Efficiency-based index tracking optimization model

Abstract

We used asset efficiency as an approach for portfolio selection in a classical index tracking (IT) optimization model. The stock efficiency levels are measured based on the Multifractal-Detrended Fluctuation Analysis (MF-DFA). We form portfolios that seek to replicate the S&P500, the Nikkei 225 and the Ibovespa index, such that our empirical analysis covers three financial markets worldwide (US, Japan and Brazil), using daily stock returns from 2012 to 2021. Our results indicate that using efficiency as a way of selecting the components for tracking portfolios provides good solutions when compared to portfolios without constraint on their size, as well as when compared to another study that was based on a robust numerical approach (a metaheuristic defined as genetic algorithm) to solve the IT problem. We evidenced that using efficiency as a way to form tracking portfolios is an interesting alternative to reduce the number of stocks in the porrtfolios without the need to design a robust methodology to solve the IT optimization especially in developed markets, in which case where the overall level of efficiency tends to be larger (such as US and Japan).

Keywords: Index tracking, Portfolio optimization, Stock efficiency
JEL Codes: C10, C44, C58, G11

1. Introduction

Index tracking (IT) is a passive investment strategy that seeks to replicate the performance of a market benchmark, such as the S&P 500 in the US stock market or the FTSE 100 in the British market. Through tracking, the investor's goal is to earn the overall market return (represented by the benchmark), especially in the long run, instead of pursuing alpha to beat the market. The Efficient Market Hypothesis (EMH, proposed by Fama, 1965a,b; Samuelson, 1965) provides the theoretical background for the choice to replicate the market returns. According to the EMH, prices fluctuate randomly and reflect all information available. Therefore, assets are priced correctly, thus eliminating arbitrage opportunities and abnormal returns (returns above the level befitting the systematic risk). Consequently, investors should be better off using IT to replicate market returns instead of taking additional risk to attempt to add alpha to their portfolios.

Assuming investors opt for the tracking strategy, one of the challenges in the IT problem is to select the stock components in the portfolios. In this paper, we describe an innovative strategy to make this selection: to compose the tracking portfolios based on stock efficiency. By doing so, we aim to explore two direct benefits. First, from a financial viewpoint, we avoid using relatively inefficient stocks in the IT portfolios, i.e., stocks that potentialy have either lower liquidity or higher volatility (or both). If we presume inneficient stocks tend to present increasing volatility, then we would prefer to exclude such stocks from the tracking portfolios, as they would likely damage the portfolios' tracking performance. Second, from a computational viewpoint, as a result of stock selection based on their efficiency level, we eliminate the cardinality constraint in the IT optimization model. Often, IT models use a cardinality constraint to limit the size of each portfolio to a certain number of stocks (for instance, to select a portfolio composed by 20 stocks to track a benchmark composed by 100 stocks). However, such constraint adds computational complexity to the model (because it relies on binary variables), thus making it more difficult to solve the model quickly without implementing a more robust numerical method (such as a heuristic approach). So, eliminating the cardinality constraint represents an advantage regarding the difficulty and time to solve the optimization.

The EMH argues that efficient markets are such that asset prices reflect all information available on each asset and the entire market (Fama, 1970). As corroborated by studies such as Fama and French (2010), one of the consequences of the EMH is the challenge managers face to obtain returns superior to the market in the long run. Therefore, investors could choose the IT strategy to follow the market instead of attempting to overcome the aggregate market return.

The most straightforward approach to select a tracking portfolio would be to make a replica of the index, i.e. a portfolio with the same index components and their respective weights in the index. However, this approach might become unviable, especially if we consider broader indexes such as the S&P 500 (composed by approximately 500 stocks), since buying all index components would lead not only to increasing transaction costs, but also larger operating costs related to the management of a larger portfolio.

Thus, a common choice for IT is to compose tracking portfolios using only a subset of the stocks that comprise the index. In this case, the IT problem becomes how to make the portfolio selection, i.e. how to select the stock components and define their weights in the tracking portfolio. To address the IT problem, the literature contains a variety of studies encompassing distinct IT optimization methods to form portfolios with a reduced number of stocks relative to the total number of index components. In this case, the problem often consists of the minimization of the tracking error volatility (TEV), where the tracking error (TE) is the difference between the portfolio and index returns.

A popular approach to solve the IT problem is the use of heuristics. Beasley et al. (2003) presented an evolutionary heuristic named genetic algorithm to solve the problem, including transaction costs, limit on the number of assets to compose the tracking portfolios, as well as rebalancing controls in the model. The use of the genetic algorithm approach allowed the authors to solve the optimization model relatively quickly, and their findings (based on stock market data for five countries) demonstrated the quality of the heuristic to solve the TEV minimization.

Gaivoronski et al. (2005) investigated the role of the number of assets in IT portfolios, as well as the impact of adjustments to new information available in the market (the rebalancing process of those portfolios), analyzing static and dynamic IT strategies. Guastaroba and Speranza (2012) explored a heuristic defined as Kernel Search to solve the problem, Scozzari et al. (2013) used Differential Evolution, and Sant'Anna et al. (2017) applied a hybrid solution method combining genetic algorithm and nonlinear mathematical programming.

The literature also shows the development of other approaches. Dunis et al. (2005) presented the use of cointegration, which is a method that relies on the application of ordinary least squares (OLS) regression. In this case, the stocks to compose the tracking portfolios must be selected ex-ante, since the OLS regression does not make variable selection. To assess this issue, we also find the use of lasso regression, such as in Wu et al. (2014) and Yang and Wu (2016), which is justified by the fact that the lasso regression allows one to determine the maximum number of independet variables that should present coefficients different from zero in the regression (which works as a constraint on the number of stocks in the portfolio). Alternatively, we also see the use of factor models (for instance, Corielli and Marcellino, 2006; Jiang and Perez, 2021).

Among most of the studies cited above, the common point is a constraint in the optimization model to set the maximum number of stocks to compose each tracking portfolio. In this paper, we present an altenative approach, which is to select the portfolio components based on stock efficiency. Based on a long range of data from both well developed and emerging markets, Tiwari et al. (2019) show that the efficiency of equity markets varies over time and that markets are more efficient in the long term than in the short term. The authors argue that the lower efficiency can be justified by the lack of liquidity in the markets, and that the level of efficiency can be improved with greater transparency of information to investors, greater activity of active investment strategies based on variations in efficiency in the markets, and better trading technologies, among other factors. In the literature, we find the use of stock efficiency for portfolio selection in Maciel (2021), who makes use of efficiency considerations to solve the minimum variance problem. The author optimizes portfolios based on using subsets composed by the most efficient stocks and least efficient stocks in his database. The findings show a good performance for the portfolio using most efficient set of stocks, when compared to the least efficient.

Similarly to Maciel (2021), in this study, we propose using stock efficiency as a parameter for portfolio selection. Instead of estimating an optimization model that includes a cardinality constraint to restrict the size of the portfolio to n stocks out of the N of stocks that compose the database (all index components), first we evaluate the efficiency level of all stocks and select the n most efficient to compose the tracking portfolio (n being the number of stocks chosen by the investor for the portfolio). Then, we solve the index tracking optimization using only the n most efficient stocks to set their weights in the portfolio. By considering asset-level efficiency, our goal is to restrict the space of solutions to some efficiency level, i.e., we reduce the number of assets that are part of the set of feasible solutions to the problem by considering only the most efficient to solve the optimization. In this way, the higher (smaller) the number of stocks to compose the portfolio, the smaller (higher) the minimum stock efficiency level required. Thus, instead of regulating the size of the portfolio via the use of constraints in the optimization model, we use efficiency stock levels for this purpose. As a result, we employ a standard index tracking optimization model to minimize TEV, without the inclusion of constraints to define the maximum number of stocks to compose the portfolio. Thus, the optimization model can be solved promptly without the necessity of using some robust numerical techniques such as an heuristic or metaheuristic.

To test our proposition for the use of stock efficiency for portfolio selection in the index tracking problem, first we select three stock markets: United States, Japan, and Brazil. We base our choice on the FTSE Equity Country Classification[1] from September 2022 (see Table 1 in Section 4.2). According to this classification, United States and Japan are two examples of developed countries, where stock market efficiency levels tend to be larger, and Brazil is an advanced emerging country, where markets might be less efficient than in developed countries. The stock efficient level is measured using the Market Deficiency Measure (MDM) approach as in Maciel (2021), and a variety of portfolios are built using a sample of data from 2012 to 2021, three different sizes of tracking portfolios, and three distinct time windows for portfolio rebalancing overtime. As a result, we are able to obtain a considerable amount of results for different markets and distinct conditions (such as strong bull markets, consistent bear markets, and the crash related to the covid pandemic in 2020).

Our tests are carried out initially based on the construction of tracking portfolios using stock efficiency levels to select the portfolio components. Then, in order to assess the quality of our approach, we apply our methodology using a dataset from another study in the index tracking literature, who used a robust numerical approach based on a metaheuristic defined as genetic algorithm for the IT problem. Our results indicate a good performance of tracking portfolios with efficiency constraints at the asset level. We identified that there is a trade-off between efficiency and tracking error: as we insert efficiency restrictions into the model, and, consequently, restrict the average number of assets in the portfolio, the tracking error increases. This effect is more pronounced for emerging markets, as is the case of Brazil in our sample. Also, when comparing with another method of solving the IT problem in the literature, where it makes use of a more complex approach, we verified the results of models with extremely close efficiency constraints, demonstrating the effectiveness of the proposed model in terms of tracking error.

This study is organized as follows. First, Section 2 describes the methodology used to estimate stock efficiency. Second, Section 3 presents the index tracking optimization model. Then, Section 4 shows the results for the empirical tests, and Section 5 concludes our study.

[1]Provided by the FTSE Russell – a market data provider company affiliated to the London Stock Exchange

2. Efficiency measurement

In this section, we discuss the Multifractal-detrended Fluctuation Analysis (MF-DFA) and the Market Deficiency Measure (MDM), which we use in our study to evaluate asset efficiency level.

Several studies have already been made to test the efficient market hypothesis in its three theoretical distinctions: weak, semi-strong and strong efficiency. Concerning the semi-strong and strong market efficiency (for instance, Keown and Pinkerton, 1981; Patell and Wolfson, 1984; Bardos, 2011), testing these forms is a somewhat laborious task, as they require higher levels of information in order to analyze informational efficiency. On the other hand, the weak form has been extensively tested in the literature due to its focus on using past trading data to analyze if stock prices follow a random walk (Holderness and Sheehan, 1985; Lin and Howe, 1990; Brio et al., 2002). There are numerous methods for testing the weak form, such as serial correlation, variance ratio, unit root and spectral analysis, among others (Lim, 2007).

One of the methods that has been explored in the literature is the Multifractal-Detrended Fluctuation Analysis (MF-DFA), proposed by Kantelhardt et al. (2002). A positive point of this approach is the ability to detect long-term correlations in non-stationary time series. Al-Yahyaee et al. (2018) uses MF-DFA to measure the efficiency for three classes: bitcoin, the gold market, and stock market. In this case, the results show that bitcoin is the least efficient among the three. Al-Yahyaee et al. (2020) use MF-DFA to check efficiency in the cryptocurrency market, as well as to search for the determinants of efficiency. The results indicate that higher liquidity combined with lower volatility help to eliminate existing arbitrage opportunities, thus raising the efficiency level. Several other studies use the MF-DFA to analyze the efficiency of markets (for instance, Tiwari et al., 2019; Zhu and Bao, 2019; Choi, 2021).

Based on the relevance of the MF-DFA to evaluate market and asset efficiency in the literature, we use this method to evaluate stock efficiency in this study. Essentially, MF-DFA collects the volatility of the time series in each time interval, as a statistical point that is used to calculate volatility functions. Then, the Hurst exponents are determined based on the power law of volatility functions. According to Kantelhardt et al. (2002), the methodology considers the following steps.

Let $x(i)$, $i = 1, ..., N$ be a time series of log asset returns, where N is its length. The first step is to determine the profile function, $y(i)$, which can be obtained by the difference between $x(i)$ and its mean, $\bar{x}(i)$, for $i = 1, .., N$:

Step 1. Profile Function:

$$y(i) = \sum_{k=1}^{i} [x(k) - \bar{x}], \tag{1}$$

where $\bar{x}$ comprises the mean of the time series.

Step 2: The profile function $(y(i))$ is divided into $N_s \equiv int(N/s)$ non-overlapping segments, of equal length s. The number of segments N will not necessarily be an integer that is a multiple of the segments s. Thus, a small part at the end of the series can be "left over". In order not to disregard this part of the time series, we repeated the same procedure, this time starting from the opposite end of the series, until its beginning. The result is two N_s segments, so we have $2N_s$.

Step 3: The local trend is calculated for each of the $2N_s$ segments by a least squares fit of the series. From there, the variance is obtained:

$$F^2(s,v) = \frac{1}{s}\sum_{i=1}^{s}\{y[(v-1)s+i] - y_v(i)\}^2, \tag{2}$$

for each segment v, $v = 1, ..., N_s$, and

$$F^2(s,v) = \frac{1}{s}\sum_{i=1}^{s}\{y[N-(v-N_s)s+i] - y_v(i)\}^2 \tag{3}$$

for each segment v, $v = 1, ..., 2N_s$. Here, $y_v(i)$ is the polynomial fitted in the segment v.

Step 4: The q-th order fluctuation function $F_q(s)$ is obtained by averaging all segments (subsets):

$$F_q(s) = \left[\frac{1}{2N_s}\sum_{v=1}^{2N_s}[F^2(s,v)]^{q/2}\right]^{\frac{1}{q}} \tag{4}$$

where $q \neq 0$. For $q = 0$, the value $h(0)$ cannot be determined directly because of the divergent exponent. Instead, a logarithmic averaging procedure should be employed. For $q = 2$, we have a standard DFA procedure (Tiwari et al., 2019).

Step 5: Determine the scaling behavior of the fluctuation functions by analyzing log-log plots of $F_q(s)$ versus each value of q. If the series $x(i)$ is correlated to the power law over a long interval, $F_q(s)$ increases to large values of s, similar to a power law:

$$F_q(s)s^{h(q)} \tag{5}$$

In general, the exponent $h(q)$ will depend on q. If $h(q)$ does not depend on q, the time series is monofractal, otherwise it is multifractal, meaning that the behavior of scaling of small fluctuations ($q < 0$) is different from that of large variations ($q > 0$). We adopt a range from -4 to 4 for q (Hurst exponent), in the same way as Maciel (2021). If $0 < h(q) < 0.5$, the series has anti-persistence. If $0.5 < h(q) < 1$, the time series has persistence. If $h(q) = 0.5$, the stochastic process corresponds to an uncorrelated geometric Brownian motion – a random walk (Maciel, 2021). To determine the asset deficiency level, we used the Market Deficiency Measure (MDM), according to Tiwari et al. (2019) and Maciel (2021):

$$MDM = \frac{1}{2}(|h(q_{min}) - 0.5| + |h(q_{max}) - 0.5|) \tag{6}$$

In this study, $q_{min} = -4$ and $q_{max} = 4$. Thus, the interpretation of the MDM is such that the larger (lower) the MDM value, the lower (larger) the asset efficiency. If MDM $= 0$, the asset or market in question can be considered efficient.

3. Index tracking optimization model

In this section, we describe the index tracking optimization model adopted in our study. We use an optimization model according to which the objective is to minimize the tracking error variance (TEV), the tracking error being the difference between portfolio and index returns. The model is described in Equations 7-9.

Let I be a set of assets $i = 1 : N$ that are part of the composition of the market index that we are attempting to replicate. Let R_t be the index return in t, and $x_{i,t}$ be the weight of asset i in the portfolio in t, and $r_{i,t}$ be the return on asset $i \in I$ in t. Let X_t^* be the portfolio used to replicate the market index in the period t. The objective is to form a portfolio $X_t^* = \{x_{i,t}, i \in I\}$, in each period t, that minimizes the average difference of return between the index and the portfolio. Let ψ be the set of portfolio projection periods. Then, we must build portfolios X_t^*, $\forall\ t \in \psi$, such that the return distance relative to the index is minimized. The frequency by which we restate the weights of assets in X_t^* is determined by the rebalancing interval.

The objective function in Equation 7 is associated with the formulation made by Gaivoronski et al. (2005) and used in several other studies such as in Sant'Anna et al. (2017). It consists of minimizing the mean squared difference between the portfolio return and the index return.

$$\min_{x} \ \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N} x_i r_{it} - R_t\right)^2 \tag{7}$$

s.t.

$$\sum_{i=1}^{N} x_i = 1 \tag{8}$$

$$x_i \geq 0 \qquad \forall\ i \in \{1, \cdots, N\} \tag{9}$$

Constraints 8 and 9 complement the model, such that Constraint 8 defines that the total wealth must be allocated to the portfolio, and Constraint 9 gives the weights a non-negativity property, so that no short positions are allowed in the portfolio.

The quadratic programming (QP) optimization model in Equations 7-9 defines that the tracking portfolio will contain all stocks in the database used for the optimization, i.e., there is no constraint on the size of the portfolio desired by the investor. Naturally, it is possible some assets receive weight equal to zero; nonetheless, the model does not impose some assets must receive weight equal to zero in order to limit the size of the portfolio. To impose a limitation on the size of the portfolio, one could add Constraints 10-12, where C is a parameter defined by the investor to set the maximum number of stocks in the portfolio, and z_i is a binary variable that receives 1 if the stock is included in the portfolio, and 0 otherwise. As a result, when combined, Equations 10-12 form a cardinality constraint to limit the number of portfolio components in the optimization.

$$\sum_{i=1}^{N} z_i \leq C \tag{10}$$

$$z_i \leq x_i \qquad \forall\, i \in \{1, \cdots, N\} \tag{11}$$

$$z_i \in \{0, 1\} \qquad \forall\, i \in \{1, \cdots, N\} \tag{12}$$

Thus, an optimization model using Equations 7-12 would define a portfolio constrained in its size (such as in Sant'Anna et al., 2017). However, when adding Constraints 10-12, we have a QP model with binary variables, which results in a NP-hard problem – a class of problems that presents high computational complexity and increasing time necessary to be solved. Thus, in the context of the index tracking problem, the consequence of this formulation is that if we consider broader indexes, such as the S&P 500 (composed by roughly 500 stocks) or the Russell 1000 (about 1,000 stocks), and we desire to form reduced portfolios, with a small number of stocks relative to the index composition, then there would be an increase in the computing time necessary to solve the optimization. For this reason, a variety of numerical approaches have already been used in the literature, such as heuristics and metaheuristics, as described in the Introduction.

3.1. Efficiency-constrained model

In this paper, we propose to define tracking portfolios based on stock level efficiency. To do so, our approach is to consider the optimization model defined in Equations 7-9, without including the constraints to limit the size of the portfolio. Instead, we perform the stock selection ex-ante based on the efficiency level computed for each stock using the MF-DFA approach described in Section 2.

First, we estimate the efficiency level for each stock in our database. Second, we rank the stocks from best (highest efficient) to worst (lowest efficient). Third, supposing our goal is to compose a portfolio using 10 stocks for instance, then we select the top 10 stocks with the highest efficiency level. Last, we solve the optimization defined by Equations 7-9 using only the subset composed by the highest efficient stocks. By doing so, we eliminate the cardinality constraint in the optimization, thus being able to solve the optimization instantly.

As described in detali in Section 4.1, we selected three market benchmarks for our empirical analysis. For each of the three indexes, we compose four portfolios.

The first one is defined as the benchmark model (M1-B), which is done using the entire dataset of index components, i.e., we do not restrict the size of the tracking portfolios. Thus, we use these portfolios as references for comparison with the restricted portfolios. Second, we compose three restricted portfolios, each one with a different size. Since the choice for the size of the portfolios would be based on the investor's preferences, we use three distinct sizes for the purpose of stressing our selection approach based on stock efficiency.

i. portfolios using half of the number the stocks that compose the index (M2-E-M – the median of the list of stocks ranked according to their efficiency);

ii. portfolios using 35% of the number of stocks that compose the market benchmark (M2-E-P35 – the 35th percentile of the ranking);

iii. portfolios using 25% of the number of index components (M2-E-1Q – the first quartile of the ranking).

As mentioned in the Section 2, the lower the MDM for a given asset (i.e. its efficiency level as defined in Equation 6), the more efficient it is. Thus, selecting assets that have an MDM below the median of the MDM distribution, comprises selecting the most efficient half of the distribution. When the constraint changes to the 35th percentile, in the M2-E-P35 model, we are being more rigorous relative to the level of efficiency that the stocks need to be included in the portfolios. The M2-E-1Q model is the most demanding, as it only considers assets that have an efficiency level that are in the first quartile of the MDM distribution.

As we make the efficiency constraint more severe in the model, i.e., we move from the median (M2-E-M model) to the first quartile (M2-E-1Q model), the number of assets is reduced considerably, and the tracking error (TE) increases. This cost of maintaining a tracking portfolio with a low number of assets, the TE, should be more pronounced in emerging markets (which is the case of Brazil in our set of three countries), where overall efficiency levels tend to be lower. Tracking portolios in markets with lower efficiency levels would be more penalized by the efficiency constraint, presenting a more restricted set of assets, which do not necessarily present a good relationship with the target index.

4. Empirical results

This section is concerned with presenting the results of the study. In Section 4.1.1, the data used are described, and the procedures conducted in the empirical tests are presented in Section 4.1.2;

in 4.2, we comment on the results obtained in measuring the efficiency of assets in the different markets analyzed; then, Section 4.3 presents the main results for the efficiency constrained portfolios; last, in Section 4.4, we compare our results with the findings from another study (Sant'Anna et al., 2017), who used a metaheuristic approach defined as genetic algorithm.

4.1. Data and empirical strategy

4.1.1. Data

To carry out the empirical research procedures, we selected three markets: the US market, the Japanese market and the Brazilian market. Thus, we have two developed markets (USA and Japan) and an emerging one (Brazil), which will help to identify differences in the behavior of projected portfolios in markets with different patterns of efficiency, liquidity and volatility. Table 1 presents the FTSE 2022 annual classification of equity markets, showing the classes in which the mentioned markets show up.

Among all countries listed in Table 1, we chose the North American and Japanese markets as two examples of developed markets, considering the large availability of historical data for both of them and their relevance in the financial markets worldwide (roughly USD 23 trillion and USD 5 trillion in trading volume, respectively[2]). Additionally, we selected the Brazilian market as an example of emerging market, considering Brazil is one of the most representative among the emerging markets in terms of trading volume (roughly USD 1 trillion in 2019).

TABLE 1 HERE

For the American market, the portfolios' target index is the S&P500, which is one of the most famous indices in the world. For this market, our dataset comprises the daily returns of the 505 assets that were part of this index in February 2022, plus the index itself, from January 2010 to December 2021. For the Japanese market, we selected as a target index the Nikkei 225, which is one of the most popular in that market, and the sample of its components in February 2022, comprising 224 assets, in the same data range (Jan/10 to Dec/21). Finally, for the Brazilian market, the Ibovespa was selected as the target index for the portfolios, since this is the most representative benchmark in Brazil. As a dataset, we selected the assets that make up the Ibovespa in February 2022, comprising 93 assets, for the same mentioned interval (Jan/10 to Dec/21). The daily returns of the datasets are adjusted for splits, mergers and dividend payments. Therefore, our tests are conducted considering a large range of data, covering periods of bull and bear markets, and in portfolios that seek to replicate both smaller indices, such as the Ibovespa, and broader indices, such as the S&P500.

[2] Information obtained from the World Bank database <https://data.worldbank.org/>

4.1.2. Empirical strategy

The portfolio projection approach used in this study is based on rolling windows. To compose each portfolio, we use a data interval (in-sample data) of $t = 120$ trading days immediately prior to the initial portfolio projection date (as in Filomena and Lejeune, 2012, 2014) to compute the optimization model. After the portfolio is formed, its result is analyzed during the subsequent period of 120 trading days for portfolios using semiannual updates, 240 trading days for portfolios using annual updates, and 480 trading days for portfolios using a time window of roughly two years before the next update. The first portfolio is estimated using data from July 2011 to December 2011, and its projection starts in January 2012, moving in a rolling window until December 2021, thus comprising a 10-year projection.

Regarding efficiency, we also adopt a dynamic approach, differently from Maciel (2021), which performs the estimation of efficiency only before the beginning of the out-of-sample period (projection period for the portfolios), and considers that the assets have a static efficiency throughout rebalancing. This study makes use of three years of data before the out-of-sample period to estimate asset efficiency. In a different way, our study uses one year of daily returns immediately prior to each rebalancing to estimate efficiency, instead of three years. This means that, for each rebalancing, we use a new, more up-to-date set of asset efficiency data, unlike the one used by Maciel (2021), which measures asset efficiency only once for portfolio projection. Considering all markets and the entire range, we have a total of 798 portfolios formed in this study, demonstrating the extent of the empirical procedures that will be dealt with in this section.

To measure the tracking performance of the portfolios in the projection interval, we use the tracking error (TE), that we define as the variance of the difference between index and portfolio daily tracking returns, as in Beasley et al. (2003):

$$TE^* = \frac{1}{T^*} \left[\sum_{i=1}^{T^*} |r_t^p - R_t^*|^2 \right]^{\frac{1}{2}} \tag{13}$$

where T^* represents the total range of out-of-sample periods; $r_{i,t}^*$ represents the return on asset i for the out-of-sample period $t \in T^*$; therefore, $r_t^{p^*} = \sum_{i=1}^{I} x_i r_{i,t}^*$ represents the return of the portfolio formed in the out-of-sample interval; and R_t^* is the index return in $t \in T^*$.

The turnover of portfolios cannot be ignored either, being extremely important when choosing suitable models. Turnover is a proxy for the amount of trades carried out by the portfolios over time. Thus, the greater the turnover, the greater the amount of asset trades in the time interval. In this way, we measure the average monthly turnover of the portfolios based on the following formulation, used by Sant'Anna et al. (2020):

$$\left[\sum_{p=2}^{np}\left(\frac{\sum_{i=n}^{N}|x_i^p - x_i^{p-1}|}{2}\right)\right] \times \frac{1}{f} \tag{14}$$

where np is the number of portfolios formed for each model; p and $p-1$ are the rebalancing time instants; and $f = 6$ for the semiannual rebalancing, 12 for the annual rebalancing, and 24 for de two years interval of rebalancing.

4.2. Efficiency levels

The FTSE Russell classifies capital markets according to their level of development, as can be seen in Table 1, which shows the Equity Country Classification's September 2022 classification. The level of efficiency also varies across markets, where emerging markets tend to be less efficient in terms of risk-return pricing than developed markets.

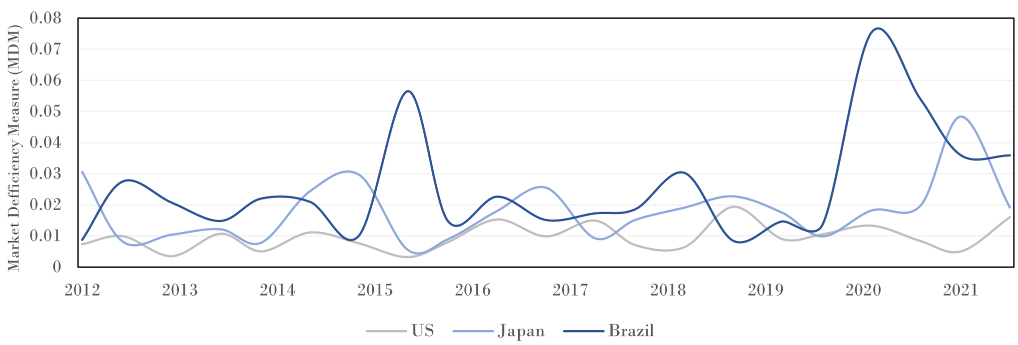

Figure 1: Minimum of the Market Efficiency Measure (MDM) per market.

Figure 1 shows the trajectory of the least efficiency level (i.e. the least Market Deficiency Measure) among all stocks per country in our databse, along the rebalancing intervals. For each market, and for the portfolios that are rebalanced at an interval of 120 days (semi-annual). As a result, we notice predominance of US market in terms of efficiency, followed by Japan and, finally, Brazil. The Japanese market has shown a trajectory of volatile efficiency over time, in the same way as the Brazilian market, where the latter is the most affected in periods of high volatility, such as the period of the covid-19 crisis. In this period, we observe an increase in MDM, demonstrating a loss of efficiency for all markets, especially for Brazil.

4.3. Optimization results

In this section, we will address the main results obtained with the empirical procedures. After estimating all portfolios per market, for each of the three rebalancing windows, Table 2 presents the descriptive results for all efficiency based portfolios.

For the US market, in the 120-day (semi-annual) rebalancing strategy, we observed that all models, including the benchmark model (M1-B), have an accumulated return in the out-of-sample period that is higher than the accumulated return of the S&P500 index. In terms of correlation, we observe that this is slightly lower for the models with efficiency constraints (M2 models), when compared to the benchmark model, and the correlation decreases as the efficiency constraint becomes more severe, i.e., for the M2-E-P35 and M2-E-1Q models, which have an efficiency constraint at the asset level that comprises the 35th percentile and first quartile of the efficiency distribution in each rebalancing, respectively (remembering that, the lower the MDM, more efficient is the asset).

However, it is worth mentioning that the M2 models also show a considerable reduction in the number of assets that are part of the portfolio, having, on average, 119 assets in the M2-E-1Q model, with a more severe restriction of efficiency, compared to 474 in the benchmark model (M1-B). As the model requires that only more efficient assets be part of the feasible solutions space, we expect fewer assets to compose the portfolios of these models. As will be seen below, when we look at the tracking error (TE), we expect it to increase as we reduce the number of assets that are part of the portfolios.

For the Japanese market, in a slightly different way from the US market, we observe a greater proximity of the accumulated return in the period of the models with efficiency restrictions to the accumulated return of the Nikkei 225 index. For example, in the case of rebalancing in an interval of 120 days, for the M2-E-P35 model, we have a cumulative return of 267%, while the Nikkei's cumulative return was 241%, and that of the M1-B model was 335%. It is also worth mentioning the difference between the average number of assets that make up the portfolios of each model. In the 120-day rebalancing strategy, while the M2-E-P35 model uses, on average, 75 assets, the M1-B model uses, on average, 201. Correlation of returns with the return of the index presents, in the same way as the US market, a downward trend as the models demand more efficient assets in their composition.

For the Brazilian market, we noticed some differences from the other markets analyzed, precisely because of the difference in efficiency and volatility standards existing between developed and emerging markets. In Brazil, dealing with an emerging market and with higher volatility patterns, we have lower correlations between the returns of the Ibovespa index and the models, including for the benchmark model (M1-B). The correlation of the model returns with the index return also decreases as the number of assets decreases, i.e., moving from the M1-B model (unrestricted model) to the M2-E-1Q model (with severe efficiency restriction).

TABLE 2 HERE

Table 3 shows the cumulative tracking error per year for each of the three countries (Brazil, Japan, and United States). The TE described in the table is defined in equation 13. For the US market, we observed that the best average TE is always the M1-B model, which presents, for all rebalancing strategies, a lower TE than the models with efficiency constraints. The results show that, as the efficiency constraint becomes more severe in the model, i.e., we require more efficient assets, the average number of assets that are part of the portfolios decreases, and the TE increases slightly. For example, considering the semiannual rebalancing strategy in the US market, the benchmark model (M1-B) has an average number of assets of 474, and an average TE of 0.10% over the entire period. On the other hand, the model with efficiency constraint on the median of the Market Deficiency Measure (MDM) distribution, the M2-E-M, has an average number of assets in the period of 237, about half the number used by the M1-B model, with an average TE of 0.16%. Thus, the M2-E-M model demonstrates that the cost of reducing the average amount of assets by 50% is a 0.06% increment in the average tracking error.

Also, comparing the same two models, and looking at the year 2020, period of the covid-19 shock in the markets, we notice that the two models have the same TE. This means that, in this specific case, portfolios formed by the M2-E-M model showed the same level of vulnerability to crisis, in terms of TE, as the portfolios formed by the M1-B model.

For the Japanese market, we also see the same movement in terms of the average number of assets and TE. The efficiency constraint helps to reduce the average number of assets that are part of the portfolios, making them more manageable than larger portfolios, however, this reduction comes with a cost, which is the TE. In the 120-day rebalancing strategy, for example, the average TE for the entire out-of-sample period (projection period) was 0.18%, 0.26%, 0.31% and 0.35%, for the M1-B, M2-E-M, M2-E-P35 and M2-E-1Q models, respectively, while the average number of assets was 201, 108, 75 and 54, respectively. A highlight for the volatility of the models with efficiency restriction, which does not present a substantial increase in relation to the benchmark model and the Nikkei 225.

Regarding the Brazilian market, which is the emerging market in our set of analyzed markets, we observed a general level of tracking error (TE) higher than the other markets, precisely because of the difference in terms of volatility and structure in these markets. Looking at the 120-day (semi-annual) rebalancing strategy, the Brazilian market has an average TE for all models of 0.74%, compared to 0.17% for the US market and 0.28% for the Japanese market. As evidenced by several studies, as in Sant'Anna et al. (2017), high volatility environments make tracking strategies difficult, increasing the tracking error of models more severely than in more developed markets, where the level of volatility tends to be lower. In the same way as for the US and Japanese markets, we observed a lower TE for the unrestricted model in all rebalancing strategies, which is expected, given that it has an average number of assets considerably higher than the efficiency-constrained models. Looking specifically at the 120-day rebalancing strategy, the average TE of the benchmark model (M1-B) is 0.62% in the period, with an average number of assets of 70. Meanwhile, the model with the most severe efficiency constraint (M2-E-1Q), has an average TE of 0.84%, with an average of 18 assets. As an example, Figure 2 illustrates the trajectory of the value of portfolios formed in the US market with the M1-B and M2-E-M models, with the semiannual update strategy, and with the S&P500 index.

TABLE 3 HERE

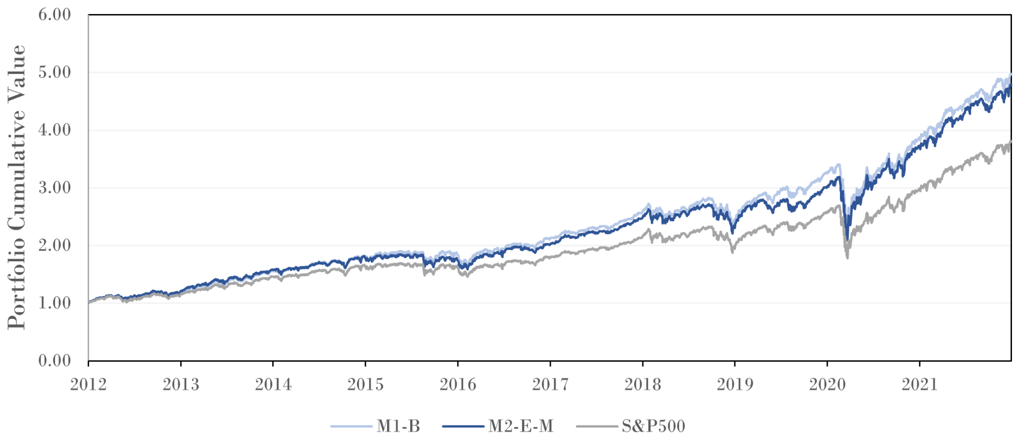

Figure 2: Minimum of the Market Efficiency Measure (MDM) per market.

4.4. Method comparison

As we mentioned in the Introduction, the nature of the index tracking problem is to replicate a market index with a limited number of assets. The mathematical formulation of this problem is often treated as an mixed-integer quadratic programming problem (MIQP), making it an NP-hard problem as mentioned in Section 3 (Coleman et al., 2006). Thus, in order to find a solution to the problem in a reasonable time, several methods have been formulated, however, many end up with high computational complexity. Sant'Anna et al. (2017) uses a hybrid approach with a genetic algorithm and nonlinear mathematical programming, obtaining good tracking results for the Brazilian market with a very small number of assets – 5 and 10 assets, out of a set of 67 assets, and also for developed markets (USA, UK and Germany). These results are achieved in less than 10 minutes of computational processing.

In order to compare the performance of models with efficiency constraint with some method already used in the literature, we selected the study by Sant'Anna et al. (2017) for this purpose. To have an adequate basis for comparison, we need to use the same number of assets in the portfolios obtained by Sant'Anna et al. (2017). Then, we raise the model's efficiency constraint until the number of assets in that study is reached, i.e., 5 and 10 assets. The portfolio projection range (out-of-sample) starts in January 2010 and runs until July 2012, using daily returns from the 67 assets that make up the Ibovespa index (target index), during the same period. This study made use of rebalancing strategies of 20, 60, 120 and 240 trading days (monthly, quarterly, semiannually and annually, respectively).

Therefore, to make the comparison with the aforementioned study, first, we performed the efficiency estimations of each of the 67 assets that are part of the assets set, using a year to estimate the Market Deficiency Measure (MDM) immediately prior to the initial date of portfolio projection. As for the portfolios designed and analyzed in the 4.3 section, we use a dynamic efficiency approach, measuring the MDM at each rebalancing, in order to have an updated measure to be used in the optimization process.

To obtain the same number of assets as the portfolios obtained by Sant'Anna et al. (2017), we analyze, at each rebalancing, the distribution of the MDM of the assets, in order to place an efficiency constraint that limits the space of feasible solutions to just the number of assets from the compared portfolios, i.e., 5 and 10 assets. Thus, we have extremely severe efficiency constraints in this approach, limiting the solution space to the 5 and 10 most efficient assets in each rebalancing interval.

We list the comparison of the results obtained by the hybrid approach of Sant'Anna et al. (2017), which makes use of a genetic algorithm and nonlinear mathematical programming, with the results of the efficiency-constrained models addressed in the present study. Table 4 shows the results obtained by the study of Sant'Anna et al. (2017), the results obtained by the models with efficiency restrictions, and the difference between the average TE obtained between the aforementioned study and the model formulated in the present research.

The results demonstrate a slight superiority of the hybrid approach between genetic algorithm and nonlinear mathematical programming, in terms of average tracking error (TE), where the hybrid solution approach has a slightly lower average TE than the model with efficiency constraints. However, as mentioned, this approach has a high computational complexity, and its results, despite taking a low computational processing time (less than 10 minutes), are not instantaneous. Although the efficiency constrained model presents a slightly worse result in terms of average TE, this problem is simply formulated through quadratic programming (QP), where the efficiency constraints present a cardinality constraint function, limiting the number of assets that are part of the set of feasible solutions as the efficiency requirement becomes more severe. Since this is a QP problem, the efficiency constrained approach presents an instantaneous result and relatively low computational complexity. Thus, we obtained similar results, in terms of average TE, employing a relatively simpler approach, through a QP formulation, when comparing the results of a more laborious method, which involves high computational complexity and does not provide an instant solution, despite get results in less than 10 minutes of processing.

TABLE 4 HERE

It is worth noting that when we limit the space of feasible solutions to only extremely efficient assets, as in the case of 5 assets, for example, we are not necessarily choosing the assets with the best relation to the index over time, but rather limiting our set of solutions to the most efficient assets in the period. Thus, the benefit of considering efficiency constraints (reduction in the number of assets) for cases of severe constraints, may end up being the main disadvantage in terms of TE, thus showing the trade-off between efficiency and tracking error.

5. Conclusions

In general, given the results presented in the previous subsections, we argue for the relevance of considering efficiency constraints in the formation of tracking portfolios. As mentioned, the index tracking (IT) problem consists of replicating the returns of a market index with a limited amount of assets. Normally, it is formulated through a problem of minimizing the quadratic difference between the return of the portfolio and the index in a given time interval. And, to limit the amount of assets, integer constraints are normally used, having, finally, a mixed-integer quadratic programming problem (MIQP), which has high computational complexity, and normally requires considerable processing time, also being characterized as an NP-hard problem.

In this study, we used an alternative formulation to solve the index tracking (IT) problem, reducing the level of computational complexity and obtaining instant solutions to the optimization problem. In particular, we explore the Market Deficiency Measure (MDM), measured by the Multifractal-detrended Fluctuation Analysis (MF-DFA), as a measure for stock efficiency, so that we can perform the selection of the portfolio components based on their efficiency levels.

We project portfolios with efficiency constraints comparing the results with an unrestricted model (benchmark) in three markets with different structures: US (developed), Japanese (developed) and Brazilian (emerging), from January 2012 to December 2021, comprising 10 years of projection. The results demonstrate the existence of a trade-off between tracking error (TE) and the average number of assets in the portfolios: insofar as we demand more efficient assets to compose the portfolios, and, consequently, we generate portfolios with a smaller number of assets, the TE increases. We demonstrate that this effect is greater in markets with lower levels of efficiency, such as the Brazilian market (emerging market in our set). For more efficient markets, the cost of increasing TE for a more severe constraint on efficiency, and reducing the number of assets, becomes lower, as demonstrated by the tracking error cost (TE Cost).

We compared our results with a method already used in the literature, which uses a hybrid approach with genetic algorithm and nonlinear mathematical programming. We demonstrate that, although our results are not absolutely better, they are very close to those obtained with the aforementioned method, which is substantially more complex from a computational point of view, and does not provide an instant solution, despite taking less than 10 minutes of computational processing time. Thus, with efficiency constraints, we achieve good results and instant solution, with a problem of low computational complexity.

We also note that efficiency restrictions have a greater impact on the tracking error (TE) of portfolios in emerging markets, which tend to have lower levels of efficiency than developed markets, where the inclusion of restrictions penalizes TE more severely than in developed markets with high levels of efficiency. Thus, the use of efficiency constraints appears to be an interesting alternative to reduce the size of tracking portfolios with a lower cost in terms of tracking error (TE cost).

References

Al-Yahyaee, K. H., Mensi, W., Ko, H. U., Yoon, S. M., Kang, S. H., 2020. Why cryptocurrency markets are inefficient: The impact of liquidity and volatility. North American Journal of Economics and Finance 52.

Al-Yahyaee, K. H., Mensi, W., Yoon, S. M., 2018. Efficiency, multifractality, and the long-memory property of the bitcoin market: A comparative analysis with stock, currency, and gold markets. Finance Research Letters 27, 228–234.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Appel, I. R., Gormley, T. A., Keim, D. B., 2016. Passive investors, not passive owners. Journal of Financial Economics 121, 111–141.

Bardos, K. S., 2011. Quality of financial information and liquidity. Review of Financial Economics 20, 49–62.

Beasley, J. E., Meade, N., Chang, T. J., 2003. An evolutionary heuristic for the index tracking problem. European Journal of Operational Research 148, 621–643.

Brio, E. B. D., Miguel, A., Perote, J., 2002. An investigation of insider trading profits in the spanish stock market. The Quarterly Review of Economics and Finance 42, 73–94.

Busse, J. A., Goyal, A., Wahal, S., 2010. Performance and persistence in institutional investment management. The Journal of Finance 65 (2), 765–790.

Cajueiro, D. O., Tabak, B. M., 2004. The hurst exponent over time: Testing the assertion that emerging markets are becoming more efficient. Physica A: Statistical Mechanics and its Applications 336, 521–537.

Choi, S. Y., 2021. Analysis of stock market efficiency during crisis periods in the us stock market: Differences between the global financial crisis and covid-19 pandemic. Physica A: Statistical Mechanics and its Applications 574.

Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and market efficiency. Journal of Financial Economics 87, 249–268.

Coleman, T. F., Li, Y., Henniger, J., 2006. Minimizing tracking error while restricting the number of assets. Journal of Risk 8 (4), 33.

Corielli, F., Marcellino, M., 2006. Factor based index tracking. Journal of Banking & Finance 30, 2215–2233.

Dunis, C. L., Ho, R., et al., 2005. Cointegration portfolios of european equities for index tracking and market neutral strategies. Journal of Asset Management 6 (1), 33–52.

Fama, E. F., 1965a. The behavior of stock-market prices. The journal of Business 38 (1), 34–105.

Fama, E. F., 1965b. Random walks in stock market prices. Financial Analysts Journal 21 (5), 55–59.

Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work.

Fama, E. F., French, K. R., 2010. Luck versus skill in the cross-section of mutual fund returns. The journal of finance 65 (5), 1915–1947.

Filomena, T. P., Lejeune, M. A., 2012. Stochastic portfolio optimization with proportional transaction costs: Convex reformulations and computational experiments. Operations Research Letters 40, 212–217.

Filomena, T. P., Lejeune, M. A., 2014. Warm-start heuristic for stochastic portfolio optimization with fixed and proportional transaction costs. Journal of Optimization Theory and Applications 161, 308–329.

French, K. R., 2008. The cost of active investing. The Journal of Finance 63, 1537–1573.

Frino, A., Gallagher, D. R., 2001. Tracking s&p 500 index funds. The Journal of Portfolio Management 28 (1), 44–55.

Gaivoronski, A. A., Krylov, S., Van der Wijst, N., 2005. Optimal portfolio selection and dynamic benchmark tracking. European Journal of operational research 163 (1), 115–131.

Goyenko, R. Y., Holden, C. W., Trzcinka, C. A., 2009. Do liquidity measures measure liquidity? Journal of Financial Economics 92, 153–181.

Guastaroba, G., Speranza, M. G., 2012. Kernel search: An application to the index tracking problem. European Journal of Operational Research 217, 54–68.

Holderness, C. G., Sheehan, D. P., 1985. Raiders or saviors? the evidence on six controversial investors. Journal of Financial Economics 14, 555.

Hshieh, S., Li, J., Tang, Y., 2021. How do passive funds act as active owners? evidence from mutual fund voting records. Journal of Corporate Finance 66, 101692.

Jiang, P., Perez, M. F., 2021. Follow the leader: Index tracking with factor models. Journal of Empirical Finance 64, 337–350.

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., Stanley, H. E., 2002. Multifractal detrended fluctuation analysis of nonstationary time series. Physica A: Statistical Mechanics and its Applications 316, 87–114.

Keown, A. J., Pinkerton, J. M., 1981. Merger announcements and insider trading activity: An empirical investigation. The Journal of Finance 36, 855–869.

Le Tran, V., Leirvik, T., 2019. A simple but powerful measure of market efficiency. Finance Research Letters 29, 141–151.

Lim, K. P., 2007. Ranking market efficiency for stock markets: A nonlinear perspective. Physica A: Statistical Mechanics and its Applications 376, 445–454.

Lin, J.-C., Howe, J. S., 1990. Insider trading in the otc market. The Journal of Finance 45, 1273–1284.

Maciel, L., 2021. A new approach to portfolio management in the brazilian equity market: Does assets efficiency level improve performance? Quarterly Review of Economics and Finance 81, 38–56.

Malkiel, B. G., 1995. Returns from investing in equity mutual funds 1971 to 1991. The Journal of Finance 50, 549–572.

Patell, J. M., Wolfson, M. A., 1984. The intraday speed of adjustment of stock prices to earnings and dividend announcements. Journal of Financial Economics 13, 223–252.

Samuelson, P. A., 1965. Rational theory of warrant pricing. Industrial Management Review 6 (2), 13–39.

Sant'Anna, L. R., Caldeira, J. F., Filomena, T. P., 2020. Lasso-based index tracking and statistical arbitrage long-short strategies. The North American Journal of Economics and Finance 51, 101055.

Sant'Anna, L. R., Filomena, T. P., Guedes, P. C., Borenstein, D., 2017. Index tracking with controlled number of assets using a hybrid heuristic combining genetic algorithm and non-linear programming. Annals of Operarions Research 258, 849–867.

Scozzari, A., Tardella, F., Paterlini, S., Krink, T., Scozzari, A., Tardella, F., Paterlini, S., Krink, T., 2013. Exact and heuristic approaches for the index tracking problem with ucits constraints. Annals of Operations Research 205, 235–250.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 425–442.

Sharpe, W. F., 1991. The arithmetic of active management. Financial Analysts Journal 47, 7–9.

Tiwari, A. K., Aye, G. C., Gupta, R., 2019. Stock market efficiency analysis using long spans of data: A multifractal detrended fluctuation approach. Finance Research Letters 28, 398–411.

Vieira, E. B. F., Filomena, T. P., Sant'Anna, L. R., Lejeune, M. A., 2021. Liquidity-constrained index tracking optimization models. Annals of Operations Research, 1–46.

Wu, L., Yang, Y., Liu, H., 2014. Nonnegative-lasso and application in index tracking. Computational Statistics & Data Analysis 70, 116–126.

Yang, Y., Wu, L., 2016. Nonnegative adaptive lasso for ultra-high dimensional regression models and a two-stage method applied in financial modeling. Journal of Statistical Planning and Inference 174, 52–67.

Zhu, X., Bao, S., 2019. Multifractality, efficiency and cross-correlations analysis of the american etf market: Evidence from spy, dia and qqq. Physica A: Statistical Mechanics and its Applications 533, 121942.

Table 1: FTSE Equity Country Classification – September 2022.

Developed	Advanced Emerging	Secondary Emerging	Frontier
Australia	Brazil	Chile	Bahrain
Austria	Czech Republic	China	Bangladesh
Belgium/Luxemburg	Greece	Colombia	Botswana
Canada	Hungary	Egypt	Bulgaria
Denmark	Malaysia	Icenland	Côte d'Ivoire
Finland	Mexico	India	Croatia
France	South Africa	Indonesia	Cyprus
Germany	Taiwan	Kuwait	Estonia
Hong Kong	Thailand	Pakistan	Ghana
Ireland	Turkey	Phillipines	Jordan
Israel		Watar	Kazakhstan
Italy		Romania	Kenya
Japan		Saudi Arabia	Latvia
Netherlands		United Arab Emirates	Lithuania
New Zeland			Malta
Norway			Mauritius
Poland			Morocco
Portugal			Nigeria
Singapore			Oman
South Korea			Palestine
Spain			Peru
Sweden			Macedonia
Switzerland			Serbia
UK			Slovak Republic
USA			Slovenia
			Sri Lanka
			Tanzania
			Vietnam

Table 2: Descriptive results for efficiency based portfolios per country

		Portfolio rebalancing frequency											
Country	Index	120 days (6 months)				240 days (1 year)				480 days (2 years)			
Brazil	Ibovespa	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q
Min	-14.78%	-14.75%	-14.72%	-16.03%	-14.07%	-14.75%	-14.72%	-16.03%	-14.07%	-15.27%	-14.92%	-16.32%	-14.07%
Max	13.91%	14.40%	14.44%	15.19%	12.93%	14.40%	14.44%	15.19%	12.93%	14.82%	14.55%	15.35%	12.93%
Annual Volatility	25.24%	22.04%	22.48%	23.10%	22.93%	22.49%	23.33%	24.20%	24.40%	23.15%	24.06%	25.23%	24.85%
Cumulative Return	85%	120%	223%	169%	76%	326%	496%	407%	177%	373%	410%	375%	164%
Correlation	1.000	0.887	0.870	0.851	0.829	0.895	0.878	0.857	0.832	0.905	0.885	0.879	0.852
Avg. Number of Assets	-	70.19	35.52	25.00	17.76	71.55	35.91	25.27	17.82	73.50	36.83	25.67	17.67
Monthly Avg. Turnover	-	9.02%	12.12%	12.71%	13.29%	4.79%	6.95%	7.64%	7.53%	2.55%	3.56%	3.84%	4.08%
Japan	Nikkei 225	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q
Min	-7.92%	-7.99%	-7.90%	-8.30%	-8.07%	-8.25%	-7.94%	-7.93%	-7.93%	-8.27%	-8.19%	-7.93%	-7.93%
Max	8.04%	7.87%	7.99%	7.49%	7.69%	8.07%	8.06%	7.31%	7.16%	8.28%	8.06%	7.31%	7.16%
Annual Volatility	20.66%	20.68%	20.61%	20.67%	20.73%	20.84%	20.74%	20.81%	21.02%	21.31%	21.29%	21.37%	21.39%
Cumulative Return	241%	335%	279%	267%	231%	330%	338%	316%	245%	348%	262%	212%	212%
Correlation	1.000	0.989	0.977	0.969	0.961	0.987	0.977	0.968	0.955	0.988	0.977	0.966	0.949
Avg. Number of Assets	-	201.19	107.57	75.43	54.05	204.36	109.18	76.45	55.00	209.17	111.50	78.33	56.00
Monthly Avg. Turnover	-	8.83%	12.21%	13.08%	13.95%	9.33%	12.81%	13.97%	14.58%	2.22%	3.28%	3.62%	3.76%
United States	S&P500	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q
Min	-11.98%	-11.78%	-12.16%	-12.11%	-13.50%	-11.74%	-13.05%	-12.89%	-13.08%	-11.75%	-13.05%	-12.89%	-13.08%
Max	9.38%	10.37%	10.99%	11.34%	11.81%	9.98%	12.13%	11.43%	11.27%	9.99%	12.13%	11.43%	11.27%
Annual Volatility	16.35%	16.35%	16.53%	16.72%	17.09%	16.32%	17.01%	17.05%	17.16%	16.36%	16.98%	17.03%	17.34%
Cumulative Return	280%	397%	379%	344%	406%	430%	416%	375%	454%	451%	465%	415%	451%
Correlation	1.000	0.994	0.987	0.981	0.976	0.994	0.981	0.976	0.973	0.990	0.975	0.968	0.964
Avg. Number of Assets	-	474.19	237.24	166.29	118.90	476.27	237.64	166.64	118.91	478.67	238.33	167.17	119.67
Monthly Avg. Turnover	-	3.53%	6.80%	6.97%	7.24%	3.45%	6.95%	7.21%	7.59%	1.73%	3.53%	3.61%	3.80%

Table 3: Tracking error per year and country

	Portfolio rebalancing frequency											
Country	120 days (6 months)				240 days (1 year)				480 days (2 years)			
Brazil	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q
2012	0.96%	0.95%	1.07%	1.20%	0.57%	0.64%	0.70%	0.80%	0.62%	0.64%	0.69%	0.79%
2013	0.97%	0.87%	0.94%	0.89%	0.86%	0.80%	0.89%	0.88%	0.64%	0.72%	0.72%	0.74%
2014	1.00%	1.07%	1.07%	1.08%	1.05%	1.07%	0.99%	0.91%	0.99%	1.04%	0.95%	0.90%
2015	1.07%	1.14%	1.28%	1.34%	1.10%	1.21%	1.46%	1.58%	0.98%	1.07%	1.06%	1.13%
2016	1.04%	1.12%	1.11%	1.15%	1.06%	1.13%	1.12%	1.24%	1.08%	1.17%	1.18%	1.29%
2017	0.44%	0.45%	0.46%	0.53%	0.64%	0.63%	0.63%	0.70%	0.68%	0.73%	0.74%	0.85%
2018	0.15%	0.43%	0.46%	0.61%	0.16%	0.31%	0.46%	0.64%	0.16%	0.29%	0.48%	0.64%
2019	0.11%	0.38%	0.41%	0.50%	0.13%	0.43%	0.44%	0.63%	0.15%	0.28%	0.40%	0.63%
2020	0.22%	0.41%	0.58%	0.57%	0.25%	0.42%	0.61%	0.67%	0.30%	0.47%	0.68%	0.72%
2021	0.19%	0.31%	0.37%	0.56%	0.17%	0.28%	0.39%	0.50%	0.26%	0.40%	0.56%	0.63%
Average	0.62%	0.71%	0.77%	0.84%	0.60%	0.69%	0.77%	0.85%	0.59%	0.68%	0.75%	0.83%
Japan	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q
2012	0.18%	0.28%	0.34%	0.36%	0.17%	0.25%	0.36%	0.39%	0.18%	0.27%	0.39%	0.44%
2013	0.20%	0.38%	0.41%	0.44%	0.21%	0.38%	0.42%	0.47%	0.19%	0.27%	0.41%	0.47%
2014	0.12%	0.17%	0.19%	0.26%	0.12%	0.18%	0.19%	0.28%	0.12%	0.18%	0.22%	0.28%
2015	0.13%	0.18%	0.26%	0.31%	0.14%	0.21%	0.31%	0.37%	0.15%	0.27%	0.31%	0.34%
2016	0.40%	0.44%	0.49%	0.49%	0.44%	0.45%	0.50%	0.49%	0.43%	0.46%	0.50%	0.48%
2017	0.18%	0.24%	0.26%	0.29%	0.24%	0.25%	0.26%	0.28%	0.12%	0.19%	0.22%	0.24%
2018	0.14%	0.20%	0.27%	0.30%	0.14%	0.24%	0.26%	0.31%	0.14%	0.27%	0.26%	0.34%
2019	0.15%	0.18%	0.25%	0.26%	0.14%	0.19%	0.30%	0.32%	0.15%	0.28%	0.31%	0.36%
2020	0.18%	0.32%	0.35%	0.46%	0.21%	0.27%	0.34%	0.58%	0.26%	0.31%	0.39%	0.66%
2021	0.13%	0.24%	0.30%	0.38%	0.15%	0.23%	0.28%	0.37%	0.19%	0.30%	0.36%	0.49%
Average	0.18%	0.26%	0.31%	0.35%	0.20%	0.27%	0.32%	0.38%	0.19%	0.28%	0.34%	0.41%
United States	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q	M1-B	M2-E-M	M2-E-P35	M2-E-1Q
2012	0.08%	0.13%	0.14%	0.20%	0.08%	0.14%	0.13%	0.20%	0.08%	0.14%	0.13%	0.19%
2013	0.07%	0.13%	0.15%	0.18%	0.07%	0.13%	0.16%	0.19%	0.07%	0.15%	0.16%	0.21%
2014	0.07%	0.14%	0.16%	0.17%	0.07%	0.13%	0.14%	0.15%	0.07%	0.13%	0.14%	0.16%
2015	0.08%	0.15%	0.17%	0.18%	0.09%	0.14%	0.16%	0.17%	0.09%	0.16%	0.18%	0.18%
2016	0.09%	0.16%	0.19%	0.20%	0.11%	0.16%	0.20%	0.21%	0.12%	0.17%	0.20%	0.20%
2017	0.07%	0.14%	0.16%	0.17%	0.07%	0.15%	0.17%	0.18%	0.08%	0.14%	0.18%	0.19%
2018	0.10%	0.17%	0.21%	0.22%	0.12%	0.16%	0.20%	0.22%	0.14%	0.17%	0.23%	0.23%
2019	0.07%	0.17%	0.20%	0.19%	0.07%	0.16%	0.17%	0.19%	0.09%	0.14%	0.17%	0.18%
2020	0.26%	0.26%	0.34%	0.42%	0.25%	0.48%	0.50%	0.53%	0.33%	0.55%	0.60%	0.65%
2021	0.11%	0.19%	0.26%	0.28%	0.13%	0.18%	0.25%	0.25%	0.17%	0.28%	0.34%	0.36%
Average	0.10%	0.16%	0.20%	0.22%	0.11%	0.18%	0.21%	0.23%	0.12%	0.20%	0.23%	0.25%

Table 4: Comparison of the hybrid approach used in Sant'Anna et al. (2017) and the efficiency-constrained models.

	10-asset portfolios				5-asset portfolios			
Portfolio rebalancing frequency	20 days	60 days	120 days	240 days	20 days	60 days	120 days	240 days
Sant'Anna et al. (2017)								
Average	0.055%	0.032%	0.024%	0.017%	0.078%	0.049%	0.034%	0.023%
Minimum	0.037%	0.025%	0.014%	0.015%	0.046%	0.036%	0.025%	0.022%
Maximum	0.088%	0.045%	0.037%	0.021%	0.131%	0.071%	0.052%	0.027%
SD	0.014%	0.006%	0.007%	0.003%	0.023%	0.010%	0.009%	0.003%
Efficiency based portfolios								
Average	0.111%	0.065%	0.040%	0.028%	0.149%	0.079%	0.049%	0.033%
Minimum	0.003%	0.042%	0.029%	0.025%	0.001%	0.060%	0.031%	0.030%
Maximum	0.189%	0.093%	0.055%	0.030%	0.294%	0.122%	0.064%	0.039%
SD	0.042%	0.016%	0.011%	0.002%	0.059%	0.018%	0.010%	0.004%
Difference in Average TE (Performance Gain/Loss)	0.056%	0.033%	0.016%	0.011%	0.071%	0.030%	0.015%	0.010%

Difference in Average tracking error (TE) comprises the difference between the average TE of the portfolios generated by the model with efficiency constraints and the average TE of the portfolios formed by the hybrid approach of Sant'Anna et al. (2017).