6-1-2017

High frequency analysis of the EWZ ETF

Nada Ragab

Master of Science in Finance

The American University in Cairo

Cairo, Egypt

High Frequency Analysis of the EWZ ETF

A Thesis Submitted to

The Faculty of the School of Business

The American University in Cairo

in partial fulfillment of the requirements for

the Degree of MASTER OF SCIENCE IN FINANCE

By

Nada Ragab

Under the supervision of

Dr. ALIAA BASSIOUNY

May/2017

The American University in Cairo

School of Business
Department of Management

By
Nada Ragab
Under the supervision of Dr. Aliaa Bassiouny

ABSTRACT

The purpose of this thesis is to conduct a high frequency study of the pricing behavior and performance of an emerging market exchange traded fund relative to its benchmark index[1]. The thesis uses a high frequency intraday data[2] set of the international EWZ ETF and its benchmark the MSCI Brazil making this high frequency analysis the first on an emerging market ETF. In testing the pricing behavior, the thesis first examines the price deviation of the ETF from its benchmark index. Second pricing behavior is analyzed using cointergration analysis and a Vector Error Correction Model (VECM) between the ETF and the index intraday movements as well as a Granger Causality test for robsutness. In testing performance differences between the ETF and index, a performance is measured and compared using Sharpe Ratio and persistence and the tracking error of the ETF are measured analysis is also conducted. Results showed that the prices of the index are higher on average that those of the ETF on both daily and intraday basis. Moreover, it was shown in the results that the ETF outperforms the index on intraday basis but the index outperforms the on daily basis. Also, the results displayed that there is an average daily tracking error on annual basis and that this error is persistent with a 0.12% rate. Furthermore, it was concluded that on intraday basis both the ETF and the index move to close the gap if a price deviation exists with a rate of 16.3% and 83.7% respectively, while on the other hand, on daily basis the results show that ETF doesn't affect the index at all

Keywords: Exchange Traded Funds, ETFs, price deviation, performance persistence, tracking error, tracking ability, co-integration, international.

[1] benchmark index is the index that the ETF replicates and mimics its performance
[2] high frequency data is 1 minute interval prices which are used in the calculation and analysis of intraday returns

TABLE OF CONTENTS

Chapter I: Introduction

1.1 Definition and History of Exchange Traded Funds …………………………....4

1.2 Thesis objectives ……………………………………………………………….5

1.3 Contribution ……………………………………………………………………5

1.4 Data ……………………………………………………………………………6

1.5 Main Methodology …………………………………………………………..6

1.6 Findings & Results …………………………………………………………..8

Chapter II: Literature Review

2.1 Overview on the Exchange Traded Funds ……………………….……......10

2.2 Price Efficiency / Deviation ……………………………………………..11

2.3 Fund Performance and Tracking Ability …………………………………...12

2.4 The Relationship between the ETF and the Index …………………..……...14

2.5 Contribution ……………………………………………………………...14

Chapter III: Data& Methodology

3.1 Data ………………………………………………….........................16

3.2 Pricing Behavior …………………………………………………………....17

3.2.1 Price Deviation……………………………………………………..17

3.2.2 ETF & Index Co-integration …………………………...……….18

3.3 Comparing ETF and Index Performance …………………………………...20

3.3.1 Performance Measure ……………………………………………..20

3.3.2 ETF Tracking Ability ………………………………………………21

3.3.3 Tracking Error Persistence ……………………...………………..22

3.3.4 Tracking Error Explanation ………………………………………22

Chapter IV: Results

4.1 Pricing Behavior …………………...……………………………………..23

4.1.1 Price Deviation ……………………………………………….....23

4.1.2 ETF & Index Co-integration ………………………………………23

4.2 Comparing ETF and Index Performance …………………………………......26

4.2.1 Performance Measure …………………………………………...26

4.2.2 ETF Tracking Ability ……………………………………………28

4.2.3 Tracking Error Persistence ……………………………………...29

4.2.4 Tracking Error Explanation ……………………………………..29

Chapter VI: Conclusion & Limitations

5.1 Conclusion ………………...………………………………………31

5.2 Limitations ………………………………………...……………...32

References ……………………………………………………………..…33

Appendices

Appendix 1: …………………………………………………………....35

Appendix 2: ……………………………..…………………………...37

Appendix 3: …………………………………………………………40

Appendix 4: …………………………………………………………41

Appendix 5: …………………………………………………………41

Chapter I

Introduction

1.1 Definition and History of Exchange Traded Funds

Exchange Traded Funds (ETFs) as indicated by their name are funds designed to mimic indices' performance and by that enabling investors to invest in index constituents – even if lacking the required financial and time resources to do so - through investing in those ETFs.

ETFs gained a lot of popularity in the latter years, now there are around 4,000 different ETFs listed on over 50 exchanges (Charupat & Miu, 2012). The popularity of the ETFs goes back to their several benefits which mainly include intraday trading, cost transparency, low expense ratios, and tax efficiency (Charupat & Miu, 2012). Due to this increase in the number and interest in ETFs, it also gained the attention of researchers. There are three main areas of research: the first area is price efficiency of ETFs which studies whether there is a price discrepancy between the ETF and its underlying index, the second area is the performance analysis which is measures performance through Sharpe ratio[3] and tracking ability through tracking error[4]), and finally the relationship between the ETF and its underlying index through measuring their co-integration; as due to the recent strength gained by ETFs, it is believed that other than the expected effect of an index on its ETF's returns, the ETF can also affect the index due to the volume of trade on ETFs

---

[3] Sharpe Ratio is defined as return per the associated risk

[4] Tracking Error is defined as the standard deviation in return differences between ETFs and benchmarks over time

1.2 Thesis objectives

This paper will study an emerging market exchange traded fund namely the EWZ Exchange Traded Fund, which tracks the MSCI Brazil Index. The MSCI Brazil Index is designed to measure the performance of the large and mid- cap segments of the Brazilian market. The index constitutes of 57 stocks which cover about 85% of the Brazilian equity universe. (MSCI website)

The purpose of this thesis is to conduct a high frequency study of the pricing behavior and performance of the EWZ relative to its benchmark index. In order to test the pricing behavior, the thesis first examines the price deviation of the ETF from its benchmark index. Second pricing behavior is analyzed using a co-integration analysis and a Vector Error Correction Model (VECM) between the ETF and the index intraday movements as well as a Granger Causality test for robustness. The second area of study is performance of the EWZ ETF against its benchmark index where performance is measured and compared using Sharpe Ratio and a tracking error is calculated to measure the tracking ability of the ETF where the persistence of the error is measured

This research will use high frequency data (1 minute interval) to analyze the above mentioned areas and will compare the results to those obtained from daily data (closing prices).

1.3 Contribution

Various studies have been conducted on ETFs in the latter years, however those studies mainly focused on US based ETFs that tracked domestic indices like the Spider that tracks the S&P 500 and the Cubes that tracks Nasdaq 100 index. Also most of the studies used the ending NAV or closing prices in their analysis of ETFs due to the unavailability of daily data. Therefore, this thesis aims at contributing to the closure of a gap in the literature from two perspectives. The first is through studying an emerging international exchange traded fund. The second is through performing a high frequency analysis on the EWZ ETF and its underlying index using 1minute interval data and comparing their results to their daily data counterparts.

1.4 Data

As discussed, this thesis will use intraday and daily EWZ and MSCI Brazil prices to examine each of the mentioned areas of interest and compare the results to be able to identify any differences in the outcomes when using different data frequencies. One minute interval and daily prices have been extracted from Thomson Reuters Tick Database for the EWZ ETF and its underlying index MISC Brazil from January 2010 to September 2014. The data has been aggregated to show all trades in a specific minute at each trading date across the period

1.5 Main Methodology

The main methodology for each of the mentioned areas will be as follows:

In order to study pricing behavior, I will start by looking at price deviation of the ETF from its bench mark index following the model of DeFusco et al, (2009); which states that price deviation is equal to the difference between the index log prices and the ETF log prices. This is calculated for both daily and intraday data for the sake of result comparison. The hypothesis is that the price deviation is zero as the ETF is supposed to track the index exactly so that there are no premiums or discounts.

The second aspect in pricing behavior is analyzing the co-integrating relationship of the index and the ETF. The ETF is designed to follow its underlying index, therefore the ETF should correct based on the movement of the index, this correction is what I am testing and quantifying. In order to achieve that, I will have to follow a model similar to that of the price discovery literature. First, I will test for non-stationarity of the EWZ index prices and MSCI Brazil prices first through the Augmented Dickey and Fuller test (ADF). The next step would be to test for co-integration using two approaches: the first approach would be to use the result of the non-stationarity of prices and test whether the price deviation between the ETF prices and the underlying index prices ($P_t^s - P_t^f$) is stationary using the ADF test statistic. In the second approach, the Johansen co-integration test is used for the null hypothesis that the number of co-integrating vectors between prices is equal to 0. After verifying the unit root and co-integration assumptions, a Granger Causality test is used to identify whether the index affects the ETF or not and whether an opposite relationship where the ETF affects the index exists. If a relationship proves to exist a Vector Error Correction Model can be used to test if the ETF follows (corrects based on) the index or vice versa and the magnitude of that correction. All the above was tested for both daily and intraday prices to compare the results.

The other area of interest in this thesis is studying the performance of the EWZ index in comparison with its MSCI Brazil index. In this area we start by comparing the performance through calculating the Sharpe ratio for both the ETF and the index on both daily and intraday basis following Rompotis's (2011) approach in measuring performance. Rompotio's uses the Sharpe Ratio in his analysis of performance, as it is an indicator of how well the ETF or the index compensates their investors per unit risk. In order to calculate the Sharpe Ratio, returns and risk (standard deviation) of the ETF and index are calculated as the Sharpe Ratio is defined as the average return minus the risk free rate divided by the standard deviation of the period of study. The higher the Sharpe ratio, the higher the return per unit risk, hence, the better the performance.

The other measure of performance is the tracking ability of the EWZ to its benchmark index which is measured by the tracking error. In practice, the tracking error is calculated as the standard deviation in return differences between ETFs and benchmarks over time (Aber et, al 2009) (Rompotis,2011). If the daily returns of a fund track the index closely, its returns should mirror the index's returns and they should have the same standard deviation (Aber et, al 2009), therefore it is assumed that the tracking error of the EWZ is zero. Moreover, the persistence of the tracking error was tested for using a single-factor cross-sectional model regressing the estimated tracking error in day t on the tracking error in day t-1(Rompotis,2011). The beta coefficient of the model is the indicator of persistence. A positive and significant beta indicates that tracking error persists between two consecutive days. On the other hand, a negative or an insignificant beta means that the tracking error is not persistent. After testing for persistence, explanations of tracking error will be studied by following Rompotis's (2011) approach in finding the operational factors affecting the tracking error. So the thesis looks at Risk and Volume as determinants of tracking error through running regression model for the mentioned independent variables on the tracking error as a dependent variable; where Risk is estimated as the standard deviation of ETFs' returns on intraday basis, and Volume is the number of intraday trades during the 5 years of the study.

1.6 Findings and Results

Findings are also presented for each of the above mentioned models. The first results pertain to the price deviation, they showed that the prices of the index are higher on average that those of the ETF on both daily and intraday basis by 0.109 cents and 0.0026 cents respectively. It was also observed that the deviation on daily basis is slightly higher than that on intraday basis. However, the maximum and minimum deviations are slightly higher for both daily and high frequency data.

The following conclusions are those for the ETF – Index co-integration relationship. As discussed in the previous section, in this part of the research there is a number of consecutive steps so results are reported for each of the steps. Results for the unit root test show that all price series contain a unit root indicating non-stationarity of the prices for both daily and intraday data, because we fail to reject the null hypothesis at 5%. Furthermore, the ADF t-statistic on the price deviation is highly significant at a 5% level of significance for both daily and intraday data indicating that a stationary combination of prices exists. Moreover, The Johansen test result rejects the null of no co-integrating vectors in favor of one co-integrating relationship which exists between the ETF and index daily and intraday prices. The results of the Granger Causality are noteworthy since they are different for intraday and daily data. When running the Granger Causality test on the intraday data, the results of the intraday data appear to be statistically significant at 5% and we reject both null hypotheses that the index doesn't affect the ETF and the ETF doesn't affect the index for intraday data, on the other hand, we fail to reject the null hypothesis that the ETF does not affect the index and the results of the VECM are statistically insignificant at 5%. The Vector Error Correction Model results show that there is a co-integration relationship between the ETF and the index as it appears that when there is a price deviation, the ETF will correct to close the gap 83.7% of the time while the index will correct 16.3% of the time.

The thesis will start by reviewing the literature on exchange traded fund research in Chapter II through summarizing the methodologies and results of different studies done on price behavior and ETF performance. Following that Chapter III will discuss the data used in that thesis and the methodology followed in order to reach the targeted results. Afterwards, Chapter IV presents the results for all the research questions and hypothesis discussed in the previous chapter. Finally the thesis will be concluded in Chapter VI and the light is shed on the limitations for future research.

Chapter II

Literature Review

2.1 Overview on the Exchange Traded Funds:

Exchange-traded funds (ETFs) are of two types; open-end mutual funds or unit investment trusts. ETFs are designed to track the performance of a certain benchmark index by investing in the constituents of that index (Aber et al., 2009). The advantages of ETFs can be summarized as follows: they are traded intraday like stocks as opposed to mutual funds that are traded at NAV[5] end of day and due to that tradability, they have low expense ratios. ETFs also do not have short selling[6] restrictions like stocks which makes them good hedging[7] tools (DeFusco et al,. 2009). They are also popular for providing cheap diversification and tax efficiency (DeFusco et al,. 2009) (Aber et al., 2009). The disadvantage of ETFs lies in the fact that traders have to pay commissions and bid/ask spreads when they trade the securities which can be high for illiquid ETFs (Charupat & Miu, 2012). Since ETFs are passively managed by their sponsors; ETF shares can be created/ redeemed at will usually by qualified institutional investors. Those investors can trade shares in blocks of 50,000 + shares. Due to the availability of this process, arbitrage opportunities can exist if there is any price discrepancy between the shares of an ETF and the index's underlying shares. For example, if the ETF is trading at a premium, the ETF shares will be sold (or short sold) and the underlying assets will be bought from the market to cover the position (Aber et al.,2009). This leads us to an important topic in the literature of ETFs; namely price efficiency or rather deviation (premiums/discounts).

---

[5] NAV stands for Net Asset Value which is the value of assets under management in a certain fund

[6] Short selling is the sale of a security that is not owned by the seller, or that the seller has borrowed

[7] Hedging is an investment to reduce the risk of adverse price movements in an asset.

2.2 Price Efficiency / Deviation:

Price efficiency is simply how close market prices are to the fund's net asset value (NAV), therefore any differences between the market prices and the funds' NAVs is defined as a price deviation and would create an arbitrage opportunity that investors can exploit by buying the underlying basket of securities from the market (if the ETF is trading at a premium) and redeem it through the ETF as enabled by the creation/redemption nature of ETFs[8] (Charupat & Miu, 2012). Many researchers have investigated price premiums and discounts and concluded that the deviation is not economically significant and that arbitrageurs exploit the opportunity quickly so it disappears within 1 day, those researchers include Ackert and Tian (2000), Elton et al. (2002) and Curcio et al. (2004). For example, Ackert and Tian (2000) and Elton et al. (2002) in their studies of pricing, they look at the SPDR which is the ETF that tracks the S&P 500 index and both conclude that there is no economically significant price deviation from the benchmark index. Furthermore, Elton et al. (2002) show that arbitrageurs trade quickly on those price discrepancies so they disappear in one day. In the case of Curcio et al. (2004), their objective was to compare price deviations of the same index; the SPDR to those of the QQQQ which tracks the Nasdaq 100 Index. They reached the conclusion that the average price deviations were small for both funds, however the standard deviation of price deviations were larger for QQQQ than those for SPDR. Also, Engle and Sarkar (2006) in their study of 21 ETFs on US based indices, including the S&P 500, the Nasdaq 100, the Dow Jones Industrial Average and the Russell 1000, and sector indices. They find that price premiums / discounts are on average minimal and within transaction costs and bid-ask spreads. They also report that the volatility of price deviations is related to the volatility of the underlying NAVs, which is consistent with the findings of Curcio et al. that price deviations depend on the underlying index as the QQQ prices have higher volatility than those of SPDR because the Nasdaq 100 index has more volatility than those of the S&P 500 index.

---

[8] Creation / Redemption of ETFs happens through an in kind trade of the underlying shares through institutional investors to either buy or sell those securities in the open market

On the other hand, there are other studies that found different results pertaining to price deviation. For example, De fusco et al. (2009) in their study of price deviation of the three major US ETFs namely the Spiders which tracks the S&P 500, the Diamonds which tracks the Dow Jones Industrial Average, and the Cubes which tracks the NASDAQ 100 using closing prices concluded that there is a persistent price deviation associated with the price forming process. Moreover, It is also concluded in the literature that price deviations are larger and more volatile for international ETFs (from a US investor point of view) due to the computation of NAVs based on prices from earlier closing times than the US market close (Engle & Sarkar, 2006) (Delcoure& Zhong, 2007). Despite the fact that NAVs are adjusted for the prevailing exchange rates, they still don't completely consider information arriving during the US market trading hours. Also, arbitrage is not effective for US investors as trading hours are not the same (DeFusco et al,. 2009). For example, Engle and Sarkar (2006) study 16 international ETFs with data up to the year 2000, and their results show that their average price deviations are 0.35% which is much higher than those of domestic ETFs with 0.01%. It is also noteworthy that they showed that international ETFs price deviations more persistent than those of domestic ETFs as deviations last for several days in international ETFs as opposed to several minutes in the domestic ones. Furthermore, Delcoure and Zhong (2007) in their study of price deviations of 20 iShares ETFs, each of which track an MSCI country-specific index, where they use data up to the year 2002, report that iShares in general trade at economically significant premiums in the range of 10 and 50 percent of the time. However, those premiums are not persistent and disappear within two days. Another study by conducted Ackert and Tian (2008) using data from the period 2002-2005 compares 21 international ETFs to seven domestic ETFs. They report that within the 21 international ETFs, those tracking emerging-market indices have bigger median price deviations and higher volatility than those tracking developed-market indices. Moreover, Ackert and Tian show first-order autocorrelations of price deviations that are statistically significant at an average of 0.20 for developed-market ETFs and 0.41 for emerging-market ETFs.

2.3 Fund Performance and Tracking Ability:

The second popular aspect of ETFs in the literature is the tracking ability of ETFs. Tracking error is used to measure the tracking ability of a given ETF and is defined as the differences between the funds' NAV returns and the returns of the underlying benchmark index (Charupat& Miu, 2012).

In theory ETFs should track the returns and risk of their underlying indices without any form of deviation, however tracking error in reality is unavoidable because the underlying portfolio is measured on paper where there are no friction and transaction costs.

This theory assumes that the weights of the securities in the ETF are changed – based on market changes - instantaneously and with no cost which is not the case as fund managers have to engage in the actual trading of the index securities (Aber et, al 2009). There are several factors that can cause a tracking error namely; the number of securities in the underlying index, the difference in market cap between the ETF portfolio and the index, index volatility, portfolio beta, management fees, transaction costs, dividends distribution, change in index composition, and the replication strategy used (Gastineau, 2002), (BlackRock, 2012) (Aber et, al 2009) (Frino et al., 2004) (Canakgoz &Beasley, 2008) (Corielli & Marcellino, 2006).Various studies aimed at measuring the tracking error of an ETF to measure its tracking ability. One of the early researches conducted on tracking error was that of Elton et al. (2002) on the Spider ETF and they reported that the ETF underperforms its bench mark index the S&P 500 by an average of 28 basis points (bps) per year from 1993 to 1998. They attribute the tracking error to the expense ratio of SPDR and the delay in the reinvestment of dividends from its constituents. Another study is that of Rompotis (2006) which studies the tracking errors of 73 domestic, international, and sector ETFs in the period of 2005-2006 and concludes that significant tracking error exist especially for international ETFs. Moreover, in investigating a list of 42 equity ETFs that track various industry sectors in the period 2006 - 2008, Qadan and Yagil (2012) conclude that tracking errors are positively correlated with the daily volatility of the ETF. Another area of research is measuring the tracking error of ETFs mimicking foreign equity indices (from a US investor perspective). For example, Shin and Soydemir (2010) performed tracking errors analysis for a list of iShares MSCI country ETFs. Using a sample 20 iShares MSCI country ETFs for the the period from 2004 to 2007, they found that tracking errors are persistent for country ETFS especially for the ones tracking Asian markets and that the change in the exchange rate is a substantial source of tracking errors. Although the study and its likes cover a number of emerging equity markets, the majority of the foreign countries being studies are the ones with developed markets

2.4 The Relationship between the ETF and the Index:

The third topic discussed is the relationship between the ETFs and indices which is supposed to be similar to the price discovery structures of comparable or different securities linked by arbitrage trading on different markets (De Fusco, 2009). The topic has not been studied extensively yet. There are studies that look at the changes in pricing through a Vector Error Correction Model between the index and the fund (Ivanov, 2013). Ivanov (2013) in his study of DIA, SPY and QQQQ concluded that the change in the ETF price is negatively related to the change in past ETF prices and positively related to changes in past index levels, which is the opposite of what DeFusco et al. (2011) concluded using daily closing price data. DeFusco et al. (2011) in their study of SPY, DIA, and QQQQ reported - through conducting a price deviation stationarity test using the Augmented Dickey Fuller model –that the ETF price and the index price are co-integrated of order (1,1) except for the QQQQ price deviation. In the same study, the researchers looked at the co-integration between the ETFs under study and their benchmark indices and concluded that a co-integrating relationship exist through using a Vector Error Correction Model. The study continues to investigate the reasons of the autocorrelation using impulse responses and lead lag relationships and reports that the price deviation responds to its own shocks and not to shocks from the benchmark index. Therefore the impact of the benchmark index volatility on the volatility of the ETF price deviation is minor which implies that ETF specific factors can be important.

2.5 Contribution

This paper aims at filling a void in the literature from two perspectives. The first is through studying an international exchange traded fund namely the EWZ ETF which is benchmarked against the MSCI Brazil index is designed to measure the performance of the large and mid- cap segments of the Brazilian market. The index constitutes of 57 stocks which cover about 85% of the Brazilian equity universe (MSCI website). The second is through performing a high frequency analysis on the EWZ ETF and its underlying index using 1minute interval data and comparing their results to their daily data counterparts. The analysis looks at three areas commonly available in the literature, which are: price deviation (premium/discount) of ETF prices, tracking ability of the fund, and co-integration between the ETF and its underlying index.

DeFusco et al. (2009) stated in the recommendation of their paper studying ETF price deviations that "The daily data used in this study is just a snapshot of a daily trading activity. Intraday data and more precise identification prices might shed additional light on the interpretation and study of pricing deviation." Moreover, Ivanov (2013) in his study of ETF price deviation using high frequency data recommended the study is extended to international ETFs. It was also suggested by Aber et al. (2009) to use "real time intraday prices to obtain more information on the persistence of premiums and/or discounts" as daily data only give a snapshot and not a full analysis. Therefore the contribution of this study to the literature is the use of intraday data to study the EWZ ETF which is an international fund mimicking the EAFE Brazil index.
The conclusions reached from the comparison of daily and intraday results are noteworthy, due to the large number of observations studied which equates to 438,152 and 438785 when studying ETF tracking error and co-integration respectively. Moreover, the use of intraday data allows us to reach results that were not possible using daily data for example; daily tracking error was calculated instead of annual tracking error. Also, I was able to compare intraday to daily Sharpe ratio results which showed that there are opportunities to be exploited on during the trading hours that don't appear when we look at daily data. Another interesting contribution is in the co-integration area, where when daily data is used it appears that the ETF doesn't appear to impact the index and it appears as a one way relation where the ETF always corrects to close the gap in prices between the ETF and the index, however, when intraday data is used, it appears that there is a 2 way relationship and that 16.3% of the time the index moves to close that gap in prices. Hence, we can conclude that studying intraday data brings some additions to the literature that could not be reached using daily data.

Chapter III

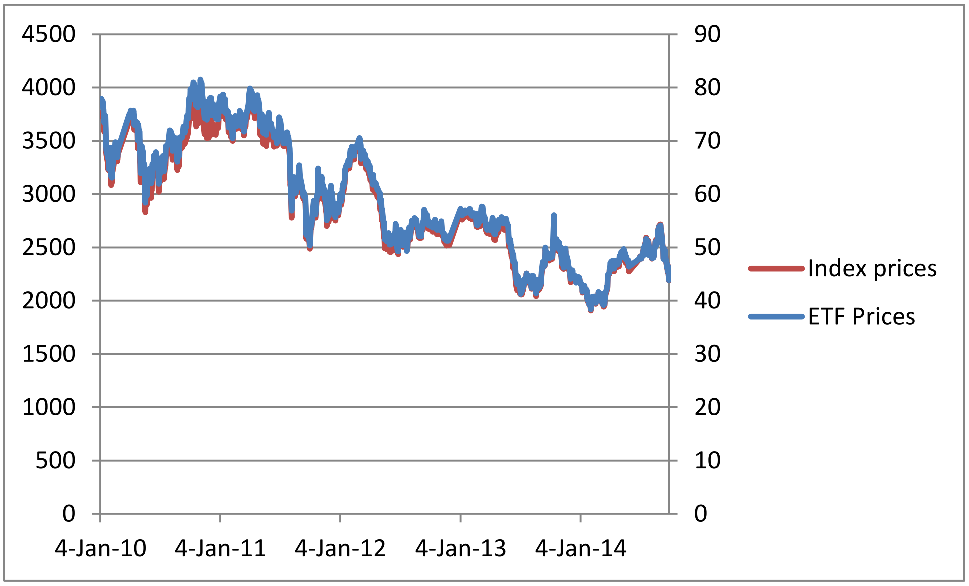
Data & Methodology

In this section, I will replicate industry standard models to study three prominent areas in the pricing and performance of ETFs using high frequency (1 minute interval data) by looking at EWZ ETF against its benchmark MSCI Brazil Index and comparing the results to those of daily data. These areas namely are (1) pricing behavior of ETFs which includes price deviation that shows whether the ETF trades at a premium or a discount from its underlying index. It also includes studying the presence/absence of a co-integrated relationship between the ETF and the index. The second is (2) Computing the index and ETF performance through comparing the Sharpe ratios of the ETF and the index and measuring tracking ability of the ETF to its benchmark and the persistence of the tracking error if it exists. I will start by explaining the data structure and then I will be exploring each of the above mentioned areas successively.

3.1 Data:

This paper will use intraday and daily EWZ and MSCI Brazil prices to examine each of the mentioned areas of interest and compare the results to be able to identify any differences in the outcomes when using different data frequencies. One minute interval and daily prices have been extracted from Thomson Reuters Tick Database for the EWZ ETF and its underlying index MISC Brazil from January 2010 to September 2014. The data has been aggregated to show all trades in a specific minute at each trading date across the period. Figures 1 below represents the daily index and ETF prices respectively.

Figure 1: Daily ETF and Index Prices over the Study Period



3.2 Pricing Behavior

3.2.1 Price Deviation

Price deviation is defined in the literature as the difference between the ETF NAV and the index price (Engle & Sarkar, 2006), in reality however, most trading of ETFs occurs intraday at market determined prices (De Fusco et al., 2009) not at the ending NAVs. Therefore due to the lack of NAV data and for a more realistic representation of price deviation, it is agreed in the literature to calculate price deviation as the difference between the benchmark (log) price and the ETF's (log) price (Ivanov, 2013) (DeFusco et al,. 2009) (DeFusco et al,. 2007). This difference is expected to be zero because "The Law of One Price and the "no-arbitrage" argument suggest that the price of a basket of securities, such as an ETF, should be equal to the sum of its components' prices" (DeFusco et, al. 2009), which means that since the ETF is composed of the stocks of the underlying index, we should expect no difference between the price of the ETF and the price of the index.

The measure for price deviation is the below equation (DeFusco et al,. 2009):

$$PD_t = S_t - F_t \quad (1)$$

Where, PD_t is Price Deviation (premium or discount), S_t is price of the market index, and F_t is price of the ETF.

The above leads the hypothesis that the expected value of pricing deviation is zero. DeFusco et al,. (2009) concluded that when daily price levels are studied, it appears that ETFs prices are higher on average than the indices they are tracking. Therefore in this paper we will compare the results of intraday to daily data.

3.2.2 ETF & Index Co-integration:

"By design the ETFs track their underlying indexes. Therefore, a co-integrating relationship must exist between ETFs and their indexes" (Ivanov, 2013). The ETF is designed to follow its underlying index, therefore the ETF should correct based on the movement of the index, this correction is what I want to test and quantify, and in order to do that I have to test for non-stationarity[9] of the EWZ index prices and MSCI Brazil prices first through the Augmented Dickey and Fuller test (ADF). The next step would be to test for co-integration using two approaches: the first approach (1) would be to use the result of the non-stationarity of prices and test whether the price deviation between the ETF prices and the underlying index prices ($P_t^s - P_t^f$) is stationary using the ADF test statistic[10]. In the second approach (2), we rely on the Johansen co-integration test for the null hypothesis that the number of co-integrating vectors between prices is equal to 0. Once the unit root and co-integration assumptions are tested, a Granger Causality test is run in order to test for the existence of a causal relationship between the ETF and the index (whether the ETF affects the index and vice versa). After verifying the existence of a relationship, the Vector Error Correction Model can be used to test if the ETF follows (corrects based on) the index or vice versa and the magnitude of that correction. All the above was tested for both daily and intraday prices to compare the results. Therefore, I will follow the stated standard literature methodology through the following steps:

---

[9] Appendix 1

[10] Appendix 2

a) Unit Root Test

First, I will use the Augmented Dickey and Fuller (ADF) approach to identify whether or not each of the EWZ ETF prices and the MISC Brazil index prices has a unit root and thus non-stationary root of 1. We run the ADF test using the random walk regression. The ADF test examines difference from zero of α coefficient in the following regression using standard t-test, with Xt a vector of lagged differences in Yt (DeFusco et, al. 2009).

$$\Delta Y_t = \alpha Y_{t-1} + \beta x_t + \varepsilon_t \tag{2}$$

The hypothesis tested is that the time series of EFA and MISC EAFE have a unit root of 1.

b) Co-integration

As discussed I will test for the existence of a co-integration relationship following two methods: The first is to test for the non-stationarity of price deviation between the ETF prices and the underlying index prices $(P_t^s - P_t^f)$ using the ADF test statistic.

The second method will rely on the Johansen co-integration test for the null hypothesis that the number of co-integrating vectors between prices is equal to 0.

I will follow the price discovery literature and use the PT common factor model of Gonzalo and Granger (1995) to determine the amount of adjustment of the ETF or its underlying index as a response to a change in the other (i.e. the coefficient of the co-integrating equation in the VECM). It is assumed that the EWZ ETF will follow the MISC Brazil index due to the nature of the ETF setup. I estimate the following generalized vector error correction model, which includes the lagged changes of prices:

---

[11] Appendix 3

$$\Delta P_t^s = \alpha_0^s + \alpha^s\left(P_{t-1}^s - \beta^f P_{t-1}^f\right) + \sum_{i=1} \gamma_i \, \Delta P_{t-i}^s + \sum_{i=1} \delta_i \, \Delta P_{t-j}^f + \varepsilon_t^s \quad (3)$$

$$\Delta P_t^f = \alpha_0^f + \alpha^f\left(P_{t-1}^s - \beta^f P_{t-1}^f\right) + \sum_{i=1} \gamma_j \, \Delta P_{t-j}^s + \sum_{j=1} \delta_j \, \Delta P_{t-j}^f + \varepsilon_t^f \quad (4)$$

The co-integrating equation $\left(P_{t-1}^s - \beta^f P_{t-1}^f\right)$ is estimated with the Johansen co-integration test above, where β^f is normalized to 1. The coefficients of main interest in determining the index/ ETF price relation using Equations (3) and (4) are α^s and α^f of the co-integrating equation. They indicate the extent to which the price of the ETF or the index will adjust to a deviation from the parity equilibrium. For example, if a price deviation is observed whereas $P_t^s > P_t^f$, parity will be restored when either or both the ETF and / or the index move to respond to the magnitude of the departure. Since the ETF is a replica of the index, I expect the ETF price to increase to close the gap, and therefore α^f is expected to be positive. On the other hand, due to the increased trade on ETFs, this volume of trade can affect the index in which its price will also adjust to the gap due to the creation and redemption process, which will be reflected in a significant and negativeα^s. The larger and more significant the sign of alpha, the greater the adjustment of the price of one variable to the change in the other (Ansoteguir et al., 2015).

3.1 Comparing ETF and Index Performance

3.3.1 Performance Measure

Although ETFs are passively managed, their performance is an important area of study because investors must be sure that the performance of their ETF is at least similar to that of the underlying index. In this section of the research I will follow Rompotis's (2011) approach in measuring performance. Rompotio's uses the Sharpe Ratio in his analysis of performance, as it is an indicator of how well the ETF or the index compensates their investors per unit risk. In order to calculate the Sharpe Ratio, returns and risk (standard deviation) of the ETF and index are calculated and the below equation (5) is used once to calculate the daily Sharpe ratio and once to calculate the intraday one:

$$Ss,f = \frac{\overline{Rs},f - \overline{RF}}{\sigma s,f} \quad (5)$$

Where $\overline{Rs}, f$ denotes the average intraday/ daily returns on the ETF, $\overline{RF}$ denotes the average daily St. Louis federal reserve rate (only used in the case of daily Sharpe Ratio as there is no intraday federal reserve rates) While $\sigma s, f$ is the intraday/ daily standard deviation of ETF's f or Index's s return. The higher the Sharpe ratio, the better the performance of the fund or the ETF.

3.3.2 ETF Tracking Ability

The tracking ability of an ETF is defined as "the deviation of the return on the NAV of an ETF from the corresponding return on its underlying benchmark index" (Charupat & Miu, 2012). Any deviations of the returns on NAV from those of their underlying benchmarks accumulation over time could significantly affect the long-term performance of the ETFs (Charupat & Miu, 2012).

I will measure the tracking ability of the EWZ ETF using Frino and Gallagher (2001) method:

In practice, the tracking error is calculated as the standard deviation in return differences between ETFs and benchmarks over time (Aber et, al 2009) (Rompotis,2011). If the daily returns of a fund track the index closely, its returns should mirror the index's returns and they should have the same standard deviation (Aber et, al 2009). The below equation (6) is used to compute the tracking error of the ETF:

$$TE = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(e_{dt} - \bar{e}_d)^2} \qquad (6)$$

Where e_{dt} is the difference of ETF and index returns during day t, $\bar{e}_d$ is the average return's difference over the day and n is the number of trades during day t.
This leads to the hypothesis that the tracking Error (TE) is equal to zero.
In this section, the data contained 438,152 observations resembling the number of differences in returns calculated using intraday data over the period of the study averaging 405 trades a day.

3.3.3 Tracking Error Persistence

After concluding that a tracking error exists, the next step is to test the persistence of that error and explaining the reasons behind it. In order to test for tracking error persistence, we perform a single-factor cross-sectional model regressing[12] the estimated tracking error in day t on the tracking error in day t-1(Rompotis,2011) using the below equation(7):

$$TE_t = \alpha + \beta TE_{t-1} + u \qquad (7)$$

where, TE_t is the tracking error coefficients on day t calculated by equation (6). The beta coefficient of the model is the indicator of persistence. If the beta is positive and statistically significant, then that tracking error persists between two consecutive days. While if the beta is negative or statistically insignificant then that tracking error is not persistent (Rompotis,2011).

3.3.4 Tracking Error Explanation

Various scholars have explained the reasons behind ETF tracking errors and attributed them to different factors including but not limited to; transaction costs (Rompotis, 2011), management fees (BlackRock, 2012), changes in index composition (Gastineau, 2002), issuance/repurchase of shares (Frino et al.,2004), distribution of dividends (Frino et al., 2004) and (Elton et al., 2002). However, in this study, I follow Rompotis's (2011) approach in finding the operational factors affecting the tracking error. Therefore, in this study we looked at Risk and Volume as determinants of tracking error through running equation (8) below following Rompotis's model:

$$TE_t = \alpha + \beta_1 Risk + \beta_2 Volume + u \qquad (8)$$

Where TE_t is the tracking error estimated in day t, Risk is estimated as the standard deviation of ETFs' returns on intraday basis, and Volume is the number of intraday trades during the 5 years of the study[13].

---

[12] Appendix 4

[13] Appendix 5

Chapter IV

Results

In this chapter of the thesis, the results for each of the above sections will be presented.

4.1 Pricing Behavior

4.1.1 Price Deviation

Table 1 below depicts the descriptive statistics of daily and intraday price deviation of the EWZ from the MSCI Brazil index. The calculations were made by scaling the data in order to make the prices of the ETF and the index comparable. It can be concluded here that as opposed to DeFusco's results that the prices of the index are higher on average that those of the ETF on both daily and intraday basis by 1.09 cents and 0.26 cents respectively. It was also observed that the deviation on daily basis is slightly higher than that on intraday basis. However, the maximum and minimum deviations are slightly higher on daily and intraday basis

Table 1: ETF Price Deviation

	Daily	Intraday
Mean	0.0109	0.0026
Min	-0.0606	-0.8043
Max	0.0297	0.0389
Median	0.0131	0.0069
Skewness	-1.9442	-4.5070
Kurtosis	6.1698	267.4159

This table provides descriptive statistics that summarize the results obtained from the price deviation calculated by subtracting the index and ETF prices after scaling them to $1.

4.1.2 ETF & Index Co-integration:

Results for the unit root test are presented in Table 2 and show that all price series contain a unit root for both daily and intraday data, because we fail to reject the null hypothesis at 5%.

Table 2 Augmented Dickey Fuller Test Statistic for prices

Variable	t-Statistic	Prob.*
ETF (Intraday Prices)	-1.497846	0.535
ETF (Daily Prices)	-1.606753	0.4787
Index (Intraday Prices)	-1.517983	0.5247
Index (Daily Prices)	-1.625438	0.4691

The ADF test summary in this table is a stationarity test for the index and ETF prices. The results show that both ETF and index prices are non-stationary as they have a unit root because the probabilities for both the ETF and the index on daily and intraday basis are higher than 5% which makes the results insignificant

Furthermore, the ADF t-statistic on the price deviation is highly significant at a 5% level of significance for both daily and intraday data indicating that a stationary combination of prices exists. Moreover, The Johansen test result rejects the null of no co-integrating vectors in favor of one co-integrating relationship which exists between the ETF and index daily and intraday prices. Both co-integration test results for ADF and The Johansen models are presented in Tables 3 & 4 below:

Table 3 Augmented Dickey Fuller Test Statistic

Variable	t-Statistic	Prob.*
Intraday $(\boldsymbol{P_t^s - P_t^f})$	-5.10781	0
Daily $(\boldsymbol{P_t^s - P_t^f})$	-3.02099	0.0333

The ADF test summary in this table is a stationarity test for the price difference between the index and the ETF on daily and intraday basis. The results are significant as the probability is lower than 5% which indicated that the difference is stationary showing that there is a trend when looking at the difference between prices over the period of the study

Table 4 Johansen Co-integration Test

Test	Level of Co-integration	Significant Value
Akaike Information Criteria (Intraday)	1	-24.51832*
Akaike Information Criteria (Daily)	1	-15.77132*
Schwarz Criteria (Intraday)	1	-24.51774*
Schwarz Criteria (Daily)	1	-15.68097*

The results are most significant at 1 level of co-integration determined by the Akaike Information Criteria & Schwarz Criteria

In the below section, the results of the Granger Causality and the Vector Error Correction Model are reported in Tables 5 and 6 respectively. The results are remarkable, since they are different for intraday and daily data. When running the Granger Causality test on the intraday data, the results of the intraday data appear to be statistically significant at 5% and we reject both null hypotheses that the index doesn't affect the ETF and the ETF doesn't affect the index which means that there is a co-integration relationship between both the ETF and the index. On the other hand, when running the same tests over daily data, the results are different as we fail to reject the null hypothesis that the ETF does not affect the index. Therefore, when the correction was measured by the Vector Error Correction model, in case of the intraday data it appears that the when there is a price deviation, the ETF will correct to close the gap 83.7% of the time while the index will correct 16.3% of the time. Conversely, the results of the VECM were statistically insignificant at 5%.

Table 5 Pairwise Granger Causality Model

	Daily			Intraday		
Observations	976			438786		
Null Hypothesis	F-Statistic	Probability	Result	F-Statistic	Probability	Result
LOGINDEX does not Granger Cause LOGETF	7.85618	0.0004	Reject the null hypothesis	3170.56	0	Reject the null hypothesis
LOGETF does not Granger Cause LOGINDEX	1.8718	0.1544	Fail to reject the null hypothesis	3845.81	0	Reject the null hypothesis

This table shows the Granger Causality test which tests 2 hypotheses (1) if the change in prices of the index cause a change in the prices of ETF and (2) if the prices of the ETF cause a change in the prices of the index.

Table 6 Vector Error Correction Model

	Daily		Intraday	
	Error Correction		Error Correction	
Observations	975		438785	
	D(LOGETF)	D(LOGINDEX)	D(LOGETF)	D(LOGINDEX)
CointEq1	-0.057182	0.055294	-0.00915	0.001786
t-stat	-0.90238	0.88443	-27.8949	12.2427
D(LOGETF(-1))	-0.416493	0.099681	-0.521152	0.068094
t-stat	-4.01197	0.97325	-335.558	98.5675
D(LOGETF(-2))	-0.313157	-0.058902	-0.261331	0.036581
t-stat	-3.20909	-0.6118	-168.105	52.9008
D(LOGINDEX(-1))	0.427428	-0.088412	0.400484	-0.088692
t-stat	4.01122	-0.84097	112.393	-55.9575
D(LOGINDEX(-2))	0.309626	0.047257	1.82E-01	-2.19E-03
t-stat	3.08781	0.47768	52.3606	-1.41378
C			-6.68E-07	-4.98E-07
t-stat			-0.53206	-0.89182

This table is a summary of the vector error correction model results which show that the results are significant in case of intraday data and insignificant in the case of daily data

4.2 Comparing ETF and Index Performance

4.2.1 Performance Measure

Tables 7 and 8 below show the results of calculating the intraday and daily Sharpe ratio in basis points (bps) for both the index and the fund and Table 9 compares those results. When we analyze the data, it is shown that the performance of the fund cannot be compared to the index as according to the t-stat test for both daily and intraday returns of the ETF and the index the data is statistically insignificant at 5% as shown in Table 9 below. This means that other measures of performance can be more suitable because although ETF and index returns are normally distributed on daily basis, ETF returns are not normally distributed on intraday basis.

Table 7 Average Daily Sharpe Ratio of the EWZ ETF and the MSCI Brazil Index

	ETF	Index	ETF	Index	ETF	Index	ETF	Index	ETF	Index	ETF	Index
year	2010		2011		2012		2013		2014		Period of the study	
Average Daily Return	0.0208	0.0239	-0.1326	-0.1226	-0.0362	-0.0466	-0.0465	-0.0410	0.0072	0.0071	-0.0386	-0.0367
Daily risk	2.1489	2.0912	2.3076	2.3522	1.7081	1.6722	1.9150	1.6625	1.6682	1.7363	1.9721	1.9166
Risk Free Rate	0.0005	0.0005	0.0003	0.0003	0.0004	0.0004	0.0003	0.0003	0.0002	0.0002	0.0004	0.0003
Sharpe Ratio	0.0095	0.0112	-0.0576	-0.0522	-0.0214	-0.0281	-0.0244	-0.0248	0.0042	0.0039	-0.0197	-0.0193

This table shows the average daily performance of the EWZ fund versus the MSCI Brazil on an annual basis using the Sharpe Ratio of each over the same period which is calculated by this equation: $\frac{\overline{Rs},f-\overline{RF}}{\sigma s,f}$. *The values are presented in basis points*

Table 8 Average Intraday Sharpe Ratio of the EWZ ETF and the MSCI Brazil Index

	ETF	INDEX	ETF	INDEX	ETF	INDEX	ETF	INDEX	ETF	INDEX	ETF	INDEX
Year	2010		2011		2012		2013		2014		Period of the study	
Average Intaday Return	0.0003	-0.0002	-0.0001	-0.0004	-0.0001	-0.0001	0.0002	-0.0003	0.0001	-0.0001	0.0001	-0.0002
Intraday Risk	0.3439	0.0831	0.2251	0.0785	0.0782	0.0644	0.2090	0.0661	0.0766	0.0730	0.2184	0.0735
Risk Free Rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sharpe Ratio	0.0829	-0.2187	-0.0598	-0.4855	-0.1304	-0.1462	0.0792	-0.4974	0.1291	-0.1093	0.0267	-0.3079

This table shows the average intraday performance of the EWZ fund versus the MSCI Brazil on an annual basis using the Sharpe Ratio of each over the same period which is calculated by this equation: $\frac{\overline{Rs},f}{\sigma s,f}$.. *The values are presented in basis points*

Table 9 Sharpe Ratio Comparison

Year	2010	2011	2012	2013	2014	Period
ETF - Index (Intraday)	0.3016	0.4258	0.0158	0.5766	0.2385	0.3346
t- Stat	0.3991	0.3266	-0.0229	0.6944	0.4257	0.8178
P- Value	0.6898	0.7440	0.9817	0.4874	0.6703	0.4135
ETF - Index (Daily)	-0.0017	-0.0053	0.0067	0.0004	0.0002	-0.0004
t- Stat	-0.0148	-0.0418	0.0599	-0.0337	0.0007	-0.0217
P – Value	0.9882	0.9667	0.9523	0.9732	0.9994	0.9827

This table shows the difference between the ETF and index Sharpe ratio results in order to show the performance of the ETF versus that of the index.

4.2.2 ETF Tracking Ability

The results in this part are explained in Table 10 below showing the average daily tracking error during the period from 2010 to 2014 on annual basis and average daily trades during the year.

Table 10 Tracking Ability of EW Fund to the MSCI Brazil Index

Year	Average Daily TE	Average Daily Trades
2010	0.12%	416
2011	0.10%	413
2012	0.07%	406
2013	0.12%	392
2014	0.08%	394

This table depicts the average daily tracking error calculated on annual basis by computing the standard deviation of the difference of returns on intraday basis. The average tracking error for the whole period is 0.001018743

It is worth mentioning here that the average intraday mean deviation of returns of the ETF from those of the index when annualized is almost zero percent in all years which means that this tracking error corrects over a year, however, when we look at the average tracking error on daily basis we will find a slight error.

4.2.3 *Tracking Error Persistence*

Table 11 below will show the results of running the regression model[13] illustrated by the equation in section 3.3.3:

Table 11 Tracking Error Persistence Test

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.0009	0.0001	13.3898	0.0000	0.0008	0.0010
X Variable 1	0.1263	0.0302	4.1825	0.0000	0.0670	0.1855

This table provides the results of the regression of the tracking error in day t on itself in day t-1. The value of the β coefficient determines percentage of persistence of the error tin the following day based on the TE today

As depicted in the table above, the beta coefficient of the model is positive and significant indicating that the tracking error persists between two successive days i.e.: traders can expect a deviation the next day, if there is a deviation today, for example: 1% deviation today means that a tracking error of 0.12% will persist the next day.

4.2.4 *Tracking Error Explanation*

The results of the tracking error regression[14] are presented in Table 12 below which explains the results of the model mentioned in section 3.3.4:

[13] Appendix 4

[14] Appendix 5

Table 12 Tracking Error Explanation Model

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.0001	0.0000	14.1173	0.0000	0.0001	0.0001
Volume	-2.14674E-14	0.0000	-2.1881	0.0289	0.0000	0.0000
Risk	0.9945	0.0023	428.7386	0.0000	0.9899	0.9990

This table provides the results of the regression of the factors that affect the tracking error namely volume and risk. The beta coefficients of each the variables indicates the direction and effect of those factors on the tracking error.

According to the above mentioned model, as the beta coefficient of volume is negative, it indicates that as the volume of trade increases the tracking error decreases, on the other hand as the beta coefficient of risk is positive, it indicates that as risk increases the tracking error increases as well.

Chapter VI
Conclusion & Limitations

4.3 Conclusion

This thesis follows the literature on exchange traded funds through studying the EWZ international ETF and its benchmark index the MSCI Brazil. The paper compares 1 minute frequency results to those of daily data in three prominent areas in the literature specifically, price deviation analysis, performance measure & tracking ability, and ETF and index co-integration.

In the price deviation analysis, I looked at whether there is a price difference (premium / discount) between the EWZ and its underlying index on daily and intraday basis. The results here show that as opposed to results found in some of the literature that the prices of the index are higher on average that those of the ETF on both daily and intraday basis with a price deviation of 1.692089 on intraday basis and of 1.692202 on daily basis. It was also concluded that the deviation on daily basis is slightly higher than that on intraday basis.

In the performance measure & tracking ability, the average daily and intraday performance of the ETF as compared to that of the index was calculated through the Sharpe ration. When the daily and intraday results were compared, the Sharpe ratio and hence the performance of both the index and the ETF was higher on intraday basis annually and for the whole period of the study. Also, when looking at the whole period of the study, I find that the ETF outperforms the index by 0.33% on intraday basis but the index outperforms the ETF by 0.0004% on daily basis. In the second area of this section I measured the tracking error of the ETF which is defined as the standard deviation in return differences between ETFs and benchmarks over time. The results showed that there is an average daily tracking error on annual basis and that this error is persistent with a 0.12% rate. It was also concluded that the persistence of the error increases when increasing the risk and decreases with increasing the volume of trade of the ETF.

Lastly, in the ETF – index co-integration analysis, the Pairwise Granger Causality and the Vector Error Correction models where used to determine whether a co-integrated relationship exists between the ETF and the index. The results here were interesting as it was observed that on intraday basis both the ETF and the index move to close the gap if a price deviation exists with a rate of 83.7% and 16.3% respectively. While on the other hand, on daily basis the results show that ETF doesn't affect the index at all as I failed to reject the null hypothesis that the ETF does not Granger cause a change in the index prices but rejected that the index Granger causes the change in the ETF prices, therefore, the index affects the ETF and not vice versa.

4.4 Limitations

There are two main areas that I consider limitations to this study:

Although this study tried to shed the light more on the study of international exchange traded funds using high frequency data, the study looks only at one ETF namely the EWZ Exchange Traded Fund. Therefore this study should be replicated on more international funds where the results can be compared and validated. It is also worth mentioning that intraday data was used before in the study of US based exchange traded funds like the Spider (SPY), the Diamond (DJIA), and the Cube (QQQQ) but to the best of my knowledge no similar study was conducted on international funds.

The other area of development is to study more aspects of the EWZ ETF that were not addressed in this study like the price volatility of the ETF, the effect of dividend distribution and transaction costs on the EWZ ETF performance, predicting the fund's performance, and studying more reasons of tracking error persistence like ETF expenses and the age of the ETF.

References

Aber, Jack W., Li, Dan, and Can, Luc. (2009) “Price volatility and tracking ability of ETFs”, *Journal of Asset Management*,10,210-221.

Ackert, L.F. and Tian, Y.S., (2008) “Arbitrage, liquidity, and the valuation of exchange traded Funds”, *Financial Markets, Institutions & Instruments*, 17, 331-362.

Ansotegui, C, Bassiouny, A and Tooma, E, (2013), “An investigation of intraday price discovery in crosslisted emerging market equities,” *Investment Analysts Journal*, 42:77, 55-67

BlackRock (2012), “The ETP Landscape Global Handbook Q4 2011”, BlackRock Advisors (UK) Limited, London.

Canakgoz, N.A. and Beasley, J.E. (2008), “Mixed-integer programming approaches for index tracking and enhanced indexation”, European Journal of Operation Research, Vol. 196, pp. 384-399.

Charupat, Narat, and Miu, Peter, (2013), Recent Developments in Exchange-Traded Fund Literature Pricing Efficiency, Tracking Ability, and Effects on Underlying Securities, *Managerial Finance,* 39, 427-443.

Corielli, F. and Marcellino, M. (2006), “Factor based index tracking”, Journal of Banking & Finance, Vol. 30, pp. 2215-2233.

Curcio, R.J., Lipka, J.M. and Thornton, J.H. Jr, (2004)“Cubes and the individual investor”, *Financial Services Review*,123-138.

DeFusco, Richard.A., Ivanov, Stoyu.I., and Karels, G.V., (2007)”The Exchange Traded Funds’ Pricing Deviation: Analysis and Forecasts”, *Journal of Economics and Finance*, 35, 181-197.

Delcoure, N. and Zhong, M. (2007), “On the premiums of iShares”, *Journal of Empirical Finance,* 14, 168-195.

DeFusco, Richard A., Ivanov, Stoyu I., and Karels**,** Gordon V. (2009), The exchange traded funds’ pricing deviation: analysis and forecasts, *The Journal of Financial Economics,* 35, 181–197.

DeFusco, Richard.A., Ivanov, Stoyu.I., and Karels, G.V., (2011), The Exchange Traded Funds’ Pricing Deviation: Analysis and Forecasts, *Journal of Economics and Finance*, 35, 181-197.

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

Engle, Robert, and Debojyoti Sarkar. (2006), “Premiums-Discounts and Exchange Traded Funds”. *The Journal of Derivatives*, 13-4, 27-45.

Fred St. Louis. Fred St. Louis Organisation. n.d. 25 April 2017 <https://fred.stlouisfed.org/series/FEDFUNDS#0>.

Frino, A., Gallagher, D.R., Neubert, A.S. and Oetomo, T.N. (2004), “Index Design and Implications for Index Tracking”, *Journal of Portfolio Management*, 30, 89-95.

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

Gastineau, G.L. (2002), “Equity Index Funds have Lost their Way”, *Journal of Portfolio Management*, 28, 55-64.

Gonzalo J, Granger C W J. (1995), “Estimation of common long-memory components in cointegrated systems”, *Journal of Business and Economic Statistics,* 13, 1-9.

Ivanov, Stoyu I.,(2013),” High-Frequency Analysis of Exchange Traded Funds’ Pricing Deviation”, , *Managerial Finance*, 39, 509-524

Qadan, M. and Yagil, J.,(2012), “On the dynamics of tracking indices by exchange traded funds in the presence of high volatility”, *Managerial Finance,* 38, 9, 804-832.

Rompotis, G.G., (2006), “Evaluating the Performance and the Trading Characteristics of iShares”, working paper.

Rompotis, Gerasimos G., (2011),"Predictable Patterns in ETFs' Return and Tracking Error", *Studies in Economics and Finance*, 28, 14 – 35

Appendices

Appendix 1

Null Hypothesis: LOGETF has a unit root
Exogenous: Constant
Lag Length: 10 (Automatic - based on SIC, maxlag=97)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.49785	0.535
Test critical values:	1% level	-3.4302	
	5% level	-2.86136	
	10% level	-2.56671	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGETF)
Method: Least Squares
Date: 04/21/17 Time: 16:15
Sample (adjusted): 12 438788
Included observations: 438777 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGETF(-1)	-2.28E-05	1.52E-05	-1.49785	0.1342
D(LOGETF(-1))	-0.50107	0.00151	-331.901	0
D(LOGETF(-2))	-0.28383	0.001688	-168.094	0
D(LOGETF(-3))	-0.16393	0.001742	-94.1064	0
D(LOGETF(-4))	-0.07508	0.001759	-42.6753	0
D(LOGETF(-5))	-0.04096	0.001762	-23.2428	0
D(LOGETF(-6))	-0.02663	0.001762	-15.1087	0
D(LOGETF(-7))	-0.01584	0.001759	-9.00415	0
D(LOGETF(-8))	-0.01242	0.001742	-7.12911	0
D(LOGETF(-9))	-0.01224	0.001689	-7.24915	0
D(LOGETF(-10))	-0.00656	0.00151	-4.34388	0
C	3.91E-05	2.69E-05	1.453196	0.1462

R-squared	0.2017	Mean dependent var	-5.40E-07
Adjusted R-squared	0.20168	S.D. dependent var	0.000938
S.E. of regression	0.000838	Akaike info criterion	-11.331
Sum squared resid	0.308136	Schwarz criterion	-11.3307
Log likelihood	2485910	Hannan-Quinn criter.	-11.3309
F-statistic	10078.12	Durbin-Watson stat	2.00005
Prob(F-statistic)	0		

Null Hypothesis: LOGINDEX has a unit root
Exogenous: Constant
Lag Length: 5 (Automatic - based on SIC, maxlag=97)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.51798	0.5247
Test critical values:	1% level	-3.4302	
	5% level	-2.86136	
	10% level	-2.56671	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGINDEX)
Method: Least Squares
Date: 04/21/17 Time: 16:15
Sample (adjusted): 7 438788
Included observations: 438782 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGINDEX(-1)	-1.07E-05	7.04E-06	-1.51798	0.129
D(LOGINDEX(-1))	-0.03836	0.00151	-25.4075	0
D(LOGINDEX(-2))	0.022331	0.001511	14.78311	0
D(LOGINDEX(-3))	0.018806	0.001511	12.44865	0
D(LOGINDEX(-4))	0.012623	0.001511	8.356263	0
D(LOGINDEX(-5))	0.009964	0.00151	6.600708	0
C	3.65E-05	2.44E-05	1.497123	0.1344

R-squared	0.002563	Mean dependent var	-5.11E-07
Adjusted R-squared	0.00255	S.D. dependent var	0.000375
S.E. of regression	0.000374	Akaike info criterion	-12.9426
Sum squared resid	0.0615	Schwarz criterion	-12.9424
Log likelihood	2839488	Hannan-Quinn criter.	-12.9425
F-statistic	187.9346	Durbin-Watson stat	2.000031
Prob(F-statistic)	0		

Appendix 2

Null Hypothesis: LOGPD has a unit root
Exogenous: Constant cointegration
Lag Length: 84 (Automatic - based on SIC, maxlag=97)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.10781	0
Test critical values:	1% level	-3.4302	
	5% level	-2.86136	
	10% level	-2.56671	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGPD)
Method: Least Squares
Date: 04/21/17 Time: 16:15
Sample (adjusted): 86 438788
Included observations: 438703 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGPD(-1)	-0.00115	0.000224	-5.10781	0
D(LOGPD(-1))	-0.75993	0.001525	-498.403	0
D(LOGPD(-2))	-0.63296	0.001908	-331.709	0
D(LOGPD(-3))	-0.53669	0.002134	-251.49	0
D(LOGPD(-4))	-0.44148	0.002283	-193.414	0
D(LOGPD(-5))	-0.38066	0.002378	-160.097	0
D(LOGPD(-6))	-0.33654	0.002446	-137.596	0
D(LOGPD(-7))	-0.29972	0.002498	-120.003	0
D(LOGPD(-8))	-0.27398	0.002538	-107.948	0
D(LOGPD(-9))	-0.25445	0.002571	-98.9541	0
D(LOGPD(-10))	-0.23432	0.0026	-90.1344	0
D(LOGPD(-11))	-0.21548	0.002623	-82.1392	0
D(LOGPD(-12))	-0.20142	0.002643	-76.1987	0
D(LOGPD(-13))	-0.18681	0.002661	-70.2124	0
D(LOGPD(-14))	-0.17178	0.002675	-64.2059	0
D(LOGPD(-15))	-0.15465	0.002688	-57.5378	0
D(LOGPD(-16))	-0.14613	0.002698	-54.1653	0
D(LOGPD(-17))	-0.1361	0.002707	-50.2823	0

D(LOGPD(-18))	-0.12807	0.002714	-47.1831	0
D(LOGPD(-19))	-0.1223	0.002721	-44.9503	0
D(LOGPD(-20))	-0.11414	0.002727	-41.8567	0
D(LOGPD(-21))	-0.11081	0.002732	-40.5631	0
D(LOGPD(-22))	-0.10892	0.002737	-39.8013	0
D(LOGPD(-23))	-0.10515	0.002741	-38.3625	0
D(LOGPD(-24))	-0.09988	0.002745	-36.3821	0
D(LOGPD(-25))	-0.09249	0.002749	-33.6491	0
D(LOGPD(-26))	-0.08549	0.002752	-31.0669	0
D(LOGPD(-27))	-0.0827	0.002754	-30.0272	0
D(LOGPD(-28))	-0.0752	0.002757	-27.2802	0
D(LOGPD(-29))	-0.07302	0.002758	-26.472	0
D(LOGPD(-30))	-0.07145	0.00276	-25.8863	0
D(LOGPD(-31))	-0.07214	0.002762	-26.1215	0
D(LOGPD(-32))	-0.06996	0.002763	-25.3175	0
D(LOGPD(-33))	-0.07011	0.002764	-25.363	0
D(LOGPD(-34))	-0.06936	0.002766	-25.0809	0
D(LOGPD(-35))	-0.06351	0.002767	-22.9541	0
D(LOGPD(-36))	-0.05717	0.002768	-20.6553	0
D(LOGPD(-37))	-0.05898	0.002768	-21.3051	0
D(LOGPD(-38))	-0.05979	0.002769	-21.593	0
D(LOGPD(-39))	-0.05738	0.002769	-20.7187	0
D(LOGPD(-40))	-0.05483	0.00277	-19.7973	0
D(LOGPD(-41))	-0.05415	0.00277	-19.5499	0
D(LOGPD(-42))	-0.05313	0.00277	-19.1818	0
D(LOGPD(-43))	-0.05153	0.00277	-18.6052	0
D(LOGPD(-44))	-0.04938	0.00277	-17.8297	0
D(LOGPD(-45))	-0.04882	0.002769	-17.6265	0
D(LOGPD(-46))	-0.04601	0.002769	-16.6172	0
D(LOGPD(-47))	-0.04575	0.002768	-16.5278	0
D(LOGPD(-48))	-0.04272	0.002768	-15.435	0
D(LOGPD(-49))	-0.04221	0.002767	-15.2554	0
D(LOGPD(-50))	-0.04025	0.002766	-14.5526	0
D(LOGPD(-51))	-0.04317	0.002765	-15.6159	0
D(LOGPD(-52))	-0.04648	0.002763	-16.82	0
D(LOGPD(-53))	-0.04228	0.002762	-15.3079	0
D(LOGPD(-54))	-0.03128	0.002761	-11.33	0
D(LOGPD(-55))	-0.03131	0.002759	-11.3474	0
D(LOGPD(-56))	-0.03296	0.002757	-11.9542	0
D(LOGPD(-57))	-0.03068	0.002755	-11.1374	0
D(LOGPD(-58))	-0.03507	0.002752	-12.7422	0
D(LOGPD(-59))	-0.03763	0.00275	-13.6828	0
D(LOGPD(-60))	-0.0365	0.002747	-13.2869	0
D(LOGPD(-61))	-0.0296	0.002743	-10.789	0
D(LOGPD(-62))	-0.03259	0.002739	-11.9008	0

D(LOGPD(-63))	-0.03262	0.002734	-11.9306	0
D(LOGPD(-64))	-0.02834	0.002729	-10.3847	0
D(LOGPD(-65))	-0.02321	0.002724	-8.5205	0
D(LOGPD(-66))	-0.02591	0.002718	-9.53158	0
D(LOGPD(-67))	-0.01998	0.002711	-7.371	0
D(LOGPD(-68))	-0.01793	0.002703	-6.63085	0
D(LOGPD(-69))	-0.01578	0.002694	-5.85691	0
D(LOGPD(-70))	-0.02009	0.002684	-7.48277	0
D(LOGPD(-71))	-0.0185	0.002672	-6.92259	0
D(LOGPD(-72))	-0.01199	0.002657	-4.51377	0
D(LOGPD(-73))	-0.01746	0.002639	-6.61745	0
D(LOGPD(-74))	-0.02401	0.002619	-9.16911	0
D(LOGPD(-75))	-0.02323	0.002595	-8.95164	0
D(LOGPD(-76))	-0.01712	0.002566	-6.66916	0
D(LOGPD(-77))	-0.02212	0.002533	-8.73379	0
D(LOGPD(-78))	-0.0305	0.002492	-12.2396	0
D(LOGPD(-79))	-0.02588	0.002439	-10.6092	0
D(LOGPD(-80))	-0.02031	0.002371	-8.56848	0
D(LOGPD(-81))	-0.01608	0.002275	-7.07038	0
D(LOGPD(-82))	-0.01386	0.002125	-6.52309	0
D(LOGPD(-83))	-0.01112	0.001897	-5.86142	0
D(LOGPD(-84))	-0.00829	0.00151	-5.49308	0
C	0.00194	0.00038	5.108586	0

R-squared	0.368822	Mean dependent var	3.00E-08
Adjusted R-squared	0.368699	S.D. dependent var	0.000926
S.E. of regression	0.000736	Akaike info criterion	-11.5918
Sum squared resid	0.237291	Schwarz criterion	-11.5896
Log likelihood	2542760	Hannan-Quinn criter.	-11.5912
F-statistic	3015.303	Durbin-Watson stat	1.99997
Prob(F-statistic)	0		

Appendix 3

Date: 04/21/17 Time: 16:16
Sample: 1 438788
Included observations: 438783
Series: LOGETF LOGINDEX
Lags interval: 1 to 4

Selected (0.05 level*) Number of Cointegrating Relations by Model

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept	Intercept	Intercept	Intercept	Intercept
	No Trend	No Trend	No Trend	Trend	Trend
Trace	0	1	1	1	2
Max-Eig	0	1	1	1	2

*Critical values based on MacKinnon-Haug-Michelis (1999)

Information Criteria by Rank and Model

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend
	Log Likelihood by Rank (rows) and Model (columns)				
0	5378834	5378834	5378834	5378834	5378834
1	5378837	5379091	5379092	5379133	5379133
2	5378837	5379093	5379093	5379138	5379138
	Akaike Information Criteria by Rank (rows) and Model (columns)				
0	-24.517	-24.517	-24.517	-24.51698	-24.51697
1	-24.517	-24.5181	-24.5181	-24.51832*	-24.51831
2	-24.517	-24.5181	-24.5181	-24.51832	-24.51832
	Schwarz Criteria by Rank (rows) and Model (columns)				
0	-24.5166	-24.5166	-24.5165	-24.51653	-24.51647
1	-24.5165	-24.5176	-24.5176	-24.51774*	-24.51771
2	-24.5164	-24.5175	-24.5175	-24.51761	-24.51761

Appendix 4

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.126249947
R Square	0.015939049
Adjusted R Square	0.015027882
Standard Error	0.001938092
Observations	1082

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	6.57072E-05	6.57072E-05	17.4929948	3.11747E-05
Residual	1080	0.004056695	3.7562E-06		
Total	1081	0.004122402			

Appendix 5

Regression Statistics	
Multiple R	0.9970827
R Square	0.994173911
Adjusted R Square	0.994163122
Standard Error	0.000149128
Observations	1083

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0.004098546	0.002049273	92146.52948	0
Residual	1080	2.40184E-05	2.22393E-08		
Total	1082	0.004122565			