e no valor médio dessas aquisições entre empresas dos países que assinaram a TDT. Além disso, analisamos o impacto da assinatura de um TDT nos prémios pagos pelo adquirente face ao valor de mercado da empresa alvo, com o objetivo de verificar se as empresas adquirente estão dispostas a pagar prémios mais elevados após a assinatura da TDT e se o impacto, a existir, é imediato ou gradual. No geral, os nossos resultados sugerem que os TDT efetivamente promovem o IDE.
Palavras-chave: Tratados de dupla tributação económica; aquisições transfronteiriças; prémios de aquisição; investimento direto estrangeiro.

Acknowledgements: This research has been financed by the European Regional Development Fund through COMPETE 2020 – Programa Operacional Competitividade e Internacionalização (POCI) and by Portuguese public funds through FCT (Fundação para a Ciência e a Tecnologia) in the framework of the project POCI-01-0145-FEDER-006890.

1. Introduction

In the introduction to its model tax convention, Organization for Economic Co-operation and Development (OECD) emphasises the harmful effects that double taxation has on the movement of capital in the development of inter-country economic relations and the importance of removing obstacles resulting from double taxation (OECD, 2014). In order to solve this problem, OECD developed a model for country-pairs to use in negotiating DTTs which is widely used all around the world.

Despite the efforts made to solve double taxation issues, it is not clear that Double Taxation Treaties (DTTs) have a positive effect on Foreign Direct Investment (FDI). Actually, regarding the effect of DTTs on FDI, the literature is not consensual, as there are empirical studies showing that the effect can be either positive, negative, or null. A possible explanation for the inconsistencies shown by previous literature is the fact that DTTs not only aim to eliminate the double taxation problem in order to facilitate the movement of capital between countries but also intend to prevent tax evasion. The coexistence of different goals may lead to different results in what concerns the effect of DTTs on FDI.

It is also important to notice that DTTs are costly given the duration, labour intensity of the negotiation process, and the effort required to match treaty versions in different languages. Additionally, the provisions in the treaty may conflict with domestic tax law, which has to be adapted as a consequence. Moreover, the potential loss of tax revenue resulting from a DTT must be considered.

Considering jointly DTTs' wide usage and the high costs associated with their implementation, it is very pertinent to examine if DTTs actually fulfil their ultimate goal – stimulate the FDI. Therefore, this study aims to investigate if DTTs have contributed to creating an attractive scenario for a specific type of FDI, the cross-border acquisitions.

To evaluate if DTTs are effectively promoting FDI, four main hypotheses are tested. Firstly, a large sample of deals between companies from countries that signed a DTT was collected to access if the number and the average value of the deals changes after the implementation of a DTT. Assuming that DTTs are effectively promoting FDI, we expected to find an increase in the number and in the value of the deals made after the signature of a DTT. Therefore, we set the following hypotheses:

[H1] The number of cross-border deals, between countries that signed a DTT, increases after the signature of a DTT.

[H2] The average value of the cross-border deals is higher after the signature of a DTT.

Secondly, following Huizinga et al. (2012), that provide empirical evidence showing that additional taxation is fully capitalized into lower takeover bid premiums,[1] and assuming a tax burden resulting from double international taxation, it is expected that less value is created in a cross-border acquisition when a DDT is not implemented. Given that with the implementation of a DTT this tax burden disappears, the present study aims to access if there are observable changes in a cross-border takeover bid premiums after a DTT is implemented. The introduction of bid premiums in our analysis is especially relevant since

[1] The takeover bid premium represents the difference between the offering price and the estimated value of a company (proxy by the share price before the announcement).

it is the first time that bid premiums are used to measure DTTs' effectiveness. Additionally, this study aims to access if the impact of DTTs on bid premium is immediate or gradual. Therefore, the following hypotheses are tested:

[H3] Companies are willing to pay higher premiums after a DTT is implemented.

[H4] Companies are willing to pay higher premiums if a DTT is in force for a longer period of time.

Despite the fact that, to our best knowledge, there is not any other study analysing the impact of DTTs on takeover bid premiums, previous literature already studied the effect of tax treaties on FDI and there is not any agreement among the literature regarding this effect. Some reasons can justify the diversity of conclusions such as the use of different samples, time frames, and estimation methods.

Blonigen and Davies (2002) use ordinary least squares and fixed effects strategy to estimate the effect of DTTs on FDI and find evidence of a negative effect. Also, Egger et al. (2006) find a significant negative impact of newly implemented tax treaties on outward FDI stocks.

Blonigen and Davies (2004) revisit the same research question and focus on U.S. FDI activity between 1980 and 1999. They use a fixed effect strategy and find that the average of new treaty effect is not statistically different from zero, for both inward and outward United States' FDI. Also, Baker (2014) and Coupé et al. (2008) find no evidence of a relationship between DTTs and FDI. Louie and Rousslang (2008) focus on the required rate of returns rather than FDI and find either a negative or no evidence of a relationship between tax treaties and required rate of returns.

In contrast, Barthel et al. (2010) and di Giovanni (2005) find out that DTTs do lead to higher FDI stocks. Barthel et al. (2010) reach the conclusion that DTTs increase the bilateral FDI stock between 27% and 31% using a sample in which both developed and developing countries are broadly represented over a long period of time. Also, di Giovanni (2005) indicates the increasing number of DTTs as an explanatory factor of the increase of mergers and acquisition activity between 1900 and 1999. By focusing on developing countries, Neumayer (2007) concludes that DTTs are only effective in the group of middle income developing countries.

Although our results suggest that the implementation of a DTT does not change the number of cross-border deals within companies from the countries that signed the DTT, the deals carried out after the implementation of a DTT are on average larger than those carried out before and that companies are willing to pay a higher premium – up to 20 p.p. – after a DTT is implemented. These results suggest that indeed that the implementation of a DTTs effectively promote FDI.

After this introduction, this work proceeds as follows. In section 2, the methodology for the current study is presented. Section 3 comprises all aspects related to the data used and the results are shown in section 4. Finally, the main conclusions are presented in section 5.

2. Methodology

This study uses four different models in order to test the four hypotheses previously presented.

The first two models aim to test if the number [H1] and the average size [H2] of deals increase after the implementation of a DTTs, respectively.

$$N.^{o}\ of\ Deals_{it} = \beta_0 + \beta_1 DTT\_dummy_{it} + \beta_2 SumGDP_{it} + \varepsilon_{it} \quad (1)$$
$$Average\ Deal\ Value_{it} = \beta_0 + \beta_1 DTT\_dummy_{it} + \beta_2 SumGDP_{it} + \varepsilon_{it} \quad (2)$$

These models are similar to models used in previous studies that measure the impact of DTTs on FDI flows (Blonigen and Davies, 2002; Egger et al., 2006; Blonigen and Davies, 2004; Baker, 2014; Coupé et al., 2008; Louie and Rousslang, 2008; Barthel et al., 2010; di Giovanni, 2005; Neumayer, 2007). The number of deals and the deal average value are used as proxies for FDI. Panel data, with fixed effects,[2] was used to estimate the models, using a sample of cross-border deals occurred from 1996 to 2017, between companies from countries that signed a DTT from 2000 to 2015.

The models differ in the dependent variable. In the first model, the dependent variable represents the number of cross-border acquisitions made between the countries of the pair *i* during the year *t*, while in the second model the average value of the deals made between the countries of the pair *i* during the year *t* is set as the dependent variable.

To estimate the first model, the negative binomial regression is used since we are dealing with count data. The poison regression is also an option to deal with this type of data, however, given the high dispersion of the data, the negative binomial regression works better. In the remaining models estimated, linear regressions were used.

For both models, the explanatory variable of main interest is the existence of a DTT – *DTT_dummy*. This variable is a dummy and it assumes the value 1 when there is a DTT in force between the countries where the deal occurred, and 0 otherwise. The effective date was taken as the reference date rather than the signature date because effectiveness is what matters most to the investors.[3] Our hypotheses [H1] and [H2] are confirmed if both coefficients associated with the variable *DTT_dummy* come out positive.

The control variable $Sum\ GDP_{it}$ represents the sum of the nominal GDP per capita of both countries involved in the transaction in the year the transaction occurred and allow to control for macroeconomics condition that can have an impact in the M&A activity. In model (2) we control for serial correlation, and report standard errors that are robust to heteroskedasticity and are clustered at country-pair level.

Regarding the hypotheses [H3] and [H4] – companies are willing to pay higher premiums when there is a DTT implemented and if the change happens immediately or gradually after the DTT is implemented – the model developed by Huizinga et al. (2012) is taken as the

[2] In the case of model (1) was not possible to perform the Hausman test, but the results were pretty similar if random effects were used. In the case of model (2), according to the Hausman test, fixed effects were not the prefer method only in the regression without any control variable, but the results were very similar if random effects were used. All other cases, Hausman test considered that fixed effects should be used.

[3] As the investors could advance the acquisition to be already installed in a country right after the DTT is implemented, for the sake of robustness, we, alternatively, used in models (1) and (2) a 'DTT-1' dummy that assumes the value 1 if the acquisition occurred after one year before the DTT is implemented, and 0 otherwise.

departure point. By introducing takeover bid premiums in our analysis, we are contributing to the introduction of a new way to measure DTTs' effectiveness.

The following linear regressions models are set:

$$BP_i = \beta_0 + \beta_1 DTT\_dummy_i + \beta_2 SumGDP_i + \beta_3 Mkt\_cap_i + \beta_4 MtB_i + \beta_5 DtM_i + \beta_6 \%ofaq_i + \varepsilon_t \quad (3)$$

$$BP_i = \beta_0 + \beta_1 DTT\_age_i + \beta_2 SumGDP_i + \beta_3 Mkt\_cap_i + \beta_4 MtB_i + \beta_5 DtM_i + \beta_6 \%ofaq_i + \varepsilon_t \quad (4)$$

In both models, *Bid Premium* at the rumoured date is set as the dependent variable. By considering the rumoured date rather than the announced date, the possible effects resulting from speculation before the transaction are mitigated.[4]

The explanatory variable of main interest in the model (3) is the *DTT_dummy* – a dummy and it assumes the value 1 when there is a DTT implemented between the countries where the deal occurred, and 0 otherwise – of a DTT as it is in the models (1) and (2). Considering the hypothesis 3, a positive coefficient indicates that DTTs are promoting FDI while a negative coefficient might indicate that different purposes of DTTs affect negatively FDI.

In model (4), *DTTage* is the variable of main interest and it represents the number of years elapsed between the effective date of the DTT and the transaction date. To compute this regression, only the deals made after the signature of a DTT are taken into account. Considering hypothesis 4, a positive coefficient may indicate that tax reductions and renegotiations of existing treaties have a positive effect on the promotion of FDI.

Regarding the control variables, they do not differ between models and can be divided into three groups: country, target firm and deal variables. The country variable is $Sum\ GDP_i$ as in model (1) and (2).

Regarding target variables – *Mkt_cap*, *MtB* and *DtM* – all use the year before the transaction as the reference. Market Capitalization (*Mkt_cap)* controls for the target size. In most observations, this value was taken directly from Zephyr.[5]

Market to book value (*MtB*), also known as price to book ratio, is used to compare a company's current market price to its book value indicating whether a company is over or under evaluated. A relatively small market-to-book ratio suggests that the target is undervalued and, consequently, it could result in a higher premium. The formula used to obtain it is:

$$MtB = \frac{Market\ Capitalization}{Total\ Assets - Total\ Liabilities}$$

[4] When rumour date was not available it was assumed that no information was revealed, and the announced date was taken as a proxy.

[5] For observations that Zephyr did not provide this information directly, the following formula was used $Market\ Capitalization = \frac{Deal\ equity\ value/\%\ of\ acquisitions}{1 + Bid\ Premium}$.

DtM is used to access the target financial situation and it measures the total amount of outstanding company debt as a percentage of the firm's total assets. This ratio is an indicator of the company's leverage, which is defined as using debt to purchase assets. In the collection of the values, Amadeus was used when Zephyr did not provide the values.

Finally, *%ofaq* represents the percentage of acquisition.

3. Data

In order to estimate the models presented in the previous section, information regarding several deals was collected. Due to the different variables used in our models, not all the deals were considered in the estimation of all models. As a result, four subsamples are considered according to each model. The selection of the deals was made through several steps, where the 3 first ones are common to all the subsamples. The data, needed to estimate the models, was collected from Zephyr and Amadeus (financial data) and World Bank (nominal GDP data).

In this section, the common steps are presented firstly, followed by a description of each of the subsamples used.

3.1. Sample selection

The first step consisted in the choice of the 10 countries with the highest levels of FDI in the last 20 years. After analysing the data provided by the World Bank regarding FDI, the countries selected were: China, United States, India, Japan, Germany, Russian Federation, Brazil, United Kingdom, France and Indonesia.

The second step was to find out which and how many DTTs were signed by each one of these countries between 2000 and 2015. In order to do this, IBFD Tax Treaties database was used and 308 DTTs were selected. Each of these DTTs corresponds to a pair of countries that signed a treaty during the period considered.

Finally, we selected 45.687 cross-border acquisitions occurred from 1996 to 2017, within the countries of pairs chosen in the second step. Information regarding the deals was collected from Zephyr.

3.2. Number of deals before and after a dtt implementation [h1]

To test if the number of deals changes after the signature of a DTT [H1], the pairs of countries that had either only observations before or after the signature of a DTT were eliminated. This result in a subsample of 155 pairs of countries and 44.913 deals. Among the deals selected, 56% were made after the signature of a DTT.

The deals in our sample occurred from June 4, 1996, to March 30, 2017, and follow the temporal evolution illustrated in Figure 1.

Figure 1: Cross-borders deals before and after a DTT is implemented

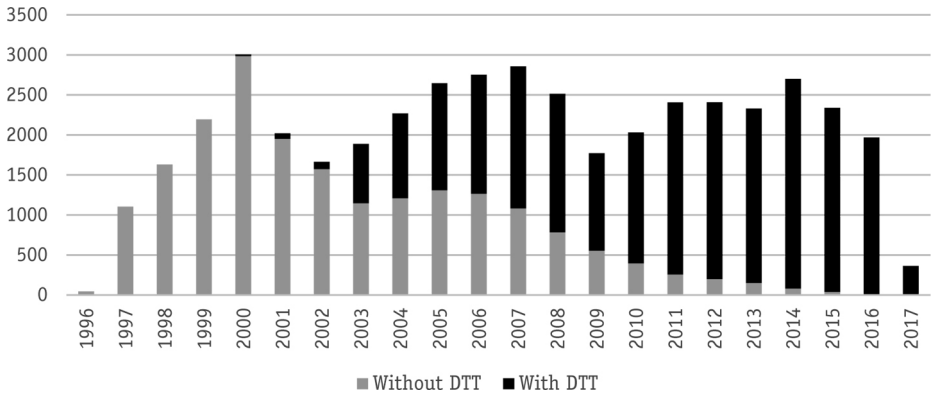
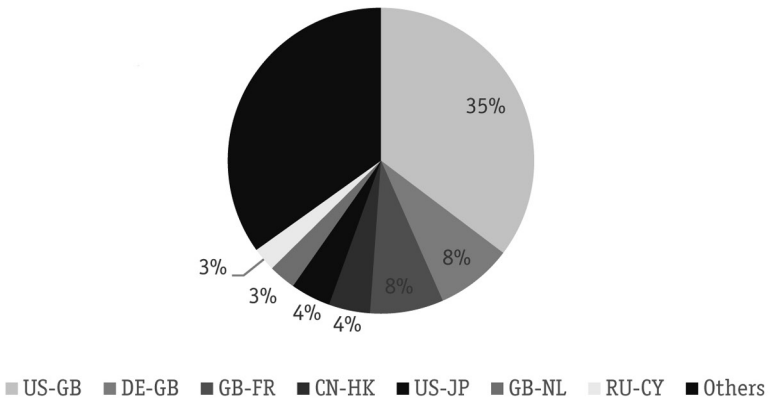


Figure 2: Distribution per pairs of countries of the sample



3.3. Deal value (H2)

The subsample used to test the hypothesis (2) only includes cross-border acquisitions with a known deal value, in a total of 20.982 deals made within 131 pairs of countries. 51% of the deals in this subsample happened after a DTT is implemented. Regarding the deal value, as shown in Table 1, the average (median) deal value made after the implementation of a DTT is around € 235 million (24 million), while before the signature of the DTT is around € 300 million (21 million).

Although the median is lower before the implementation of a DTT, the mean follows the opposite pattern. This discrepancy is due to the presence of 5 (outliers) deals that occurred before a DTT was implemented. By eliminating the 5 deals with the highest values, which corresponds to 0.02% of our sample, we verify that mean becomes higher in the group of deals made after a DTT is implemented. Nonparametric equality-of-medians test and t- test for testing mean differences were used.

Table 1: Descriptive statistics [H2]

	Deal Value (M€)				
	Mean	Median	S.D.	Min	Max
Total	267	22	2 454	0	204 730
Without DTT	300	21	3 296	0	204 730
With DTT	235	24	1 164	0	41 174
Difference	-64*	2***			

Note: ***, **, and * denote statistical significance at the 0.01, 0.0 5, and 0.1 levels, respectively.
Source: Own calculations considering information of Zephyr.

3.4. Acquisition premium [H3]

To test if companies are willing to pay higher premiums after a DTT is implemented, only the deals with the bid premium, at the announced date, known were kept. Additionally, the cross-border deals that involved acquisition of less than 5% of the shares were eliminated due to its lack of relevance. The final subsample comprises 137 deals where 74% were made after a DTT is implemented.

When comparing the average deal value of the previous subsample [H2] with the actual subsample [H3], it can be seen that the average deal value of the actual subsample [H3] – € 1.888 million – is higher than the average deal value of the previous subsample [H2] – € 267 million. This is not surprising since it is easier to obtain data for large deals.

Table 2 shows the average and medians of the bid premium at the rumoured date, market capitalization, market to book value and leverage (debt divided by market capitalization).

When testing for the differences on bid premiums between the deals that occurred after DTT is implemented and the deals made before the implementation of a DTT, we can see a positive difference (statistically significant) for both mean and median. These results suggest that after the DTT is implemented the size of deals increase. Nonparametric equality-of-medians test and t-test for testing mean differences were used.

Table 2 – Descriptive Statistics [H3]

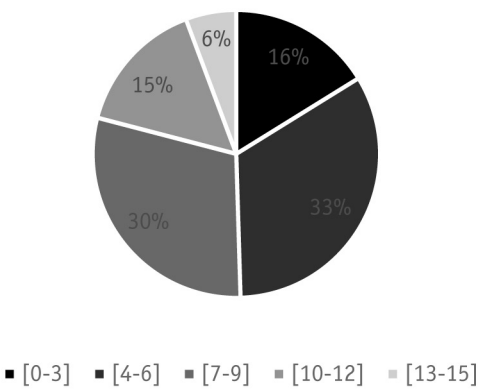
	Total		Without DTT		With DTT		Difference	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Deal Value (M€)	1 888	215	711	54	2 296	308	1 585	254***
Bid Premium (%)	54%	36%	30%	18%	62%	42%	32%	24%***
Market Capitalization (M€)	1 693	248	748	72	2 029	409	1 281**	337***
Market to Book Value	7.07	2.37	5.54	1.57	7.61	2.84	2.07	1.27*
Debt/Market Capitalization	3.43	0.30	2.25	0.51	3.84	0.24	1.59	-0.27**
% of acquisition	82%	100%	71%	79%	85%	100%	14%**	21%***

Note: ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.1 levels, respectively.
Source: Own calculations considering information of Zephyr.

3.5. Bid premium evolution after the DTT [H4]

Finally, in order to access if the bid premiums change immediately or gradually after the DTT implementation, only the 101 deals made after a DTT is implemented were kept from the previous subsample. Regarding these deals, they happened on average (median) 6.69 (7.00) years after the DTT is implemented and follow the distribution shown in Figure 3.

Figure 3: Number of years after the DTT is implemented



4. Results

In this section, the results are presented according to the hypotheses previously stated.

4.1. Impact of dtt implementation on the number of cross-border deals

To access if the number of deals increases after the signature of a DTT, model (1) is estimated as shown in Table 3. All regressions use panel data, where 155 different pairs of countries that had signed a DTT recently (2000-2015) are observed for the period between 1996 and 2017. The dependent variable is the number of deals made by each of the pairs considered for a given year and the independent variable of main interest is a dummy indicating whether there is a DTT in force. As the investors could advance the acquisition to be already installed in a country right after the DTT is implemented, alternatively we used a 'DTT-1' dummy that assumes the value 1 if the acquisition occurred after one year before the DTT is implemented, and 0 otherwise. The regressions differ regarding the controls used. In the first two, no controls are used, in the third and fourth, the $Sum\ GDP_{it}$ is used as the control variable and finally in the last two the natural logarithm of $Sum\ GDP_{it}$ is set as the control variables.

Table 3: Impact of DTTs on the Number of Deals

	I	II	III	IV	V	VI
Dependent Variable						
Nº of Deals						
Independent Variables						
DTT_dummy	0.38***		0.05		0.04	
	(0.04)		(0.04)		(0.04)	
DTT-1_dummy		0.08**		-0.14		0.02
		(0.32)		(0.04)		(0.04)
Sum GDP (M$)			15.17***	5.24***		
			(1.03)	(0.93)		
Ln (Sum GDP (M$))					0.78***	0.18***
					(0.05)	(0.05)
Constant	0.24***	1.36***	-0.39***	1.14***	2.89***	1.95***
	(0.05)	(0.00)	(0.07)	(0.06)	(0.18)	(0.16)
N	3 410	2.263	3 410	2 263	3 410	2 263
Wald chi2	97.16***	6.25**	320.83***	38.16***	324.19***	21.19***

Notes: This table reports the negative binomial regression used to test H1. The dependent variable is the annual number of cross-borders deals occurred from 1996 to 2017, between companies from countries that signed a DTT

from 2000 to 2015. The independent variables of main interest are the DTT_dummy and the DTT-1_dummy and these variables assume the value 1 when there is a DTT in force between the countries or after one year before the DTT is in force, respectively, and 0 otherwise. The control variable Sum GDP (and ln (Sum GDP)) represents the sum (and the log of the sum) of the nominal GDP per capita of both countries involved in the transaction in the year the transaction occurred and allow to control for macroeconomics condition that can have impact in the M&A activity. Panel data with fixed effects was used to estimate the model. *, **, and *** denotes statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses.

As shown in Table 3, there are no strong evidence that the implementation of a DTT increases the number of cross-border deals. The coefficients associated with *DTT_dummy* and *DTT-1_dummy* in the first two regressions are positive statistically significant (at a 5% level), which may suggest a positive relationship between the implementation of a DTT and the number of cross-border acquisitions. However, when controlled by the macroeconomic variable (sum of nominal GDP per capita of both countries) the relationship disappears as the coefficients associated with our variables of main interest (*DTT_dummy* and *DTT-1_dummy*) are not anymore statistically significant. Consequently, the results suggest that the implementation of a DTT does not influence the number of cross-border acquisitions. This contradicts di Giovanni (2005) who indicates the increasing number of DTTs as an explanatory factor of the increase of mergers and acquisition activity in the 90's.

4.2. Impact of dtts on the average deal value

Although our previous results suggest that the number of cross-border acquisitions made before and after the implementation of a DTT does not change, if the deals are different in size, the DTT may still effectively promote FDI. To test the hypothesis (H2), 6 new regressions are estimated using as the dependent variable the average deal value.

The results presented in Table 4, show that the coefficients associated with *DTT_dummy* and *DTT-1_dummy*, (our independent variables of main interest) are always positive and statistically significant, which strongly suggest that DTTs do have a positive impact on the average value of the deals practised. The average deal value is estimated to increase up to € 53 million after a DTT is implemented (or € 70 million if take into consideration the year before the DTT implementation).

Table 4: Impact of DTTs on the average deal value

	I	II	III	IV	V	VI
Dependent Variable						
Average Deal Value (M€)						
Independent Variables						
DTT_dummy	29.61*		52.57**		39.78*	
	(15.16)		(24.87)		(34.10)	
DTT-1_dummy		40.73***		70.30**		60.05*
		(15.56)		(35.83)		(31.12)
Sum GDP (M$)			-1 283.54	-1 644.33		
			(1293.00)	(1 527.59)		
Ln (Sum GDP) (Mln US$)					-20.95	-39.56
					(32.60)	(41.38)
Constant	88.32***	81.56***	139.85***	146.16**	14.48	-58.81
	(6.70)	(7.58)	(53.70)	(58.01)	(116.51)	(150.80)
N	2 882	2 882	2 882	2 882	2 882	2 882
F-test	3.81*	6.85***	2.93*	3.33**	1.94	3.49**

Notes: This table reports the ordinary least square regression used to test H2. The dependent variable is the average value of cross-borders deals occurred from 1996 to 2017, between companies from countries that signed a DTT from 2000 to 2015. The independent variable of main interest is the DTT_dummy, and this variable assumes the value 1 when there is a DTT in force between the countries and 0 otherwise. The control variable Sum GDP (and ln (Sum GDP)) represents the sum (and the log of the sum) of the nominal GDP per capita of both countries involved in the transaction in the year the transaction occurred and allow to control for macroeconomics condition that can have impact in the M&A activity. Panel data with fixed effects was used to estimate the model. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Reported standard errors in parentheses are robust to heteroskedasticity and are clustered at country-pair level.

Our results suggest that in spite of the fact that DTTs are not stimulating the number of deals made, these treaties are actually promoting FDI since larger deals are made after a DTT is implemented and so larger amount of money is being invested in cross-border deals between two countries after they signed a DTT. A possible explanation is that smaller deals are more likely to be justified by tax evasion purposes that tend to more difficult to occur after the implementation of a DTT since tax avoidance is one of the objectives of a DTT implementation. However, larger deals, those that are worth to stimulate, are the ones that tend to rely mostly on synergies and tend to occur more often after a DTT is implemented.

Our results are consistent with Barthel et al. (2010) who conclude that DTTs do lead to higher FDI stocks.

4.3. Impact of dtts on bid premiumS

To access if companies pay higher premiums after a DTT is implemented, model (3) was estimated.[6] All the regressions shown in table 5 have in common the dependent variable – *Bid Premium*. However, different explanatory variables are used.

All the regressions estimate a positive and statistically significant (for a significance level of at least 5%) relationship between the implementation of a DTT and the bid premium.

Table 5: Impact of DTTs on bid premium (1)

	I	II	III	IV	V	VI
Dependent Variable						
BP (+)						
Independent Variables						
DTT_dummy	15.83***	14.60**	19.65***	11.83**	15.67***	16.86**
	(5.62)	(5.96)	(5.63)	(5.51)	(5.62)	(6.91)
Ln (Sum GDP (M$))		4.39			1.08	2.85
		(7.00)			(7.09)	(7.15)
Ln (Mkt_cap (T€))			-3.91***		-3.74***	-3.56***
			(1.28)		(1.24)	(1.29)
MtB (+)			0.99		0.43	0.07
			(0.65)		(0.63)	(0.65)
DtM (+)			0.87		3.20	3.24
			(2.85)		(2.78)	(2.91)
%ofaq				27.61***	31.82***	34.12***
				(7.86)	(8.56)	(9.52)
Year fixed effects	Not Included	Not Included	Not Included	Not Included	Not Included	Included
Constant	29.16***	41.31***	70.35***	9.57	48.52	70.73
	(4.82)	(19.98)	(15.96)	(7.25)	(29.08)	(41.22)
N	137	137	137	137	137	137
R^2	0.06	0.06	0.12	0.14	0.22	0.33

Notes: This table reports the ordinary least square regression used to test H3. Bid premium at the rumour date of cross border deals occurred from 1996 to 2017, between companies from countries that signed a DTT from 2000 to 2015, is the dependent variable. The independent variables of main interest are the DTT_dummy and the DTT-1_dummy

[6] In order to estimate a more robust model, the variables were either winsorized (variables Bid Premium, MtB and DtM) or logarithmized (GDP and MarketCap variables). These transformations limit extreme values of the variables and so reduce the effect of possibly spurious outliers.

and these variables assume the value 1 when there is a DTT in force between the countries or after one year before the DTT is in force, respectively, and 0 otherwise. The control variable ln (Sum GDP) represents the log of the sum of the nominal GDP per capita of both countries involved in the transaction in the year the transaction occurred. Ln (Mkt_cap) is the log of the market capitalization of target company, MtB, represents the market to book value (also known as price to book ratio) of the target company, DtM measures the total amount of outstanding company debt as a percentage of the firm's total assets %ofaq represents the percentage of acquisition. All these variables are in the year before the deal and variables signalling with (+) are winsorized between 0.01 and 0.99 percentiles. Finally, Year_dummy controls for year fixed effects. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses.

The regression coefficients associated with *DTT_dummy* vary between 12 to 20 percentage points (p.p.). This strongly suggests that companies are willing to pay higher premiums after a DTT is implemented. High premiums as a result of a DTT are a strong indicator that tax treaties are being effective on the promotion of the FDI. It is possible to argue that due to the elimination of the double taxation burden, companies become willing to pay more to invest abroad.

Our results are also consistent with Huizinga et al. (2012) who provide empirical evidence showing that additional taxation resulting from international double taxation is fully capitalized into lower takeover bid premiums.

Finally, in order to understand if tax reductions and renegotiations of existing DTTs stimulate FDI gradually, only the deals occurred after a DTT is implemented are used in the regressions estimated in Table 6.

Table 6: Impact of DTTs on bid premiums (2)

	I	II	III	IV	V	VI
Dependent Variable						
BP (1)						
Independent Variables						
DTTage	0.51	0.03	0.71	-0.07	0.21	-0.60
	(0.83)	(0.92)	(0.85)	(0.83)	(0.89)	(1.17)
Ln (Sum GDP (M$))		11.03			2.24	12.73
		(9.30)			(10.53)	(11.96)
Ln (Mkt_cap (T€))			-3.71**		-3.82***	-3.9***
			(1.43)		(1.39)	(1.45)
MtB (1)			0.55		0.14	-0.40
			(0.77)		(0.76)	(0.79)
DtM (1)			-1.42		0.25	1.54
			(3.46)		(3.40)	(3.56)
%ofaq				27.33***	30.01**	3.56**
				(9.89)	(11.48)	(11.95)

Year fixed effects	Not Included	Not Included	Not Included	Not Included	Not Included	Included
Constant	41.55***	72.27***	86.24***	22.12***	71.50	79.03
	(6.22)	(26.63)	(19.01)	(9.26)	(39.62)	(79.03)
N	101	101	101	101	101	101
R^2	0.00	0.02	19.01	0.08	0.15	0.25

Notes: This table reports the ordinary least square regression used to test H4. Bid premium at the rumour date of cross border deals occurred from 1996 to 2017, between companies from countries that signed a DTT from 2000 to 2015 is the dependent variable. The independent variable of main interest is the DTTage and it represents the number of years elapsed between the effective date of the DTT and the transaction date. The control variable ln (Sum GDP) represents the log of the sum of the nominal GDP per capita of both countries involved in the transaction in the year the transaction occurred. Ln (Mkt_cap) is the log of the market capitalization of target company, MtB, represents the market to book value (also known as price to book ratio) of the target company, DtM measures the total amount of outstanding company debt as a percentage of the firm's total assets %ofaq represents the percentage of acquisition. All these variables are in the year before the deal and variables signalling with $^{(+)}$ are winsorized between 0.01 and 0.99 percentiles. Finally, Year_dummy controls for year fixed effects. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses.

Contrary to our expectations, our results do not show any evidence of a positive relationship between our explanatory variable of main interest – *DTTage* – and the bid premiums practised. Actually, the coefficients associated with *DTTage* are not statistically significant in all the regressions estimated. This might suggest that tax reductions and renegotiations do not have any gradual effect on FDI but instead do have an immediate effect which is consistent with Davies (2003a). The author concludes that renegotiations do not have a robust positive impact on FDI.

Nevertheless, our results lead us to conclude that companies are willing to make larger deals and to pay higher premiums after the signature of a DTT which indicates that DTTs are an effective tool on the stimulation of FDI.

5. Conclusions

The objective of this research was to study the impact of DTTs on FDI, more specifically on cross-border acquisitions. There are 3 main reasons for justifying the importance of studying the real impact of DTTs. First, DTTs are widely used all around the world and its importance has increased in the last decades. Second, the implementation of DTT forces countries to incur in various costs. And, finally, there is a lack of agreement among the literature regarding their effectiveness in promoting FDI.

The current study starts by replicating previous studies that analysed the impact of DTTs on the FDI flows, using both the number and the average value of deals practised between two countries as proxies for FDI flows. Moreover, it introduces a new way to measure the impact of DTTs – through takeover bid premiums. Jointly considering double taxation as a driver of lower takeover bid premiums and DTTs as a powerful tool in solving the double

taxation problem, we expected to find positive impact of DTTs on the premiums practised. Furthermore, assuming FDI is stimulated by sequential tax reductions either specified in the treaties or achieved through renegotiations, we aimed to find a positive and gradual relationship between the treaties' age and the premiums practised.

When replicating previous studies, our results suggest that the implementation of a DTT does not change the number of cross-border deals within companies from the countries that signed the DTT. However, our results suggest that deals carried out after the implementation of a DTT is on average larger than those carried out before. Consequently, our findings suggest that DTTs effectively promote FDI. Lower value deals may be (at least partially) justified by tax evasion purposes are more difficult to occur after the implementation of a DTT. However, larger value deals, those that are worth to stimulate, occur more often.

Regarding the bid premiums, the results suggest that companies are willing to pay a higher premium – up to 20 p.p. – after a DTT is implemented. This reinforces the idea that DTTs are effectively promoting FDI. Nevertheless, we conclude that the years elapsed between the effective date of a DTT and the transaction date do not affect the bid premiums.

References

Baker, P. L. (2014) An analysis of double taxation treaties and their effect on foreign direct investment. *International Journal of the Economics of Business*, *21*(3), 341-377.

Barthel, F.; Busse, M.; Neumayer, E. (2010) The impact of double taxation treaties on foreign direct investment: evidence from large dyadic panel data. *Contemporary Econom ic Policy*, *28*(3), 366-377.

Blonigen, B. A.; Davies, R. B. (2002) Do bilateral tax treaties promote Foreign Direct Investment? National Bureau of Economic Research (NBER), WP no. 8834.

Blonigen, B. A.; Davies, R. B. (2004) The effects of bilateral tax treaties on U.S. FDI activity. *International Tax and Public Finance*, *11*, 601-622.

Coupé, T.; Orlova, I.; Skiba, A. (2008) The effect of tax and investment treaties on bilateral FDI flows to transition countries. 9th Annual Global Development Conference, Brisbane.

Davies, R. B. (2003) Tax Treaties, Renegotiations, and Foreign Direct Investment. *Economic Analysis and Policy*, *33*(2), 251-273.

di Giovanni, J. (2005) What drives capital flows? The case of cross-border M&A activity and financial deepening. *Journal of International Economics*, *65*, 127-149.

Egger, P.; Pfaffermayr, M.; Larch, M.; Winner, H. (2006) The impact of bilateral investment treaties on foreign direct investment. *Canadian Journal of Economics*, *39*(3), 901-931.

Huizinga, H.; Voget, J.; Wagner, W. (2012) Who bears the burden of international taxation? Evidence from cross-border M&As. *Journal of International Economics*, *88*, 186-197.

Louie, H. J.; Rousslang, D. J. (2008) Host-country governance , tax treaties and US direct investment abroad. *International Tax and Public Finance*, *15*, 256-273.

Neumayer, E. (2007) Do Double Taxation Treaties Increase Foreign Direct Investment to Developing Countries? *Journal of Development Studies*, *43*(8), 1501-1519.

OECD (2014) *Model Tax Convention on Income and on Capital*. Paris, OECD Publishing.

Pedro Lopes Ferreira
Faculdade de Economia
Universidade de Coimbra
e CEISUC
pedrof@fe.uc.pt

DOI: https://doi.org/10.14195/2183-203X_48_3

A morte saiu à rua

Death stepped out on the street

Pedro Lopes Ferreira

Received for publication: November 2, 2018
Revision accepted for publication: February 1, 2019

ABSTRACT
The main purpose of this paper is to demonstrate the evidence that already exists in the Portuguese context of evaluation of health states worse than death. The aim is to also make a personal reflection on the current state of the literature on the measurement of these health states and list some existing problems of methodological nature. This type of health states, in fact, exist and it is known and understood by patients and by the general population. Several are the techniques used to determine the cardinal preferences for health states worse than death and its use is usual in cost-effectiveness or cost-utility assessments. However, the measurement of preferences for health states worse than death is still at an early stage of development, keeping the need of elicitation techniques probably more valid and reliable. It is essential to deepen the discussion on the concept of health states worse than death.
Keywords: Quality of life; patient preference; health status; death.

JEL Classification: J17; I12; I18

RESUMO
O objetivo principal deste artigo é demonstrar a evidência que já existe no contexto português da avaliação de estados de saúde piores que morte. Pretende-se também fazer uma reflexão pessoal sobre o estado atual da literatura sobre a medição destes estados de saúde e elencar alguns problemas existentes de natureza metodológica. Este tipo de estados de saúde existe de facto e é apercebido e compreendido pelos doentes e pela população em geral. Várias são as técnicas utilizadas para a determinação das preferências cardinais para estados de saúde piores que morte e a sua utilização é hoje corrente em avaliações custo-efetividade ou custo-utilidade.No entanto, a medição de preferências para estados de saúde piores que morte está ainda numa fase inicial de desenvolvimento, mantendo-se a necessidade de técnicas de elicitação provavelmente mais válidas e fiáveis. É essencial aprofundar a discussão sobre o conceito de estado de saúde pior que morte.
Palavras-chave: Qualidade de vida; preferência do paciente; nível de saúde; morte

A morte saiu à rua num dia assim
Naquele lugar sem nome pra qualquer fim
José Afonso, 1972

1. Introdução

Neste texto é abordado um tema que tem preocupado alguns autores nos últimos anos: o conceito e a medição de estados de saúde piores que morte. O que quer isto dizer? Será que faz sentido? Não será apenas um daqueles conceitos interessantes que os economistas da saúde criaram, que não são facilmente compreendidos nem têm reflexo no dia-a-dia dos cidadãos? Ou, pelo contrário, será que só recentemente os economistas da saúde foram finalmente capazes de operacionalizar este conceito tão natural na mente dos cidadãos, principalmente quando passam pela experiência de terem um familiar ou amigo próximo a sofrer, de uma forma que consideram demasiada, injusta e desumana? E, já agora, será que estamos a medir corretamente este estado de saúde?

É difícil reconstituir o que vai na cabeça de uma pessoa quando é confrontada com a morte e com decisões sobre a morte. Há mesmo quem defenda que o medo da morte é psicologicamente tão potente que poderá justificar respostas menos racionais e que é difícil acreditar que as pessoas, face a situações piores que morte, possam desenvolver preferências racionalmente perversas (De Charro et al., 2005). Outros, como Torrance e Feeney (1989) reconheceram há quase 30 anos a existência de estados de saúde piores que morte, propondo que lhes não fossem atribuídos valores negativos, e sim que todos tivessem o valor 0, igual ao da morte.

Ao medir as preferências cardinais dos indivíduos para diferentes estados de saúde, vários investigadores têm encontrado condições crónicas que podem apelidar de piores que morte (Ferreira et al., 2014a). Estas descobertas não são isoladas e verificam-se com a aplicação de vários instrumentos de medição.

O objetivo principal deste artigo é demonstrar a evidência que já existe no contexto português da avaliação de estados de saúde piores que morte. Além disso, pretende-se fazer uma reflexão pessoal sobre o estado atual da literatura sobre a medição destes estados de saúde e elencar alguns problemas existentes de natureza metodológica.

Este texto começa por abordar a evidência destes estados de saúde piores que morte e de que forma as preferências cardinais dos indivíduos têm sido medidas. Por fim, partilhar-se-ão algumas reflexões e serão lançadas algumas discussões sobre este tema.

2. Estudos Empíricos

Vários estudos têm sido conduzidos que demonstram a evidência de avaliação de estados de saúde piores que morte. Começaremos por brevemente apresentar dois estudos pioneiros realizados nas últimas décadas do século xx e, de seguida, outros três realizados mais recentemente em contexto português na população em geral, em doentes oncológicos, e em jovens.

Estudos pioneiros

Tradicionalmente, os indicadores de estado de saúde assumiam a morte como o pior estado, num contínuo de bem-estar. George Torrance, Rachel Rosser e Paul Kind foram, de facto, os primeiros em 1978 a referir medições de preferência em estados piores que morte (Torrance, 1984). Numa pequena amostra de 70 ingleses, usando a técnica de estimação da magnitude (razão entre medições) – na altura muito em moda – e analisando 29 estados crónicos de saúde, identificaram dois considerados como piores que morte. Foram eles a "inconsciência crónica", considerada duas vezes pior que morte, e "estar confinado a uma cama, com dores extremas e com prescrição de heroína", estado de saúde considerado entre duas vezes a uma vez e meia pior que morte (Kind e Rosser, 1979; Rosser e Kind, 1978).

Mais tarde, em 1981, num outro estudo realizado em Boston com 338 futuros pais numa consulta de aconselhamento genético, foi-lhes perguntado qual a probabilidade que aceitavam como sendo suficiente para uma decisão de aborto (caso tivesse sido detetada durante a gravidez uma grave deformação no feto), face à possibilidade de virem a ter esse filho com aquela grave deformidade. Pouco mais de um terço dos casais aceitariam tomar a decisão de interromper a gravidez com probabilidades de deficiência inferiores a 10% (Pauker et al., 1981). De notar que esta forma de perguntar pode ser encarada como uma implementação do jogo padrão (Torrance, 1982), definido adiante no texto, tendo os autores conseguido atribuir um valor, provavelmente não generalizável, de utilidade de uma criança deformada em relação a não ter essa criança ou ter uma criança saudável.

Alguns autores se seguiram, sugerindo que a morte nem sempre é olhada como o resultado mais indesejável, e que alguns estados de saúde podem mesmo ser encarados como piores que morte. Outros recomendam a inclusão sistemática de estados piores que morte de forma a descrever um leque mais vasto de valores de preferência (Patrick et al., 1994).

Estudo de preferências de cidadãos em geral

Tentando contextualizar este conceito de estado de saúde pior que morte, no âmbito de um estudo para determinar o sistema português de valores do EQ-5D-5L, questionou-se uma amostra de 1.451 indivíduos sobre opiniões relativas à vida e à morte. Esta amostra, representativa da população portuguesa em termos de idade e género, englobou áreas urbanas e rurais. Perguntou-se, em primeiro lugar, o que consideravam mais importante, se (i) uma vida saudável, independentemente do tempo de vida; (ii) uma vida longa, independentemente da qualidade de vida; ou (iii) eram incapazes de responder, porque dependia [da situação].

Para tentar compreender um pouco melhor as valorações que estavam a ser fornecidas pelos respondentes, perguntou-se também qual das seguintes frases descrevia melhor o que acreditavam que acontecia após a morte; (i) uma vida maravilhosa após a morte; (ii) há vida após a morte, mas não estou certo se serei feliz nessa vida; (iii) não sei se há vida após a morte; (iv) não há vida após a morte; ou (v) recuso-me a responder.

Por fim, perguntámos, se tivessem uma doença grave, o que mais os preocupava: (i) não conseguir continuar a gozar a vida; (ii) não conseguir realizar os planos para o futuro; (iii) sofrer com a doença; (iv) não ter capacidades para trabalhar; (v) não ter dinheiro para

pagar as dívidas com a saúde; (vi) não conseguir mais tomar conta da família; (vii) ser um fardo para os outros; ou (viii) não haver ninguém para cuidar de si. As respostas estão apresentadas na tabela 1.

Tabela 1: Preferências dos portugueses perante a vida e a morte (N=1451)

	N	%
O que é mais importante		
Uma vida saudável, independentemente do tempo de vida	960	66,1%
Uma vida longa, independentemente da qualidade de vida	85	5,9%
Não consigo responder. Depende.	406	28,0%
O que acredita que acontece após a morte		
Terei uma vida maravilhosa após a morte	172	11,9%
Há vida após a morte, mas não estou certo/a se serei feliz nessa vida	205	14,1%
Não sei se há vida após a morte	604	41,6%
Não há vida após a morte	303	20,9%
Recuso-me a responder	167	11,5%
O que o/a preocupa se tivesse uma doença grave		
Não conseguir continuar a gozar a vida	114	7,9%
Não conseguir realizar os meus planos para o futuro	84	5,8%
Sofrer com a minha doença	353	24.3%
Não ter capacidades para trabalhar	26	1,8%
Não ter dinheiro para pagar as minhas dívidas com a saúde	45	3,1%
Não conseguir mais tomar conta da minha família	143	9,9%
Ser um fardo para os outros	586	40,3%
Não haver ninguém para cuidar de mim	100	6,9%

Pouco mais de um quarto (28%) dos inquiridos não conseguiram ou quiseram responder à primeira pergunta. No entanto, de entre os restantes 1.045 respondentes – mantendo-se a representatividade – cerca de 92% preferiram a primeira opção, o que parece ser extremamente importante para permitir influenciar muitos dos procedimentos de tomada de decisão clínica ou mesmo de política de saúde. Ter em conta o equilíbrio entre quantidade e qualidade de vida nestas decisões é também uma forma de respeitar os valores dos principais interessados, os doentes envolvidos.

Por outro lado, ao contrário do que se esperava, apenas 167 (11,5%) dos respondentes consideraram que esta pergunta não lhes deveria ter sido feita. Dos restantes 1.284, e apesar do forte sentimento religioso que tradicionalmente é associado à população portuguesa, quase metade dos inquiridos (47,0%) afirmou que não sabia se há vida após a morte e 23% foram mesmo perentórios a dizer que, para eles, não há vida após a morte.

A terminar, para além do receio de sofrer ter sido selecionado por cerca de um quarto das pessoas, destacam-se dois outros aspetos. O primeiro, é o receio de a sua doença ser um fardo para os outros, o que evidencia a sua preocupação para com os que os rodeiam. Num outro estudo sobre cuidados paliativos na Europa, resultados semelhantes foram encontrados em países como a Espanha e a Itália, o que parece definir uma característica dos cidadãos dos países de origem latina (Bausewein et al., 2013).

O segundo aspeto a salientar são os escassos 3,1% dos respondentes que afirmam temer não terem dinheiro suficiente para pagarem as dívidas com a saúde, resultado revelador da confiança e da proteção social e financeira que o Serviço Nacional de Saúde português dá às pessoas, aliás na sequência das orientações da própria OMS (WHO, 2000).

Estudo em doentes oncológicos

Recentemente, foi também encontrada evidência empírica sobre a forma como as pessoas relacionam a morte com estados de saúde. Num estudo intitulado 'valor em oncologia' realizado no norte de Portugal em mulheres com cancro de mama, aplicando o EQ-5D-3L foram obtidos valores de utilidade correspondentes a avaliações piores que morte, em especial quando o tempo de duração do cancro se situava entre os nove meses e os quatro anos e estas doentes estavam a começar a sentir os efeitos de formas mais violentas de tratamento, nomeadamente sessões de quimioterapia (Sequeira et al., 2015).

Estudo de preferências de jovens

Por outro lado, numa pequena experiência de início de ano, numa unidade curricular de Economia da Saúde perguntou-se a alunos do 3.º ano da licenciatura em Economia o que consideravam poder ser encarado como pior que morte. A razão de ser deste estudo teve essencialmente a ver com o facto de se recolher opiniões de indivíduos jovens, na sua generalidade saudáveis e com perspetivas de (e para o) futuro potencialmente diferentes de estudos anteriores em que foram inquiridos indivíduos da população geral, com outra distribuição etária, ou doentes com uma esperança de vida mais limitada.

Tratou-se de uma amostra por conveniência de 67 alunos com uma média etária de 21 anos e 56,7% mulheres. Os resultados são apresentados na tabela 2.

Tabela 2: Preferências de jovens perante a vida e a morte (N=67)

	N	%
O que é para si um sofrimento considerado demasiado		
Dependente, incapaz de realizar tarefas básicas do dia-a-dia, em estado vegetativo	23	37,7%
Demasiada dor	17	27,9%
Incapaz mentalmente de decidir por si ou de realizar os seus sonhos, sem ânimo de viver, sem qualidade de vida	12	19,7%
Doença muito grave ou terminal, sem cura	9	14,7%
Situações de saúde piores que morte		
Incapaz de ver, comunicar ou decidir e estar em dependência física sem perda de consciência, em dependência mental ou perda de autonomia	20	33,3%
Estado vegetativo, estado terminal ou morte cerebral	16	26,7%
Sofrimento exagerado e constante, intolerável e dor extrema sem cura possível, nem alívio	15	25,0%
Quando não se é feliz, não se faz o que se gostaria de fazer, em solidão extrema, fazer sofrer os outros	9	15,0%

O que é mais importante		
Uma vida saudável, independentemente do tempo de vida	57	85,1%
Uma vida longa, independentemente da qualidade de vida	1	1,5%
Não consigo responder. Depende.	9	13,4%
O que acredita que acontece após a morte		
Terei uma vida maravilhosa após a morte	3	4,5%
Há vida após a morte, mas não estou certo/a se serei feliz nessa vida	1	1,5%
Não sei se há vida após a morte	37	55,2%
Não há vida após a morte	19	28,4%
Recuso-me a responder.	7	10,4%
O que o/a preocupa se tivesse uma doença grave		
Não conseguir continuar a gozar a vida	8	11,9%
Não conseguir realizar os meus planos para o futuro	14	20,9%
Sofrer com a minha doença	25	37,3%
Não ter capacidades para trabalhar	1	1,5%
Não ter dinheiro para pagar as minhas dívidas com a saúde	0	0,0%
Não conseguir mais tomar conta da minha família	3	4,5%
Ser um fardo para os outros	16	23,9%
Não haver ninguém para cuidar de mim	0	0,0%

Quando questionados em que situações consideravam que o sofrimento de uma pessoa poderia ser considerado demasiado, estes jovens afirmaram que, para eles, o maior sofrimento reside na falta de funcionalidade física ou mental ou na dor extrema. Apenas três alunos deixaram esta pergunta em branco e outros três disseram que não sabiam responder. Indagou-se também que situações em vida consideravam como piores que morte e verificou-se então que, para além da dependência física ou mental, antes da dor, surgem as situações clínicas de estado vegetativo, terminal ou de morte cerebral. Aqui, já cinco não responderam e outros cinco declararam são saber responder.

Por fim, fizeram-se as mesmas perguntas de opinião anteriormente questionadas à amostra representativa da população portuguesa em geral e, comparando com os resultados obtidos nessa outra amostra, estes jovens reforçaram a opinião de que o importante é termos uma vida saudável, independentemente do tempo de vida. Relativamente ao que acreditam que sucede após a morte, reforçou-se o desconhecimento sobre se há vida após a morte, mas diminuiu drasticamente a convicção de que haverá uma vida após a morte. Se tivessem uma doença grave, para estes estudantes o receio de ser um fardo para os outros mantém-se, mas é ultrapassado pelo receio de sofrer com a doença e quase iguala a preocupação de não conseguir realizar os seus planos para o futuro.

3. Medição

Atendendo agora à própria medição dos estados de saúde, vários métodos de valoração têm sido propostos, incluindo curvas de indiferença e medições de utilidade (Culyer et al., 1971; Grossman, 1972). Neste contexto, a valoração de diversos estados de saúde pelos cidadãos tem demonstrado ser de uma enorme relevância para a formulação de políticas de saúde e na avaliação clínica e económica dos cuidados de saúde. E também são várias as formas de elicitar estados de saúde piores que morte. Dependendo da

técnica utilizada, todas elas, no entanto, são apresentadas como alterações ou extensões de técnicas criadas para medir estados de saúde melhores que morte.

Os métodos de ordenação, as escalas categóricas visuais analógicas (VAS, do inglês *visual analogue scale*) entre 0 e 100, a equivalência em tempo (TTO, do inglês *time-tradeoff*) e o jogo padrão (SG, do inglês *standard gamble*) têm sido técnicas adaptadas para quantificar estados de saúde piores que morte (Tabela 3). De uma forma geral, são usados estados de saúde hipotéticos baseados num número pequeno de atributos e elicitadas preferências para esses estados, normalmente entre um padrão de saúde perfeita ao qual é atribuído o valor 1 e a morte à qual é atribuído o valor 0.

Tabela 3: Técnicas de quantificação de estados de saúde

Característica	Método		
	Escala Categórica VAS	Equivalência em tempo TTO	Jogo padrão SG
Duração no estado de saúde	Implícito: *Para o resto da sua vida*	Explícito: *Número de anos*	Implícito: *Para o resto da sua vida*
Risco	Resultado: certo	Resultado: certo	Resultado: incerto
Juízo	Direto: Atribuição de um valor específico na escala	Indireto: Escolha entre duas opções	Indireto: Escolha entre duas opções

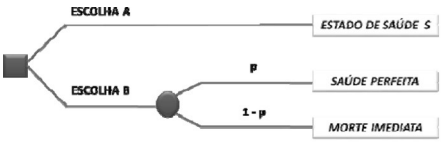
Fonte: Adaptado de Patrick *et al.* (1994).

Na implementação destas técnicas, a duração ou o tempo passado num determinado estado de saúde pode estar explícita (como na equivalência em tempo) ou implícita (como no jogo padrão quando se afirma que estará num determinado estado de saúde para o resto da sua vida).

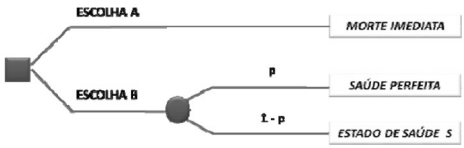
Nos métodos que incorporam o risco é solicitado aos indivíduos que escolham entre uma opção incerta e uma opção certa; os que não incorporam risco requerem normalmente avaliações de resultados certos. É o caso do jogo padrão (Figura 1A), em que o respondente tem uma escolha entre uma alternativa A com o resultado certo do estado de saúde em análise, i.e., viver o resto da sua vida neste estado de saúde S, e uma alternativa B com dois possíveis resultados, um melhor do que o estado de saúde certo, normalmente denominado 'saúde perfeita', com uma probabilidade p de ocorrência, e outro bem pior, designado por 'morte' (Drummond *et al.*, 2005). No processo de elicitação, a probabilidade p varia até que o respondente esteja numa situação de indiferença na escolha entre as duas alternativas. O valor final de probabilidade é exatamente igual ao valor de utilidade associado a este estado de saúde, U(S)=p.

Figura 1: Jogo padrão

A: Estados de saúde melhores que morte



B: Estados de saúde piores que morte

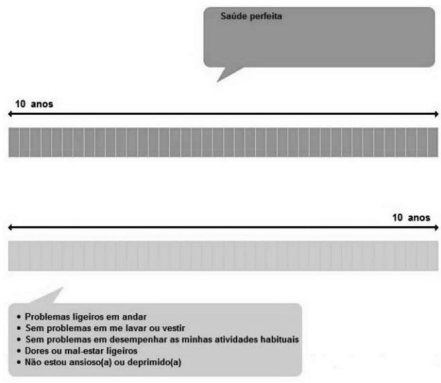


Para situações piores que morte (Figura 1B), no entanto, a alternativa B passa a ser entre 'saúde perfeita' e o estado de saúde considerado pior que morte. E esta torna-se o resultado certo correspondente à alternativa A (Ferreira et al., 2010).

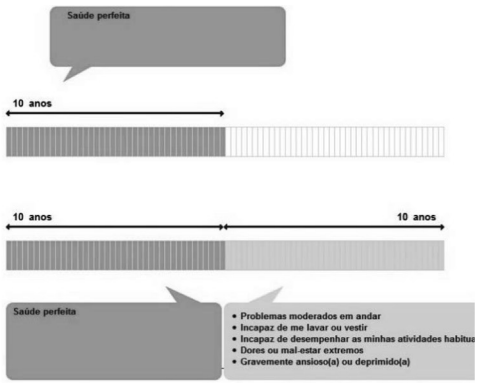
Na equivalência em tempo, desenvolvida por Torrance e colegas (1982), são apresentadas duas alternativas certas entre viver um número x certo de anos em saúde perfeita e viver um número fixo t de anos (e.g., t=10 anos) num estado de saúde S a valorar (Figura 2A). Varia-se o número de anos da vida A até o respondente se sentir indiferente perante as alternativas. O valor final de utilidade para um estado de saúde inferior à saúde perfeita é calculado e dado por U(S)=x/t, adotado pelo Grupo EuroQol no estudo Measurement and Valuation of Health (Dolan, 1997).

Figura 2: Equivalência em tempo

A: Estados de saúde melhores que morte



B: Estados de saúde piores que morte



No entanto, quando o respondente prefere viver 0 anos com a doença (i.e., morrer imediatamente), estamos perante uma situação considerada pior do que morte e, assim, tal como no jogo padrão, para estados de saúde piores que morte, a vida A passa a ser a morte imediata e a vida B passa a representar um número variável (t-x) de anos em saúde perfeita seguidos dos restantes anos de vida no estado de saúde descrito, i.e., é pedido ao respondente que equilibre mais anos em pior saúde, com menos anos em melhor saúde. O valor para S é dado por U(S)=-x/(t-x).

Ou, utilizando a versão 'lead time', a vida A corresponde a viver apenas x anos em saúde perfeita comparada com a alternativa de viver t=10 anos em saúde perfeita seguidos de igual tempo no estado de saúde em análise (Ferreira *et al.*, 2014b; Robinson e Spencer, 2006), de modo a evitar valores negativos extremos (Devlin *et al.*, 2011) (Figura 2B). O valor final para S é dado por U(S)=(x-t)/t.

É fácil de verificar que estas técnicas produzem valores que podem ser interpretados como preferências numa escala intervalar.

De salientar que no estudo de preferências de cidadãos em geral atrás referido, e porque por vezes este problema se pode levantar, avaliou-se também o esforço cognitivo envolvido na obtenção das utilidades para determinar se os respondentes tinham compreendido perfeitamente as tarefas por que tiveram de passar para fornecer as suas preferências sobre estados de saúde. E isto é particularmente importante para estados de saúde próximos de ou piores que morte, um estado desconhecido e completamente irreversível. Com esta preocupação, avaliou-se concomitantemente a possível sobrecarga cognitiva das perguntas e procedimentos utilizados.

Assim, perguntou-se aos 1.451 respondentes se tinham tido facilidade em perceber as perguntas formuladas e se lhes tinha sido fácil distinguir e decidir entre vidas mais ou menos semelhantes. Numa escala de concordância de 5 pontos, quase 90% acharam que lhes tinha sido fácil compreender as questões colocadas, quase 75% acharam fácil distinguir entre vidas, embora quase metade tenha sentido alguma dificuldade ao decidir entre vidas semelhantes, o que aliás parece natural. Esta experiência é confirmada pela muita literatura desta área da economia da saúde, em que se torna patente que adultos são perfeitamente capazes de compreender e executar as tarefas que lhes são pedidas de elicitação de utilidades, mesmo que extensas no tempo. A percentagem de inconsistências, por outro lado, é normalmente diminuta.

4. Questões metodológicas

A terminar, há ainda algumas questões metodológicas que merecem ser salientadas e discutidas. Comecemos por duas primeiras considerações mais gerais.

Uma primeira questão que nos deve importunar é se a gravidade dos estados de saúde pode ou não ser quantificada num índice, questão aliás já muito antiga e levantada por Paul Kind e Rachel Rosser nos anos 80 do século passado (Kind e Rosser, 1988; Rosser, 1983), secundando a classificação proposta por Stevens (1946; 1951) de medição em vários níveis, incluindo o descritivo, o ordinal e o quantitativo. Mais tarde, Khanna e Tsevat (2007) apresentaram uma outra classificação baseada em métodos diretos e métodos indiretos de utilidade.

De qualquer forma, trata-se de um passo essencial para uma análise custo-efetividade ou custo-utilidade de intervenções em saúde. Os avanços tecnológicos alcançados e os critérios de qualidade existentes associados à quantificação em saúde, permitem uma grande dose de segurança quando se utilizam estes índices.

Uma outra questão que se pode levantar é qual o significado do valor de saúde zero. Todos nós reconhecemos a vantagem e a utilidade de existir um índice de estado de saúde que nos permita determinar o indicador QALY (do inglês Quality Adjuted Life Years) a variar de 0 (morte) a 1 (saúde perfeita), permitindo valores negativos. De acordo com este índice, um ano em saúde perfeita corresponde a 1 QALY, isto é, o máximo de saúde que uma pessoa pode ter num ano. Logo aqui há quem considere que o valor zero (estar morto ou morte) não devia ser considerado como estado de saúde, exceto eventualmente para zombies ou vampiros, tal como referido de uma forma algo provocatória por David Parkin no Academic Health Economists' Blog (Dr. Panik, 2011). Para os defensores desta tese, estar morto é um estado, embora não possa ser considerado um estado de saúde.

Olhando agora com um pouco mais de detalhe para as técnicas e para os resultados, podemos acrescentar mais alguns pontos de reflexão. Um primeiro poderá questionar: será que as utilidades de estados de saúde piores que morte devem ser medidas com técnicas que são derivadas de outras concebidas para medir utilidade de estados de saúde melhores que morte? Será que isto faz sentido, tendo em conta a existência de contextos e de enquadramentos completamente diferentes (Torrance, 1986)? Não seria necessária outra abordagem? Até porque já é sabido há algum tempo que a aversão ao risco apresentada pelos indivíduos em situações melhores que morte se transforma, muitas vezes, em propensão para o risco em situações piores que morte (Gafni e Torrance, 1984). Será que o 'lead time' minimiza este possível enviesamento?

E, acrescentando, será que faz sentido tentarmos obter um valor que represente a preferência cardinal para estados de saúde piores que morte? Não seria melhor mantermo-nos com valores qualitativos representando a gravidade da situação?

E a qualidade da medição? Poucos são os estudos que apontam para a evidência da fiabilidade e da validade destas medições de utilidade de estados de saúde piores que morte. Como referido anteriormente, quando se medem estados de saúde melhores que morte, esta medição é realizada numa escala padronizada que vai de 0 (morte) a 1,0 (saúde perfeita). No entanto, para estados de saúde piores que morte não existe limite inferior.

Ilustrando, nas versões portuguesas do SF-6D e do EQ-5D-5L os valores mínimos obtidos não são muito grandes em valor absoluto (u[64565]=-0,50 e u[55555]=-0,53, respetivamente), provavelmente devido a terem sido submetidos a transformações lineares para uma escala simétrica. De facto, a transformação

$$\begin{aligned} U > 0 &\quad \Rightarrow \quad U' = U \\ U < 0 &\quad \Rightarrow \quad U' = U \,/\, (1-U) \end{aligned}$$

em que U corresponde ao valor de utilidade obtido diretamente pela técnica, produz valores U' num intervalo -1 a +1 com extremos igualmente distantes da morte, numa distribuição muito menos enviesada do que a da variável original U.

Mas, se não houver anos de vida para viver, continuará a fazer sentido determinar QALYs? E se também questionarmos o limite inferior –1? Por exemplo, a frase 'mais valia nunca ter nascido' pode aplicar-se a uma pessoa que passa por tortura ou outro tipo de sofrimento extremo. Não existe ainda base teórica para impor um limite ao nível de desutilidade associada a estes casos extremos (Devlin et al., 2011).

E isto para não falar na agregação dos valores obtidos melhores e piores que morte. Será que a determinação das médias tem validade? Provavelmente são necessários mais estudos para abordar esta questão.

Por outro lado, qual o tempo máximo para que uma situação grave, mas ainda encarada como melhor que morte, passe a ser considerada pior que morte? Segundo alguns autores, à medida que aumenta a duração de um estado de saúde disfuncional, aumenta a probabilidade de se encontrarem estados de saúde piores que morte (Sutherland et al., 1982). Isto leva a um outro conceito – o conceito de duração – não muito explorado pelos economistas da saúde. Este conceito de duração é usado na elicitação do TTO mas subsequentemente é ignorado quando esses valores são utilizados. Os próprios QALYs parece não se preocuparem muito com esta questão.

Um último aspeto de reflexão, de certo modo relacionado com o anterior, é o facto de algumas técnicas apresentarem maior tendência para valores negativos do que outras. Será isto dependente das amostras utilizadas para a determinação de sistema de valores? Talvez não, pois pelo lado positivo há maior coincidência. Será da própria técnica? Será da forma como a técnica lida com estados de saúde piores que morte?

Estas e outras são reflexões que certamente nos permitem garantir uma melhor medição de estados de saúde melhores e piores que morte.

5. Conclusões

Neste artigo de reflexão, baseado em cinco estudos, ficou demonstrado que é possível encontrar-se evidência de avaliação de estados de saúde piores que morte, particularmente no contexto português. Foi também possível descrever as principais técnicas de elicitação deste tipo de estados de saúde e fez-se uma reflexão mais pessoal relativa à sua medição, elencando-se alguns problemas de natureza metodológica existentes.

Por fim, fez-se uma referência à aplicação de estados de saúde piores que morte em avaliações económicas, como parte da mala de ferramentas do decisor clínico e do decisor de políticas públicas.

Na sua base está o conceito de valor enquadrado por Alan Williams (1987) como um dos aspetos de investigação e aplicação da economia da saúde. A questão de como valorar a saúde é central para a prática dos economistas da saúde, pois é difícil ver a utilidade do seu papel se apenas estiverem a medir custos (Kind, 2014).

REFERÊNCIAS

Bausewein, C.; Calanzani, N.; Daveson, B.A.; Simon, S.T.; Ferreira, P.L.; Higginson, I.H.; Bechinger--English, D.; Deliens, L.; Gysels, M.; Toscani, F.; Ceulemans, L.; Harding, R.; Gomes, B. (2013) 'Burden to others' as a public concern in advanced cancer: A comparative survey in seven European countries, BMC Cancer, 13, 105.

Culyer, A.J.; Lavers, R.J, Williams, A. (1971) Social indicators: Health, Social Trends, 2, 31-42.

De Charro, F.; Busschbach. J, Essink-Bot, M.-L.; van Hout, B.; Krabbe P. (2005) Some considerations concerning negative values for EQ-5D health states, in: P. Kind, R. Brooks, R. Rabin R (eds.), EQ-5D Concepts and Methods: A Development History, Dordrecht, Springer, 171-179.

Devlin, N.J.; Tsuchiya, A.; Buckingham, K.; Tilling, C. (2011) A uniform time trade off method for states better and worse than dead: Feasibility study of the 'lead time' approach, Health Economics, 20(3), 348-61.

Dolan, P. (1997) Modelling valuations for Euroqol health states, Medical Care, 35(11), 1095-108.

Dr. Panik. (2011, jul 13). What does a health value of zero mean? [Blog post]. Retrieved from https://aheblog.com/2011/07/13/what-does-a-health-value-of-zero-mean/

Drummond, M.F.; Sculpher, M.J.; Torrance, G.W.; O'Brien, B.J.; Stoddart, G.L. (2005) Methods for the economic evaluation of health care programme, Third edition, Oxford, Oxford University Press.

Ferreira, L.N.; Ferreira, P.L.; Pereira, L.N.; Brazier, J.; Rowen, D. (2010) A Portuguese value set for the SF-6D, Value Health, 13(5), 624-3.

Ferreira, L.N.; Ferreira, P.L.; Pereira, L.N.; Oppe, M. (2014b) The valuation of the EQ-5D in Portugal, Quality Life Research, 23(2), 413-423.

Ferreira, L.N.; Ferreira, P.L.; Pereira, L.N. (2014a) Comparing the performance of the SF-6D and the EQ-5D in different patient groups, Acta Médica Portuguesa, 27(2), 236-245.

Gafni, A.; Torrance, G.W. (1984) Risk attitude and time preference in health, Management Science, 30(4), 440-451.

Grossman, M. (1972) The demand for health: A theoretical and empirical study. National Bureau for Economic Research, Columbia University.

Khanna, D.; Tsevat, J. (2007) Health-related quality of life: An introduction, American Journal of Managed Care, 13, S218-S223.

Kind, P.; Rosser, R. (1979) Death and dying: Scaling of death for health status indices, in: D.A.B. Lindberg, P.L. Reichertz, (eds.), Lecture Notes in Medical Information, Berlin, Spring-Verlag, 28-36.

Kind, P.; Rosser, R. (1988) The quantification of health, European Journal of Social Psychology, 18, 63-77.

Kind, P. (2014) Valuing EQ-5D health states: Methods, myths and myopia, Leeds Institute of Health Sciences, Working Paper 14_09.

Patrick, D.L.; Starks, H.E.; Cain, K.C.; Uhlmann, R.F.; Pearlman, R.A. (1994) Measuring preferences for health states worse than death, Medical Decision Making, 14(1), 9-18.

Pauker, S.G.; Pauker, S.P.; McNeil. B.J. (1981) The effect of private attitudes on public policy: prenatal screening for neural tube defects as a prototype, Medical Decision Making, 1, 103-114.

Robinson, A.; Spencer, A. (2006) Exploring challenges to TTO utilities: valuing states worse than dead, Health Economics, 15(4), 393-402.

Rosser, R.; Kind, P. (1978) A scale of valuation of states of illness: Is there a social consensus? International Journal of Epidemiology, 7, 347-358.

Rosser, R.M. (1983) A history of the development of health indicators, in: G. Teelong-Smith, (ed.) Measuring the Social Benefits of Medicine, Office of Health Economics, 50-62.

Sequeira, T.; Ferreira, P.L.; Teixeira, J.; Peres, I.; Oliveira, J.; Silveira, A. (2015) Patient-reported outcomes in prostate cancer: Prospective changes analysis for prognosis prediction, Journal of Cancer Therapy, 6(15), 1238-1248.

Stevens SS. (1951) Handbook of Experimental Psychology, New York, John Wiley and Sons.

Stevens, S.S. (1946) On the theory of scales of measurement, Science, 103, 677-680.

Sutherland, H.J.; Llewellyn, T.H.; Boyd, D; Till, J.E. (1982) Attitudes towards quality of survival: The concept of maximum endurable time, Medical Decision Making, 2, 299-309.

Torrance, G.W. (1976) Social preferences for health states: An empirical evaluation of three measurement techniques, Socio-economic Planning Sciences, 10, 129-136.

Torrance, G.W. (1982) Preferences for health states: A review of measurement methods. In: JC Sindair (ed.), Clinical and Economic Evaluation of Perinatal Programs: Mead Johnson Symposium on Perinatal Medicine No. 20, Evansville, Indiana, Mead Johnson Co., 37-46.

Torrance, G.W. (1984) Health states worse than death, in W. van Eimeren, R. Engelbrecht, C.D. Flagle (Eds.), Third International Conference on System Science in Health Care, Health Systems Research, Berlin, Heidelberg, Springer.

Torrance, G.W. (1986) Measurement of health state utilities for economic appraisal, Journal Health Economics, 5, 1-30.

Torrance, G.W.; Feeney, D. (1989) Utilities and quality-adjusted life years, International Journal of Technology Assessment in Health Care, 5, 559-575.

WHO, World Health Organization (2000) The World Health Report 2000, Health Systems: Improving Performance, Geneva, WHO.

Williams, A. (1987) Health economics: The cheerful face of a dismal science, in A. Williams, Health and Economics, British Association for the Advancement of Science, London, Palgrave Macmillan.

IMPRENSA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS