Josip Arnerić
University of Zagreb
Faculty of Economics and Business
10000 Zagreb, Croatia
jarneric@net.efzg.hr

Luka Osojnik
University of Zagreb
Faculty of Economics and Business
10000 Zagreb, Croatia
losojnik@net.efzg.hr

JEL: C58, G11, G51
Original scientific article
https://doi.org/10.51680/ev.37.2.3

Received: November 28, 2023
Revision received: April 1, 2024
Accepted for publishing: April 1, 2024

EXPLANATORY FACTORS OF EUROZONE EQUITY ETF TRACKING ERROR

Abstract

Purpose: According to the existing literature, it remains unclear whether a specific ETF outperforms or underperforms its benchmark index in terms of tracking error, especially during crisis periods. Therefore, this study concentrates on the largest and most liquid Eurozone equity ETF, iShares Euro Stoxx 50, which tracks the Euro STOXX 50 index, with the fundamental objective of identifying explanatory factors of tracking errors during crisis periods, encompassing the COVID-19 pandemic and the onset of the Ukrainian war.

Methodology: The added value of current research lies in the utilization of Markov regime switching regression with two-state variables. This approach is supported by the idea that the influence of explanatory factors on tracking error may vary between bearish and bullish regimes, which typically align with non-crisis and crisis periods, respectively.

Results: Empirical findings indicate that an increase in volatility led to a stronger decrease in tracking error during periods of market stress than in a bullish regime, while a negative impact of illiquidity on tracking error is similar for both regimes. Unlike a single-regime model, Markov switching exhibits a negative relationship between the net flows and tracking error, as expected. The effect of premium/discount seems to be both positive and negative, but a weaker influence was found during a bearish regime due to herding behavior of investors or higher trading costs.

Conclusion: This study relies on an ex-ante approach with its main advantage of providing a forward-looking estimate of tracking error that takes into account changes in market conditions and the ETF's underlying holdings, unlike historical or ex-post tracking error.

Keywords: ETF tracking error, Eurozone equity market, crisis, Markov switching regression

1. Introduction

Active investing in a highly efficient market requires a complex understanding of available information, which can be challenging and time-consuming for an average investor. Fortunately, the evolution of technology has revolutionized the investment landscape, enabling investors to efficiently allocate their assets into the stock market without the constant need to interpret current market information and react accordingly. This paradigm shift has given rise to the concept of passive investing, where investors seek to replicate the performance of an index or a specific set of assets. Thus, exchange-traded funds (ETFs) have become increasingly popular passive investments in Europe in recent years (Le Sourd & Safaee, 2021). Beginning with a modest $100 billion in assets under management in 2000, the ETF market skyrocketed to $1 trillion by 2010, and almost achieved a $10 trillion market cap in 2020. In contrast, US-based mutual funds, which have been present in the market for nearly a century, held approximately $27 trillion in assets as of 2020. One of the reasons for such an extraordinary increase in popularity are the benefits it offered. Some of these are minimal fees associated with owning and trading securities, high liquidity while managing a large basket of stocks, diversification benefits and simple tradability (Mussavian & Hirsch, 2002; Madhavan, 2014). As a result, measuring the performance of ETFs with respect to tracking error has attracted great attention among scientists and practitioners (Johnson, 2009; Dorocáková, 2017; Tsalikis & Papadopoulos, 2019; Feder-Sempach & Miziołek, 2023). It measures how closely an ETF tracks its benchmark index. A low tracking error indicates the ability of an ETF to replicate its target portfolio or the target index almost perfectly, while a high tracking error suggests that the ETF deviates from its benchmark. Several studies attempted to examine the performance of European ETFs with respect to tracking error and identify the most influenced factors. Feder-Sempach & Miziołek (2023) concluded that tracking error of eurozone ETFs over a period of ten years was relatively low, on average 0.3%, although ETFs with distributing income had a higher tracking error against accumulation income ETFs.

However, according to the existing literature, it is still unclear if specific ETF outperforms or underperforms its benchmark index in terms of tracking error, particularly in a crisis period. In general, during a crisis period, market volatility tends to increase, which can make it more difficult for an ETF to perfectly track its underlying benchmark (Johnson et al., 2013). This is because the prices of the ETF's constituent securities may fluctuate more widely than usual, and the ETF's portfolio manager may not be able to rebalance the portfolio efficiently due to market conditions such as liquidity constraints and transaction costs. Additionally, during a crisis period, there may be significant changes in the composition of the underlying benchmark as some securities may become more or less important to the benchmark compared to the ETF's holdings, which can also contribute to tracking error (Vardharaj et al., 2004). By contrast, in some cases, the tracking error of an ETF may decrease during a crisis period if the ETF's portfolio manager is able to identify undervalued securities and add them to the portfolio. A decreased tracking error implies that ETF outperforms the benchmark index and *vice versa*.

In that context, the fundamental objective of the current study is to find out if the tracking error of the Eurozone ETF relative to its benchmark index decreased or increased during crisis periods, specifically covering COVID-19 pandemic and the Ukrainian war. Additionally, it offers empirical evidence for widely used market-based measures that may influence tracking error, such as market volatility, illiquidity proxy (bid-ask spread), net flow, premium or discount and trading volume. For the same reason, daily observations from May 27, 2019 to May 26, 2023 are provided by Refinitiv Eikon service[1]. According to different market regimes, the influence of the aforementioned factors on tracking error might be distinctive, and thus this paper tries not only to provide empirical evidence and comprehensive explanations of those distinctive influences, but also to fill the gap in eurozone ETF performance analysis with respect to regime switching methodology.

This study relies on an ex-ante approach with its main advantage of providing a forward-looking estimate of tracking error that takes into account changes in market conditions and the ETF's underlying holdings, unlike historical or ex-post tracking error. It uses the root of the squared residual from a simple regression of net asset value (NAV) returns on the benchmark returns as an indicator of a daily tracking error. Entire research focuses on the Eurozone equity ETF, which tracks the Euro STOXX 50 index. The STOXX 50 index is a widely followed benchmark for the Eurozone equity market, representing the performance of fifty blue-chip companies from 18 Eurozone countries. There are several exchange-traded funds that track the STOXX 50 index, but the most popular one is the iShares Euro Stoxx 50 ETF as the largest and most liquid one, with over 10 billion USD in assets under management as of May 2023. This ETF is particularly attractive to investors who want to reinvest the dividends into the fund, and not to pay them out, which is unique to accumulating income ETFs. Unlike distributing income ETFs, the accumulating ones maximize future returns. This was yet another reason, except the size and its liquidity, for selecting iShares Euro Stoxx 50 ETF.

Added value of the research consists of the employed methodology considering regime switching regression with two-state variables that follow the Markov chain. The major reason for this is that the

1 Refinitiv Eikon is a financial data platform available at: https://eikon.refinitiv.com/ [accessed April 17, 2023].

influence of tracking error explaining factors may deviate over time depending on whether bearish or bullish state is in the market. Those two market regimes usually coincide with non-crisis and crisis periods, and according to the span of time-series data, it captures the COVID-19 pandemic and the Ukrainian war. This is particularly important as the prices of securities tend to fluctuate more during times of stress and market turmoil, which can hinder the ability of the portfolio manager to efficiently manage portfolios of the constituent ETF. The inability of the portfolio manager to balance portfolios strays the price of the ETF away from its benchmark index and hence decreases the ETF tracking error. Conversely, actively managed ETFs aiming to outperform their benchmark index may experience a decrease in tracking error during stressful periods. This occurs when the portfolio manager successfully identifies undervalued securities and incorporates them into the portfolio, leading to superior performance compared to the index. In addition, a Markov switching approach is appropriate for nonlinear time-series models with regimes determined by unobserved states which must be inferred from the data and the parameters are expanded to include the transition probabilities. Unlike current literature, which mainly revolves around the American market and the ETFs, this paper shifts attention and contributes to the literature which focuses on the European market.

The rest of the paper is organized as follows. Section 2 explains the theoretical framework of ETFs, including a review of previous studies. Section 3 presents data and methodology. Section 4 provides empirical findings, while Section 5 offers a comprehensive discussion of the results. Finally, Section 6 provides a conclusion.

2. Theoretical framework and a review of previous studies

Both mutual funds and ETFs serve to reduce risk through diversification. However, there are key distinctions between the two. ETFs are pooled investment vehicles that track specific indexes, mostly passively managed. With a focus on risk reduction rather than returns, ETFs have seen a significant increase in total market cap in the last two decades. This simplicity makes them attractive to retail investors. By contrast, mutual funds lack the ability for investors to sell shares at any time and are actively managed. The research conducted by Kaminsky (2001) suggests that mutual funds play a significant role among institutional investors as the primary channel for financial flows into emerging markets. On the other hand, the findings of Sy and Ong (2004) indicate that this phenomenon is more pronounced in the European market compared to the United States. This active management of the fund comes with greater initial and management costs as well as higher transaction costs compared to ETFs which drive retail investors toward cheaper alternatives. Broman (2016) says that higher liquidity ETF shares attract investors who are not willing to invest directly into illiquid assets such as commodities, emerging markets, etc. When buying ETFs from the broker, one will find that there are many different options of an ETF that they want to buy from different markets. For example, an ETF from the London stock exchange will show prices in pounds, whereas on some exchanges there will be a limited number of ETFs available to be bought. Investors also want to avoid buying from multiple exchanges as there are annual charging fees. One important factor to consider is the volume that certain instruments have on different exchanges. This is particularly important when buying ETFs because of their tracking error. If the volume and trading frequency of the ETF is low, it can exhibit higher spreads and therefore increase tracking error. Therefore, it is wise to choose a stock exchange with higher trading volume compared to other exchanges.

There are several approaches to obtaining ETF tracking error (a historical approach, an ex-post or an ex-ante approach), each with its own pros and cons (De Rossi, 2015). Most studies have used regression analysis after obtaining tracking error, indicating that European ETFs generally exhibit good performance in terms of tracking their benchmarks. However, there is significant variation depending on the specific ETF under analysis, the measurement of tracking error, the observed period, and the approaches employed to examine the factors influencing tracking error. Understanding these factors can assist investors in making informed decisions when selecting ETFs for their portfolios. Because of the nature of ETFs and their mechanics, the movements of tracking error can be expected to move in directions which align with the theory behind the variables. The deviation of ETF prices from their NAV is primarily maintained through the arbitrage process. Theoretically, an increase in premium/discount should invite arbitrageurs, which in turn should align ETF prices better with its NAV. Moreover, ETFs that include international stocks are expected to exhibit greater deviation due to the continued trading of their shares on the domestic exchange, while the market for the underlying securities in the creation basket is closed. Similarly, in theory, ETFs containing illiquid securities should experience higher deviations as the arbitrage process would require a larger deviation to compensate for the higher transaction costs associated with trading those less liquid securities. Increased trading volume positively impacts liquidity and bid-ask spread, while market volatility increases bid-ask spreads. Net inflows and their effect on the tracking error highly depend on the state of the market as market participants have different behaviors during each regime. To expand the theory, the examination also includes a review of what the literature suggests about variable relationships. ETFs which invest in less liquid assets may experience difficulty in replicating their benchmark index leading to higher tracking errors. Bae & Kim (2020) have documented a positive relation between illiquidity and tracking error.

Hillard & Le (2022) found that emerging European markets have higher tracking error in comparison to developed Europe, and it was around 0.67% and 0.33%, respectively. Rompotis (2011) states that tracking error for ETFs with higher expense ratios was higher. Additionally, Tsalikis & Papadopoulos (2019) confirmed that tracking error for European ETFs was, on average, higher than that of US ETFs, while a possible explanation for the aforementioned could lie in the economies of scale and thus lower costs. Chu (2011) also found that economies of scale will improve tracking ability, while their research suggests that expense, delay in receiving dividends, the trading cost and the market risk increase tracking error. Additionally, Elton et al. (2019), as well as Chu & Xu (2021), suggested that tracking error is significantly influenced by delayed reinvestments of dividend. Regardless of the tracking error measurement, higher assets under management (AUM) positively affect tracking ability. The study also found that higher expense ratios are associated with higher tracking errors, although statistical significance is observed only for one measurement. Another study by Frino & Gallagher (2001) presented evidence that tracking error is positively and significantly correlated with dividend payments, and also that there were seasonal patterns with higher error rates in January and May, and a lower error rate in the quarters ending in March, June, September, and December. Aber et al. (2009) stated that the range of daily price fluctuations was very large, which indicated that active traders or arbitrageurs were more likely to profit than passive traders. Blitz et al. (2012) revealed in their study that index funds and ETFs in Europe underperform their benchmarks by larger amounts than their reported expenses, with dividend taxes explaining a significant portion of underperformance. This highlights the need to account for dividend taxes in evaluating fund performance and measuring fund costs accurately.

Other well-known factors explaining ETFs tracking error are market volatility, trading volume, the net flow as well as premium or discount. Higher market volatility and trading volumes can lead to wider bid-ask spreads, which can increase the cost of trading and result in higher tracking errors, as observed in several studies, including Ben-David et al. (2019). In a study on Hong-Kong ETFs, Chu (2011) demonstrated that trading volume increases tracking error; however, it is not significant, while Yiannaki (2015) suggested that there is a weak correlation between tracking error and trading volumes. Dorocáková (2017) found that fluctuations in the underlying index can have a relative influence on tracking error. In the case of bid-ask spreads, Meinhardt et al. (2015) indicated a positive relation to tracking error for the German ETF market.

On the demand side, the net flow ETF may affect its tracking error. The ETF net flow tends to increase during a bulish period when investors are more optimistic and confident about the future of financial markets. Conversely, during a bearish period in the market, the ETF net flow tends to decrease as investors become more risk-averse and seek to reduce their exposure to equities. According to research of Ben-David et al. (2017), tracking error is negatively related to the ETF net flow. Another study by Osterhoff, & Kaserer (2016) confirmed that the net flow had a significant negative effect on tracking error for small ETFs.

Divergence of ETF market prices from their net asset value, reported as premium (or discount), is yet another explaining factor of ETF tracking error. A study by Wong & Shum (2010) found that tracking error of the examined ETFs is consistently positive in both bullish and bearish markets. This suggests that investors are willing to pay a premium for ETF investments, as ETFs provide positive returns that can cover transaction costs and potentially yield returns in different market conditions. Rompotis (2010) found tracking error to be positively affected by premium/discount, while Li and Zhao (2014) found that premiums can lead to increased tracking error in ETFs that hold illiquid securities.

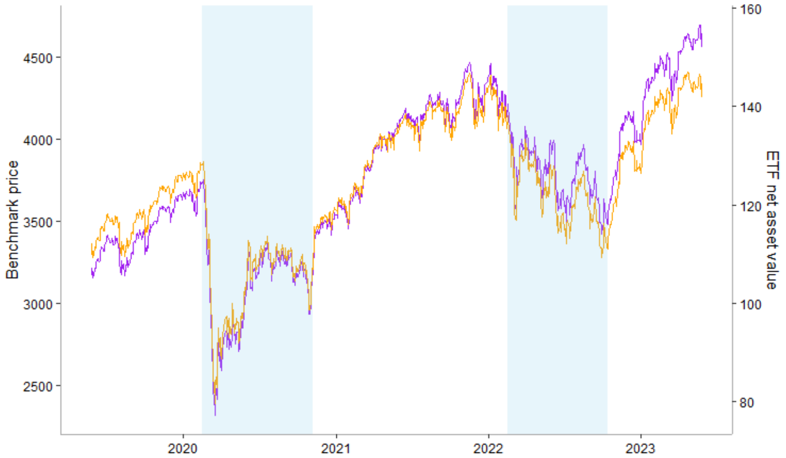
Aber et al. (2009) suggested that ETFs traded more at a premium than at a discount, indicating that the market tended to overvalue ETFs compared to their NAV. Additionally, premiums have shown to be higher for newly created ETFs, as documented in a study by Piccotti (2018), which indicates that investors are willing to pay a premium in order to access the liquidity benefits provided by ETFs, which allow indirect availability to less accessible underlying securities.

3. *Data and methodology*

In previous studies, researchers have used both NAV returns and closing market price returns to evaluate the tracking error of ETFs compared to their benchmark index returns (Zawadzki, 2020). However, due to its advantages, the NAV-based measurement of tracking error is widely recognized as the preferred approach. NAV returns consider dividends or any income generated by the underlying assets, providing a more accurate and reliable measure of the ETF's performance in accordance with GIPS - Global Investment Performance Standards (CFA, 2020). Additionally, changes in the net asset value reflect what an investor would actually receive from holding the ETF (Osojnik, 2023). In contrast, closing price ETF returns may be influenced by short-term price fluctuations that do not necessarily reflect the underlying performance of the ETF. Therefore, using closing price returns to assess tracking error can be misleading. Furthermore, the difference between ETF market prices and their respective net asset values introduces another variable known as the premium or discount. This variable will be utilized to explain ETF tracking error. Mispricing of an ETF in relation to its net asset value creates arbitrage opportunities through the creation and redemption mechanism, which can be advantageous for investors (Osojnik, 2023).

The first impression of tracking error can be made by visual inspection of iShares Euro Stoxx 50 ETF net asset values and market closing prices of a benchmark Euro STOXX 50. Figure 1 uses a dual scale axis for comparison and highlights the shaded area covering turbulent periods of the COVID-19 crisis and the Ukrainian war. Both net asset values and closing prices are expressed in the same currency (EUR), but with different scales. Although Figure 1 clearly indicates that the ETF tracks its benchmark quite well with few disparities during a bullish regime, commenting on the ETF performance solely based on price differences is not possible; instead, log returns are considered.

Figure 1 ETF net asset values vs. benchmark index prices



Source: Authors' construction using data provided by Refinitiv Eikon

Before analysis continues, all variables of interest are derived from the raw data. Firstly, tracking error is estimated following the ex-ante approach by regressing NAV returns of the ETF (Ret_t^{NAV}) on benchmark returns (Ret_t^{BEN}). The root of squared regression residual for each trading day resulted in tracking error (Osojnik, 2023):

$$track\_error_t = \sqrt{(Ret_t^{NAV} - 0.0109 - 0.9949Ret_t^{BEN})^2}\,. \quad (1)$$

In the above expression, -0.0109 and -0.9949 are the constant term and the slope coefficient, respectively.

Daily NAV returns of ETF and benchmark returns, used in the regression, are obtained following the same formulation:

$$Ret_t^{NAV} = \frac{NAV_t - NAV_{t-1}}{NAV_{t-1}} 100\%$$
$$Ret_t^{BEN} = \frac{C_t^{BEN} - C_{t-1}^{BEN}}{C_{t-1}^{BEN}} 100\%, \quad (2)$$

where NAV_t and NAV_{t-1} are ETFs net asset values on the current and previous trading day, while C_t^{BEN} and C_{t-1}^{BEN} are closing prices of a benchmark index.

Next, an illiquidity proxy measure is obtained as bid-ask spread with end-of-day ETF quotes toward its mid-quote, by the following expression:

$$iliquidity_t = \frac{2(A_t^{ETF} - B_t^{ETF})}{A_t^{ETF} + B_t^{ETF}} 100\%. \quad (3)$$

ETF daily premium is also expressed as a percentage like all other variables according to:

$$premium_t = \frac{C_t^{ETF} - NAV_{t-1}}{NAV_{t-1}} 100\%, \quad (4)$$

where C_t^{ETF} are closing (market) prices of the ETF on day t. The same indicator (4) exhibits a discount (negative values) when the ETF market price is lower than its NAV. This means that investors buy the ETF at a price cheaper than the underlying value of its assets. Conversely, an ETF is trading at a premium when its market price is higher than its NAV (positive values of the aforementioned indicator).

The daily ETF net flow, which represents the inflow and outflow of ETF, is given by the formula:

$$net\_flow_t = \frac{TNAV_t - \left(1 + \frac{Ret_{t-1}^{ETF}}{100}\right) TNAV_{t-1}}{TNAV_{t-1}} 100\%, \quad (5)$$

where the total net asset value $TNAV_t$ represents a product of NAV per share and the number of outstanding shares on the current day. The previous day total net asset value $TNAV_{t-1}$ is adjusted with the respective ETF daily return to account for the performance effect of the change, which is independent of capital flows.

Market volatility is measured by the official Euro Stoxx 50 volatility index (VSTOXX), which is the European version of VIX, reflecting investor's sentiment as expectations of future volatility.

Summary statistics of variables of interest are reported in Table 1. All values of variables are expressed in percentages, except the volatility index and trading volume. Only trading volume is transformed into logs due to a large scale and extreme variations of trading across days.

Table 1 Descriptive statistics of iShares Euro Stoxx 50 ETF tracking error and its explanatory predictors along with ADF unit root test

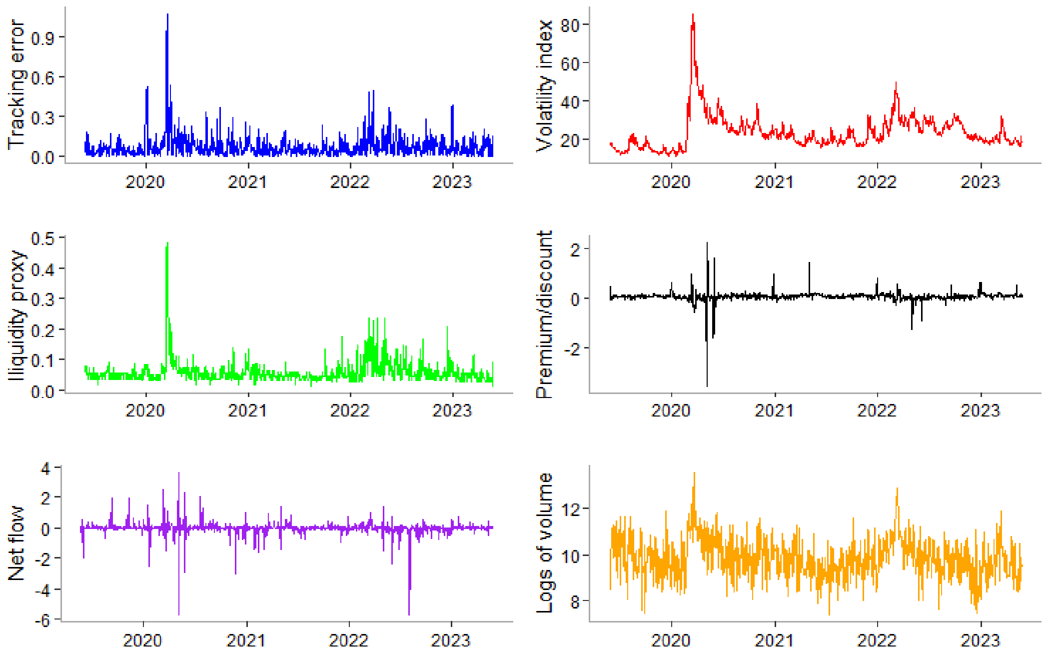
Variable	Min	Max	Mean	SD	Median	ADF test
Tracking error	0.00	1.08	0.07	0.09	0.05	-15.1558***
Illiquidity proxy	0.01	0.48	0.06	0.04	0.05	-11.5989***
Volatility index	10.69	85.62	23.21	8.85	21.36	-13.8413***
Net flow	-5.74	3.59	-0.03	0.48	-0.01	-20.9258***
Premium/discount	-3.55	2.19	0.04	0.21	0.05	-18.7641***
Logs of volume	7.37	13.56	9.84	0.85	9.82	-13.4245***

Note: significance levels * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Source: Authors' calculation using data provided by Refinitiv Eikon

It can be noticed in Table 1 that the mean and median tracking errors are 0.07% and 0.05%, respectively. The maximum value of 1.08% can be expected during high market volatility and stress, when assets in the portfolio become less liquid and more difficult to allocate. The null hypothesis of the Augmented Dickey-Fuller (ADF) unit root test is rejected at the significance level of 1%, indicating that all considered variables are stationary. ADF in the levels is performed without trend and without drift, except for the net flow, and the premium/discount as their mean is approximately zero, and thus a drift term is not omitted for those two variables (Osojnik, 2023). Stationarity of all variables is preferred to eliminate possible suspicion of the results in the post-estimation phase caused by the non-stationarity issue.

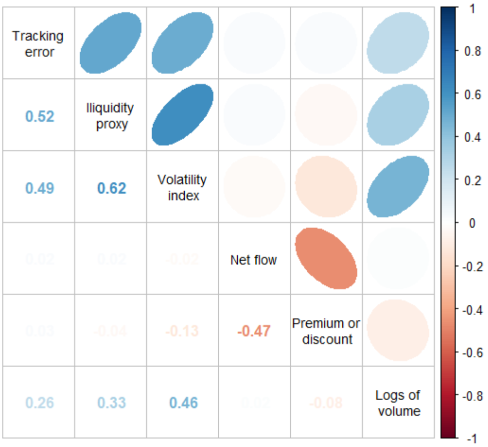
Figure 2 shows the clustering of tracking error, volatility and illiquidity, particularly in crisis periods which can be identified as bearish states of the ETF regime. Therefore, it is not surprising that these three variables are more correlated than other variables, indicating that illiquidity and volatility contribute positively to tracking error (Figure 3).

Figure 2 Time-series of variables observed from May 27, 2019 to May 26, 2023



Source: Authors' construction using data provided by Refinitiv Eikon

Figure 3 Correlation matrix between the observed variables



Source: Authors' construction using data provided by Refinitiv Eikon

An application of Markov regime-switching (MRS) models has attracted great interest in capturing dynamics of financial time-series, primarily due to the nonlinear dependence between considered variables as well as their nonstationary property (time-varying moments). In these circumstances, the main advantage of MRS is that it allows regression parameters to switch across multiple states or regimes, with the probabilities of switching between these states being dependent on the current state (Peovski et al., 2022). For example, the MRS model can capture changes of the dependence between two or more variables during different economic cycles or market regimes, such as high and low volatility regimes or bearish and bullish regimes, which usually coincides with crisis and non-crisis periods (Osojnik, 2023).

A simple Markov switching bivariate regression model, which considers two states of regime, can be formalized as follows:

$$y_t = \alpha_{St} + \beta_{St} \cdot x_{St} + u_{St}$$

$$u_{St} \sim WN(0, \sigma_{St}^2)$$

$$\alpha_{St} = \alpha_1(2 - S_t) + \alpha_2(S_t - 1) \qquad (6)$$

$$\beta_{St} = \beta_1(2 - S_t) + \beta_2(S_t - 1)$$

$$\sigma_{St}^2 = \sigma_1^2(2 - S_t) + \sigma_2^2(S_t - 1),$$

where $S_t=j$ is a discrete state variable that indicates in which regime the Markov process is $j=1,2,..., k$. Consequently, if the process is in the first regime state, then $S_t=1$, with parameters α_1, β_1 and σ_1^2, but if the process is in the second regime state, then $S_t=2$, with parameters α_2, β_2 and σ_2^2. Assuming that the conditional probability density function is Gaussian:

$$f(S_t) = \frac{1}{\sqrt{2\pi\sigma_{St}^2}} exp\left\{-\frac{(y_t-\alpha_{St}-\beta_{St}\cdot x_{St})^2}{2\sigma_{St}^2}\right\}, \qquad (7)$$

then a log-likelihood function $lnL = \sum_{t=1}^{T} ln\,\{f(S_t)\}$ can be maximized with respect to parameters α_1, α_2, β_1, β_2, σ_1^2 and σ_2^2. However, the state variable is usually unobserved in practical applications, but it is commonly assumed that it follows a Markov chain process with a k-dimensional state space (Hamilton, 1989). The specificity of a Markov chain process is the first-order dependence, implying that a state variable at the moment t depends only on the previous state of the process at the moment t-1 (Goldfeld & Quandt, 1973). Thus, for k=2, the log-likelihood function takes the form:

$$lnL = \sum_{t=1}^{T} \ln\left[\left(\sum_{j=1}^{2} \frac{1}{\sqrt{2\pi\sigma_{St}^2}} \exp\left\{-\frac{(y_t - \alpha_{St} - \beta_{St} \cdot x_{St})^2}{2\sigma_{St}^2}\right\}\right) \Pr(S_t = j|I_{t-1})\right] \cdot (8)$$

The probability density function (8) for each observation $t=1,2,..., T$ is presented as a weighted sum of conditional probability density functions for both regime states $j=1,2.$ The associated weights $\Pr(S_t = j|I_{t-1})$ are interpreted as conditional probabilities that the process is in the state j at the moment t, conditioned on all information from previous periods up to and including the moment t-1. These conditional probabilities are called ex ante probabilities (Kim & Nelson, 2017). In order to maximize the log-likelihood function it is necessary to assume a priori the behavior of a discrete state variable S_t. It is assumed that the state variable is generated by a first-order Markov process:

$$Pr\,(S_{t-1}, S_{t-2}, \ldots, S_1, I_{t-1}) = Pr(S_t|S_{t-1}). \qquad (9)$$

Ex ante probabilities $Pr\,(S_t = j|I_{t-1})$ are generated by a matrix of transitional probabilities, the so-called stochastic matrix:

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} p & (1-p) \\ (1-q) & q \end{bmatrix}. \qquad (10)$$

The matrix of transition probabilities P is an irreducible and primitive matrix (Hamilton, 1989). This means that all states of the Markov chain communicate with each other, i.e. that there is a probability of transition from state *i* to state *j*, as well as a probability of transition from state *j* to state *i*. Therefore, it is assumed that all elements of the stochastic matrix are greater than zero (a primitive matrix). In the matrix *P* the probability $p_{ij} = Pr\ (S_t = j|S_{t-1} = i)$ is the conditional probability that the process is in the state *j* at the moment *t* if it was in the state *i* at the previous moment *t-1*. For example, p_{12} is interpreted as the probability of transition from the first state to the second state of the regime, and p_{22} as the probability that the process will remain in the second state. The probabilities p_{11} and p_{12} are complementary, just like the probabilities in p_{21} and p_{22}. The transition probabilities *p* and *q* i1 the stochastic matrix *P* are commonly parameterized using inverse logit transformation:

$$p = \frac{e^{p_0}}{1 + e^{p_0}}\ ; \quad q = \frac{e^{q_0}}{1 + e^{q_0}}. \tag{11}$$

Upon transitional probabilities $Pr\ (S_t = j|S_{t-1} = i)$, conditional probabilities $Pr\ (S_t = j|I_{t-1})$ can be generated and then the log-likelihood function can be maximized by the parameters α_1, α_2, β_1, β_2, σ_1^2, σ_2^2 and q_0. Since the process of maximizing the log-likelihood function is iterative, in each new iteration conditional probabilities are updated using Kim's smoothing algorithm, the so-called Kim's filter. Kim's smoothing algorithm can be described in two steps (Kim & Nelson, 2017). In the first step, at the beginning of iteration, *ex ante* probabilities are calculated as follows:

$$Pr\ (I_{t-1}) = \sum_{j=1}^{2} Pr(S_t = j|S_{t-1} = i) Pr(S_{t-1} = i|I_{t-1})$$
$$j = 1, 2. \tag{12}$$

In the second step, according to the Bayes rule, for the observed values of response variable y_t, the so-called filtered probabilities are obtained:

$$Pr\ (I_{t-1}, y_t) = \frac{f(y_t|S_t = j, I_{t-1})\ Pr\ (I_{t-1})}{\sum_{j=1}^{2} f(y_t|S_t = j, I_{t-1})\ Pr\ Pr\ (I_{t-1})}. \tag{13}$$

However, initial probabilities need to be determined before the iterative procedure of maximizing the likelihood function can begin. For the initial probabilities, Hamilton (1989) proposed unconditional probabilities of the state of the regime, i.e. steady state probabilities:

$$\mu_1 = \frac{1-p}{2-p-q}\ ;\ \mu_2 = \frac{1-q}{2-p-q}. \tag{14}$$

Based on the transitional probabilities of the regime state, the expected duration of the process in the *j*- th regime state can be calculated as:

$$d_1 = \frac{1}{1-p}\ ;\ d_2 = \frac{1}{1-q}. \tag{15}$$

In the two-state regime model, the first regime state is assumed to be a low-volatility state and the second regime state is a high-volatility state. Then the parameters μ_1 and μ_2 can be interpreted as the expected probabilities that the process is in the regime of low (high) volatility in the long term, while the parameters d_1 and d_2 show the duration of the process in low and high volatility regimes in terms of days (Osojnik, 2023). Furthermore, it is worthwhile to analyze the time it takes for the process to switch from low to high volatility states, and *vice versa*.

4. Empirical results

In accordance with the previously described methodology and research objectives, assuming two states of regime *k=2* , the following MRS regression model is estimated:

$$\begin{aligned} track\_error_t = {} & \alpha_{St} + \beta_{1,St} \cdot illiquidity_{St} \\ & + \beta_{2,St} \cdot volatility_{St} + \beta_{3,St} \\ & \cdot net\_flow_{St} + \beta_{4,St} \cdot premium_{St} \\ & + \beta_{5,St} \cdot ln\ (volume)_{St} + u_{St}. \end{aligned} \tag{16}$$

For two states of regime, 16 parameters (constant term, five coefficients, and error standard deviation for each state along with two transitional probabilities) are estimated by the approximate maximum likelihood method using the expectation–maximization (EM) algorithm due to its convenience (Perlin, 2010). For comparison purposes, a single-regime regression model is also estimated to verify the switching property of regression coefficients (Table 2). In the post-estimation phase, appropriateness of the MRS approach is supported by diagnostic checking of unobserved error term u_{st} that should follow a white noise process with zero mean and constant variance for each state of regime (Table 3).

Table 2 Estimates of a single-regime model and a two-states regime model

	Single	**Two-states regime-switching model**	
Variable	regime	Regime 1	Regime 2
(Intercept)	0.0172	0.0165***	-0.0321***
	(0.0134)	(0.0043)	(0.0006)
Volatility index	0.0012***	0.0008***	0.0029***
	(0.0001)	(0.0002)	(0.0001)
Illiquidity proxy	1.1595***	1.0975***	1.0049***
	(0.0404)	(0.0666)	(0.0096)
Net flow	0.0522***	-0.1482***	-0.0155***
	(0.0044)	(0.0067)	(0.0014)
Premium/discount	-0.0424***	-0.3891***	0.0507***
	(0.0117)	(0.0277)	(0.0030)
Logs of volume	-0.0051***	0.0033***	-0.0514***
	(0.0015)	(0.0004)	(0.0031)
Error standard deviation	0.0101	0.0089	0.0024
Transitional probability	-	0.8509	0.5762
Observations	1019	1019	-
R2	0.897	0.9318	0.9925
AIC	-6536.0	-8507.4	-
BIC	-6501.5	-8365.2	-
Log.Lik.	3274.995	4265.739	-
RMSE	0.011	0.008	0.002

Note: significance levels * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; standard errors in parenthesis
Source: Authors' estimation using data provided by Refinitiv Eikon

For 1019 observations, the results of a single-regime model and a two-states regime-switching model are compared in Table 2. All variables are statistically significant at a 0.1% significance level. An increase in volatility and illiquidity increases tracking error. The results are the same for single-regime and two-states regime models. However, tracking error increases more in the second regime (0.29%) than in the first one (0.08%) with respect to a 1% increase in volatility, while 1% change of illiquidity has approximately the same impact on tracking error in both regimes (increases by 1.09% and 1%, respectively).

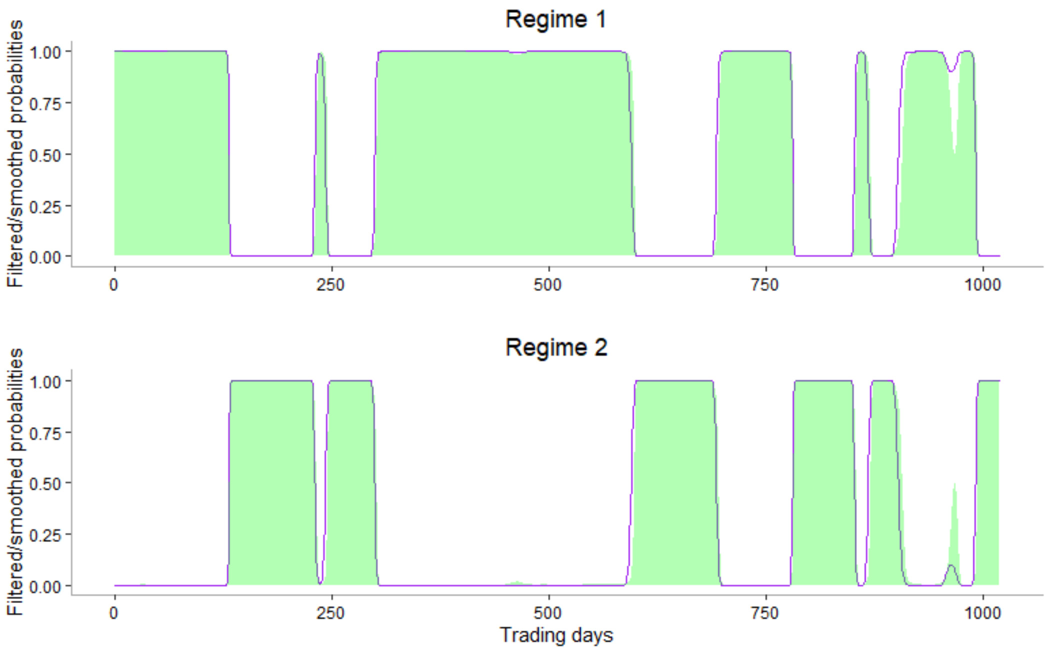
For the case of net flows, it shows a positive relationship with tracking error only in a single-regime model. The expected negative relationship was present for the two-states switching model, also indicating a steeper coefficient for a bullish period, meaning that the effect of net flows is stronger during bull periods, i.e. tracking error reduces by 0.14% with respect to a 1% increase of net flow. Furthermore, the results indicate that a premium/discount affects tracking error negatively in a single-regime model; however, their effect is both negative and positive for the bulish and bearish periods, respectively, for the two-states switching model.

Furthermore, the transition probability matrix provides information about probability transitions between two regime states. The probabilities p_{11} = 0.8509 and p_{22} = 0.5762 indicate the likelihood of remaining in the first and second states of the regime, respectively. Conversely, p_{12} =

0.1491 represents the probability of transitioning from the first state to the second state, while p_{21} = 0.4237 denotes the probability of transitioning from the second state to the first state. Accordingly, it is more likely to remain in the bullish state regime once the market gets to that state and approximately stays in that state for 7 trading days (a week and a half). In addition, the transition from a bearish to a bullish state of regime is 2.8 times more likely than the reverse, with a probability of 0.4237 compared to 0.1491.

Goodness-of-fit measures confirm the appropriateness and superiority of the two-states regime model over the single-regime model in terms of R2, information criteria, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), respectively, as well as the root mean square error (RMSE). In both regimes, R2 is substantially greater than in a single-regime case. Likewise, RMSE indicates lower regression standard errors in both states against a single state. Smaller AIC and BIC are observed within the Markov regime-switching model, supporting its preference. An R2 value close to 1 does not imply that the results from regression of time-series are spurious. Spurious regression occurs when regressing two or more independent time-series, resulting in false relationships due to nonstationary properties. To address this issue, the Augmented Dickey-Fuller (ADF) unit root test was conducted on each variable (Table 1), confirming their stationarity. Furthermore, the absence of clear trending behavior in the variables (Figure 2) eliminates the possibility of spurious results stemming from common trends. Once the Markov switching model parameters are estimated, the filtered probabilities of the regime states are obtained by Kim's filtering algorithm, which is a byproduct of the iterative log-likelihood maximization procedure (Perlin, 2010). Inspecting both the filtered and smoothed probabilities is useful for interpreting the switching regression coefficients associated with different time periods (Osojnik, 2023). In Figure 4, it is clear that regime 1 corresponds to the bullish state of the market, while regime 2 corresponds to the bearish state, and more importantly, it covers crisis periods including the COVID-19 pandemic and the onset of the Ukrainian war.

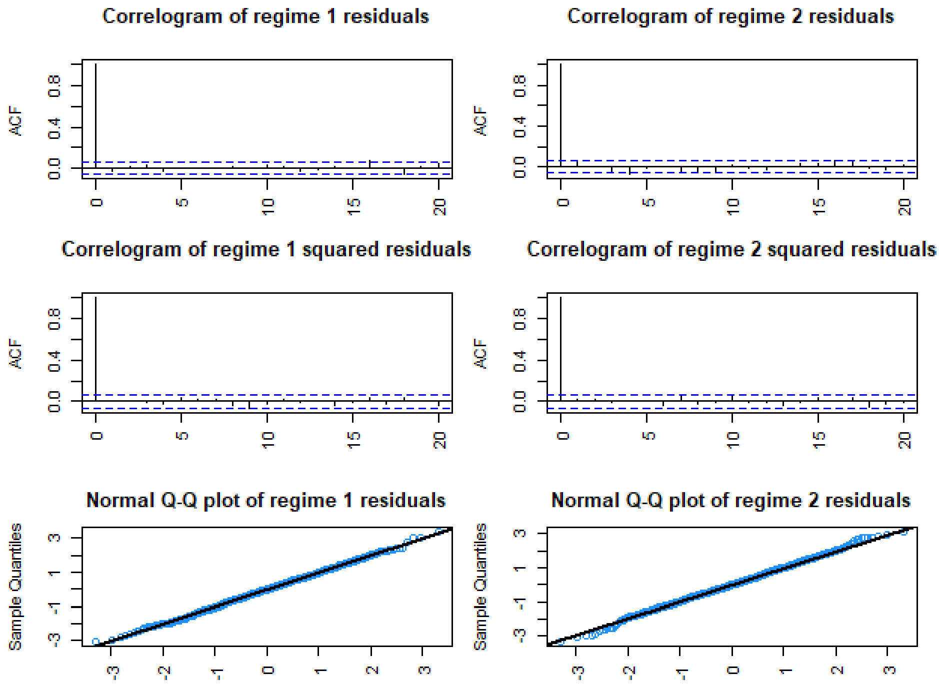
Figure 4 Filtered and smoothed probabilities of the Markov two-states switching model



Source: Authors' construction using data provided by Refinitiv Eikon

The validity of the Markov switching model requires diagnostic checking of residuals. Two series of residuals are generated in total, one for each state of regime. Three diagnostic plots are constructed for each residual series in Figure 5 (the correlogram of the residuals, the correlogram of the squared residuals, and the normal quantile-quantile plot), whereas formal diagnostic tests (the Ljung-Box test, the ARCH test and the Jarque-Bera test) are performed on the weighted residuals, i.e. a linear combination of two residual series using smoothed probabilities as the weights. In Figure 5, correlograms indicate no autocorrelation of residuals in both regimes and no autocorrelation of squared residuals, confirming the serial independence of error terms as well homoscedasticity (error terms have constant variance). The same conclusion is supported by non-rejection of the Ljung-Box test null hypothesis with 5 and 10 time lags, and by non-rejection of the ARCH test null hypothesis for autoregressive conditional heteroscedasticity at all significance levels (Table 3). According to the Jarque-Bera test, the normality assumption of weighted residuals is met.

Figure 5 Diagnostic plots of two regime residuals



Source: Authors' construction using data provided by Refinitiv Eikon

Table 3 Diagnostic checking of weighted residuals

Test	Statistic
Ljung-Box (5)	1.8469
Ljung-Box (10)	3.7356
ARCH	20.7821
Jarque-Bera	1.3569

Note: significance levels * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Source: Authors' calculation using data provided by Refinitiv Eikon

5. Discussion

The primary objective of this study was to address a critical gap in the analysis of Eurozone ETF performance by focusing specifically on the European market and investigating the impact of market-related variables on ETF tracking error. Both models, the single regime, and the two-states switching regime, confirm that volatility and illiquidity increase tracking error. This is likely because the increase is associated with higher trading costs for arbitrageurs, reducing their ability to create and redeem ETF shares and underlying assets.

When comparing results obtained by both models, it becomes evident that parameters in the single-regime case are overestimated or exhibit unexpected signs. For instance, the impact of illiquidity is overestimated, and the net flow shows a misleading direction of influence. According to Ben-David et al. (2019), increases in trading volume can lead to wider bid-ask spreads, which should, in turn, increase tracking error. Our results confirm this relationship, but only for the bullish period in the market (the first state of the regime). In contrast, during the bearish regime (the second state), trading volume reduces tracking error more than it does in the bullish regime.

Theoretically, an increase in premium should attract authorized participants, hedge funds, and arbitrageurs, thereby decreasing tracking error. This effect is confirmed only for bullish periods (-0.38%). The results regarding the effect of premium/discount on tracking error align with those documented by Rompotis (2012). Interestingly, our study found similar results for bearish periods, contrary to previous findings. One possible explanation for this positive relationship during bearish periods could be attributed to herding behavior among investors, as documented by Ferreruela & Mallor (2021). Shum & Kang (2013) also noted higher premiums/discounts in ETFs during crisis periods, suggesting reduced arbitrage activity due to heightened trading costs. Our analysis supports these findings, indicating consistency with the economic literature for most variables. Notably, in the two-states switching-regime model, the coefficients for net flow, premium, and volume change vary, providing a more comprehensive understanding of the variables influencing tracking error.

These findings underscore the importance of utilizing the two-states switching methodology for researchers. The introduction of the two-states switching methodology clearly demonstrates how the effects of variables such as net flow, premium, and volume can change. In addition to the current variables, supplementary variables like ETF provider's rebalancing frequency, benchmark index composition, and expense ratios could help explain tracking error. However, obtaining this information from publicly available data is not straightforward, and the significance of these variables may be questionable due to their daily time-invariance. For example, expense ratios are typically reported as annual fees, not on a daily basis.

6. Conclusion

The research explored how the tracking error of a Eurozone ETF, concerning its benchmark index, is influenced during crisis periods, including the COVID-19 pandemic and the Ukrainian war. By carefully examining market periods or regimes, the paper offers empirical evidence on several market-based measures, such as market volatility, liquidity proxy, net flow, premium or discount, and trading volume, to comprehensively understand their influence on tracking error. In addition to presenting in-depth explanations and empirical evidence, this study contributes to the literature by employing a regime-switching methodology. The findings support existing economic literature to some extent and highlight the importance of considering different market regimes. While all variables are statistically significant, it was found that an increase in volatility and illiquidity led to a decrease in tracking error. However, the method of switching regimes has shown that the influence of volatility on tracking error is stronger during periods of market stress than in bullish periods (0.29% and 0.08%, respectively), while in terms of illiquidity, the influence is the same for both regimes. When considering trading volume, the results confirm the findings of Ben-David et al. (2019) that an increase in volume does increase tracking error, but only slightly (0.003%). The relationship holds true only for the first regime (the bullish period). On top of that, the study found the influence of volume to be both negative and stronger for periods of market stress (-0.05%). Regarding net flows, the results surprisingly show a positive relationship with tracking error in a single-regime model. However, the use of switching regimes yields the excepted negative relationship between net flows and tracking error. To be more precise, 1% of net flow reduces tracking error for 0.14% during bullish periods and 0.01% during bearish periods. Lastly, the results show that premium/discount negatively affects tracking error using a single-regime model. With a two-states switching model, the results yield interesting findings. The effect of premium/discount seems to be both positive and negative. A negative influence was expected, and it was found only during the bullish period of the market (-0.38%). A positive but weaker (0.05%) influence of premium/discount on tracking error was found during periods of stress and it was also documented by Rompotis (2012). One of the explanations for a positive influence could lie in herding behavior exhibited by investors during periods of market stress. Another reason could be higher trading costs which arise in the periods of market stress. Higher trading costs tend to make arbitration more expensive. Hence, it keeps authorized participants waiting for the price between ETF and its NAV to be further and further away, explaining the positive relationship between premium/discount and tracking error. One limitation of the current research is the fact that only a single ETF is used in the analysis. It would be of great value if future researchers take into account ETFs with different liquidities and sizes.

Acknowledgements

This article originated from the undergraduate thesis of the BDiB study program at the University of Zagreb, Faculty of Economics and Business.

References

1. Aber, J. W., Li, D. & Can, L. (2009). Price volatility and tracking ability of ETFs. *Journal of Asset Management*, *10*(4), 210-221. https://doi.org/10.1057/jam.2009.13
2. Bae, K. & Kim, D. (2020). Liquidity risk and exchange-traded fund returns, variances, and tracking errors. *Journal of Financial Economics*, *138*(1), 222-253. https://doi.org/10.1016/j.jfineco.2019.02.012
3. Ben-David, I., Franzoni, F. & Moussawi, R. (2017). Exchange-Traded Funds. *Annual Review of Financial Economics*, *9*(1), 169-189. https://doi.org/10.1146/annurev-financial-110716-032538
4. Ben-David, I., Franzoni, F. & Moussawi, R. (2019). Do ETFs increase volatility? *Journal of Finance*, *73*(6), 2471-2535. https://doi.org/10.1111/jofi.12727
5. Blitz, D., Huij, J. & Swinkels, L. (2012). The Performance of European Index Funds and Exchange-Traded Funds. *European Financial Management*, *18*(4), 649-662. https://doi.org/10.1111/j.1468-036x.2010.00550.x
6. Broman, M. S. (2016). Liquidity, Style Investing, and Excess Comovement of Exchange-Traded Fund Returns. *SSRN Electronic Journal*, 1-53. http://dx.doi.org/10.2139/ssrn.2307229
7. CFA (2020). *Global Investment Performance Standards (GIBS)*. Chartered Financial Analyst Institute. https://www.cfainstitute.org/-/media/documents/code/gips/2020-gips-standards-firms.ashx
8. Chu, P. K. K. (2011). Study on the tracking errors and their determinants: Evidence from Hong Kong exchange traded funds. *Applied Financial Economics*, *21*(5), 309-315. https://doi.org/10.1080/09603107.2010.530215
9. Chu, P. K. K. & Xu, D. (2021). Tracking Errors and Their Determinants. *Journal of Prediction Markets*, *15*(1), 67-96. https://doi.org/10.5750/jpm.v15i1.1851
10. De Rossi, G. (2015). Measuring the tracking error of exchange traded funds. In Koopman, S. J. & Shephard, N. (Eds.), *Unobserved Components and Time Series Econometrics* (pp. 30-44). Oxford Academic. https://doi.org/10.1093/acprof:oso/9780199683666.003.0004
11. Dorocáková, M. (2017). Comparison of ETF's performance related to the tracking error. *Journal of International Studies*, *10*(4), 154-165. https://doi.org/10.14254/2071-8330.2017/10-4/12
12. Elton, E. J., Gruber, M. J. & de Souza, A. (2019). Are Passive Funds Really Superior Investments? An Investor Perspective. *Financial Analysts Journal*, *75*(3), 7-19. https://doi.org/10.1080/0015198x.2019.1618097

13. Feder-Sempach, E. & Miziołek, T. (2023). How precisely European equity ETFs mirror their flagship benchmarks? Evidence from funds replicating performance of Euro Stoxx 50 Index. *Journal of Asset Management*, *24*(2), 121-135. https://doi.org/10.1057/s41260-022-00287-9
14. Ferreruela, S. & Mallor T. (2021). Herding in the bad times: The 2008 and COVID-19 crises. *The North American Journal of Economics and Finance*, *58*, 101531. https://doi.org/10.1016/j.najef.2021.101531
15. Frino, A. & Gallagher, D. R. (2001). Tracking S&P 500 Index Funds. *Journal of Portfolio Management*, *28*(1), 44-55. https://doi.org/10.3905/jpm.2001.319822
16. Goldfeld, S. M. & Quandt, R. E. (1973). A Markov Model for Switching Regressions. *Journal of Econometrics*, *1*(1), 3-15. https://doi.org/10.1016/0304-4076(73)90002-X
17. Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, *57*(2), 357-384. https://doi.org/10.2307/1912559
18. Hillard, J. & Le, T. D. (2022). Exchange-Traded Funds Investing in the European Emerging Markets. *Journal of Eastern European and Central Asian Research*, 9(2), 260-270. http://dx.doi.org/10.15549/jeecar.v9i2.804
19. Johnson, W. F. (2009). Tracking errors of exchange traded funds. *Journal of Asset Management*, *10*(4), 253-262. https://doi.org/10.1057/jam.2009.10
20. Johnson, B., Bioy, H., Kellet, A. & Davidson, L. (2013). On The Right Track: Measuring Tracking Efficiency in ETFs. *The Journal of Index Investing*, *4*(3), 35-41. https://doi.org/10.3905/jii.2013.4.3.035
21. Kaminsky, G. L. (2001). Mutual Fund Investment in Emerging Markets: An Overview. *The World Bank Economic Review*, *15*(2), 315-340. https://doi.org/10.1093/wber/15.2.315
22. Kim, C. J. & Nelson, C. R. (2017). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* The MIT Press. https://doi.org/10.7551/mitpress/6444.001.0001
23. Le Sourd, V. & Safaee, S. (2021). The European ETF Market: Growth, Trends, and Impact on Underlying Instruments. *Journal of Portfolio Management*, *47*(7), 95-111. https://doi.org/10.3905/jpm.2021.1.252
24. Li, M. & Zhao, X. (2014). Impact of leveraged ETF trading on the market quality of component stocks. *The North American Journal of Economics and Finance*, *28*, 90-108. https://doi.org/10.1016/j.najef.2014.02.001
25. Madhavan, A. (2014), Exchange-Traded Funds: An Overview of Institutions, Trading, and Impacts. *Annual Review of Financial Economics*, *6*(1), 311-341. https://doi.org/10.1146/annurev-financial-110613-034316
26. Meinhardt, C., Mueller, S. & Schoene, S. (2015). Physical and Synthetic Exchange-Traded Funds: The Good, the Bad, or the Ugly?. *The Journal of Investing*, *24*(2), 35-44. https://doi.org/10.3905/joi.2015.24.2.035
27. Mussavian, M. & Hirsch, J. (2002). European Exchange-Traded Funds. *Journal of Alternative Investments*, *5*(2), 63-77. https://doi.org/10.3905/jai.2002.319055
28. Osojnik, L. (2023). *Market predictors of European equity ETF performance* [Undergraduate thesis, University of Zagreb]. University of Zagreb. https://urn.nsk.hr/urn:nbn:hr:148:360888
29. Osterhoff, F. & Kaserer, C. (2016). Determinants of tracking error in German ETFs - the role of market liquidity. *Managerial Finance*, *42*(5), 417-437. https://doi.org/10.1108/MF-04-2015-0105
30. Peovski, F., Cvetkoska, V., Trpeski, P. & Ivanovski, I. (2022). Monitoring stock market returns: A stochastic approach. *Croatian Operational Research Review*, *13*(1), 65-76. https://doi.org/10.17535/crorr.2022.0005
31. Perlin, M. (2010). MS_Regress - The MATLAB Package for Markov Regime Switching Models. *SSRN Electronic Journal*, 1-40. https://doi.org/10.2139/ssrn.1714016
32. Piccotti, L. R. (2018), ETF Premiums and Liquidity Segmentation. *Financial Review*, *53*(1), 117-152. https://doi.org/10.1111/fire.12148

33. Rompotis, G. G. (2010). Does premium impact Exchange-Traded Funds' returns?. *Journal of Asset Management*, *11*(4), 298-308. https://doi.org/10.1057/jam.2009.23

34. Rompotis, G. G. (2011). Predictable patterns in ETFs' return and tracking error. *Studies in Economics and Finance*, *28*(1), 14-35. https://doi.org/10.1108/10867371111110534

35. Rompotis, G. G. (2012). The German Exchange Traded Funds. *The IUP Journal of Applied Finance*, *18*(4), 62-82. https://ssrn.com/abstract=2184748

36. Shum, P. M. & Kang J. (2013). Leveraged and inverse ETF performance during the financial crisis. *Managerial Finance*, *39*(5), 476-508. https://doi.org/10.1108/03074351311313825

37. Sy, A. N. & Ong, L. L. (2004). The Role of Mature Market Mutual Funds in Emerging Markets: Myth or Mayhem? *IMF Working Papers*, *4*(133), 1-24. https://doi.org/10.5089/9781451855524.001

38. Tsalikis, G. & Papadopoulos, S. (2019). ETFS – performance, tracking errors and their determinants in Europe and the USA. *Risk Governance and Control: Financial Markets and Institutions*, *9*(4), 67-76. https://doi.org/10.22495/rgcv9i4p6

39. Vardharaj, R., Fabozzi, F. & Jones, J. (2004). Determinants of Tracking Error for Equity Portfolios. *Journal of Investing*, *13*(2), 37-47. https://doi.org/10.3905/joi.2004.412305

40. Wong, K. H. Y. & Shum, W. C. (2010). Exchange-traded funds in bullish and bearish markets. *Applied Economics Letters*, *17*(16), 1615-1624. https://doi.org/10.1080/13504850903085035

41. Yiannaki, S. M. (2015). ETFs Performance Europe - A Good Start or Not? *Procedia Economics and Finance*, *30*, 955-966. https://doi.org/10.1016/S2212-5671(15)01346-5

42. Zawadzki, K. (2020). The performance of ETFs on developed and emerging markets with consideration of regional diversity. *Quantitative Finance and Economics*, *4*(3), 515-525. https://doi.org/10.3934/qfe.2020024