Master Finance – Grade: 8.5/10

Comparative analysis of ESMA and LASSO BIC Machine Learning models for predicting one-year future European passive stock ETFs and the impact of ETF characteristics on forecast performance

Author: Sven Peter Kooi

Date: June 28, 2024

Supervisor: Dr. E. Karmaziene

Abstract

This study evaluates the ESMA methodology against the LASSO BIC Machine Learning model for predicting one-year future returns of 42 European passive stock ETFs during the out-of-sample period from 2019 to 2023, focusing on forecast unbiasedness, accuracy and efficiency. The results show that the ESMA model produces unbiased but inaccurate and inefficient forecasts, thereby accepting the weak form of market efficiency regarding accuracy and efficiency but rejecting it regarding unbiasedness. The LASSO BIC model generates accurate but biased and inefficient forecasts, supporting semi-strong market efficiency regarding unbiasedness and efficiency but rejecting it regarding accuracy. Comparative analyses demonstrate that the LASSO BIC model is significantly more accurate and less inefficient than the ESMA model, despite higher bias. Additionally, this study explores the impact of ETF characteristics on forecast performance. The results suggest that larger ETFs initially provide more accurate forecasts, but this relationship turns negative beyond a certain size threshold. Moreover, high-liquidity-risk ETFs generate less accurate and more unbiased forecasts, whereas high-market-risk ETFs yield more biased forecasts. The findings indicate that investors should exercise caution with interpretating ESMA forecasts and consider the LASSO BIC model for better forecasting. Investors can use size and risk characteristics to select ETFs based on their preference for ETFs generating unbiased or accurate forecasts. For regulators, this study advocates for new forecasting models that incorporate macroeconomic variables and are tailored to ETF characteristics rather than a one-size-fits approach. For ETF managers, insights are provided for optimising asset allocation strategies to enhance forecast performance.

Keywords: ESMA | LASSO BIC | European passive stock ETFs | EMH | Macroeconomic variables | Forecast unbiasedness | Forecast accuracy | Forecast efficiency

Disclaimer: I used Generative Artificial Intelligence tools to resolve coding errors in STATA and Rstudio, efficiently search for literature and as a general tool to gather relevant information.

1 Introduction

Since January 1, 2023, the European Securities & Markets Authority (ESMA) has mandated that fund managers include Performance Scenario analyses in their Key Information Documents (KIDs). These analyses predict one-year holding period returns based solely on past return data, providing crucial information for investors across the European Union. Reliable forecasts are vital as investors rely on the information provided in the KIDs. This study evaluates the ESMA forecasting methodology and introduces an alternative approach using the LASSO BIC Machine Learning technique, which incorporates forward-looking macroeconomic variables. By integrating these predictors, the LASSO BIC model aims to improve the forecasts by selecting variables while minimising the Bayesian Information Criterion (BIC) (Angrist & Frandsen, 2022). Additionally, to understand the underlying processes that influence the LASSO BIC forecasts, this study examines how the size, risk and cost characteristics of the ETFs affect their respective forecasts. Therefore, the research question is:

How does the ESMA methodology on predicting one-year future returns for European passive stock ETFs deviate from a more sophisticated LASSO BIC Machine Learning prediction model and which ETF characteristics are related to higher return predictability?

Using a dataset of 42 European passive stock ETFs with monthly returns from 2008 to 2023, and focusing on the period from 2019 to 2023 for out-of-sample forecasts, I compare the ESMA and LASSO BIC methodologies. The ESMA forecasts are based on a ten-year historical return series and generate median one-year holding period returns, as the guidelines state in Commission Delegated Regulation (EU) 2017/653[1] and Commission Delegated Regulation (EU) 2021/2268[2]. In contrast, the LASSO BIC methodology incorporates macroeconomic variables, such as European unemployment, inflation, interest, yield, term spread, GDP and implied volatility rates. The LASSO BIC method only incorporates the macroeconomic variables with the highest explanatory power to ensure optimal predictive accuracy and parsimonious regression models.

[1] Commission Delegated Regulation (EU) 2017/653 is the primary regulatory document that outlines the requirements for KIDs under the Packaged Retail and Insurance-based Investment Products (PRIIPs) Regulation, providing the framework and guidelines for performance scenarios and risk assessments in the EU.

[2] Commission Delegated Regulation (EU) 2021/2268 consists of adjustments to the original 2017/653 regulation.

To compare the forecasts, I evaluate the forecast errors for unbiasedness using a t-test, accuracy using the Mean Squared Prediction Errors (MSPE) and efficiency using the Mincer-Zarnowitz regression. Moreover, to statistically compare the ESMA and LASSO BIC models, I perform comparative analyses using a t-test for unbiasedness, the Diebold-Mariano test for accuracy and an F-test for efficiency. Although ESMA regulations do not mandate robustness tests for their forecasts, I conduct several robustness tests for the LASSO BIC forecasts. The descriptive statistics and Augmented Dickey-Fuller tests confirm the absence of anomalies and unit roots, respectively. To ensure valid regression results, the study addresses multicollinearity using correlation matrices and VIF tests, autocorrelation in residuals via the Portmanteau Q test, heteroskedasticity through the Breusch-Pagan/Cook-Weisberg test, normality with the Shapiro-Wilk W test and linearity and mis-specification using the Ramsey RESET test.

This study finds that while the ESMA methodology produces unbiased forecasts, it falls short in forecast accuracy and efficiency. This implies that the ESMA methodology generates forecast errors that do not significantly deviate from zero but fails to capture the dynamics of actual returns and incorporate all relevant information. In contrast, the LASSO BIC methodology demonstrates accurate but biased and inefficient forecasts. This study identifies implied volatility and GDP as accurate predictors of future ETF returns, but not as unbiased and efficient predictors. Implied volatility is significantly negatively related to future ETF returns across all out-of-sample forecasting years, supporting previous findings regarding accurate forecasting but contradicting them in terms of unbiased and efficient forecasts (Bekaert & Wu, 2000; Li et al., 2005; Giot, 2005). GDP shows a significant positive relationship with future ETF returns for the first three out-of-sample years. This relationship aligns with prior studies in relation to accurate forecasting but contradicting them in terms of forecast unbiasedness and efficiency (Alexius & Sp, 2018; Somoye et al., 2019; Ogutu, 2011; Al-Tamimi et al., 2011; Österholm, 2016).

Comparative analysis results show that the ESMA model produces significantly more unbiased forecasts compared to the LASSO BIC model, while the LASSO BIC model generates significantly more accurate and less unbiased forecasts. However, as the LASSO BIC model still produces inefficient forecasts, the results show that the LASSO BIC model yet faces challenges in fully optimising the inclusion of macroeconomic variables.

The implications of these findings are significant for both investors and regulators. For investors, the ESMA results indicate the importance of carefully interpreting the current ETF forecasts presented in EIDs, given that the ESMA model is applied in practice. The LASSO BIC model offers an alternative, providing more accurate and efficient forecasts, which can enable more informed decision-making. For regulators, this study proposes that they should enhance regulatory frameworks by incorporating macroeconomic variables, as demonstrated by the LASSO BIC methodology, leading to more reliable forecasts.

This study extends the existing literature by thoroughly examining the weak and semi-strong form of Fama's Efficient Market Hypothesis (EMH) on the three forecast evaluation techniques through the application of ESMA and LASSO BIC forecasts. This approach provides a detailed evaluation of different forms of market efficiency for specific forecasting assessments, offering valuable and precise insights into conclusions regarding the EMH. According to the weak form, markets are efficient if future returns cannot be predicted solely based on past return data (Fama, 1970). This study examines the weak form by employing the ESMA model. The semi-strong form posits that prices reflect not only past return data but also all publicly available information, making future returns unpredictable (Fama, 1970). I examine the semi-strong form by using the LASSO BIC model, which incorporates macroeconomic variables in its forecasts. Despite extensive testing, the validity of the EMH remains a subject of debate (Lekovic, 2018; Malkiel, 2003; Grossman et al., 1980).

This study documents support for the weak form of market efficiency in terms of forecast accuracy and efficiency, consistent with prior studies. However, the results provide evidence to reject the weak form of market efficiency in terms of forecast unbiasedness, contradicting prior studies (Fama, 1970; Getmansky et al., 2004; Jensen, 1978). In contrast, the LASSO BIC forecasts partially challenge the semi-strong form of efficiency. Although these forecasts are biased and inefficient, they demonstrate accuracy. Therefore, the findings suggest rejecting the semi-strong from of efficiency in terms of forecast accuracy, but accepting it regarding forecast unbiasedness and efficiency. The accuracy outcome aligns previous research showing the superior out-of-sample performance of the LASSO technique in predicting stock prices (Li & Chen, 2014; Zhang et al., 2019; Roy et al., 2015; Sermpinis et al., 2018).
Nevertheless, this study also challenges the literature on the predictive power of the LASSO BIC model concerning forecasting unbiasedness and efficiency, suggesting that incorporating

macroeconomic variables does not fully overcome the limitations of forecasting in semi-strong efficient markets.

The next step is to identify the ETF characteristics associated with higher LASSO BIC forecast unbiasedness, accuracy and efficiency. The assets under management (AUM) and number of holdings measure the ETF size effect on return predictability. Additionally, the three-year volatility and concentration of the ten largest holdings capture the market and liquidity risk, respectively. Lastly, the management fee measures the cost effect on return predictability. I also use the asset allocation of the ETF to small cap firms, large cap firms and cash and derivatives as control variables to enhance model fit. To identify the relationships between the forecasts and ETF characteristics, the regression models employ the unbiasedness, accuracy and efficiency outcomes of the 42 ETFs as dependent variables, with the ETF characteristics serving as independent variables. Similar to the prediction regressions, robustness tests confirm the validity of the regressions regarding multicollinearity, autocorrelation, heteroskedasticity, normality, linearity and mis-specification issues.

This study finds a significant positive relationship between the number of holdings and LASSO BIC forecast accuracy, suggesting that more diversified ETFs generate more accurate forecasts. However, this relationship turns negative at a certain point, indicating a non-linear effect where initial increases in holdings improve forecast accuracy, but benefits diminish beyond a certain point. This suggests that there is an optimal number of holdings for an ETF to achieve the most accurate forecasts, beyond which the forecast accuracy diminishes. Additionally, the results regarding ETF risk indicate that ETFs with higher liquidity risk produce significantly more inaccurate but less biased forecasts. Conversely, higher market risk is significantly associated with more biased forecasts.

The implications of these findings are crucial for investors, regulators and ETF managers. For investors, the results suggest to include ETF size and risk characteristics in their decision-making processes for trading ETFs, dependent on their preference for less biased or more accurate forecasts. For regulators, the results imply that they should develop tailored forecasting methodologies rather than a one-size-fits all methodology. For ETF managers, the results indicate that they can optimise their asset allocation strategies to improve their forecasts.

While the first part of this study enriches the literature by examining the weak and semi-strong forms of market efficiency in terms of unbiasedness, accuracy and efficiency, the subsequent analysis advances the existing literature further by exploring the semi-strong form of market efficiency and identifying specific ETF characteristics that influence its acceptance or rejection. While previous research only explores the relationships between returns and asset characteristics, like size (e.g. Xiong et al., 2009; Yan, 2008), liquidity risk (e.g. Pastor & Stambaugh, 2003; Sadka, 2010), market risk (e.g. Blitz & Van Vliet, 2007; Ang et al., 2006) and costs (e.g. Gil-Bazo & Ruiz-Verdú, 2009), this study uniquely evaluates how forecast unbiasedness, accuracy and efficiency relate to asset characteristics. This approach provides significant insights into the underlying mechanisms that affect market predictability, thereby enriching the literature with valuable evidence on the drivers of forecast performance.

2 Data

2.1 Data and variables

The analysis relies on two main datasets.

First dataset. I use the data from the first dataset, provided by iShares and Yahoo Finance, for the ESMA and LASSO BIC forecasts. This dataset includes monthly percentage returns of 42 European passive stock ETFs from January 31, 2008 to December 31, 2023. The dataset starts on January 31, 2008, in compliance with the ESMA methodology, which necessitates a ten-year historical return dataset to generate the first forecast for the out-of-sample period from January 31, 2019 to December 31, 2023. I select the 42 ETFs randomly from those ETFs that meet two criteria: having access to monthly return data from January 31, 2008 onwards and complying with ESMA regulations. This allows me to make a fair comparison of the ESMA and LASSO BIC forecasting methodologies, as including non-compliant ETFs would not provide a meaningful analysis. Moreover, by selecting 42 ETFs randomly, the dataset encompasses various ETF characteristics (e.g., different sectors, asset classes and market capitalisations) which reduce potential sample selection bias. For the comparison of the ESMA and LASSO BIC methodologies, this study uses the average monthly percentage returns of the 42 ETFs, which are aggregated into a single Pooled ETF.

The ESMA method only uses the Pooled ETF data in its forecasts. However, the LASSO BIC forecasts use the Pooled ETF data as the dependent variable and incorporate forward-looking macroeconomic variables as independent variables. This study uses eight macroeconomic variables in the LASSO BIC OLS regression models, which are selected for their predictive power, as outlined in Section 3.3. I use several data sources for the macroeconomic variables. First, I employ the European unemployment rates, inflation rates, interest rates and yield curve rates provided by the European Central Bank Data Portal. Second, I use the Industrial Production rates provided by Eurostat. Lastly, I use investing.com to gather VSTOXX rates.

The unemployment rate represents the percentage proportion of individuals aged 15 to 74 who are unemployed relative to the total number of both employed and unemployed individuals in the Euro area. The inflation rate is the Harmonized Index of Consumer Prices

(HICP) overall index in the Euro area, which measures percentage changes in the prices of goods and services. Moreover, this study uses the one-year Euribor variable as proxy for the interest rates. This measures the percentage interest rate at which Eurozone banks offer to lend unsecured funds to other banks for a one-year period. Since the ESMA and LASSO BIC methods produce one-year holding period forecasts, I use the one-year Euribor rate to match this time horizon. In addition, the yield curve rates consist of the percentage one-year yield curve rates, 30-year yield curve rates and the term spread rates. The one-year and 30-year yield curve rates are the continuously compounded government bond rates for issuers with a triple-A rating in the Euro area, with maturities of one year and 30 years, respectively. The ECB uses the Svensson model to estimate these yield rates. The term spread is the difference between the 30-year yield rate and one-year yield rate. This spread is an indicator of the yield curve's shape and provides insights into market expectations for future economic activity. Similar to prior studies (Ascari & Haber, 2022; Bocchio et al., 2023), I use the Industrial Production as proxy for GDP growth. Industrial Production measures the percentage output of the industrial sector, including manufacturing and utilities, which indicates overall economic activity. While the regular GDP data provided by the ECB measures GDP on an annual basis, this benchmark offers monthly data, making it more suitable for this study. Lastly, the VSTOXX data measures the percentage European implied volatility. This indicates the market expectations of future price fluctuations and reflects investor sentiment and risk levels.

These macroeconomic variables allow me to create multiple LASSO BIC regression models, where I incorporate macroeconomic variables into the predictions rather than relying solely on past return data. Additionally, I include dummy variables for the Global Financial Crisis (GFC) and Covid-19 to control for their short-term effects on ETF returns (Chowdhury et al., 2022).

Second dataset. The second dataset contains the size, risk and cost characteristics of the 42 ETFs, provided by the iShares Factsheets, PRIIP KIDs and Morningstar. Moreover, the second dataset contains the ETF's asset composition in terms of small cap firms, large cap firms and allocation to cash and derivatives. In contrast to the first dataset, the second dataset merely contains cross-sectional data. I retrieved the data on June 5, 2024. This study uses the Assets Under Management (AUM) and the number of holdings to reflect the size of the ETF, where the AUM indicates the total market value of the assets managed by each ETF expressed in millions of euros. Second, I asses the risk through two key measures: the three-year volatility and the ten largest stake holdings in an ETF. The three-year volatility captures the percentage variability in the ETF's returns over a three-year period and it indicates market risk. The ten largest stake holdings indicates the percentage contribution of the ten highest market value assets relative to the total value of the ETF. This is a measure of concentration and liquidity risk. The management fees measure the costs of the ETFs, which implies that the management fee is the percentual annual fee that a fund manager charges for managing the ETF. Lastly, I include the percentage allocation to small cap firms, large cap firms and cash and derivatives as control variables in the regression analysis. The ETF characteristics allow me to discover which characteristics relate to higher LASSO BIC forecast predictability.

2.2 Descriptive statistics

Table 1 shows the descriptive statistics of the variables used in this study. I apply first-difference transformations to the macroeconomic variables to eliminate unit roots. The Pooled ETF returns are inherently in first-differences, as they represent the percentage monthly returns. The ETF characteristics do not require first-difference transformations, as this is cross-sectional data. I use the monthly ETF returns and macroeconomic variables from 2008 to 2023 for the ESMA and LASSO BIC forecasts, resulting in 192 observations for each variable. The ETF characteristics include 42 observations corresponding to the 42 different ETFs in the dataset. Notably, the three-year volatility variable has two missing observations and the control variables have 4 missing observations.

The monthly Pooled ETF return demonstrates a mean (median) of 0.48% (0.95%) with a standard deviation of 4.54%. This implies that, on average, one receives a monthly return of 0.48% when investing in the Pooled ETF. The lowest (highest) monthly return equals -15.49% (15.94%). The lowest monthly return was in March 2020, coinciding with the onset of the Covid-19 pandemic. The summary statistics indicate moderate negative skewness (-0.33) and low (excess) kurtosis (1.65), suggesting a distribution with a slight left tail and more extreme values than a normal distribution.

Analysing the average first-differenced macroeconomic variables, the interest rate demonstrates the most positive average change (12.40%) and standard deviation (152.98%), indicating substantial variability in the short-term interest rates. The maximum delta of 2087.79% further emphasises extreme fluctuations. The difference between the mean (12.40%) and median (-0.58%) suggests significant positive skewness (13.01), driven by extreme outliers.

This aligns with periods of economic uncertainty (e.g. Global Financial Crisis, Covid-19), where central banks apply rapid interest rate changes (Mishkin, 2019).

In contrast, the 30-year yield rate shows the most negative average change (-16.85%) and minimum delta (-608.58%), reflecting significant downward movements in long-term bond yields. The mean (-16.85%) is smaller than the median (-1.00%), implying that the mean is lower due to outliers and the distribution is negatively skewed (-4.00). The negative skewness indicates frequent large decreases, which are often associated with economic downturns when long-term rates fall due to increased demand for safer investments (Goda et al., 2011).

The unemployment rate and GDP demonstrate relatively modest average changes of -0.06% and 0.01%, respectively. This corresponds with their small standard deviations, 1.25% for the unemployment rate and 2.41% for GDP, indicating a limited variability in these macroeconomic variables. The unemployment rate has the narrowest range among all variables, spanning from -2.74% to 6.02%, which results in the lowest kurtosis value of 4.15. This suggests a distribution with fewer extreme values and relatively stable labour market conditions. In contrast, GDP shows a wider range of -18.70% to 13.70%, leading to a significantly higher kurtosis (25.09). This indicates more frequent extreme changes in economic output, reflecting periods of economic downturn and recovery. The mean of the unemployment rate is slightly higher than its median (-0.30%), resulting in a positively skewed distribution (1.40). On the other hand, GDP has a mean that is lower than its median (0.10%), producing a negatively skewed distribution (-1.76). Nevertheless, both distributions are not heavily skewed.

The inflation and one-year yield rates have a moderate negative average delta of -3.84% and -5.57%, respectively. Both macroeconomic variables have significant negative and positive outliers, resulting in high kurtosis values. This implies that the inflation and one-year yield variables are volatile, aligning with the high standard deviations of 80.92% and 86.14%, respectively. The average inflation and one-year yield rates are smaller than their median values, resulting in slightly negatively skewed distributions.

On the other hand, the summary statistics show a moderate positive average delta for the yield spread (6.45%) and implied volatility (2.19%). The delta yield spread shows a range of -158% to 1197%, which is extremely large. The high standard deviation (89.50%) and kurtosis (160.00) confirm its high variability, which is often influenced by changing market expectations of economic growth. However, the implied volatility demonstrates a narrower range of -40.30% to 146.30%, though this still indicates significant variability. The standard deviation (23.51%) and kurtosis (7.26) confirm the variability in the implied volatility. The average yield spread and implied volatility rates are larger than their median values, indicating positively skewed distributions.

The sample also includes eight variables from which five variables measure the size, risk and cost characteristics for each of the 42 ETFs and three variables – small cap, large cap and cash and derivatives – serve as control variables (Table 1). The average (median) ETF has 1.654 (495) million euros Assets Under Management (AUM), divided over 114 (42) holdings, from which 56.27% (58.28%) of the AUM is tied up in the ten largest holdings. Moreover, the average (median) ETF has a three-year volatility of 17.22% (16.46%) and annual management fee of 0.36% (0.44%). The large differences between the mean and median of the AUM and number of holdings indicate the presence of outliers. The minimum (maximum) values confirm this with 13 (14.393) million euros AUM and 11 (606) holdings. Furthermore, the distributions of the AUM and number of holdings are both positively skewed and leptokurtic. The summary statistics also show that the ten largest stake holdings in an ETF has a wide range of 7.82% to 98.21%. The maximum ten largest stake holdings value of 98.21% corresponds to the minimum number of holdings (11), indicating that almost the entire AUM is tied up in the ten largest holdings in this ETF. Furthermore, the skewness and kurtosis of the ten largest stake holdings, three-year volatility and management fee demonstrate close-to-Gaussian distributions. The standard deviations and narrow ranges of minimum and maximum values of these characteristics confirm the close-to-Gaussian distributions. Lastly, the control variables indicate that the average (median) ETF allocates 0.05% (0.00%) to small cap firms, 84.07% (88.70%) to large cap firms and 0.95% (0.83%) to cash and derivatives. The minimum and maximum values demonstrate significant variation in the allocation to large cap firms, as evidenced by a relatively high standard deviation of 17.62%. Conversely, the allocations to small cap firms and cash and derivatives show minimal variation, reflecting relatively low proportions invested in these categories.

3 Methodology

3.1 Formulation of hypotheses

I expect the LASSO BIC methodology to perform better in terms of forecasting unbiasedness, accuracy and efficiency compared to the ESMA methodology. The LASSO BIC methodology accounts for macroeconomic events, such as movements in interest rates, unemployment rates and inflation rates, which are often correlated to movements in the stock market. In contrast, the ESMA methodology relies solely on past return data and does not include these correlations in its forecasts. Therefore, by incorporating the macroeconomic variables, the LASSO BIC model likely offers higher explanatory power, leading to superior forecasts. This expectation forms the basis for the following hypothesis:

Hypothesis 1: *The LASSO BIC methodology demonstrates greater predictive unbiasedness, accuracy and efficiency compared to the ESMA methodology, aligning with the weak form of efficiency and contradicting the semi-strong form of efficiency.*

Additionally, this study examines how ETF size, risk and costs influence forecast predictability. Firstly, I hypothesise that ETF size significantly positively influences the excellence of the forecasts. Larger ETFs tend to be more diversified, leading to more stable and predictable returns. This diversification mitigates the impact of individual asset volatility, thereby enhancing the overall reliability of return forecasts. Secondly, I expect that ETF risk, as measured by the market and liquidity risk, significantly negatively impacts the forecast performance. Risky ETFs, whether characterised by high market or liquidity risk, tend to be more volatile than safe ETFs, such as government bond ETFs. I expect that high volatility increases the unpredictability of returns, making unbiased, accurate and efficient forecasting more challenging. Lastly, I hypothesise that ETF costs significantly negatively impacts the excellence of the forecasts. Higher management fees often indicate more active trading compared to passive index tracking. Active trading involves frequent buying and selling of assets, leading to increased transaction costs and higher volatility in returns. This introduces additional uncertainties regarding ETF predictions. These expectations lead to the following hypotheses:

Hypothesis 2: *Large ETFs demonstrate significantly more unbiased, accurate and efficient forecasts compared to small ETFs.*

Hypothesis 3: *Risky ETFs demonstrate significantly less unbiased, accurate and efficient forecasts compared to less risky ETFs.*

Hypothesis 4: *Expensive ETFs demonstrate significantly less unbiased, accurate and efficient forecasts compared to less expensive ETFs.*

3.2 ESMA forecast methodology

Commission Delegated Regulation (EU) 2017/653 and Commission Delegated Regulation (EU) 2021/2268 prescribe the methodology for the ESMA forecasts. I comply with these regulations by using a ten-year lookback period of monthly Pooled ETF returns to forecast one-year holding period returns. Initially, with a ten-year lookback period of monthly returns, I observe 120 returns for each forecast. However, since this study calculates one-year holding period returns, the calculation begins from the 12^{th} observation. This leads to 109 different one-year holding period returns per forecast. The ESMA regulations stipulate that I use the median return from the series of 109 one-year holding period returns to generate the Pooled ETF one-year holding period return prediction. According to the ESMA regulations, I do not perform any robustness tests.

Additionally, I adjust the ten-year lookback period for each new monthly forecast by adding the most recent return to the in-sample dataset and removing the return that exceeds the ten-year threshold, as required by the ESMA regulations. Given that the out-of-sample period in this study spans from January 31, 2019 to December 31, 2023, I repeat this process 60 times to generate the Pooled ETF one-year holding period returns on a monthly basis from 2019 to 2023.

3.3 LASSO BIC forecast methodology

For the LASSO BIC forecasts, I start by verifying the stationarity and overall quality of my dataset. Subsequently, I apply the LASSO BIC technique to the in-sample dataset to identify significant macroeconomic variables for the regression models. I ensure robustness in the in-sample regression models and estimate the coefficients of the macroeconomic variables. Finally, I generate the Pooled ETF one-year holding period forecasts for each out-of-sample year. The remainder of this section provides a detailed four-step description of the LASSO BIC forecast methodology, as it is a comprehensive methodology for forecasting.

Step 1.

As outlined in Section 2.2, I use first-difference transformations for the macroeconomic variables to eliminate unit roots. I conduct the Augmented Dickey-Fuller (ADF) test to confirm the effectiveness of the first-difference transformations. The ADF results indicate that each (in)dependent variable is stationary at a confidence level of at least 95%. Moreover, I examine the outliers identified in the descriptive statistics using quantile-quantile plots to ensure the robustness of the dataset. These plots confirm that the outliers are either vertical or horizontal with good leverage points, indicating that they do not significantly change the slope of the regression line when removed. Consequently, I do not remove or winsorize the outliers. Lastly, given the robustness of the LASSO technique against non-normality, I do not perform any other adjustments to the data prior to conducting the LASSO BIC regressions.

Step 2.

I apply the LASSO BIC technique for each out-of-sample forecasting year. Each LASSO BIC model uses a different in-sample dataset, as forecasts progress sequentially by one year. Specifically, the in-sample dataset spans 2008-2017 for the 2019 predictions, 2008-2018 for 2020, 2008-2019 for 2021, 2008-2020 for 2022 and 2008-2021 for 2023. This approach ensures that each forecast is based on the most recent available data. Based on the different in-sample periods, I apply the LASSO BIC technique five times to the following regression model:

$$Pooled\ ETF_t = \alpha + \beta_1\, \Delta Unemp_t + \beta_2\, \Delta Infl_t + \beta_3\, \Delta TYMY_t + \beta_4\, \Delta OYMY_t + \beta_5\, \Delta YS_t + \beta_6\, \Delta GDP_t + \beta_7\, \Delta IV_t + \beta_8\, \Delta IR_t + \beta_9\, \text{GFCDummy}_t + \beta_{10}\, \text{CovidDummy}_t + \varepsilon_t \tag{1}$$

Where $Pooled\ ETF_t$ is the percentage monthly return of the Pooled ETF at time t, $Unemp_t$ is the percentage monthly European unemployment rate at time t, $Infl_t$ is the percentage monthly European inflation rate at time t, $TYMY_t$ is the percentage monthly European 30-year maturity yield rate at time t, $OYMY_t$ is the percentage monthly European one-year maturity yield rate at time t, YS_t is the percentage monthly European yield spread at time t, GDP_t is the percentage monthly European gross domestic product at time t, IV_t is the percentage monthly European implied volatility rate at time t, IR_t is the percentage monthly European interest rate at time t, GFCDummy_t is a dummy variable for the global financial crisis and CovidDummy_t is a dummy variable for the Covid-19 pandemic.

Next to two dummy variables to account for the short-term effects of the global financial crisis and the Covid-19 pandemic, I incorporate eight macroeconomic variables in equation (1). The macroeconomic variables are selected for their demonstrated predictive power in prior studies. Research shows that unemployment rates have predictive power for stock prices and economic growth, exhibiting a negative relationship between the two (Wojdylo, 2009; Shah et al., 2022; Elorhor, 2019). Contrarily, while maintaining the predictive power of unemployment rates, several studies find positive relationships (Arnott et al., 2016; Li & Suominen, 2020; Shiblee, 2009; Gonzalo & Taamouti, 2017). Some studies find differing results for the relationships, but still highlight the importance of unemployment rates in prediction models (Boyd et al., 2005; Boyd et al., 2006; Pan, 2018). For instance, according to Boyd et al. (2005) and Boyd et al. (2006), an increase in unemployment rate is positively associated with stock prices during economic expansions while it is negatively related to stock prices during economic recessions.

Similarly, inflation rates have significant predictive power for stock prices and returns. While studies predominantly find a negative relationship (Adrangi et al., 2002; Wu, 2000; Kaul, 1990; Spyrou, 2001; Eldomiaty et al., 2020), some evidence suggests a positive relationship (Choudhry, 2001). Regardless of the direction of the relationship, inflation rates are crucial in prediction models (Chen, 2009; Nelson, 1976; Reddy, 2012; Rapach et al., 2005). For instance, Chen (2009) shows that inflation rates are the most useful predictors of recessions.

The third macroeconomic variable with significant predictive power for stock prices and returns is the interest rate. Numerous studies establish a negative relationship (Alam & Uddin, 2009; Lei, 2007; Tursoy, 2019), while others find a positive relationship (Katechos, 2011; Eldomiaty et al., 2020). Interest rates play a crucial role in stock prediction models (Chen, 2009; Hjalmarsson, 2010; Pimentel & Choudhry, 2014; Lioui & Maio, 2014). Rapach et al. (2005) even argue it is the most consistent and reliable predictor of stock returns.

The next macroeconomic variables with significant predictive power are the one-year maturity yield rate, 30-year maturity yield rate and the term spread. Long-term maturities are particularly appealing to institutional investors, such as pension funds, aligning with the preferred habitat theory (Greenwood & Vissing-Jorgensen, 2018). The preference for long-term yield assets lead to higher demand of these assets, which influences their yields. Hence, it is crucial to consider the yield differences between short-term and long-term assets, which is the term spread (Greenwood & Vissing-Jorgensen, 2018). Prior studies confirm that term spread is a good predictor of stock prices (Hjalmarsson, 2010; Faria & Verona, 2020; Bluwstein et al., 2023; Chen, 2009; Croushore & Marsten, 2016).

GDP shows both positive (Alexius & Sp, 2018; Somoye et al., 2019; Ogutu, 2011; Al-Tamimi et al., 2011; Österholm, 2016) and negative relationships (Dimitrova, 2005; Hakim & Tursoy, 2021) with stock prices and returns. While studies often use GDP as a control variable (Chen, 2009; Lawrenz & Zorn, 2017), it remains essential in prediction models. Lastly, volatility is a predictive variable for stock price and return forecasting, demonstrating a negative relationship (Bekaert & Wu, 2000; Li et al., 2005; Giot, 2005).

Step 3

By applying equation (1) to each distinct in-sample period, the LASSO BIC method selects the most predictive macroeconomic variables for each Pooled ETF out-of-sample forecasting year. Consequently, I develop five different regression models, each using different in-sample periods to estimate the coefficients of the LASSO BIC selected variables for the Pooled ETF prediction models. Table 2 presents the in-sample coefficient results used in the prediction models. I discuss the implications of these results in Section 4.1.1. This leads to the following final Pooled ETF prediction models:

2019

$$Pooled\ ETF_t = 0.008 - 0.437\ \Delta Unemp_t + 0.507\ \Delta GDP_t - 0.151\ \Delta IV_t - 0.012\ \mathrm{GFCDummy}_t + \varepsilon_t \tag{2}$$

2020

$$Pooled\ ETF_t = 0.007 - 0.351\ \Delta Unemp_t + 0.538\ \Delta GDP_t - 0.149\ \Delta IV_t - 0.012\ \mathrm{GFCDummy}_t + \varepsilon_t \tag{3}$$

2021

$$Pooled\ ETF_t = 0.007 - 0.384\ \Delta Unemp_t + 0.467\ \Delta GDP_t - 0.150\ \Delta IV_t - 0.013\ \mathrm{GFCDummy}_t + \varepsilon_t \tag{4}$$

2022

$$Pooled\ ETF_t = 0.004 + 0.189\ \Delta GDP_t + 1.470\ \Delta GDP_t^{\ 2} - 0.169\ \Delta IV_t + 0.099\ \Delta IV_t^{\ 2} - 0.046\ \text{GFCDummy}_t - 0.070\ \text{CovidDummy}_t + \varepsilon_t \quad (5)$$

2023

$$Pooled\ ETF_t = 0.005 + 0.181\ \Delta GDP_t + 1.634\ \Delta GDP_t^{\ 2} - 0.161\ \Delta IV_t + 0.096\ \Delta IV_t^{\ 2} - 0.049\ \text{GFCDummy}_t - 0.075\ \text{CovidDummy}_t + \varepsilon_t \quad (6)$$

Where $Pooled\ ETF_t$ is the percentage monthly return of the Pooled ETF in month *t*, $Unemp_t$ is the percentage monthly European unemployment rate in month *t*, $GDP_t^{\ (2)}$ is the (squared) percentage monthly European gross domestic product in month *t*, $IV_t^{\ (2)}$ is the (squared) percentage monthly European implied volatility in month *t*, GFCDummy_t is a dummy variable for the global financial crisis and CovidDummy_t is a dummy variable for the Covid-19 pandemic.

I conduct several robustness tests to ensure the validity of the LASSO BIC regression models, presented in equation (2) to (6). First, I perform the Variance Inflation Factor (VIF) tests for each regression model and analyse the correlations among the macroeconomic variables to confirm that no multicollinearity issues exist. Additionally, I address autocorrelation in the residuals using the Portmanteau Q test (Ljung-Box test), Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). These tests confirm the absence of autocorrelation in all regression models. Moreover, I perform the Breusch-Pagan/Cook-Weisberg test for heteroskedasticity issues. These tests indicate that there is no heteroskedasticity in the 2019-2021 models, which allow me to use normal standard errors. However, the 2022 and 2023 models demonstrate heteroskedasticity, significant at the 99% confidence level. Consequently, these models use Newey-West (HAC) standard errors. I use the Shapiro-Wilk W test to test for normality. The tests show that the data in each regression model is not normally distributed, significant at the 99% confidence level. Nevertheless, since the sample size exceeds 30 observations, I assume normality by accepting the Central Limit Theorem. Lastly, I address non-linearity and mis-specification issues via the Ramsey RESET test. The 2019-2021 models do not show such issues. However, the 2022 and 2023 models contain these issues, significant at the 99% confidence level. Consequently, I include squared macroeconomic variables in equations (5) and (6) to account for non-linear and mis-specified patterns.

Step 4

I use equations (2) to (6) to predict the twelve monthly returns of the Pooled ETF for each out-of-sample forecasting year. I use the 'predict' command in STATA to perform this prediction. This yields a series of monthly forecasted returns from January 31, 2019 to December 31, 2023. Consequently, I compute the rolling annual returns from the series of monthly returns to generate the one-year holding period forecasts.

3.4 ESMA and LASSO BIC forecast errors & evaluations

I use equation (7) to compare the predicted ESMA (Section 3.2) and LASSO BIC (Section 3.3) one-year holding period returns to the realised one-year holding period returns of the Pooled ETF. This comparison results in a monthly series of ESMA and LASSO BIC forecast errors from January 31, 2019 to December 31, 2023. To examine the forecast errors, I evaluate them on unbiasedness, accuracy and efficiency. Additionally, I assess the relative unbiasedness, accuracy and efficiency between the ESMA and LASSO BIC forecasts.

$$e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t} \quad (7)$$

Where $e_{t+1|t}$ is the percentage forecast error in month t, y_{t+1} is the percentage one-year holding period realised return in month t and $\hat{y}_{t+1|t}$ is the ESMA or LASSO BIC percentage one-year holding period predicted return in month t.

Unbiasedness

I evaluate the absolute unbiasedness of both forecasting methodologies by performing a t-test on the forecast errors from equation (7). The t-test determines whether the average forecast errors significantly deviate from zero, indicating if the forecasts systematically underestimate or overestimate realised returns. I use Newey-West (HAC) standard errors in the t-tests to ensure robustness, as the Portmanteau Q test demonstrates autocorrelations in the forecast errors of both ESMA and LASSO BIC methodologies.

In addition, to evaluate the relative unbiasedness between the ESMA and LASSO BIC methodologies, I conduct a t-test on the differences in their forecast errors. To ensure robustness, I apply Newey-West (HAC) standard errors.

Accuracy

I assess the absolute accuracy of both forecasting methodologies by comparing the Mean Squared Prediction Error (MSPE) to the natural benchmark of the model, which corresponds to the variance of the Pooled ETF in the out-of-sample window. A forecast is considered accurate when the 'VarMinusMSPE' is positive, where the 'VarMinusMSPE' is the variance in the out-of-sample window minus the MSPE of the Pooled ETF. I use the following formula for the MSPE:

$$MSPE = \frac{1}{p}\sum_{t=T_1}^{T-1}(e_{t+1|t})^2 \tag{8}$$

Where p is the number of observed forecast errors, T is the end of the forecast window and $e_{t+1|t}$ is the percentage forecast error from equation (7) in month *t*.

Moreover, I evaluate the relative accuracy between the ESMA and LASSO BIC methodologies by employing the Diebold-Mariano test. This involves performing a t-test on the differences in the squared prediction errors of the two methodologies. I use Newey-West (HAC) standard errors to ensure robust results.

Efficiency

This study assesses the absolute efficiency of both forecasting methodologies with the Mincer-Zarnowitz regression (equation 9). I perform an F-test to examine if $\beta_0 = 0$ and $\beta_1 = 1$. These conditions must be met for an efficient forecast, as any deviation suggests that not all available information is incorporated into the forecast, implying inefficiency.

$$y_{t+1} = \beta_0 + \beta_1\hat{y}_{t+1|t} + \eta_{t+1} \tag{9}$$

Where y_t is the percentage one-year holding period realised return in month *t* and $\hat{y}_t$ is the ESMA or LASSO BIC percentage one-year holding period predicted return in month *t*.

Furthermore, I examine the relative efficiency between the ESMA and LASSO BIC methodologies by comparing their F-statistics from the absolute efficiency tests. I compute the F-rate by dividing the F-statistics of the two methodologies and test the statistical significance of this ratio by an F-test.

3.5 ETF characteristic analysis

Initially, I use the Pooled ETF for a uniform comparison of ESMA and LASSO BIC methodologies. However, to assess the impact of unique ETF characteristics on their respective forecast performance, I need the forecasts for each individual ETF. These forecasts include the most predictive macroeconomic variables selected by LASSO BIC for each, rather than using a uniform set of macroeconomic variables. I conduct the ETF characteristic analysis exclusively to the LASSO BIC forecasts, as they provide superior overall performance compared to the ESMA forecasts.

Step 1

This entails repeating the steps in Sections 3.3 and 3.4 for each ETF individually. Consequently, I apply the LASSO BIC technique to equation (1) across five out-of-sample years for 42 different ETFs, resulting in 210 regression models. I ensure the robustness, estimate in-sample coefficients and predict out-of-sample forecasts for each of the 210 models. Then, I evaluate the forecasts on unbiasedness, accuracy and efficiency, yielding 42 t-statistics for unbiasedness, VarMinusMSPE statistics for accuracy and F-statistics for efficiency. Higher absolute t-statistics indicate greater likelihood of biased forecasts, positive VarMinusMSPE values indicate greater accuracy and larger F-statistics suggest greater inefficiency.

Step 2

To evaluate the relationship between return predictability and ETF characteristics, I use the outcomes of unbiasedness, accuracy and efficiency as dependent variables and five ETF characteristics – representing ETF size, risk and costs – as independent variables in the regression models. Moreover, I include the asset composition of the ETFs, specifically the percentages allocated to small cap firms, large cap firms and cash & derivatives, as control variable. Including these control variables improves the model's fit and allows for the isolation of the size, risk and cost characteristics that this study examines. Prior studies find relationships between stock returns and size (e.g. Xiong et al., 2009; Yan, 2008), liquidity risk (e.g. Pastor & Stambaugh, 2003; Sadka, 2010), market risk (e.g. Blitz & Van Vliet, 2007; Ang et al., 2006) and costs (e.g. Gil-Bazo & Ruiz-Verdú, 2009). This study builds on these findings by examining whether these characteristics also influence the unbiasedness, accuracy and efficiency of return predictions.

This study ensures robustness by testing for multicollinearity, autocorrelations, heteroskedasticity, normality, linearity and mis-specification issues. I use white robust standard errors[3] in the accuracy and efficiency models due to heteroskedasticity. The Ramsey-RESET test indicates significant linearity and mis-specification issues in the accuracy and efficiency models, leading to the use of logarithmic[4] variable transformations. However, I only use logarithmic variable transformations for AUM and number of holdings, as the logarithmic values of the other characteristics exhibit multicollinearity issues. Despite the unbiasedness model showing no linearity or mis-specification issues, I include the logarithmic AUM to prevent coefficients of zero. Hence, I use the following three regression models:

$$Unbiasedness_i = \beta_0 + \beta_1\text{AUM}_i + \beta_2\text{NOH}_i + \beta_3\text{TLSH}_i + \beta_4\text{TYV}_i + \beta_5\text{MF}_i + \beta_6\text{logAUM}_i + \beta_7\text{SC}_i + \beta_8\text{LC}_i + \beta_9\text{CD}_i + \varepsilon_i \quad (10)$$

$$Accuracy_i = \beta_0 + \beta_1\text{AUM}_i + \beta_2\text{NOH}_i + \beta_3\text{TLSH}_i + \beta_4\text{TYV}_i + \beta_5\text{MF}_i + \beta_6\text{logAUM}_i + \beta_7\text{logNOH}_i + \beta_8\text{SC}_i + \beta_9\text{LC}_i + \beta_{10}\text{CD}_i + \varepsilon_i \quad (11)$$

$$Efficiency_i = \beta_0 + \beta_1\text{AUM}_i + \beta_2\text{NOH}_i + \beta_3\text{TLSH}_i + \beta_4\text{TYV}_i + \beta_5\text{MF}_i + \beta_6\text{logAUM}_i + \beta_7\text{logNOH}_i + \beta_8\text{SC}_i + \beta_9\text{LC}_i + \beta_{10}\text{CD}_i + \varepsilon_i \quad (12)$$

Where $Unbiasedness_i$ is the absolute t-statistic of ETF i, $Accuracy_i$ is the VarMinusMSPE value of ETF i, $Efficiency_i$ is the F-statistic of ETF i, AUM_i is the assets under management in millions of euros of ETF i, NOH_i is the number of holdings of ETF i, TLSH_i is the percentage proportion of the ten largest stake holdings in ETF i, TYV_i is the percentage three-year volatility of ETF i, MF_i is the percentage annual management fee of ETF i, logAUM_i is the logarithmic assets under management of ETF i, logNOH_i is the logarithmic number of holdings in ETF i, SC_i is the percentage asset allocation to small cap firms in ETF i, LC_i is the percentage asset allocation to large cap firms in ETF i and CD_i is the percentage asset allocation to cash and derivatives in ETF i,

.

[3] I use Newey-West HAC standard errors in the LASSO BIC forecasts due to the time series nature of the data. Conversely, I use white robust standard errors in the ETF characteristic regressions, as this dataset comprises only cross-sectional data.

[4] I use squared variable transformations in the LASSO BIC forecasting procedure to address issues of linearity and model mis-specification. However, squared transformations are ineffective in the ETF characteristic regressions. Therefore, I tested logarithmic, square root and inverse variables transformations and found that logarithmic transformations provide the most robust results. Consequently, I use logarithmic transformations in the ETF characteristic regression models.

4 Results

4.1 Comparative performance of ESMA and LASSO BIC forecast models

4.1.1 Empirical results

Table 3 shows that the LASSO BIC forecast models outperform the ESMA forecast models in terms of accuracy and efficiency for the 60 out-of-sample observations, although they fall short in terms of unbiasedness. Table 4 confirms that these performance differences are statistically significant across all measures. Figures 1 and 2 graphically substantiate the findings presented in Tables 3 and 4.

Table 3 demonstrates that the ESMA forecasts show unbiasedness with a t-statistic of -0.45, indicating that the average ESMA forecast error of -2.02% does not significantly deviate from zero. In contrast, the LASSO BIC forecasts show biasedness with a t-statistic of 1.82, significant at the 90% confidence level. This implies that the average LASSO BIC forecast error of 5.04% significantly deviates from zero. Consequently, the LASSO BIC forecasts significantly underestimate the realised returns, leading to significant positive forecast errors. This indicates that users of the LASSO BIC model systematically expect lower returns compared to the realised returns. Conversely, the ESMA forecasts slightly overestimate the realised returns, though this effect is not significant. This means that users of the ESMA model obtain higher expected returns compared to the realised returns. Since the ESMA regulations mandate fund managers to use the ESMA model, this study provides empirical evidence that the current ESMA model produces unbiased forecasts, confirming its adequacy in terms of unbiasedness for practical use.

Furthermore, Table 3 demonstrates that the LASSO BIC (ESMA) model produces accurate (inaccurate) forecasts. The MSPE of the LASSO BIC forecasts (0.0116) is smaller than the variance in the benchmark model (0.025). This implies that the LASSO BIC model has a smaller variance than the actual results in the out-of-sample window, resulting in a positive VarMinusMSPE statistic, indicating accuracy. Conversely, the MSPE of the ESMA forecasts (0.0275) is larger than the variance in the benchmark model, resulting in a negative VarMinusMSPE statistic, indicating inaccuracy. The summary statistics (Table 3) of both forecast errors substantiate the results of the accuracy tests, as the standard deviation, minimum and maximum values of the LASSO BIC forecast errors are closer to zero than those of the ESMA forecast errors. The results imply that users of the ESMA model expect returns that deviate more from the actual returns compared to users of the LASSO BIC model. Therefore, the LASSO BIC model better captures the dynamics of the actual returns than the ESMA model. These findings provide empirical evidence of the ESMA model's inadequacy in forecasting accuracy and suggests that the LASSO BIC model serves as a valid alternative.

Additionally, the efficiency results in Table 3 demonstrate that both methodologies produce inefficient forecasts at the 99% confidence level. This means that both models do not incorporate all relevant available information. In other words, it is possible to forecast the ESMA and LASSO BIC forecast errors with the available information since this information is not included in the forecasting models. Nevertheless, since the F-statistic of the LASSO BIC model (8.84) is smaller than the F-statistic of the ESMA model (14.59), it implies that the LASSO BIC model produces less inefficient forecasts compared to the ESMA model. The findings provide empirical evidence for the inadequacy of the ESMA model in terms of forecasting efficiency. I also find that the LASSO BIC model can be used as an alternative to the ESMA model for less inefficient forecasts, though the LASSO BIC model still provides significant inefficient forecasts.

Although Table 3 shows that the LASSO BIC model produces more biased, accurate and less inefficient forecasts compared to the ESMA model, Table 4 confirms that these differences are statistically significant across all measures. The results show a statistically significant difference in unbiasedness at the 90% confidence level, with a t-statistic of -1.84. This suggests that the ESMA model produces significantly more unbiased forecasts than the LASSO BIC model. Moreover, the Diebold-Mariano test reveals that the difference in accuracy is significant at the 99% confidence level, with a Diebold-Mariano test-statistic of 3.29. This implies that the LASSO BIC model produces significantly more accurate forecasts than the ESMA model. Also, the F-stat of 1.65 demonstrates that the difference in efficiency is significant at the 95% confidence level. This confirms that the LASSO BIC model generates significantly less inefficient forecasts compared to the ESMA model.

Figures 1 and 2 provide a graphical confirmation of the results presented in Tables 3 and 4. Figure 1 illustrates the generated ESMA and LASSO BIC forecasts, as outlined in sections 3.2 and 3.3, alongside the realised returns. The LASSO BIC model closely tracks the actual returns, suggesting that the inclusion of macroeconomic variables allows the model to accurately explain the observed returns. In contrast, the ESMA model demonstrates a rigid horizontal line, failing to capture the dynamics of the actual returns, thereby producing inaccurate forecasts. Figure 2 depicts the forecast errors, calculated using equation (7), which represent the differences between the forecasts and the actual returns from Figure 1. The graphical representation of the forecast errors shows significant deviations from zero for the ESMA model, confirming its inaccuracy. This also confirms the inefficiency of the forecast, as an efficient forecast should have forecast errors equal to zero. Despite the substantial deviation from zero at various points in time, the average forecast error remains close to zero, supporting its unbiased nature. Conversely, the LASSO BIC model shows smaller deviations from zero, indicating higher forecast accuracy. While accurate forecasts can still contain some non-zero forecast errors, the LASSO BIC model's forecast errors are not efficient as the forecast errors are not equal to zero. Additionally, Figure 2 shows that the average LASSO BIC forecast error is primarily positive, confirming the systematic bias in the LASSO BIC forecasts.

I provide a couple of explanations for the unbiasedness, accuracy and efficiency outcomes. The inclusion of additional variables in the LASSO BIC model increases the exposure to biases inherent to these macroeconomic variables, resulting in systematic forecast biases. Table 2 shows that implied volatility is highly significant across all out-of-sample years, highlighting its critical role in producing the LASSO BIC forecasts but also potentially introducing bias. The significance of GDP in the prediction models for 2019 to 2021 could also explain the forecast biases. In addition, Table 1 shows high kurtosis values for implied volatility and GDP, which may have impacted the biasedness of the LASSO BIC model. In contrast, the ESMA model, which is less prone to such external biases, generates more unbiased forecasts. The summary statistics of the Pooled ETF returns in Table 1 demonstrate low skewness and kurtosis, which could have contributed to the unbiased forecasts. Furthermore, the simplicity of the ESMA model minimises the risk of overfitting, whereas the moderate complexity of the LASSO BIC model increases the risk by potentially fitting to specific training data that do not generalise well to out-of-sample data.

On the other hand, additional variables improve forecast accuracy. The ESMA model lacks accuracy due to the limited predictive power of historical return data. In contrast, the LASSO BIC model's macroeconomic variables, especially implied volatility and GDP, explain future returns. This is evidenced by their significance in Table 2 and prior studies (Section 3.3). Therefore, LASSO BIC models produce accurate forecasts. Moreover, the summary statistics of the implied volatility in Table 1 show high standard deviation, minimum and maximum values. Models that handle these extreme values well will likely be more accurate. As the implied volatility exhibits outliers and given that implied volatility is a highly significant predictor, the LASSO BIC model handles the implied volatility data well, resulting in accurate forecasts.

Despite accurate LASSO BIC forecasts, efficiency is lacking. Two main reasons explain this inefficiency. First, the significant constant terms in the first three out-of-sample years in Table 2 suggest missing macroeconomic variables with explanatory power. Second, the LASSO BIC methodology penalises additional macroeconomic variables to maintain model parsimony. Consequently, while the LASSO BIC model achieve accurate forecasts, it reduces efficiency by not incorporating a broader set of explanatory variables. The R-squared values in Table 2 confirm adequate model fit, which yields accurate forecasts. However, they also indicate room for improvement, which highlights the inefficiency. This suggests that the LASSO BIC method may over-penalise additional variables, resulting in lower R-squared values and reduced efficiency.

4.1.2 Economic significance of ESMA and LASSO BIC model results

For investors, the ESMA model's ability to provide significantly (more) unbiased forecasts enhances decision-making in financial contexts. Unbiased forecasts are crucial for sound investment decisions, as they prevent systematic overestimation or underestimation of returns. This reliability is important for effective risk management because biased forecasts can lead to misestimation of financial risks. However, the inaccuracy and inefficiency of the ESMA forecasts increase uncertainty, which negatively impacts the decision-making processes of investors. Given that the ESMA model is currently applied in practice, the findings suggest that investors might be making decisions based on inaccurate and inefficient forecasts. This could lead to suboptimal investment strategies, financial losses and unexpected risk exposures. These results underscore that relying solely on past returns is insufficient for accurate and efficient forecasting. The LASSO BIC model could enhance the decision-making processes of the investors, as this model process accurate and less inefficient forecasts. The accurate LASSO BIC forecasts reduce uncertainty for investors regarding future returns and risk exposures.

For regulators, the accuracy and reduced inefficiency of the LASSO BIC model make it an alternative for implementation in new regulatory frameworks. The critical role of implied volatility and GDP in these forecasts suggests that regulators should introduce new prediction models incorporating these two variables rather than relying solely on past return data. However, as the LASSO BIC model still produces inefficient forecasts, it highlights the need for regulators to explore and include a broader set of macroeconomic variables in forecast models to enhance efficiency.

4.1.3 Evaluation of Hypothesis 1

Hypothesis 1: *The LASSO BIC methodology demonstrates greater predictive unbiasedness, accuracy and efficiency compared to the ESMA methodology, aligning with the weak form of efficiency and contradicting the semi-strong form of efficiency.*

The findings suggest to partially accept and reject Hypothesis 1. Specially, I accept Hypothesis 1 based on the findings that the LASSO BIC model produces significantly more accurate and less inefficient forecasts than the ESMA model. However, I reject Hypothesis 1 regarding unbiasedness, as the ESMA model generates significantly more unbiased forecasts than the LASSO BIC model.

Furthermore, the results support the weak form of market efficiency in terms of forecast accuracy and efficiency. The ESMA model's inability to predict future returns accurately and efficiently confirms that using only past return data does not yield accurate and efficient forecasts. However, since the ESMA model generates unbiased forecasts, it implies that past return data can produce unbiased forecasts, which should not occur in a weakly efficient market. Therefore, I reject the weak form of market efficiency regarding unbiasedness.

Additionally, I document support for the semi-strong form of market efficiency regarding forecast unbiasedness and efficiency. While the LASSO BIC forecasts produce less inefficient forecasts compared to the ESMA model, the forecasts still demonstrate inefficiency. This indicates that using publicly available information in the form of macroeconomic variables does not result in unbiased and efficient forecasts, aligning with semi-strong market efficiency. Nonetheless, I reject the semi-strong form of market efficiency regarding forecast accuracy, as the LASSO BIC model produces accurate forecasts. This contradiction challenges the notion of semi-strong market efficiency and suggests that incorporating publicly available information in the form of macroeconomic variables leads to accurate forecasts.

4.2 Relationships between ETF characteristics and forecasting performance

4.2.1 Empirical results of size characteristics

Table 5 shows a significant negative relationship between logarithmic AUM and forecast accuracy (-0.232*), indicating that ETFs with higher logarithmic AUMs generate significantly less accurate forecasts. The insignificance and zero-coefficient results of the normal AUM suggest that the negative relationship only holds when AUM reaches a certain high threshold. The number of holdings demonstrate comparable results. While the normal number of holdings shows a slightly significant positive relationship with forecast accuracy (0.002*), the logarithmic number of holdings demonstrates a significant negative relationship (-1.014*). This pattern suggests that an initial increase in the number of holdings significantly enhances ETF forecast accuracy. However, as the number of holdings continues to increase, the effect diminishes and eventually becomes negative. Therefore, the results consistently indicate that while larger ETFs and those with more holdings initially lead to more accurate forecasts, there exists a threshold beyond which the relationship turns significantly negative. Additionally, the analysis does not reveal significant relationships between AUM, number of holdings and forecast unbiasedness or efficiency.

I provide a few explanations for the size results. Initially, an increase in the number of holdings significantly improves accuracy, likely due to diversification. Previous studies indicate that portfolios with 30 to 40 assets achieve diversification, where idiosyncratic risk is minimised and only systemic risk remains (Statman, 1987; Leković, 2018). This early diversification effect enhances forecast accuracy, as diversified portfolios are less risky and thus more predictable. However, the diversification benefit does not continue to grow with an increasing number of holdings beyond the initial 30 to 40 assets, as the portfolio is already adequately diversified. Larger ETFs introduce greater complexity and operational challenges. These factors could outweigh the diversification benefits, leading to less stable ETFs and decreased forecast accuracy. The minimum (11), mean (114) and maximum (606) number of holdings in Table 1 confirm that the ETFs in the dataset show different levels of diversification, supporting the finding that initial increases in holdings could improve forecast accuracy. Nevertheless, high skewness and kurtosis values for AUM and number of holdings in Table 1 suggest that the negative relationship for logarithmic AUM and number of holdings may be influenced by outliers. Although the accuracy models use Newey-West (HAC) standard error and rely on the Central Limit Theorem for normality, it remains crucial to consider that extreme values might drive the observed negative relationship, rather than it being a generalisable effect across the dataset.

4.2.2 Empirical results of risk characteristics

Table 5 demonstrates that liquidity risk, measured by the percentage proportion invested in the ten largest stake holdings, has a significant negative relationship with both forecast biasedness (-4.266**) and accuracy (-1.469*). This implies that ETFs with higher liquidity risk produce forecasts that are significantly less biased and accurate. The results do not reveal significant relationships between liquidity risk and forecast efficiency. The market risk, measured by three-year volatility, reveals a significant positive relationship with unbiasedness (18.973*), suggesting that more volatile ETFs produce more biased forecasts. The results do not show significant relationships between market risk and forecast accuracy or efficiency.

Several factors could explain the observed risk results. When a large proportion of the ETF's market value is tied up in the ten largest assets, it could imply that it is challenging to sell these assets for a stable price, especially during stress markets. This increases the price volatility of the ETF, reducing the predictability of the ETF's returns, leading to more inaccurate forecasts. However, because the forecasts of these ETFs rely less on macroeconomic variables and more on the performance of a few assets, systemic biases inherent to macroeconomic variables are reduced, resulting in more unbiased forecasts. Additionally, ETFs with high market risk produce significantly more biased forecasts. An explanation of this finding could be that these ETFs with greater volatility are more susceptible to macroeconomic variables, as they have less stable cash flows. Consequently, the inherent macroeconomic biases influence the forecasts of these ETFs, leading to biased forecasts.

However, these explanations present a notable contradiction, as it remains unclear whether higher volatility in ETFs increases or decreases their susceptibility to the macroeconomic variables used in this study. In this study, the LASSO BIC model's accuracy is the best measure of susceptibility to the macroeconomic variables. Table 5 presents a positive relationship between market risk and accuracy (2.96), implying that higher volatility could lead to more susceptibility to macroeconomic variables. However, as this relationship is insignificant, no strong conclusions can be drawn. Moreover, it is based on the assumption that the accuracy of the model measures susceptibility to macroeconomic variables, which might not be true. This highlights an important area for further research, detailed in Section 5.2.

4.2.3 Empirical results of cost characteristics

The results indicate that ETFs with higher management fees tend to produce more biased, less accurate and less efficient forecasts (Table 5). However, since these relationships are not statistically significant, I cannot draw strong conclusions regarding the impact of ETF costs and forecast performance. The summary statistics from Table 1 and the conducted robustness tests confirm the reliability of the analysis. Hence, the insignificance of the results may be attributed to other factors offsetting the effects of management fees, an insufficient sample size or the possibility that management fees do not impact forecast performance.

4.2.4 Economic significance of size, risk and cost results

For investors, the results suggest considering the size and risk characteristics of an ETF in their decision-making process, based on whether they prioritise ETFs that produce unbiased or accurate forecasts. Investors seeking accurate ETF forecasts could focus on ETFs with relatively low AUM, moderate number of holdings and low liquidity risk. However, the findings suggest to avoid overly small ETFs due to the initial benefits of diversification. Additionally, investors seeking ETFs that produce unbiased forecasts could narrow their pool of ETFs by only considering ETFs with high liquidity risk and low market risk. The relationship between liquidity risk and forecast unbiasedness is particularly robust, significant at the 95% confidence level, while market risk shows significance at the 90% confidence level. The results indicate there is a trade-off for investors regarding liquidity risk, as ETFs with higher liquidity risk produce less accurate but more unbiased forecasts. This implies that investors need to decide whether they prefer unbiased or accurate forecasts based on their individual preferences. Nevertheless, it is important to recognise that forecasts are inherently uncertain and relying solely on the size and risk effects is often not sufficient for making well-considered investment decisions.

For regulators, the findings suggest the need for differentiated forecasting methodologies tailored to specific ETF characteristics. Section 4.1 provides statistically significant evidence that the LASSO BIC model produces more accurate and efficient forecasts compared to the ESMA model, recommending the LASSO BIC model as an alternative forecasting model. However, Section 4.2 indicates that a one-size-fits all approach is suboptimal. Regulators could consider ETFs' AUM, number of holdings and liquidity risk when developing forecasting models to enhance accuracy. Additionally, regulators could differentiate forecasting models based on liquidity and market risk profiles for the unbiasedness of forecasts.

The results show that ETFs with higher liquidity risk produce less accurate but more unbiased forecasts. This trade-off highlights the importance of the relationships between ETF characteristics and forecasting performance and the need for a nuanced regulatory framework that can adapt to varying ETF profiles.

For ETF managers, the relationships of the size and risk characteristics with forecasting performance hold significant implications for their asset allocation process. The results of this study can help to optimise asset allocation to produce unbiased and accurate forecasts. ETF managers might develop strategies to maintain a balanced and diversified portfolio without generating inaccurate forecasts. Furthermore, the ETF managers could choose to invest a high percentage of the fund's value in a few low market risk assets to generate more unbiased forecasts. Nevertheless, by allocating more money in a few assets, the diversification of the ETF decreases, which results in less accurate forecasts.

4.2.5 Control variables and model robustness

The analysis includes the ETF's asset allocation to small cap firms, large cap firms and cash and derivatives to control for these characteristics and enhance model fit. The results show a significant positive relationship between the proportions invested in large cap firms (6.604***) and cash and derivatives (90.854*) with forecast biases. Asset allocations to large cap firms and cash and derivatives typically indicate safer, less volatile investments. This finding contributes to the contradiction between volatility and susceptibility to macroeconomic variables (Section 4.2.2), as the results suggest that large cap ETFs produce more biased forecasts. However, the result could also indicate that large cap ETFs are volatile as well. Hence, this topic needs further research, as detailed in Section 5.2. Additionally, the constant term in the accuracy model is significant (3.438*), indicating a baseline level of forecast accuracy that is not explained by the included ETF characteristics and control variables. This suggests that the accuracy model does not capture enough important ETF characteristics. Lastly, the R-squared values indicate that the models explain a moderate portion of the variance in the forecast performance variables, leaving a significant amount of variable unexplained. Incorporating additional ETF characteristics into the regression models could improve model fit and uncover new key characteristics driving forecast unbiasedness, accuracy or efficiency.

4.2.6 Hypothesis testing

This section tests the hypothesis formulated in Section 3.1 by evaluating each hypothesis separately for forecast unbiasedness, accuracy and efficiency to ensure clarity.

Hypothesis 2: *Large ETFs demonstrate significantly more unbiased, accurate and efficient forecasts compared to small ETFs.*

Forecast accuracy. The results partially support and contradict Hypothesis 2. A higher number of holdings significantly improves forecast accuracy, suggesting acceptance of Hypothesis 2. However, this relationship becomes negative beyond a certain threshold, which indicates that larger size significantly impairs forecast accuracy, leading to a partial rejection.

Forecast unbiasedness. While larger ETFs tend to produce more unbiased forecasts, the relationships are not statistically significant. Therefore, I cannot accept Hypothesis 2.

Forecast efficiency. I cannot accept nor reject Hypothesis 2. Although a higher number of holdings initially relates to more efficient forecasts, this effect reverses beyond a certain threshold. However, given the insignificance of these results, this study cannot accept nor reject Hypothesis 2.

Hypothesis 3: *Risky ETFs demonstrate significantly less unbiased, accurate and efficient forecasts compared to less risky ETFs.*

Forecast accuracy. The results partially support Hypothesis 3. Liquidity risk significantly reduces forecast accuracy, supporting the hypothesis. However, while market risk appears to enhance accuracy, this relationship is insignificant, which prevents to reject the hypothesis. Thus, I accept Hypothesis 3 with respect to liquidity risk and I cannot reject Hypothesis 3 in relation to market risk because of insignificance.

Forecast unbiasedness. The results partially support and contradict Hypothesis 3. Higher liquidity risk relates significantly to less biased forecasts, which contradicts the hypothesis. Higher market risk relates significantly to more biased forecasts, supporting the hypothesis. Thus, I reject the hypothesis concerning liquidity risk and accept it regarding market risk.

Forecast efficiency. I cannot accept nor reject Hypothesis 3. Higher liquidity (market) risk relates to more inefficient (efficient) forecasts, implying to accept (reject) Hypothesis 3. However, as the relationships are insignificant, I cannot accept nor reject Hypothesis 3.

Hypothesis 4: *Expensive ETFs demonstrate significantly less unbiased, accurate and efficient forecasts compared to less expensive ETFs.*

The results show that more expensive ETFs tend to produce less unbiased, accurate and efficient forecasts. However, the lack of statistical significance in these relationships means that I cannot accept Hypothesis 4.

Chapter 5. Conclusion, discussion and limitations

5.1 Conclusion

This study exploits how the ESMA methodology for predicting one-year future returns of European passive stock ETFs deviates from a more sophisticated LASSO BIC Machine Learning prediction model and which ETF characteristics relate to higher return predictability. The findings indicate that the ESMA model produces unbiased but inaccurate and inefficient forecasts. In contrast, the LASSO BIC model generates accurate but biased and inefficient forecasts. To statistically evaluate the differences between these two models, comparative analyses reveal that the LASSO BIC model provides significantly more accurate and less inefficient forecasts than the ESMA model. Conversely, the ESMA model generates significantly more unbiased forecasts than the LASSO BIC model.

In the next step, this study examines the impact of ETF characteristics on forecast performance. The results indicate that larger ETFs initially generate significantly more accurate forecasts, but this relationship turns significantly negative beyond a certain threshold. Additionally, the results show that the liquidity and market risk of ETFs significantly impacts forecasts. High-liquidity-risk ETFs generate significantly less accurate and biased forecasts, whereas high-market-risk ETFs produce significantly less unbiased forecasts. Lastly, this study does not find significant relationships between ETF costs and forecast performance.

For investors, the results suggest caution when interpreting ETF forecasts, as the ESMA model – currently mandated by regulators – produces inaccurate and inefficient forecasts. Investors might consider using the LASSO BIC model for forecasting. Additionally, they might factor in ETF size and risk characteristics in their decision-making processes for buying or selling ETFs, depending on their preference for less biased or more accurate forecasts.

For regulators, the findings highlight the need for a new regulatory forecasting model, including macroeconomic variables such as implied volatility and GDP. While the LASSO BIC model offers a potential alternative with significantly more accurate and less inefficient forecasts, it still produces biased results. Moreover, the findings suggest that a one-size-fits all forecasting methodology is not adequate and that regulators could develop forecasting models tailored to different ETFs, based on their size and risk characteristics.

For ETF managers, this study provides insights for optimising asset allocation strategies to improve forecast performance. ETF managers might focus on maintaining a diversified portfolio and balancing the accuracy trade-offs as the ETF grows. Additionally, they might consider the trade-offs between liquidity and market risk to produce more unbiased and accurate forecasts.

5.2 Discussion

In terms of forecast unbiasedness, this study enlarges the literature by rejecting the weak form of market efficiency and accepting the semi-strong form. This study connects the EMH findings to the inclusion of macroeconomic variables in the forecast models. The results suggest that while the implied volatility and GDP are valuable for accurate predictions, they also introduce systematic errors. This finding refines the claims of prior studies regarding the predictive power of implied volatility and GDP by demonstrating that these variables enhance forecast accuracy, but not unbiasedness. The results show that biases inherent to macroeconomic variables are particularly strong for ETFs with low liquidity risk and high market risk. This indicates that macroeconomic variables could still be useful for unbiased predictions of low liquidity and high market risk ETFs. Therefore, further research could explore these relationships in more detail to evaluate which macroeconomic variables introduce biases and how these forecast biases relate to ETFs with different characteristics.

Regarding forecast accuracy, the findings enrich the literature by supporting the weak form of market efficiency and rejecting the semi-strong form. This result corroborates prior studies affirming the predictive power of the LASSO BIC model (Li & Chen, 2014; Zhang et al., 2019; Roy et al., 2015; Sermpinis et al., 2018) and links these findings to the rejection of the semi-strong form of market efficiency in terms of forecasting accuracy. Additionally, unlike prior studies that only validate the overall predictive power of implied volatility and GDP for stock returns through regression analysis (Bekaert & Wu, 2000; Li et al., 2005; Giot, 2005; Alexius & Sp, 2018; Somoye et al., 2019; Ogutu, 2011; Al-Tamimi et al., 2011; Österholm, 2016), this research refines the relationship by demonstrating their direct impact specially on forecast accuracy. This distinction adds substantial value to the literature by providing a more precise understanding of how these variables contribute to accurate forecasts. Moreover, this study refines the relationships of macroeconomic variables further, showing that improved accuracy is particularly evident in large ETFs with low liquidity risk, although the relationship of size with forecast accuracy diminishes beyond a certain threshold due to increased complexity and systemic risk. This finding introduces new questions that need to be addressed. Further research could explore the optimal size of ETFs that yield the most accurate forecasts.

Regarding efficiency, this study expands the literature by accepting the weak and semi-strong form of market efficiency. This finding might seem to contradict prior studies that highlight the predictive power of macroeconomic variables. However, such a conclusion regarding forecast efficiency is too severe. Given the moderate R-squared values and significant constant terms in the LASSO BIC regression models, it is more plausible that the LASSO BIC model may penalises additional macroeconomic variables to heavily in an effort to maintain model parsimony. Moreover, this study does not find significant relationships between the efficiency of the forecasts and ETF characteristics. Therefore, further research could investigate the efficiency of a forecasting model that penalises additional macroeconomic variables less heavily than the LASSO BIC model. In addition, future studies could deepen the research by investigating the relationship between the efficiency of new prediction models and ETF characteristics, as this study does not find significant relationships between forecast efficiency and ETF characteristics. By exploring these relationships, researchers can identify which ETFs generate the most efficient forecasts.

The proposed future research on forecast unbiasedness, accuracy and efficiency holds significant academical and practical relevance. Academically, these investigations will refine the understanding of how macroeconomic variables affect forecast unbiasedness and accuracy, offering a more nuanced view of their predictive power. Additionally, by exploring the efficiency of alternative forecasting models, this research could challenge the current understanding and application of the LASSO BIC model. Practically, the insights gained from these studies will enhance decision-making processes for investors, enable regulators to develop more robust regulatory frameworks and help ETF managers optimise their asset allocation strategies.

Lastly, further research could explore the relationship between ETF volatility and susceptibility to macroeconomic variables. Section 4.2.2 shows contradicting explanations for the unbiased (biased) forecasts of ETFs with high liquidity (market) risk. Examining the relationship between ETF volatility and susceptibility to macroeconomic variables could clarify the observed relationships.

5.3 Limitations

This study has several limitations that should be acknowledged. First, the dataset comprises 42 European passive stock ETFs, primarily sources from iShares. This selection may not fully represent the broader ETF market, limiting the generalisability of the findings. A larger and more diverse sample could yield more comprehensive results.

Second, the reliance on cross-sectional data for the ETF characteristics means that the analyses do not account for the temporal instability of these characteristics. Over time, ETF characteristics such as assets under management, number of holdings and management fees may fluctuate. This potentially influences the robustness of the findings.

Third, while this study incorporates several key macroeconomic variables, it may not capture all factors that influence ETF returns. For instance, the LASSO BIC model excludes variables such as technological innovations, investor sentiment and environmental factors, which could also impact the performance of the ETF forecasts.

Fourth, the findings of this study are based on a specific time frame (2008-2023), which may not capture longer-term trends in the financial markets. Future research could extend the analysis to different time periods to verify robustness of the results. Moreover, this study employs an expanding in-sample period for each new forecast. For instance, the in-sample coefficient estimates for the 2019 predictions are based on 120 observations, while those for 2020 are based on 132 observations, and so forth. The varying lengths of the in-sample periods could potentially influence the coefficient estimates. Future research could examine the effects of differing in-sample periods on the model's performance.

Fifth, this study focuses solely on passive stock ETFs. I anticipated that ETF costs would significantly impact forecast performance, as higher costs could indicate more active trading. Although the summary statistics show some cost variability, it is relatively limited. Including active ETFs could better capture the cost effect. Consequently, the insignificant cost relationships found in this study may be due to the exclusive focus on passive ETFs. Future research could investigate the relationship between costs and forecast performance, incorporating both passive and active ETFs.

References

Adrangi, B., Chatrath, A., & Sanvicente, A. Z. (2002). Inflation, output, and stock prices: Evidence from Brazil. *Journal of Applied Business Research (JABR)*, *18*(1).

Alam, M. M., & Uddin, G. (2009). Relationship between interest rate and stock price: empirical evidence from developed and developing countries. *International Journal of Business and Management (ISSN 1833-3850)*, *4*(3), 43-51.

Alexius, A., & Sp, D. (2018). Stock prices and GDP in the long run. *Journal of Applied Finance and Banking*, *8*(4), 107-126.

Al-Tamimi, H. A. H., Alwan, A. A., & Abdel Rahman, A. A. (2011). Factors affecting stock prices in the UAE financial markets. *Journal of transnational management*, *16*(1), 3-19.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The journal of finance*, *61*(1), 259-299.

Angrist, J. D., & Frandsen, B. (2022). Machine labor. *Journal of Labor Economics, 40(S1)*, S97-S140.

Arnott, R., Li, F., & Liu, X. (2016). Labor Conditions and Future Capital Market Performance. *The Journal of Portfolio Management*, *43*(1), 54-71.

Ascari, G., & Haber, T. (2022). Non-linearities, state-dependent prices and the transmission mechanism of monetary policy. *The Economic Journal, 132(641)*, 37-57.

Atanasov, V. (2021). Unemployment and aggregate stock returns. *Journal of Banking & Finance*, *129*, 106159.

Bekaert, G., & Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The review of financial studies*, *13*(1), 1-42.

Blitz, D., & Van Vliet, P. (2007). The volatility effect: Lower risk without lower return. *Journal of portfolio management*, 102-113.

Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S., & Şimşek, Ö. (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics*, *145*, 103773.

Bocchio, C., Crook, J., & Andreeva, G. (2023). The impact of macroeconomic scenarios on recurrent delinquency: A stress testing framework of multi-state models for mortgages. *International Journal of Forecasting, 39(4)*, 1655-1677.

Boyd, J. H., Hu, J., & Jagannathan, R. (2005). The stock market's reaction to unemployment news: Why bad news is usually good for stocks. *The Journal of Finance*, *60*(2), 649-672.

Boyd, J. H., Jagannathan, R., & Liu, Q. (2006). The stock market's reaction to unemployment news, stock-bond return correlations, and the state of the economy. *Journal of Investment Management*, *4*(4), 73.

Chang, S. H., & Huang, L. C. (2010). The nexus of finance and GDP growth in Japan: Do real interest rates matter?. *Japan and the World Economy*, *22*(4), 235-242.

Chen, S. S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance, 33(2)*, 211-223.

Choudhry, T. (2001). Inflation and rates of return on stocks: evidence from high inflation countries. *Journal of International Financial Markets, Institutions and Money*, *11*(1), 75-96.

Chowdhury, E. K., Khan, I. I., & Dhar, B. K. (2022). Catastrophic impact of Covid-19 on the global stock markets and economic activities. *Business and Society Review, 127(2)*, 437-460.

Croushore, D., & Marsten, K. (2016). Reassessing the relative power of the yield spread in forecasting recessions. *Journal of Applied Econometrics*, *31*(6), 1183-1191.

Dimitrova, D. (2005). The relationship between exchange rates and stock prices: Studied in a multivariate model. *Issues in political Economy*, *14*(1), 3-9.

Eldomiaty, T., Saeed, Y., Hammam, R., & AboulSoud, S. (2020). The associations between stock prices, inflation rates, interest rates are still persistent: Empirical evidence from stock duration model. *Journal of Economics, Finance and Administrative Science, 25(49)*, 149-161.

Elorhor, O. T. (2019). Unemployment and its effect on the growth of Nigeria economy. *Journal of Economics, Management and Trade*, *1*(9), 23-43.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, *25*(2), 383-417.

Faria, G., & Verona, F. (2020). The yield curve and the stock market: Mind the long run. *Journal of financial markets*, *50*, 100508.

Getmansky, M., Lo, A. W., & Makarov, I. (2004). An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics, 74(3)*, 529-609.

Gil-Bazo, J., & Ruiz-Verdú, P. A. B. L. O. (2009). The relation between price and performance in the mutual fund industry. *The Journal of Finance*, *64*(5), 2153-2183.

Giot, P. (2005). Relationships between implied volatility indices and stock index returns. *Journal of Portfolio Management*, *31*(3), 92-100.

Goda, T., Lysandrou, P., & Stewart, C. (2011). The Contribution of US Bond Demand to the US Bond Yield Conundrum of 2004 to 2007: *An Empirical Investigation*.

Gonzalo, J., & Taamouti, A. (2017). The reaction of stock market returns to unemployment. *Studies in Nonlinear Dynamics & Econometrics*, *21*(4), 20150078.

Greenwood, R. M., & Vissing-Jorgensen, A. (2018). The impact of pensions and insurance on global yield curves. *Harvard Business School Finance Working Paper*, (18-109), 19-59.

Grossman, Sanford J. and Joseph E. Stiglitz. 1980. “On the Impossibility of Informationally Efficient Markets.” *American Economic Review. 70:3*, pp. 393–408

Hakim, S., & Tursoy, T. (2021). The impact of oil price on stock market: empirical evidence from US.

Hjalmarsson, E. (2010). Predicting global stock returns. *Journal of Financial and Quantitative Analysis*, *45*(1), 49-80.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of financial economics, 6(2/3)*, 95-101.

Katechos, G. (2011). On the relationship between exchange rates and equity returns: A new approach. *Journal of International Financial Markets, Institutions and Money*, *21*(4), 550-559.

Kaul, G. (1990). Monetary Regimes and the relation between stock returns and inflationary expectations. *Journal of Financial and Quantitative Analysis, 15*, 307- 321.

Lawrenz, J., & Zorn, J. (2017). Predicting international stock returns with conditional price-to-fundamental ratios. *Journal of Empirical Finance*, *43*, 159-184.

Lei, Z. (2007). Dynamic Conditional Correlation Analysis between Stock Price and Interest Rate. *Journal of Business Economics*.

Leković, M. (2018). Investment diversification as a strategy for reducing investment risk.

Leković, M. (2018). Evidence for and Against the Validity of Efficient Market Hypothesis. *Economic Themes, 56*(3), 369-387

Li, J., & Chen, W. (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, *30*(4), 996-1015.

Li, Q., Yang, J., Hsiao, C., & Chang, Y. J. (2005). The relationship between stock returns and volatility in international stock markets. *Journal of Empirical Finance*, *12*(5), 650-665.

Li, Y., & Suominen, M. (2020). Equity returns, unemployment, and monetary policy. *Unemployment, and Monetary Policy (November 19, 2020)*.

Lioui, A., & Maio, P. (2014). Interest rate risk and the cross section of stock returns. *Journal of Financial and Quantitative Analysis*, *49*(2), 483-511.

Ma, F., Lu, X., Liu, J., & Huang, D. (2022). Macroeconomic attention and stock market return predictability. *Journal of International Financial Markets, Institutions and Money, 79*, 101603.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of economic perspectives, 17(1)*, 59-82.

Mishkin, F. S. (2019). *The Economics of Money, Banking and Financial Markets*. Pearson.

Nelson, C. R. (1976). Inflation and rates of return on common stocks. *The journal of Finance, 31(2)*, 471-483.

Ogutu, C. O. (2011). *The relationship between the GDP and share price movement at the Nairobi Stocks Exchange* (Doctoral dissertation, University of Nairobi, Kenya).

Österholm, P. (2016). The Long-Run Relationship between Stock Prices and GDP in Sweden. *Economic Notes: Review of Banking, Finance and Monetary Economics*, *45*(2), 283-297.

Pan, W. F. (2018). Does the stock market really cause unemployment? A cross-country analysis. *The North American Journal of Economics and Finance, 44*, 34-43.

Pastor, L., & Stambaugh, R. (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*.

Pimentel, R. C., & Choudhry, T. (2014). Stock returns under high inflation and interest rates: Evidence from the Brazilian market. *Emerging Markets Finance and Trade*, *50*(1), 71-92.

Rapach, D. E., Wohar, M. E., & Rangvid, J. (2005). Macro variables and international stock return predictability. *International journal of forecasting, 21(1)*, 137-166.

Reddy, D. L. (2012). Impact of inflation and GDP on stock market returns in India. *International journal of advanced research in management and social sciences, 1(6)*, 120-136.

Roy, S. S., Mittal, D., Basu, A., & Abraham, A. (2015). Stock market forecasting using LASSO linear regression model. In *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014* (pp. 371-381). Springer International Publishing.

Sadka, R. (2010). Liquidity risk and the cross-section of hedge-fund returns. *Journal of Financial Economics*, *98*(1), 54-71.

Sermpinis, G., Tsoukas, S., & Zhang, P. (2018). Modelling market implied ratings using LASSO variable selection techniques. *Journal of Empirical Finance*, *48*, 19-35.

Shah, S. Z. A., Shabbir, M. R., & Parveen, S. (2022). The impact of unemployment on economic growth in Pakistan: An empirical investigation. *iRASD Journal of Economics*, *4*(1), 78-87.

Shiblee, L. S. (2009). The impact of inflation, GDP, unemployment, and money supply on stock prices. *GDP, Unemployment, and Money Supply On Stock Prices (December 29, 2009)*.

Somoye, R. O., Akintoye, I. R., & Oseni, J. E. (2009). Determinants of equity price in the stock market. *International Research Journal of Finance and Economics*, (30).

Spyrou, I. S. (2001). Stock returns and inflation: evidence from an emerging market. *Applied Economics Letters, 8*, 447-450.

Statman, M. (1987). How many stocks make a diversified portfolio?. *Journal of financial and quantitative analysis*, *22*(3), 353-363.

Tursoy, T. (2019). The interaction between stock prices and interest rates in Turkey: empirical evidence from ARDL bounds test cointegration. *Financial Innovation*, *5*(1), 1-12.

Wojdylo, J. (2009). Best prediction of the stock market and economy-unemployment rate. *Subprime Blogger*.

Wu, Ying, 2000. "Stock prices and exchange rates in a VEC model-the case of Singapore in the 1990s", *Journal of Economics and Finance 24(3)*, pp. 260-274.

Xiong, J., Idzorek, T., Chen, P., & Ibbotson, R. (2009). Impact of Size and Flows on Performance for Funds of Hedge Funds. *The Journal of Portfolio Management*, 35, 118 - 130.

Yan, X. S. (2008). Liquidity, investment style, and the relation between fund size and fund performance. *Journal of Financial and Quantitative Analysis*, *43*(3), 741-767.

Zhang, Y., Ma, F., & Wang, Y. (2019). Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors?. *Journal of Empirical Finance*, *54*, 97-117.

Appendix A: Tables

Table 1 Descriptive Statistics.[a]

Variable	*N*	Mean	Std. dev.	Min	Median	Max	Skew	Kurt
Return variables								
Pooled ETF returns	192	0.48	4.54	-15.49	0.95	15.94	-0.33	1.65
Macroeconomic variables								
Unemployment rate	192	-0.06	1.25	-2.74	-0.30	6.02	1.40	4.15
Inflation rate	192	-3.84	80.92	-600.00	0.00	500.00	-1.44	25.22
30-year yield rate	192	-16.85	92.22	-608.58	-1.00	246.90	-4.00	20.49
1-year yield rate	192	-5.57	86.14	-541.48	-1.04	322.03	-2.84	19.32
Yield spread	192	6.45	89.50	-158.00	0.31	1197.45	12.25	160.44
GDP	192	0.01	2.41	-18.7	0.10	13.70	-1.76	25.09
Implied volatility	192	2.19	23.51	-40.30	-2.77	146.30	1.77	7.26
Interest rate	192	12.40	152.98	-118.63	-0.58	2087.79	13.01	173.43
ETF characteristics								
AUM (m)	42	1654	2994	13	495	14393	2.68	7.67
Number of holdings	42	114	151	11	42	606	1.95	3.11
10 largest stake holdings	42	56.27	27.75	7.82	58.28	98.21	-0.17	-1.38
3-year volatility	40	17.22	4.10	10.21	16.46	27.24	0.71	-0.18
Management fee	42	0.36	0.13	0.10	0.44	0.51	-0.90	-0.66
Small cap firms	38	0.05	0.17	0.00	0.00	1.03	5.35	30.59
Large cap firms	38	84.07	17.62	7.49	88.70	99.61	-2.64	9.07
Cash and derivatives	38	0.95	0.55	0.00	0.83	2.17	0.64	0.15

[a]This table presents summary statistics for variables used in the study. I use first-difference transformations for the macroeconomic variables because it eliminates unit roots. *Pooled ETF* is the average monthly returns of the 42 ETFs from 2008 to 2023, expressed in percentages. *Unemployment rate* indicates the proportion of individuals aged 15 to 74 who are unemployed relative to the total number of both employed and unemployed individuals in the Euro area, expressed in percentages. *Inflation rate* is the HICP overall index in the Euro area, measuring changes in the prices of goods and services, expressed in percentages. *30-year yield rate* and *1-year yield rate* are the continuously compounded government bond rates for issuers with a triple-A rating in the Euro area, with maturities of 30 years and 1 year, respectively, expressed in percentages. *Yield spread* is the difference between the 30-year and 1-year yield rates. *GDP* is the output of the industrial sector, used as proxy for GDP growth, expressed in percentages. *Implied volatility* is the European implied volatility rate measured by VSTOXX, expressed in percentages. *Interest rate* measures the rate at which Eurozone banks lend unsecured funds to other banks for a one-year period, measured by the 1-year Euribor and expressed in percentages. *AUM* measures the total market value of assets managed by each ETF in millions of euros. *Number of holdings* reflect the number of holdings in each ETF. *10 largest stake holdings* is the percentage proportion of the ten assets within each ETF that have the highest market value. *3-year volatility* captures the percentual variability in the ETF's returns over a three-year period. *Management fee* is the annual percentage fee charged by fund managers for managing the ETF. *Small cap firms* presents the percentual asset allocation to small cap firms in an ETF. *Large cap firms* is the percentual asset allocation to large cap firms in an ETF. *Cash and derivatives* is the percentual allocation to cash and derivatives in an ETF.

Table 2 In-sample coefficients for Pooled ETF using LASSO BIC model.[a]

LASSO BIC selected macroeconomic variables	Prediction 2019 (1)	Prediction 2020 (2)	Prediction 2021 (3)	Prediction 2022 (4)	Prediction 2023 (5)
Unemployment	-0.437	-0.351	-0.384		
	(0.287)	(0.263)	(0.249)		
GDP	0.507*	0.538**	0.467**	0.189	0.181
	(0.260)	(0.241)	(0.226)	(0.290)	(0.277)
Squared GDP				1.470	1.634
				(1.719)	(1.625)
Implied volatility	-0.151***	-0.149***	-0.150***	-0.169***	-0.161***
	(0.014)	(0.013)	(0.012)	(0.016)	(0.015)
Squared implied volatility				0.099**	0.096**
				(0.042)	(0.038)
GFC dummy	-0.012	-0.012	-0.013	-0.046**	-0.049***
	(0.012)	(0.019)	(0.018)	(0.019)	(0.019)
Covid dummy	0	0	0	-0.070	-0.075
	(omitted)	(omitted)	(omitted)	(0.084)	(0.078)
Constant	0.008**	0.007**	0.007***	0.004	0.005
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Observations	120	132	144	156	168
R2	0.545	0.544	0.555	0.564	0.561

*Denote statistical significance at the 10% level.
**Denote statistical significance at the 5% level.
***Denote statistical significance at the 1% level.

[a]This table shows the results of the in-sample OLS regressions presented in equations (2) to (6) in the text. The standard errors are presented in parentheses below the coefficients, where Column 1, 2 and 3 show normal standard errors and Colum 4 and 5 present Newey-West (HAC) standard errors to account for heteroskedasticity. Column 1 confirms the results of the 2019 in-sample coefficient estimations, presented in equation (2): $Pooled\ ETF_t = 0.008 - 0.437\ \Delta Unemp_t + 0.507\ \Delta GDP_t - 0.151\ \Delta IV_t - 0.012\ \text{GFCDummy}_t + \varepsilon_t$. Column 2 shows the results for 2020, presented in equation (3): $Pooled\ ETF_t = 0.007 - 0.351\ \Delta Unemp_t + 0.538\ \Delta GDP_t - 0.149\ \Delta IV_t - 0.012\ \text{GFCDummy}_t + \varepsilon_t$. Column 3 demonstrates the results for 2021, presented in equation (4): $Pooled\ ETF_t = 0.007 - 0.384\ \Delta Unemp_t + 0.467\ \Delta GDP_t - 0.150\ \Delta IV_t - 0.013\ \text{GFCDummy}_t + \varepsilon_t$. Column 4 shows the results for 2022, presented in equation (5): $Pooled\ ETF_t = 0.004 + 0.189\ \Delta GDP_t + 1.470\ \Delta {GDP_t}^2 - 0.169\ \Delta IV_t + 0.099\ \Delta {IV_t}^2 - 0.046\ \text{GFCDummy}_t - 0.070\ \text{CovidDummy}_t + \varepsilon_t$. Column 5 demonstrates the results for 2023, presented in equation (6): $Pooled\ ETF_t = 0.005 + 0.181\ \Delta GDP_t + 1.634\ \Delta {GDP_t}^2 - 0.161\ \Delta IV_t + 0.096\ \Delta {IV_t}^2 - 0.049\ \text{GFCDummy}_t - 0.075\ \text{CovidDummy}_t + \varepsilon_t$. In these regression models hold that $Pooled\ ETF_t$ is the percentage monthly return of the Pooled ETF in month *t*, $Unemp_t$ is the percentage monthly European unemployment rate in month *t*, ${GDP_t}^{(2)}$ is the (squared) percentage monthly European gross domestic product in month *t*, ${IV_t}^{(2)}$ is the (squared) percentage monthly European implied volatility in month *t*, GFCDummy_t is a dummy variable for the global financial crisis and CovidDummy_t is a dummy variable for the Covid-19 pandemic.

Table 3 Forecasting Performance Results for ESMA and LASSO BIC Models[a]

Statistics	ESMA Forecast Errors	LASSO BIC Forecast Errors
Summary statistics		
N	60	60
Mean	-2.02	5.04
Std. dev.	16.59	9.60
Min	-26.06	-14.30
Max	37.25	20.49
Unbiasedness		
t-stat	-0.45	1.82*
Accuracy		
MSPE	0.0275	0.0116
Variance BM	0.0250	0.0250
VarMinusMSPE	-0.0025	0.0134
Efficiency		
F-stat	14.59***	8.84***

*Denote statistical significance at the 10% level.
**Denote statistical significance at the 5% level.
***Denote statistical significance at the 1% level.

[a]This table presents the summary statistics of the Pooled ETF forecasting performance results and the unbiasedness, accuracy and efficiency results for both ESMA and LASSO BIC models. The evaluation covers the out-of-sample forecasting period from 2019 to 2023, using monthly forecasts for one-year holding periods. The ESMA and LASSO BIC forecasts are made using the steps outlined in Section 3.2 and 3.3, respectively. The forecast errors of both methodologies are performed using equation (7) in the text: $e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}$, where $e_{t+1|t}$ is the percentage forecast error in month t, y_{t+1} is the percentage one-year holding period realised return in month t and $\hat{y}_{t+1|t}$ is the ESMA or LASSO BIC percentage one-year holding period predicted return in month t. This table presents the summary statistics of the forecast errors under *Summary statistics*, where N counts the number of forecast errors, which equals 60 because I forecast on a monthly basis for the 2019-2023 out-of-sample years, Mean is the average forecast error in percentages, Std. dev. is the standard deviation of the forecast errors in percentages, Min is the most negative forecast error in percentages and Max is the most positive forecast error in percentages. The t-stat under *Unbiasedness* measures whether the average forecast error differs significantly from zero, including Newey-West (HAC) standard errors. The MSPE, Variance BM and VarMinusMSPE under *Accuracy* demonstrate the accuracy results of the forecasts. MSPE is calculated according to equation (8) in the text: $MSPE = \frac{1}{p}\sum_{t=T_1}^{T-1}(e_{t+1|t})^2$, in which p is the number of observed forecast errors, T is the end of the forecast window and $e_{t+1|t}$ is the percentage forecast error from equation (7) in month *t*. Variance BM is the variance of the Pooled ETF in the out-of-sample window. VarMinusMSPE is the MSPE subtracted from the Variance BM, where a positive (negative) VarMinusMSPE indicates an accurate (inaccurate) forecast. The F-stat under *Efficiency* measures whether all available information is incorporated into the forecast. I test whether β_0 = 0 and β_1 = 1 in the Mincer-Zarnowitz equation (equation 9): $y_{t+1} = \beta_0 + \beta_1\hat{y}_{t+1|t} + \eta_{t+1}$, in which y_t is the percentage one-year holding period realised return in month *t* and $\hat{y}_t$ is the ESMA or LASSO BIC percentage one-year holding period predicted return in month *t*. Significant efficiency results indicate inefficiency.

Table 4 Relative ESMA and LASSO BIC Model Performance.[a]

Statistics	Value
Relative Unbiasedness: t-stat	-1.84*
Relative Accuracy: Diebold-Mariano test-statistic	3.29***
Relative Efficiency: F-stat	1.65**

*Denote statistical significance at the 10% level.
**Denote statistical significance at the 5% level.
***Denote statistical significance at the 1% level.

[a]This table presents the significant differences between the ESMA and LASSO BIC models in terms of unbiasedness, accuracy and efficiency, as described in Table 3. The relative unbiasedness is examined by using a t-test with Newey -West (HAC) standard errors to determine if the average forecast errors of ESMA and LASSO BIC differ significantly. The relative accuracy is assessed by using the Diebold-Mariano test with Newey-West (HAC) standard errors, which involves a t-test on the differences between the squared prediction errors of ESMA and LASSO BIC, indicating the significance of their accuracy difference. The relative efficiency is evaluated by using an F-test on the ratio of the F-statistics from Table 3. This measures the significance of the difference in efficiency between ESMA and LASSO BIC.

Table 5 Relationships Between ETF Characteristic and Forecasting Performance.[a]

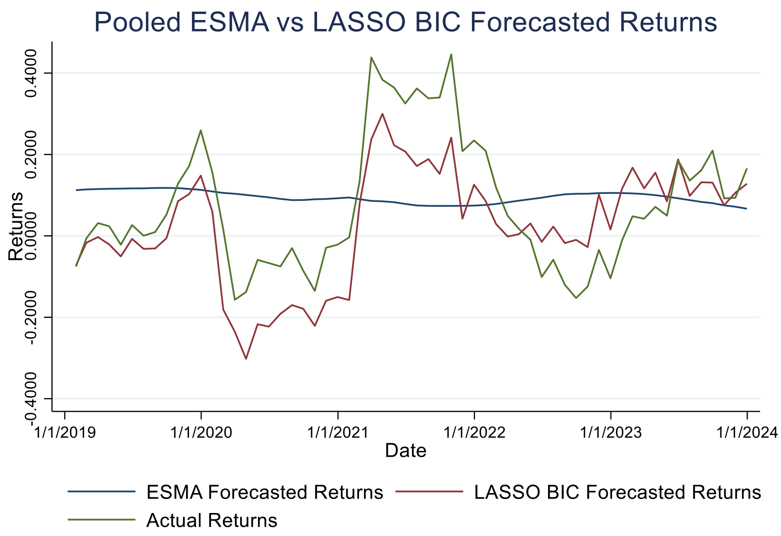
ETF characteristics	Unbiasedness (1)	Accuracy (2)	Efficiency (3)
AUM	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Log AUM	-0.055	-0.232*	504.630
	(0.618)	(0.132)	(418.411)
Number of holdings	-0.002	0.002*	-4.494
	(0.003)	(0.001)	(3.378)
Log number of holdings		-1.014*	2938.929
		(0.586)	(2078.932)
Ten largest stake holdings	-4.266**	-1.469*	4582.967
	(1.634)	(0.859)	(3055.195)
Three-year volatility	18.973*	2.960	-14349.570
	(9.409)	(2.853)	(9874.503)
Management fee	463.040	-57.805	90145.420
	(318.279)	(45.114)	(109558.700)
Small cap	166.706	30.496	-89635.180
	(169.831)	(19.376)	(65974.610)
Large cap	6.604***	0.832	-2920.508
	(2.085)	(0.579)	(1995.691)
Cash and derivatives	90.854*	4.019	-36608.540
	(52.902)	(11.311)	(33409.310)
Constant	-1.630	3.438*	-7950.355
	(5.321)	(1.764)	(6005.643)
Observations	36	36	36
R2	0.439	0.467	0.423

*Denote statistical significance at the 10% level.
**Denote statistical significance at the 5% level.
***Denote statistical significance at the 1% level.

[a]This table shows the results from the OLS regressions presented in equations (10), (11) and (12) in the text. The standard errors are presented in parentheses below the coefficients, where Column 1 shows normal standard errors and Column 2 and 3 show Newey-West (HAC) standard errors to account for heteroskedasticity. Column 1 presents the results of equation (10): $Unbiasedness_i = \beta_0 + \beta_1 \mathrm{AUM}_i + \beta_2 \mathrm{NOH}_i + \beta_3 \mathrm{TLSH}_i + \beta_4 \mathrm{TYV}_i + \beta_5 \mathrm{MF}_i + \beta_6 \mathrm{logAUM}_i + \beta_7 \mathrm{SC}_i + \beta_8 \mathrm{LC}_i + \beta_9 \mathrm{CD}_i + \varepsilon_i$, Column 2 demonstrates the results of equation (11): $Accuracy_i = \beta_0 + \beta_1 \mathrm{AUM}_i + \beta_2 \mathrm{NOH}_i + \beta_3 \mathrm{TLSH}_i + \beta_4 \mathrm{TYV}_i + \beta_5 \mathrm{MF}_i + \beta_6 \mathrm{logAUM}_i + \beta_7 \mathrm{logNOH}_i + \beta_8 \mathrm{SC}_i + \beta_9 \mathrm{LC}_i + \beta_{10} \mathrm{CD}_i + \varepsilon_i$ and Column 3 shows the results of equation (12): $Efficiency_i = \beta_0 + \beta_1 \mathrm{AUM}_i + \beta_2 \mathrm{NOH}_i + \beta_3 \mathrm{TLSH}_i + \beta_4 \mathrm{TYV}_i + \beta_5 \mathrm{MF}_i + \beta_6 \mathrm{logAUM}_i + \beta_7 \mathrm{logNOH}_i + \beta_8 \mathrm{SC}_i + \beta_9 \mathrm{LC}_i + \beta_{10} \mathrm{CD}_i + \varepsilon_i$. In these regressions hold that $Unbiasedness_i$ is the absolute t-statistic of ETF i, $Accuracy_i$ is the VarMinusMSPE value of ETF i, $Efficiency_i$ is the F-statistic of ETF i, AUM_i is the assets under management in millions of euros of ETF i, NOH_i is the number of holdings of ETF i, TLSH_i is the percentage proportion of the ten largest stake holdings in ETF i, TYV_i is the percentage three-year volatility of ETF i, MF_i is the percentage annual management fee of ETF i, logAUM_i is the logarithmic assets under management of ETF i, logNOH_i is the logarithmic number of holdings in ETF i, SC_i is the percentage asset allocation to small cap firms in ETF i, LC_i is the percentage asset allocation to large cap firms in ETF i and CD_i is the percentage asset allocation to cash and derivatives in ETF i.

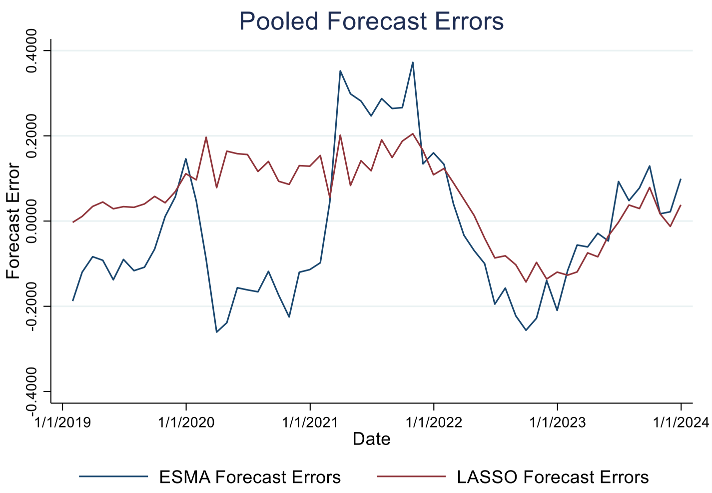
Appendix B: Figures

Figure 1 Pooled ESMA versus LASSO BIC Forecasted Returns.[a]



[a]This figure displays the Pooled ETF forecasts generated by the ESMA (blue line) and LASSO BIC (red line) methodologies, as outlined in sections 3.2 and 3.3, respectively, alongside the realised returns (green line). The x-axis represents the 60 monthly out-of-sample forecast dates from January 2019 to December 2023. The y-axis shows the Pooled ETF returns. This graphical representation highlights the comparative performance of both forecasting models against the actual returns over the out-of-sample period.

Figure 2 Pooled Forecast Errors.[a]



[a]This figure visualises the Pooled ETF forecast errors for the ESMA (blue line) and LASSO BIC (red line) methodologies. It is the difference between the forecasted returns and realised returns shown in Figure 1. The forecast errors are calculated using equation (7) in the text: $e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}$, where $e_{t+1|t}$ is the percentage forecast error in month t, y_{t+1} is the percentage one-year holding period realised return in month t and $\hat{y}_{t+1|t}$ is the ESMA or LASSO BIC percentage one-year holding period predicted return in month t. The x-axis represents the 60 monthly out-of-sample forecast dates from January 2019 to December 2023. The y-axis shows the Pooled ETF forecast errors. This graphical representation allows for a clear evaluation of the forecast unbiasedness, accuracy and efficiency of the two methodologies.