University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 20/2023 (427)

Ensembling ARIMAX Model in Algorithmic Investment Strategies on Commodities Market

Paweł Jakubowski
Robert Ślepaczuk
Franciszek Windorbski

Warsaw 2023

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

Ensembling ARIMAX Model in Algorithmic Investment Strategies on Commodities Market

Paweł Jakubowski[a], Robert Ślepaczuk[b], Franciszek Windorbski[c]

[a] *University of Warsaw, Faculty of Economic Sciences*
[b] *University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group, Department of Quantitative Finance*
[c] *University of Warsaw, Faculty of Management*
Corresponding authors: kc429229@student.wne.uw.edu.pl, rslepaczuk@wne.uw.edu.pl, franciszekwidnorbski@gmail.com

Abstract: This paper presents the results of investment strategies based on predictions from an ARIMA with exogenous variables (ARIMAX/ARIMAX-Garch) model, using the prices of selected commodities and companies from the DJIA index as explanatory variables. The explained variables are four Invesco ETF funds (DBE, DBA, DBP, DBB) corresponding to baskets of energy, agricultural, precious, and industrial metals. The models are optimized using the Walk-Forward technique, and the selection of exogenous variables is based on Granger causality tests. By analyzing the results, we conclude that ARIMAX/ARIMAX-Garch models are not useful tools for making buy or sell decisions for the selected commodity baskets. Out of the 80 estimated models, 44 outperform the Buy & Hold strategy, however, none achieved statistically significant results. Combining individual models into an investment portfolio reduced the risk without significantly reducing the profit, enabling us to consistently beat the benchmark. We also observe that using returns of commodities listed on stock exchanges is more effective than using stock returns. Sensitivity analysis shows instability in results with changes in the length of the training and testing windows. The highest annual return rate of 15.37% from 02.01.2008 to 01.12.2022 was characterized by an ARIMAX model with one commodity exogenous variable.

Keywords: ARIMA(X), GARCH, ARIMA(X)/GARCH, Algorithmic Investment Strategies, Granger Causality, Investment Performance Evaluation, Trading Systems, Forecasting Models

JEL codes: C4, C14, C45, C53, C58, G11, G13, G15

Note: The article was prepared within the framework of a non-competitive project of the Ministry of Education “School of Eagles” based on the agreement between the Ministry of Education and the University of Warsaw, contract no. MNISW/2020/142/DIR/KH, signed on 26 May 2020. Project co-financed from the European Social Fund under the Operational Programme Knowledge Education Development 2014-2020.

1 Introduction

The motivation behind this research was to deepen the knowledge of time series analysis, investigate market theories and practices that can be used when investing using financial instruments. The aim of the work was to build an investment model that creates efficient signals to buy or sell tested ETF's. We assumed that a properly programmed investment strategy should bring us benefits in the form of above-average risk adjusted returns. In the study, we focused on the correct creation of the database and the well-conducted parameterization of the model.

We analyzed time series of open-end investment funds (ETFs) investing in the most popular commodities traded on exchanges. We chose this group of assets due to its high liquidity and the possibility of diversifying the portfolio with instruments of different purposes. We were guided by the assumption that the price trends of individual commodities are strongly linked to different stages of business cycles. As a result, creating a solid model based on different groups of commodities will result in a stable return rate, regardless of market conditions.

The models we constructed are expansions of the base ARIMA model. Model extensions were performed by adding exogenous variables using the Granger causality test (ARIMAX) and by incorporating a volatility forecasting factor into the ARIMA model (ARIMA-GARCH/ARIMAX-GARCH), resulting in individual model outcomes. To create the final strategy, we combined signals from the individual models. Depending on the signal combination method, we obtained 9 and 4 final models, respectively. The base instruments were ETFs from the Invesco group covering four commodity baskets (DBE - energy basket, DBA - agricultural basket, DBP - precious metal basket, DBB - industrial metal basket). The exogenous variables are the individual series of commodities that make up the selected ETFs. In addition, we used quotes from companies listed on the popular Dow Jones Industrial Average stock index and well-known commodities that are not included in the portfolios of the selected ETFs. The results of the study were compared with each other based on the Information Ratio** (IR**). The model testing procedure was carried out using the Walk-Forward method.

In this research, we put forward the following research hypotheses:

($H1$) Based on predictions from individual ARIMA/ARIMA-Garch and ARIMAX/ARIMAX-Garch models, we are unable to generate signals for algorithmic investment strategies that outperform the market (i. e. characterized by a higher IR** than Buy&Hold).

($H2$) Based on predictions from ensemble ARIMA/ARIMA-Garch and ARIMAX/ARIMAX-Garch models, we are unable to generate signals for algorithmic investment strategies that outperform the market.

($H3$) Predictions from the ARIMA-Garch/ARIMAX/ARIMAX-Garch model are more effective in algorithmic investment strategies than predictions from the ARIMA model. Both exogenous data and volatility forecasting factor contribute to improving the informational value of the ARIMA model in the context of IR**.

($H4$) The number of exogenous variables affects the results of algorithmic investment strategies based on ARIMAX model predictions.

($H5$) The length of the training and testing window affects the results of investment strategies.

($H6$) The final strategy diversified through simultaneous investment in four ETFs (DBA, DBE, DBB, DBP) improves results in terms of risk-adjusted return (IR**) compared to individual models.

The work consists of seven chapters. After introduction in the first chapter, chapter two describes the selected literature in the field of quantitative finance. Chapter three presents the time series used, the models and their characteristics, the measures used in the analysis and a description of the study, outlining the step-by-step procedures performed by us. Chapter five presents the research results in the form of equity lines and performance metrics. This chapter also includes the results necessary for hypothesis verification. Sensitivity analysis - chapter five - examines the robustness of the model to changes in its underlying parameters. In chapter six, we approach hypothesis verification and formulate our observations.

2 Literature review

The first speculative-investment-related scientific works were published in the late 19th century (Dimson & Mussavian (1998)). The curiosity about the process of generating price changes and the possibility of using it to quickly multiply capital has increased interest for decades. Years of observation and research have led to the creation and improvement of market theories. Economists tried to explain whether the liquid stock market trade was called an efficient trade.

In the modern era, the pursuit of a market edge remains a common practice among thousands of quants worldwide. Trading strategies and techniques have evolved with advancements in technology, allowing for more sophisticated analysis and faster execution. Quantitative traders leverage mathematical models, statistical analysis, and algorithms to identify patterns, exploit market inefficiencies, and generate profits. The field of quantitative trading continues to evolve, driven by a constant quest for improved performance and profitability.

One of the most well-known works in the field that helps helps distinguish the states of market efficiency was published in 1970. Fama (1970) formulated three forms of market efficiency hypotheses related to predicting asset prices. The conclusions from his work are still used today when conducting research and testing investment strategies. The weak form hypothesis states that we are unable to predict future prices based on their past values. If the weak form is fulfilled, the use of technical analysis indicators and autoregressive models should not bring us any exceptional gains. The medium form claims that neither past quotations nor fundamental analysis were able to help us predict the future behavior of a selected instrument. The strong form states that even adding insider information (unavailable to stock exchange participants) to the medium form hypothesis would not be enough to overcome the market.

The market efficiency hypothesis has been tested on various assets (stocks and indices, commodities, currencies, and futures contracts), on different markets (emerging and developed), and using different methods (technical analysis, fundamental analysis, macro-econometric models, etc.). It can be concluded that works showing evidence that predicting prices is possible negate the weak and medium forms of the market efficiency hypothesis (Ślepaczuk, Sakowski, & Zakrzewski (2018)).

Although the literature review on time series forecasting is very rich, the main problem is that most of the articles testing even the weak form of market efficiency using algorithmic investment strategies do not base on proper testing methodology (C. W. Granger (1992)). As a result, the results of these studies cannot be treated as reliable and solid. Due to the fact that testing strategies is a complex process with many stages, each step requires attention and knowledge both theoretical and practical. In many popular scientific works on the subject of stock market price forecasting, incorrect nuances can be found, which can completely change the research results. Michańków, Sakowski, & Ślepaczuk (2022) indicate seven main mistakes most commonly made during the testing of algorithmic strategies:

- Single training and testing window causing results to be strongly dependent on the testing period (Bailey, Borwein, Lopez de Prado, & Zhu (2016), Wiecki, Campbell, Lent, & Stauth (2016), C. W. Granger (1992))
- Tests are carried out only on one underlying instrument, making the results strongly dependent on the distribution characteristics of that instrument (Castellano Gomez & Ślepaczuk (2021))
- Overoptimization of the model (Bailey, Borwein, Lopez de Prado, Salehipour, & Zhu (2016))
- Inappropriate model evaluation metrics or optimization criteria (Di Persio & Honchar (2016))

- Forward looking bias - looking into the future when creating buy or sell signals (Bailey, Borwein, Lopez de Prado, Salehipour, et al. (2016))
- Lack of sensitivity analysis weryfying the robustness of the final investment model (Di Persio & Honchar (2016))
- Data snooping bias - publication of only the best set of results (Bailey, Borwein, Lopez de Prado, Salehipour, et al. (2016), C. W. Granger (1992))

Most of the analyzed scientific works (Xiong, Li, Bao, Hu, & Zhang (2015), Patel, Shah, Thakkar, & Kotecha (2015), Kriechbaumer, Angus, Parsons, & Casado (2014), Zhang, Chu, & Shen (2021), Atsalakis & Valavanis (2009), Chen & Maher (2013), Wang, Liu, Diao, & Wu (2015), Shen, Jiang, & Zhang (2012), Kohzadi, Boyd, Kermanshahi, & Kaastra (1996), Chen & Maher (2013), Li, Ma, Wang, & Zhang (2015), Abda, Essaa, & Jassima (2021), Mondal, Shit, & Goswami (2014)) despite having interesting concepts for forecasting data prices, actually possess at least one of the above-mentioned errors. This also makes their results an unreliable source of information. In response to the shortcomings in the literature, there have been numerous works aimed at raising awareness of the importance of proper testing of strategies. Castellano Gomez & Ślepaczuk (2021) explains the key elements of backtesting, including the rolling window parameterization method and the use of appropriate model evaluation measures. Replacing Performance Metrics with Error Metrics, which is most commonly used, has been explained by the fact that although we use strictly predictive models for future return rates, we are only interested in generating a buy signal (1) or sell signal (-1). We therefore conclude that using measures such as RMSE may result in low prediction errors, but we overlook the crucial aspect of the impact of the generated signal on our portfolio result. Hypothetically, a model could predict the price with a small error in 90% of cases, but the remaining 10% of missed predictions could be responsible for the largest percentage changes in the underlying instrument.

In the generation of forecasts in algorithmic investment strategies, technical analysis indicators (Ślepaczuk (2004), Bui & Ślepaczuk (2022)), linear regression models (Vo & Ślepaczuk (2022)), machine learning models (Di Persio & Honchar (2016)), or a combination of these tools (Kijewski & Ślepaczuk (2020), Adcock & Gradojevic (2019)) are most commonly used. Kohzadi et al. (1996) believes that one-dimensional time series, such as ARIMA models (Box, Jenkins, Reinsel, & Ljung (2015)), are as accurate as more advanced linear or vector regression models. Wang et al. (2015) adds that although the applications of some nonlinear methods have achieved remarkable success, linear regression is still the most popular in forecasting. However, the results of linear model forecasts depend on whether the data generation process is linear and non-random. This is related to one of the views in financial economics, which assumes that past market prices are random and cannot be used to predict future prices. In this case, the ARIMA model is not a useful tool for estimating and forecasting prices. The above views prompted us to pose one hypothesis, which aims to verify the statement that ARIMA models are a useful forecasting tool.

On the other hand, Julián, Fernando, Maria-Dolores, & Simon (2003) investigated the profitability of non-linear trading rules based on nearest neighbour (NN) predictors. Applying this investment strategy to the New York Stock Exchange for the 1997-2002 period, their results suggest that, taking into account transaction costs, the NN-based trading rule is superior to both a risk-adjusted buy-and-hold strategy and a linear ARIMA-based strategy in terms of returns for all of the years studied, except for 2000 and 2001. Recent work illustrates how efficient method of moments and indirect inference can enhance the estimation of stationary ARMA models ("Estimating ARMA Models Efficiently"). By examining asymptotic and finite sample properties, this research introduces an improved methodology, allowing for better handling of both invertible and non-invertible ARMA models. Through Monte Carlo experiments, it offers a comparison to maximum likelihood estimators. This approach broadens the scope of quantitative finance, allowing for more precise financial predictions and robust investment models (Chumacero (2001)).

To avoid relying solely on past values of a single time series, exogenous variables that may influence the outcome of our forecast are used. In many works, the input data of the model (exogenous variables) primarily come from the same market that they relate to. Such isolation overlooks important information conveyed by other entities and makes the prediction result more susceptible to local perturbations. Globalization impacts interactions between financial markets worldwide, meaning that no financial market is isolated today. Economic data, political events, and all other foreign affairs can cause drastic fluctuations. Georgios & Theodore (2015) examined the causal relationship between crude oil and gold spot prices before and after the recent financial crisis. In the pre-crisis period, causality is linear and unidirectional, running from oil to gold. In the post-crisis period, a bidirectional nonlinear causality relationship emerges. They find that the causal linkage from gold to oil is time dependent and that the non-Granger causality null hypothesis rejection rate increased considerably in the post-financial crisis period. The probability of gold Granger causing oil in the short-run increases by more than 30 % during the recent financial and euro crisis. The mutual interaction implies a relationship between financial products and the possibility of using one or some of them to predict the movement of the rest (Shen et al. (2012), Abda et al. (2021)). Selection of additional variables should take place within a certain causal chain. C. Granger (1969) presents a linear way in which we can determine whether a variable contributed to generating parameters in the final model, meaning if adding the exogenous variable to the ARIMA model actually results in adding another linear parameter that affects the estimation in the ARIMAX model.

Further developments in quantitative forecasting models have been explored. For instance, John (2005) studied the capability of GARCH models, specifically the component GARCH models, in capturing the long-range dependence observed in volatility measures of time-series. This long-range dependence can be critical in understanding price fluctuations in various market scenarios. Through the use of sample autocorrelations, semiparametric and parametric estimations of the long-memory parameter, and the implementation of the parametric fractionally integrated GARCH (FIGARCH) model, these models are scrutinized. Notably, the findings suggest that a two-component GARCH model effectively encapsulates the autocorrelation function of volatility and aligns with long-memory based on semiparametric and parametric estimates. This suggests the potential of GARCH models to explain the long-range dependence in financial market volatility, adding another layer to the intricate process of predictive modeling in commodity markets.

The most effective way of verifying assumptions between different models is to compare their results in a single study. Alexander, Aaron, Svetlozar, Michael, & Frank (2013) examined the S&P 500 index log-returns on short intraday time scales with three different ARMA-GARCH models. In order to forecast market risk, they described the innovation process with tempered stable distributions which was compared to commonly used methods in financial modeling. Kijewski & Ślepaczuk (2020) compared six algorithmic strategies using both technical analysis indicators, classical linear models, and machine learning models (LSTM). Among the classical models, they used the ARIMA model to predict future S&P500 index prices. They pointed out two moments in which the model brought above-average returns. The first was the 2008 and 2009 crisis. The second was the Covid-related crisis in 2020. According to them, ARIMA performed well in strong downward trend conditions, while in other years the annualized return fluctuated slightly below 0. For comparison, the machine learning model throughout the period achieved results better than the ARIMA model by Information Ratio* = 0.27 to Information Ratio* = 0.13. The authors emphasized the importance of combining signals from all the strategies created. By generating a total buy or sell signal, a ensemble strategy was obtained with a result of Information Ratio* = 0.54, allowing for a doubling of returns with no increased risk. The results of this study may suggest that applying the methods described by Markowitz (1968) in our work can also improve IR* results.

In our work, we focused on commodity markets as research has shown that portfolios composed of commodity futures contracts can generate returns comparable to equities. As a result, the financial industry has developed products allowing institutions and smaller players to invest in commodities through index funds (ETFs) and liquid exchange-traded commodities and other structured products (Sanders, Irwin, & Merrin (2009)). By design, ETFs are meant to follow the movements of underlying prices and their trading resembles that of assets such as stocks (Poterba & Shoven (2002)). The use of ETF time series in place of commodity prices helps us navigate through the sensitive moment of expiration of successive futures contract series, described in detail by Ma, Mercer, & Walker (1992). Indifference during the choice of the series representing a particular instrument may result in a false set of results after testing. This stems from the fact that in order to be active in the market (trade daily) for a longer period of time (more than the duration of a single futures contract series), one must be aware of the effects of contango and backwardation, which the model treats as normal percentage changes in the instrument. Often, price gaps have a scale of several percent, which affects the estimation of parameters. The details related to the difference between futures contract series and ETF prices are presented in the further part of the work.

Creating accurate financial predictions for time series is an incredible challenge for researchers, primarily due to their non-linearity and chaos. With the development of quantitative finance and the intensification of the use of algorithmic investment strategies, accurate price predictions have become increasingly important, not only in the financial industry but also in academic circles (Zhang et al. (2021)). However, literature related to price forecasting usually focuses on simple and overly optimized single investment strategy procedures, not utilizing all the advantages of modern quantitative tools. Attention should be paid to several aspects, such as optimizing the procedure to create more robust and long-lasting investment strategies. Additionally, investors should focus on creating uncorrelated investment strategies that perform well under different market conditions. This approach should result in obtaining reasonable returns adjusted for risk (Castellano Gomez & Ślepaczuk (2021)).

3 Data and Methodology

3.1 Data description

The paper analyzes the time series of four commodity groups, namely energy commodities (DBE), agricultural commodities (DBA), precious metals (DBP), and base metals (DBB). Daily closing data were obtained from Yahoo Finance to carry out the research. Each group is represented by a separate ETF from the Invesco group, an open-end investment fund that invests in selected assets, designed to follow the movements of those asset prices. The reason for using ETFs instead of futures contract quotes is that we avoid rolling individual contract series at expiration, thereby avoiding price gap corrections. The created investment model is present in the market non-stop, so in this case, the use of a continuous time series is crucial to obtain results similar to those achieved by investing in the market in reality. For example, Figure 1 shows the difference between the quotes of two series. USO (ETF) shows the continuous time series quotes, while WTI (Futures) represents quotes without gap correction. Both instruments present the price of crude oil.

Figure 1: USO vs WTI

Note: The difference between a time series adjusted for gap prices that occurs when rolling a contract and a time series without gap correction shows how big an impact on the strategy results can have the use of inappropriate data set. The chart illustrates the price with correction (USO) and the time series without gap correction (WTI) over the last 14 years.

The data used in the study starts from 02.01.2008 and ends on 01.12.2022. It should be noted that the range of data is determined by the length of the ETFs time series. We tried to choose ETFs with the longest series in order to extend the testing window. All instruments used in the study along with descriptive statistics can be found in Table 1.

Table 1: Table with descriptive statistics of instruments used in the study from 02.01.2008 to 01.12.2022.

Instrument	Maximum	Minimum	Mean	StDev	Median	Skewness	Kurtosis	1st. Quartile	3rd. Quartile
NG=F	0.4648	-0.2595	0.0005	0.0348	-0.0003	1.0817	13.5126	-0.0185	0.0170
CL=F	**0.3766**	**-3.0597**	**-0.0007**	**0.0614**	**0.0009**	**-35.6670**	**1711.7953**	**-0.0123**	**0.0128**
BZ=F	0.3155	-0.2440	0.0003	0.0247	0.0002	0.1723	15.8934	-0.0099	0.0109
RB=F	0.2510	-0.3198	0.0004	0.0274	0.0010	-0.4289	17.0456	-0.0118	0.0135
ZW=F	0.2178	-0.1068	0.0001	0.0213	-0.0005	0.4540	4.6153	-0.0121	0.0119
ZC=F	0.1361	-0.2356	0.0002	0.0189	0.0004	-0.7452	11.2122	-0.0095	0.0099
ZS=F	0.2253	-0.2087	0.0002	0.0164	0.0008	-0.4475	20.2160	-0.0075	0.0084
SB=F	0.1395	-0.1163	0.0004	0.0208	-0.0005	0.1638	3.0398	-0.0113	0.0115
KC=F	0.1251	-0.0863	0.0003	0.0206	0.0000	0.3299	1.6654	-0.0124	0.0117
LE=F	0.0725	-0.1448	0.0002	0.0117	0.0004	-0.9670	13.0640	-0.0050	0.0059
CC=F	0.1218	-0.0924	0.0002	0.0183	0.0003	-0.0381	2.1717	-0.0104	0.0105
HE=F	0.2666	-0.2091	0.0004	0.0246	0.0004	0.2028	23.6505	-0.0082	0.0087
GF=F	0.1104	-0.0825	0.0002	0.0110	0.0002	0.5882	10.8933	-0.0046	0.0049
CT=F	0.0892	-0.2388	0.0002	0.0188	0.0001	-0.7733	9.5100	-0.0090	0.0098
GC=F	0.0903	-0.0935	0.0002	0.0113	0.0004	-0.0748	5.8830	-0.0050	0.0059
SI=F	0.1297	-0.1775	0.0003	0.0209	0.0005	-0.4507	5.7011	-0.0091	0.0103
HG=F	0.1249	-0.1104	0.0001	0.0169	0.0000	0.0724	4.0567	-0.0085	0.0088
PL=F	0.1600	-0.3856	0.0000	0.0170	0.0000	-2.8447	76.9187	-0.0078	0.0081
PA=F	0.2535	-0.2354	0.0006	0.0226	0.0003	-0.2083	14.1107	-0.0100	0.0117
HO=F	0.1315	-0.2193	0.0003	0.0229	0.0008	-0.4884	8.2850	-0.0104	0.0114
ZM=F	0.1087	-0.1855	0.0003	0.0191	0.0000	-1.2110	11.0312	-0.0087	0.0098
KE=F	0.0843	-0.0860	0.0001	0.0198	-0.0006	0.1925	1.4007	-0.0121	0.0113
USO	**0.1667**	**-0.2532**	**-0.0003**	**0.0243**	**0.0007**	**-0.6988**	**10.2298**	**-0.0118**	**0.0122**
AAPL	0.1390	-0.1792	0.0011	0.0199	0.0010	-0.1186	6.3903	-0.0080	0.0114
AMGN	0.1392	-0.0958	0.0007	0.0167	0.0004	0.6406	7.5118	-0.0076	0.0089
AXP	0.2188	-0.1759	0.0006	0.0242	0.0007	0.7737	14.8454	-0.0082	0.0097
BA	0.2432	-0.2385	0.0005	0.0238	0.0005	0.2706	15.5714	-0.0096	0.0107
CAT	0.1472	-0.1428	0.0005	0.0209	0.0004	0.0572	4.7582	-0.0092	0.0105
CRM	0.2604	-0.1845	0.0010	0.0255	0.0007	0.5997	9.3034	-0.0106	0.0129
CSCO	0.1595	-0.1621	0.0004	0.0185	0.0004	-0.1691	11.6047	-0.0072	0.0086
CVX	0.2274	-0.2212	0.0004	0.0191	0.0006	0.1698	21.3225	-0.0078	0.0086
DIS	0.1597	-0.1316	0.0005	0.0185	0.0004	0.4503	10.0219	-0.0076	0.0085
GS	0.2647	-0.1896	0.0005	0.0234	0.0003	0.7904	17.9979	-0.0096	0.0104
HD	0.1407	-0.1979	0.0008	0.0171	0.0007	-0.0467	11.7727	-0.0069	0.0086
IBM	0.1152	-0.1285	0.0002	0.0151	0.0003	-0.2246	8.2282	-0.0065	0.0074
INTC	0.1952	-0.1804	0.0003	0.0201	0.0004	-0.0312	9.5746	-0.0090	0.0098
JNJ	0.1223	-0.1004	0.0003	0.0114	0.0002	0.1691	11.8509	-0.0048	0.0058
JPM	0.2510	-0.2073	0.0006	0.0252	0.0001	0.9155	17.4665	-0.0089	0.0099
KO	0.1388	-0.0967	0.0003	0.0124	0.0005	0.0816	12.2072	-0.0051	0.0060
MCD	0.1813	-0.1588	0.0005	0.0129	0.0007	0.4180	23.0311	-0.0053	0.0064
MMM	0.1260	-0.1295	0.0002	0.0151	0.0006	-0.2496	7.6445	-0.0062	0.0074
MRK	0.1265	-0.1474	0.0004	0.0157	0.0002	-0.0906	10.4839	-0.0069	0.0079
MSFT	0.1860	-0.1474	0.0008	0.0182	0.0005	0.2849	9.5837	-0.0075	0.0094
NKE	0.1553	-0.1281	0.0007	0.0189	0.0006	0.4716	8.9826	-0.0080	0.0096
PFE	0.1086	-0.1032	0.0004	0.0151	0.0000	0.1972	5.8312	-0.0068	0.0074
PG	0.1201	-0.0874	0.0003	0.0121	0.0004	0.1197	10.5676	-0.0051	0.0060
RTX	0.1576	-0.1448	0.0004	0.0173	0.0004	0.2555	11.7654	-0.0069	0.0079
TRV	0.2556	-0.2080	0.0005	0.0181	0.0008	0.2813	28.7524	-0.0064	0.0078
UNH	0.3476	-0.1864	0.0009	0.0205	0.0008	1.2649	32.5010	-0.0079	0.0095
V	0.1500	-0.1364	0.0009	0.0190	0.0011	0.3025	8.6375	-0.0079	0.0094
VZ	0.1463	-0.0807	0.0001	0.0133	0.0002	0.4659	9.1308	-0.0064	0.0065
WBA	0.1664	-0.1434	0.0002	0.0186	0.0000	0.0509	7.6630	-0.0084	0.0093
WMT	0.1171	-0.1138	0.0004	0.0132	0.0004	0.2374	13.7674	-0.0056	0.0064
DBE	0.0927	-0.1405	0.0000	0.0181	0.0007	-0.3774	4.0647	-0.0091	0.0096
DBA	0.0667	-0.0861	-0.0001	0.0105	-0.0003	-0.2522	6.4042	-0.0053	0.0054
DBP	0.1262	-0.0978	0.0002	0.0127	0.0005	-0.1971	7.4336	-0.0057	0.0065
DBB	0.0851	-0.0837	0.0001	0.0146	0.0000	-0.0189	2.6607	-0.0078	0.0081
^DJI	0.1137	-0.1293	0.0004	0.0125	0.0006	-0.1637	14.7818	-0.0042	0.0056

Note: Descriptive statistics are based on the returns of the instruments used in the study. The CL=F series corresponds to one (usually the closest and most liquid) series of the underlying instrument contract. On the other hand, the USO ETF series (continuous) in the portfolio has only contracts for WTI with different expiration dates. Both of these series, although they describe the behavior of the same instrument, have significantly different distribution characteristics. This gives another reason why we used time series of instruments adjusted for price gaps and with exposure to more than one expiration date. The series such as USO, ^DJI were used only for this table and were not used in the forecast.

To examine the impact of an additional parameter on the model, we selected two sets of exogenous variables. The first set includes the prices of all time series in each basket of each ETF. Additionally, we included a set of popular commodities that did not belong to any ETF, but are often found in the portfolios of other funds. This decision was based on the belief that the prices of commodities are subject to similar cycles and therefore, existing trend signals are robust on a macroeconomic level. The use of these data in the ARIMAX model would result in better results than the model that does not take into account exogenous variables (ARIMA).

The second set of variables are the companies in the Dow Jones Industrial Average (DJIA) index, one of the oldest and most well-known stock indices in the world. The companies in the DJIA index are global conglomerates such as Chevron, Exxon Mobile, Microsoft, and JPMorgan Chase. The index is also sector diversified. We made this choice due to the high volume of trade and high market capitalization of the companies involved, and their popularity in the financial world and their significance in the global economy. We assumed that the quotations of the largest companies in the world may reflect expectations about the economy and future trends, which are cyclically linked to commodity quotations.

3.2 Methodology

Our goal was to create an investment strategy based on buy or sell signals from autoregressive models with moving average and exogenous variables such as commodity prices, stock quotes, or market volatility factors (ARIMAX/ARIMAX-Garch). Each model contained transactions on one of four underlying instruments, which were baskets of commodities corresponding to each ETF. Additionally, depending on the model specification, they included one to three lagged exogenous variables ranging from 1 to 5 periods..

ARIMA (p, d, q) combines AR and MA models. The AR part suggests that the series of interest is a linear regression of its own lags (p). The I (integrative) part corresponds to the number of differences (d) needed to transform the original time series into a stationary one. The MA part assumes that the regression error is a linear combination of errors at lags (q), whose values occurred simultaneously and at different periods in the past. The differentiation used in this work corresponds to the percentage changes between the closing prices of consecutive days. In general, the values of p, d, and q define the order of the ARIMA model (Castellano Gomez & Ślepaczuk (2021)).

GARCH (p, q) combines the autoregressive (AR) and moving average (MA) models. The AR part captures the conditional variance of the series by regressing it on its own past values (p). The MA part models the error term as a linear combination of past error terms (q) at different time periods. The GARCH model estimates the conditional volatility by considering both the lagged squared errors and the lagged conditional variances. The parameters p and q determine the order of the GARCH model. The differentiation used in this study involves computing the percentage changes between consecutive closing prices. The specific values of p and q are determined based on the model estimation and selection process. (Bollerslev, Chou, & Kroner (1992))

$$ARIMA(p,d,q) : \hat{y}_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i e_{t-i} + e_t \tag{1}$$

In our work, we used the ARIMA (1,1,1) model due to its reduction of overfitting (Castellano Gomez & Ślepaczuk (2021)).

$$ARIMA(1,1,1) : \hat{y}_t = \phi_1 y_{t-1} + \theta_1 e_{t-1} + e_t \tag{2}$$

$$GARCH(p,q) : \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i e_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{3}$$

$$GARCH(p,q) : g_t = \sqrt{\sigma_t} \gamma_t \tag{4}$$

$$GARCH(1,1) : \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{5}$$

Ensemble model used in the study:

$$ARIMA - GARCH : \hat{y}_t = \phi_1 y_{t-1} + \theta_1 e_{t-1} + g_t \tag{6}$$

$$ARIMAX(1,1,1)(X) : \hat{y_t} = \phi_1 y_{t-1} + \theta_1 e_{t-1} + g_t + \sum_{i=1}^{m} \psi_m x_{m,t-n} \tag{7}$$

where:

$\hat{y_t}$ - the predicted return rate on day t

$\hat{\sigma^2}$ - the predicted variance on day t

y_{t-1} - the given return rate on day $t-1$

e_{t-1} - the MA coefficient on day $t-1$ (ARIMA)

e_{t-1}^2 - the square of the error on the day $t-1$ (Garch)

σ_{t-1}^2 - the variance on the day $t-1$

g_t - the volatility factor from the Garch model

$x_{m,t-n}$ - the exogenous variable m on day $t-n$

γ_t - the white noise

$\phi_i, \theta_i, \alpha_i, \beta_j$ - the parameters from ARIMA and Garch process

ψ - the parameterof the exogenous variable

Each model used daily close price data from the previous session (t-1) along with lagged exogenous variables in the range of 1 to 5 periods (t-n). Based on this data, we created a forecast for the next day's closing price (t).

3.3 Exogenous variables

The selection of exogenous variables was a key aspect of our model. Variables along with their lags were chosen based on the Granger causality test. Granger causality refers to the relationship between the processes generating the data. A variable x is a cause of variable y if including past values of variable x in a forecasting model of variable y increases its predictive accuracy (C. W. Granger (1969)).

The use of two models is necessary to perform the causality test, the restricted model and the unrestricted model.

- The restricted model states that y is linearly dependent on its past values with a linear coefficient γ_i and a noise term that is dependent on the lag ϵ_t.

$$y_t = \gamma_0 + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t \tag{8}$$

- Unrestricted model - assumes that y is linearly dependent on past values of both x and y, determined by coefficients α_i and β_i and noise dependent on the lag u_t.

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \epsilon_t + \sum_{i=1}^{p} \beta_i x_{t-i} + u_t \tag{9}$$

The null hypothesis assumes that for each $i, \beta_i = 0$ meaning that adding the variable x did not improve the results of our model.

In order to verify whether a given variable is appropriate, we applied an ensemble F-test using the residual sum of squares (RSS). The RSS is a measure of the discrepancy between a given estimation model such as linear regression. The null hypothesis is rejected when the calculated F was greater than the critical level (in our case $p-value < 0.05$). This was equivalent to the fact that the given variable is the result of random arrangement of data insead of actual dependence.

- RSS:

$$RSS = \sum_{i=1}^{p} (y_i - f(x_i))^2 \tag{10}$$

where:

y_i - the i-th value of the variable to predict

x_i - the i-th value of the explanatory variable

$f(x_i)$ - the predicted value of y_i.

- F-test:

$$\frac{RSS_{restricted} - RRS_{unrestricted}/p}{RSS_{unresttricted}/(T-2p-1)} \sim F_{p,T-2[-1]} \tag{11}$$

where:

T - length of the time series

p - number of lags

3.4 Walk-Forward Optimization

In the model parameterization, we utilized the Walk-Forward (WF) optimization mentioned in Chapter 2. WF has several advantages compared to the traditional In Sample (IS) and Out Of Sample (OOS) single window optimization method. The Walk-Forward optimization uses rolled IS and OOS windows. In the IS window, the best model parameters (exogenous variables) were selected based on the Granger causality test and then used in trading on the OOS window. This approach reduced the chances of overfitting because all combinations were tested on different windows. The rolling window allowed the model to cycle its parameters, which was important due to changing market conditions and environments. Another positive aspect was that summing up all the individual OOS windows allowed us to create a sufficiently large testing window that presents results for different price trends of the tested instruments.

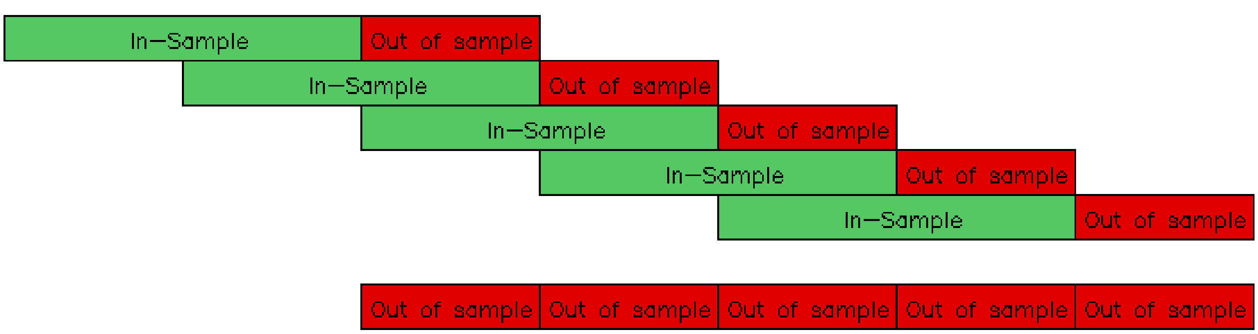

Figure 2: Walk-Forward Technique

Note: The above scheme presents the walk-forward procedure used to parameterize the model. At the end of optimization, we are able to create an equity line created from many test windows.

3.5 Terminology and metrics

In contrast to many works on price forecasting that assess the quality of forecasts by quantifying the difference between the predicted and actual prices (forecast error metrics), such as the root mean squared error (RMSE) or the mean absolute percentage error (MAPE), in our work we mainly relied on measures related to the changes in invested capital (performance metrics). Our goal was to build a model that would not only have a high annual rate of return (ARC) but also have low risk as measured by the standard deviation (aSD). By using these metrics, we could obtain the risk-adjusted rate of return (IR*). The interpretation of this measure was the same as the Sharp Ratio, assuming that the risk-free rate was 0. Another important measure for the evaluation of the strategy was the maximum percentage drawdown in capital (MD), allowing us to assess the worst-case scenario during testing. All the used measures were thoroughly described in Ryś & Ślepaczuk (2019). Below are the formulas for efficiency measures that were used in our study.

- Annualized Return Compounded (ARC):

$$ARC = \prod_{i=1}^{n}(r_i + 1)^{252/n} - 1 \tag{12}$$

where:

r_i - the daily percentage return rate, $r_i = \frac{r_i - r_{i-1}}{r_{i-1}}$

n - number of trading days, considering that there are 252 trading days in a year.

- Annualized Standard Deviation (aSD) :

$$aSD = \sqrt{252}\sqrt{*\frac{1}{n-1} * \sum_{i=1}^{n}(r_i - \overline{r})^2} \tag{13}$$

where:

$\overline{r}$ - average daily percentage return

- Information Ratio* (IR^*) - may be interpreted as a risk-weighted rate of return:

$$IR^* = \frac{ARC}{aSD} \tag{14}$$

- Information Ratio** (IR^{**})

$$IR^* * = \frac{ARC * ARC * sign(ARC)}{aSD * MD} \tag{15}$$

- Maximum Drawdown (MD) - maximum percentage loss of capital during the trading period.

$$MD = sup_{x,y \in \{[t_1,t_2]^2 : x \leq y\}} \frac{P_x - P_y}{P_x} \tag{16}$$

where:

P_t - the level of the equity line at time t

- Maximum Loss Duration (MLD) - the longest period of time (in years) required to reach the last maximum capital amount:

$$\forall_{m_j > m_i} MLD = max\frac{m_j - m_i}{S} \tag{17}$$

where:

m_j, m_i - the number of days indicating an uninterrupted local maximum of the equity line.

S - parameter equal to the number of trading periods per year for a given frequency.

3.6 Significance analysis of the model

In order to test the statistical significance of our models, we conducted a test of two means for both dependent samples used to verify (H1), H(2), H(3), and independent samples for (H6). We assumed that the return rate distributions of our strategies are close to the T-student distribution. The results of the test, including the t-statistic and p-value, are presented in the tables of performance metrics. The significance level was set at 0.05. Below are the equations describing the conducted test.

$$t_{related} = \frac{\overline{D}}{\frac{s_D}{\sqrt{n}}} \tag{18}$$

$$t_{independent} = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{19}$$

where:

$\overline{D}$ - the mean difference between related data

s_D - the standard deviation of the differences between related data

n - the sample size

$\overline{X_1}, \overline{X_2}$ - the means of samples 1 and 2

s_1, s_2 - standard deviations of samples 1 and 2

n_1, n_2 - sample sizes

The null hypothesis in our test stated that the means of the return rates for both distributions are equal, while the alternative hypothesis stated that the return rate of the analyzed strategy is greater than the benchmark (depending on the hypothesis, it took various forms). In simple terms, the hypothesis was rejected when:

$$p - value < 0.05 \text{ and } t - statistic > 0$$

3.7 Research discritpion

In this part, we presented each stage that led us to the achieved results.

- Formulation of research assumptions and hypotheses based on the literature
- selection of endogenous and exogenous variables, selection of model evaluation measures.
- Creation of a database based on daily closing prices obtained from Yahoo Finance.
- Formulation of walk-forward technique
- Formulation of causality tests in the sense of Granger's on the IS window - selection of the number of delays of variables in the model (from 1 to 5) and selection of the number of exogenous variables in the model (from 1 to 3). The decision on these parameters was made based on the F-test explained in section 3.3. The exogenous variable was used when the p-value was below the critical level (p-value < 0.05).
- Formulation of buy and sell signals

Buy signal: prediction (t+1) > current price (t)

Sell signal: prediction (t+1) < current price (t)

- Programming of the models using the above steps - conducting the Granger test for each IS window.
- Generation of buy or sell signals using the Walk-Forward procedure.
- Creation of an equity line (capital line) and tables with statistical measures (performance-metrics) to evaluate the model.
- Creation of ensemble models - combination of individual models into a model based on the number of variables (1, 2 or 3 exogenous variables) and combination of models based on their formula (ARIMA, ARIMA-GARCH, ARIMAX, ARIMAX-GARCH).
- Conducting sensitivity analysis - changing the basic parameters of the model. Changing the IS window (50, 100, 200) and changing the OOS window (10, 20, 40) to check the robustness of its results.
- Hypothesis verification based on the test of equality of two means (test for dependent or independent samples)
- Formulation of conclusions based on the results (main measure - IR**)

4 Empirical results

In this section of the paper, we present the results of strategies tested on four underlying instruments (DBA, DBB, DBE, DBP) from 02-01-2008 to 01-12-2022. For each underlying instrument, we created two panels with four charts each. A table with the model's results based on the performance metrics outlined in Chapter 3.5, and a table with signals generated by the model, is assigned to each panel. The first panel features three charts that present the results of the ARIMAX model with different sets of exogenous variables (commodity basket, DJIA basket, commodity basket + DJIA basket), and the fourth chart displays the equity line for the Buy & Hold benchmark, the ARIMA (1,1,1) model, and the ARIMA-GARCH (1,1,1) model. The second panel presents the results of the ARIMAX-GARCH model similarly for each of the baskets mentioned above.

Models that achieved negative results in terms of Information Ratio* are not interpretable due to algebraic equation reasons.

4.1 Agricultural commodities

Among the models investing in the commodity basket (DBA), the best solution was the DBA-ARIMA-GARCH (IR** = 0.038 and ARC = 5%), which was characterized by a strong upward trend since 2020. The second-best solution was ARIMAX-GARCH using two commodity variables (IR** = 0.035, ARC = 4.9%). The charts of these two models behaved similarly over the testing period. The vast majority of models achieved a positive annual return rate (19/20) and thus outperformed Buy&Hold. The DBA-ARIMA-GARCH had the highest number of days with a short position compared to ARIMAX models and the second-highest in the ARIMAX-GARCH set. Attention should be paid to the difference in short position opening signals for both groups of models, where the sell signal appeared much more frequently in the ARIMAX-GARCH models.

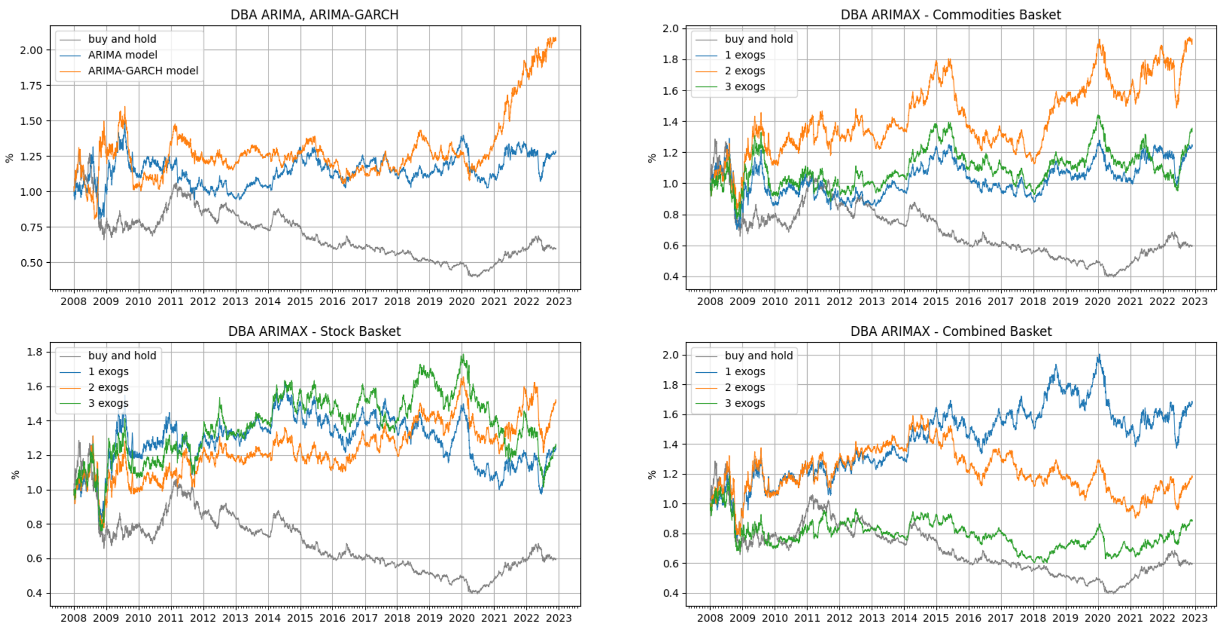

Figure 3: Results of ARIMAX models investing in agricultural commodity baskets.

Note: The chart panel shows ARIMAX model portfolios investing in agricultural commodities from 02.01.2008 to 01.12.2022. 20 out of 22 models achieved a positive annual return rate, outperforming the Buy&Hold strategy which had an average annual loss of 3.41%. Despite the different exogenous variables and different number of variables used in the model, all results were similar.

Table 2: Results of ARIMAX models investing in agricultural commodity baskets

Instruments	ARC	aSD	IR*	MD	MLD	IR**	t(H1)	p(H1)	t(H3)	p(H3)
DBA	-3.41%	17.03%	-0.200	48.88%	6.706	-0.014	-	-	-	-
DBA ARIMA	1.67%	17.03%	0.098	38.9%	1.825	0.004	0.84	0.4	-	-
DBA ARIMA_GARCH	**5.0%**	**17.03%**	**0.294**	**38.69%**	**2.246**	**0.038**	**1.29**	**0.2**	**0.61**	**0.54**
DBA ARIMAX Cmdt. B. 1	1.46%	17.03%	0.086	45.49%	2.254	0.003	0.83	0.41	-0.1	0.92
DBA ARIMAX Cmdt. B. 2	4.51%	17.03%	0.265	34.07%	2.956	0.035	1.28	0.2	0.87	0.38
DBA ARIMAX Cmdt. B. 3	2.06%	17.03%	0.121	34.07%	3.095	0.007	0.9	0.37	0.11	0.91
DBA ARIMAX Stck. B. 1	1.44%	17.03%	0.085	37.94%	2.345	0.003	0.81	0.42	-0.11	0.91
DBA ARIMAX Stck. B. 2	2.81%	17.03%	0.165	40.77%	2.254	0.011	1.04	0.3	0.39	0.69
DBA ARIMAX Stck. B. 3	1.54%	17.03%	0.090	40.12%	2.345	0.003	0.84	0.4	-0.04	0.97
DBA ARIMAX Cmb. B. 1	3.53%	17.03%	0.207	37.94%	1.397	0.019	1.15	0.25	0.85	0.4
DBA ARIMAX Cmb. B. 2	1.11%	17.03%	0.065	40.77%	2.210	0.002	0.74	0.46	-0.16	0.87
DBA ARIMAX Cmb. B. 3	-0.82%	17.03%	-0.048	42.99%	2.631	-0.001	0.43	0.67	-0.67	0.5

Note: Table of performance-metric results and the t-test for dependent groups, verifying the first and third hypotheses of ARIMAX models investing in a basket of commodities from 02.01.2008 to 01.12.2022. In the above table, we used abbreviations that will be used in subsequent tables. "Cmdt" is an abbreviation for "Commodity," "B" is an abbreviation for "Basket," and "Stck" is an abbreviation for "Stocks" which represents the basket of exogenous variables in the form of stocks. "Cmb" stands for "Combined," indicating a combined basket of stocks and commodities. The DBA-ARIMA-GARCH achieved the best result in terms of most metrics. The exceptions are Maximum Drawdown and Maximum Loss Duration, where DBA-ARIMAX with two and three commodity variables, DBA-ARIMAX with one commodity variable + DJIA, and DBA-ARIMA respectively, won.

Table 3: ARIMAX model signals for agricultural commodities

Instrument	long_signals	short_signals	neutral_signals
DBA ARIMA	1931	1828	0
DBA ARIMA_GARCH	**1715**	**2044**	**0**
DBA ARIMAX Commodities Basket 1	1941	1818	0
DBA ARIMAX Commodities Basket 2	1907	1852	0
DBA ARIMAX Commodities Basket 3	1900	1859	0
DBA ARIMAX Stock Basket 1	1909	1850	0
DBA ARIMAX Stock Basket 2	1894	1865	0
DBA ARIMAX Stock Basket 3	1910	1849	0
DBA ARIMAX Combined Basket 1	1909	1850	0
DBA ARIMAX Combined Basket 2	1881	1878	0
DBA ARIMAX Combined Basket 3	1877	1882	0

Note: The table shows the buy or sell signals generated by each model. It's worth noting that the best model had the most short positions. We can assume that having more days with a short position improved the result due to the downward trend of the underlying instrument.

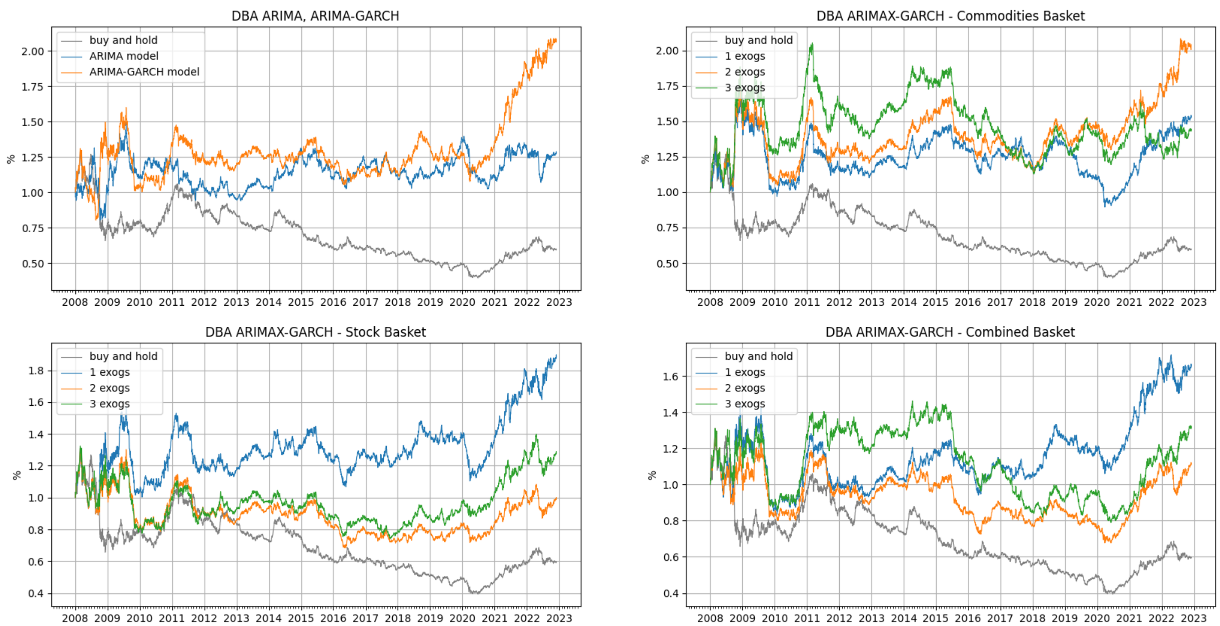

Figure 4: Results of ARIMAX-GARCH models investing in agricultural commodity baskets

Note: Chart of ARIMAX-GARCH models investing in the basket of agricultural commodities from 02.01.2008 to 01.12.2022. The models outperformed Buy&Hold for the majority of the time. The use of two exogenous variables was the best solution in most cases. For the ARIMAX set, one exogenous variable produced better results.

Table 4: Results of ARIMAX-GARCH models investing in agricultural commodity baskets

Instruments	ARC	aSD	IR*	MD	MLD	IR**	t(H1)	p(H1)	t(H3)	p(H3)
DBA	-3.41%	17.03%	-0.200	48.88%	6.706	-0.014	-	-	-	-
DBA ARIMA	1.67%	17.03%	0.098	38.9%	1.825	0.004	0.84	0.4	-	-
DBA ARIMA_GARCH	**5.0%**	**17.03%**	**0.294**	**38.69%**	**2.246**	**0.038**	**1.29**	**0.2**	**0.61**	**0.54**
DBA ARIMAX_GARCH Cmdt. B. 1	2.9%	17.03%	0.170	43.08%	2.310	0.011	0.97	0.33	0.23	0.82
DBA ARIMAX_GARCH Cmdt. B. 2	4.9%	17.03%	0.288	39.92%	2.881	0.035	1.27	0.21	0.61	0.54
DBA ARIMAX_GARCH Cmdt. B. 3	2.46%	17.03%	0.144	32.02%	4.198	0.011	0.9	0.37	0.15	0.88
DBA ARIMAX_GARCH Stck. B. 1	4.34%	17.03%	0.255	36.16%	2.321	0.031	1.19	0.24	0.5	0.62
DBA ARIMAX_GARCH Stck. B. 2	-0.04%	17.03%	-0.002	39.9%	3.115	0.000	0.53	0.6	-0.33	0.74
DBA ARIMAX_GARCH Stck. B. 3	1.68%	17.03%	0.098	37.26%	2.992	0.004	0.8	0.43	0	1
DBA ARIMAX_GARCH Cmb. B. 1	3.44%	17.03%	0.202	41.3%	2.226	0.017	1.05	0.29	0.33	0.74
DBA ARIMAX_GARCH Cmb. B. 2	0.72%	17.03%	0.043	35.7%	2.556	0.001	0.65	0.52	-0.18	0.86
DBA ARIMAX_GARCH Cmb. B. 3	1.85%	17.03%	0.108	35.59%	4.341	0.006	0.82	0.41	0.03	0.97

Note: Table with performance-metrics results and the t-test for dependent groups, verifying the first and third hypotheses of ARIMAX-GARCH models investing in a basket of agricultural commodities from 02.01.2008 to 01.12.2022. The group with commodity variables won in terms of sum of IR*, however, the best result was achieved by DBA-ARIMA-GARCH model..

Table 5: ARIMAX-GARCH model signals for agricultural commodities.

Instrument	long_signals	short_signals	neutral_signals
DBA ARIMA	1931	1828	0
DBA ARIMA_GARCH	**1715**	**2044**	**0**
DBA ARIMAX_GARCH Commodities Basket 1	1719	2040	0
DBA ARIMAX_GARCH Commodities Basket 2	1731	2028	0
DBA ARIMAX_GARCH Commodities Basket 3	1747	2012	0
DBA ARIMAX_GARCH Stock Basket 1	1721	2038	0
DBA ARIMAX_GARCH Stock Basket 2	1724	2035	0
DBA ARIMAX_GARCH Stock Basket 3	1738	2021	0
DBA ARIMAX_GARCH Combined Basket 1	1713	2046	0
DBA ARIMAX_GARCH Combined Basket 2	1718	2041	0
DBA ARIMAX_GARCH Combined Basket 3	1754	2005	0

Note: The table presents buy or sell signals generated by individual models. There is a significant advantage of short positions over long positions. The results of ARIMAX and ARIMAX-GARCH models do not differ significantly in terms of performance metrics, but there is a significant difference in the positions taken by the models.

4.2 Base metals

Definitely, the most interesting set of results falls on the models estimated on the DBB series. 8 out of 9 ARIMAX models showed solid results in terms of the risk-adjusted return rate. The best was the DBB-ARIMAX model with one commodity variable with IR** = 0.311. The DBB-ARIMAX with three exogenous variables also achieved a high IR** = 0.215. The DBB-ARIMAX with three commodity variables + DJIA, which is different from the other models, is characterized by a stable growth from the end of 2011 to 2021, its result is IR** = 0.166. Adding the GARCH model significantly impacted the worsening of the results, which were ultimately similar to Buy&Hold. This is evidenced by the ARIMA result with IR** = 0.117, which beat all the models using the GARCH model.

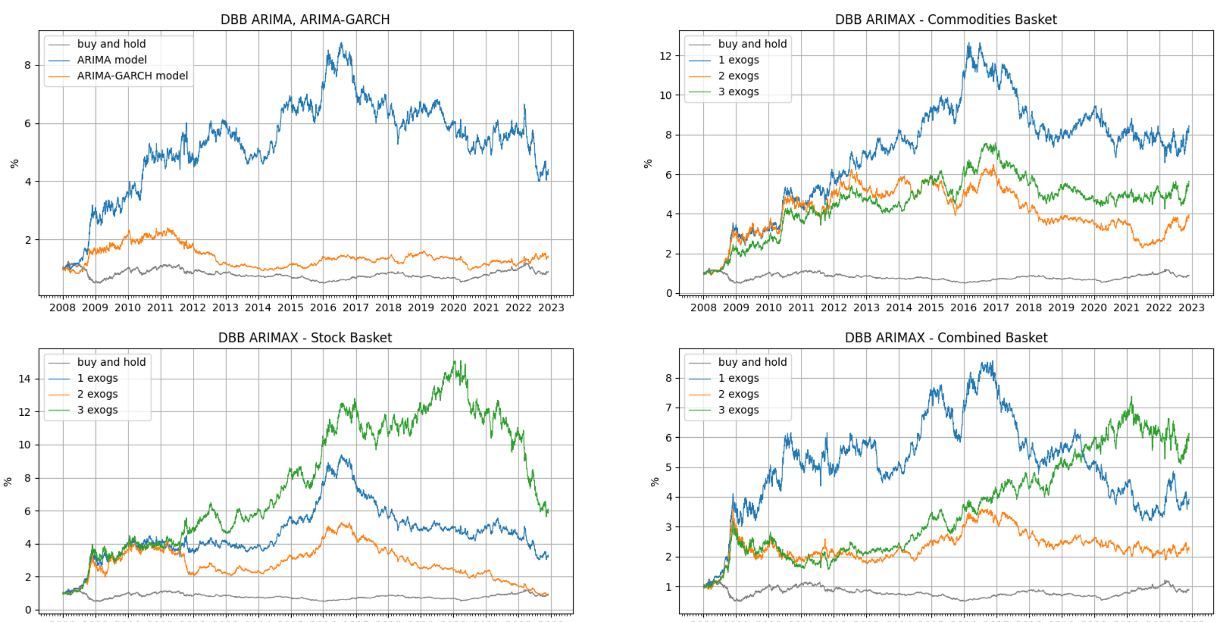

Figure 5: Results of ARIMAX models investing in base metals basket

Note: This panel shows the charts of ARIMAX models investing in a basket of industrial metals from 02.02.2008 to 01.12.2022. The best model from this set was DBB-ARIMAX with one commodity variable. Regardless of the trend in the underlying instrument, most models steadily generated profits until mid-2016. From 2016 onwards, DBB-ARIMAX with three commodity variables + DJIA turned out to be the best-performing model, which ranked second in terms of Information Ratio* (IR* = 0.552).

Table 6: Results of ARIMAX models investing in base metals basket.

Instruments	ARC	aSD	IR*	MD	MLD	IR**	t(H1)	p(H1)	t(H3)	p(H3)
DBB	-0.7%	23.42%	-0.030	60.75%	3.258	0.000	-	-	-	-
DBB ARIMA	10.44%	23.41%	0.446	39.96%	2.413	0.117	1.24	0.22	-	-
DBB ARIMA_GARCH	2.44%	23.42%	0.104	40.86%	3.679	0.006	0.36	0.72	-1.1	0.27
DBB ARIMAX Cmdt. B. 1	**15.37%**	**23.39%**	**0.657**	**32.46%**	**3.274**	**0.311**	**1.74**	**0.08**	**0.79**	**0.43**
DBB ARIMAX Cmdt. B. 2	9.63%	23.41%	0.411	40.14%	3.802	0.099	1.16	0.24	-0.14	0.89
DBB ARIMAX Cmdt. B. 3	12.3%	23.4%	0.526	30.03%	3.972	0.215	1.43	0.15	0.27	0.79
DBB ARIMAX Stck. B. 1	8.35%	23.41%	0.357	40.18%	4.409	0.074	1.01	0.31	-0.53	0.6
DBB ARIMAX Stck. B. 2	-0.29%	23.42%	-0.012	51.1%	6.365	0.000	0.05	0.96	-1.94	0.05
DBB ARIMAX Stck. B. 3	12.87%	23.4%	0.550	49.06%	2.706	0.144	1.48	0.14	0.4	0.69
DBB ARIMAX Cmb. B. 1	9.57%	23.41%	0.409	35.53%	2.813	0.110	1.13	0.26	-0.18	0.86
DBB ARIMAX Cmb. B. 2	5.78%	23.41%	0.247	47.35%	2.849	0.030	0.73	0.47	-0.72	0.47
DBB ARIMAX Cmb. B. 3	12.91%	23.4%	0.552	42.82%	1.778	0.166	1.48	0.14	0.35	0.72

Note: Table with performance metrics and the t-test for dependent groups, verifying the first and third hypotheses of ARIMAX models investing in base metals basket from 02.01.2008 to 01.12.2022. The highest result from this group was achieved by the DBB-ARIMAX model using one exogenous variable (IR* = 0.657). 10 out of 11 models have a solid risk-adjusted return. The only model with a negative result is DBB-ARIMAX with two DJIA variables.

Table 7: Signals of ARIMAX model for industrial metals

Instrument	long_signals	short_signals	neutral_signals
DBB ARIMA	1865	1894	0
DBB ARIMA_GARCH	1951	1808	0
DBB ARIMAX Commodities Basket 1	**1847**	**1912**	**0**
DBB ARIMAX Commodities Basket 2	1869	1890	0
DBB ARIMAX Commodities Basket 3	1814	1945	0
DBB ARIMAX Stock Basket 1	1871	1888	0
DBB ARIMAX Stock Basket 2	1863	1896	0
DBB ARIMAX Stock Basket 3	1818	1941	0
DBB ARIMAX Combined Basket 1	1857	1902	0
DBB ARIMAX Combined Basket 2	1833	1926	0
DBB ARIMAX Combined Basket 3	1812	1947	0

Note: The table presents buy and sell signals generated by individual models. In most cases, the number of short positions corresponds to achieving a higher IR* score.

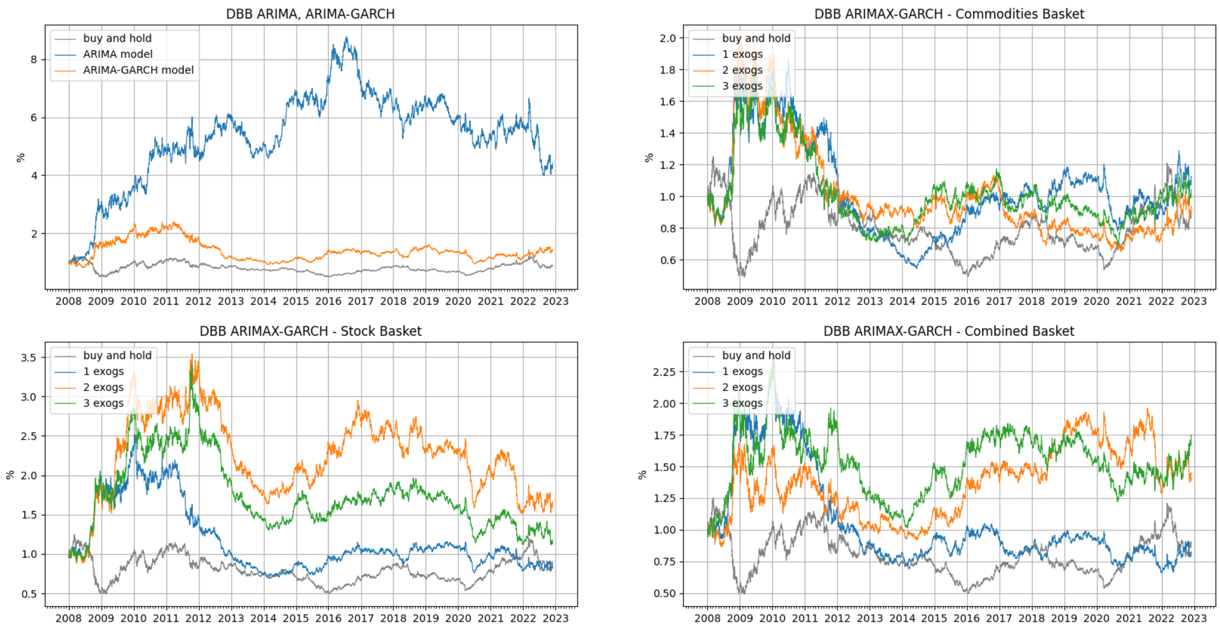

Figure 6: Results of ARIMAX-GARCH models investing in base metals basket

Note: A panel of charts of ARIMAX-GARCH models investing in the industrial metals basket from 02.01.2008 to 01.12.2022.

Table 8: Results of ARIMAX-GARCH models investing in base metals basket

Instruments	ARC	aSD	IR*	MD	MLD	IR**	t(H1)	p(H1)	t(H3)	p(H3)
DBB	-0.7%	23.42%	-0.030	60.75%	3.258	0.000	-	-	-	-
DBB ARIMA	**10.44%**	**23.41%**	**0.446**	**39.96%**	**2.413**	**0.117**	**1.24**	**0.22**	**-**	**-**
DBB ARIMA_GARCH	2.44%	23.42%	0.104	40.86%	3.679	0.006	0.36	0.72	-1.1	0.27
DBB ARIMAX_GARCH Cmdt. B. 1	0.79%	23.42%	0.034	44.63%	4.298	0.001	0.17	0.87	-1.28	0.2
DBB ARIMAX_GARCH Cmdt. B. 2	-0.44%	23.42%	-0.019	36.39%	5.766	0.000	0.03	0.98	-1.47	0.14
DBB ARIMAX_GARCH Cmdt. B. 3	0.26%	23.42%	0.011	40.98%	4.381	0.000	0.11	0.91	-1.34	0.18
DBB ARIMAX_GARCH Stck. B. 1	-1.05%	23.42%	-0.045	44.18%	4.841	-0.001	-0.04	0.97	-1.63	0.1
DBB ARIMAX_GARCH Stck. B. 2	3.48%	23.42%	0.149	36.14%	3.099	0.014	0.47	0.64	-0.91	0.36
DBB ARIMAX_GARCH Stck. B. 3	1.1%	23.42%	0.047	42.61%	3.099	0.001	0.2	0.84	-1.24	0.22
DBB ARIMAX_GARCH Cmb. B. 1	-1.29%	23.42%	-0.055	46.24%	4.536	-0.002	-0.07	0.95	-1.63	0.1
DBB ARIMAX_GARCH Cmb. B. 2	2.51%	23.42%	0.107	34.36%	5.893	0.008	0.36	0.72	-1.01	0.31
DBB ARIMAX_GARCH Cmb. B. 3	3.8%	23.42%	0.162	36.69%	2.845	0.017	0.48	0.63	-0.83	0.41

Note: The table shows the performance metrics and the t-test for dependent groups, verifying the first and third hypotheses of ARIMAX-GARCH models investing in the industrial metals basket from 02.01.2008 to 01.12.2022. DBB-ARIMAX-GARCH with three commodity variables + DJIA achieved the second-best result with IR* = 0.162 compared to DBB-ARIMA (IR* = 0.446).

Table 9: Signals of ARIMAX-GARCH model for industrial metals

Instrument	long_signals	short_signals	neutral_signals
DBB ARIMA	**1865**	**1894**	**0**
DBB ARIMA_GARCH	1951	1808	0
DBB ARIMAX_GARCH Commodities Basket 1	1857	1902	0
DBB ARIMAX_GARCH Commodities Basket 2	1975	1784	0
DBB ARIMAX_GARCH Commodities Basket 3	1867	1892	0
DBB ARIMAX_GARCH Stock Basket 1	1907	1852	0
DBB ARIMAX_GARCH Stock Basket 2	1937	1822	0
DBB ARIMAX_GARCH Stock Basket 3	1919	1840	0
DBB ARIMAX_GARCH Combined Basket 1	1924	1835	0
DBB ARIMAX_GARCH Combined Basket 2	1898	1861	0
DBB ARIMAX_GARCH Combined Basket 3	1860	1899	0

Note: The table presents buy and sell signals generated by individual models.

4.3 Energy commodities

This set was characterized by the clear advantage of the ARIMA-GARCH and ARIMAX-GARCH models. No ARIMAX model was able to beat the benchmark. 5 out of 11 models showed a double-digit loss over the year, however, the addition of the GARCH model improved the results. As a result, only one model had a negative ARC (DBE-ARIMAX-GARCH with three commodity variables + DJIA). The best result was achieved by the DBE-ARIMAX-GARCH with two DJIA variables (IR** = 0.072). Despite the downward trend of the underlying instrument, the significant advantage of long positions allowed for positive returns.

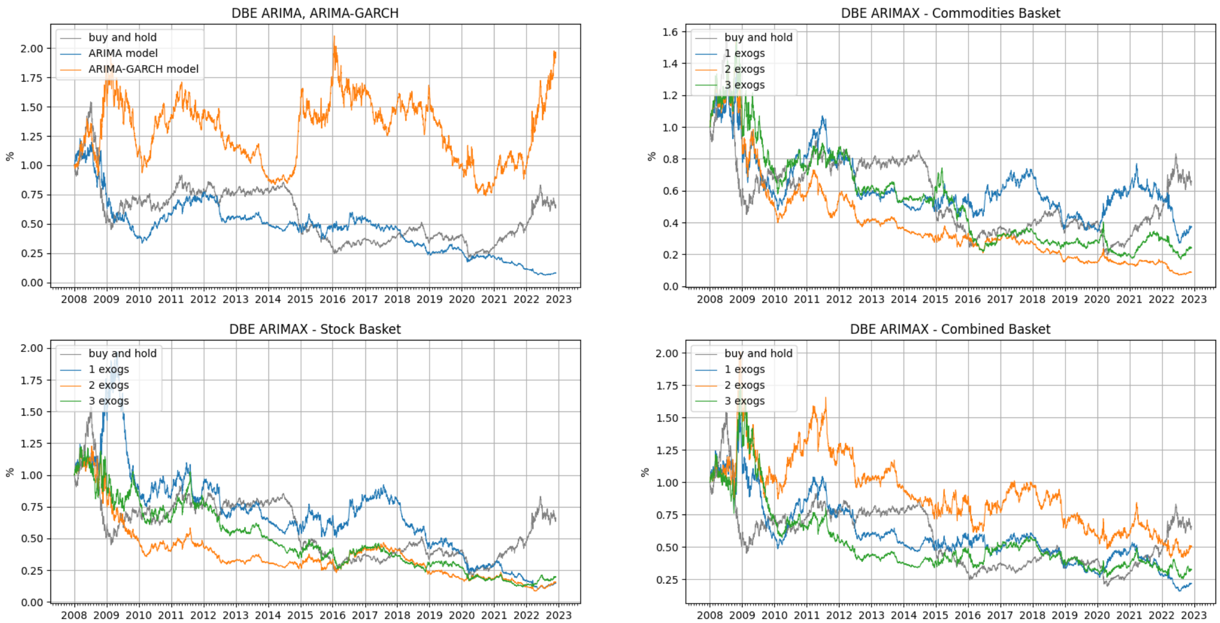

Figure 7: Results of ARIMAX models investing in a basket of energy commodities

Note: A panel with charts of ARIMAX-GARCH models investing in a basket of energy commodities from 02.01.2008 to 01.12.2022. A significant majority of the strategies performed worse than Buy & Hold.

Table 10: Results of ARIMAX models investing in a basket of energy commodities

Instruments	ARC	aSD	IR*	MD	MLD	IR**	t(H1)	p(H1)	t(H3)	p(H3)
DBE	-2.78%	28.69%	-0.097	70.79%	2.794	-0.004	-	-	-	-
DBE ARIMA	-15.6%	28.68%	-0.544	64.99%	6.310	-0.131	-1.33	0.18	-	-
DBE ARIMA_GARCH	**4.59%**	**28.69%**	**0.160**	**51.83%**	**3.504**	**0.014**	**0.67**	**0.51**	**2.49**	**0.01**
DBE ARIMAX Cmdt. B. 1	-6.38%	28.69%	-0.222	59.86%	2.357	-0.024	-0.36	0.72	1.61	0.11
DBE ARIMAX Cmdt. B. 2	-15.03%	28.68%	-0.524	64.24%	2.980	-0.123	-1.29	0.2	0.09	0.93
DBE ARIMAX Cmdt. B. 3	-9.06%	28.69%	-0.316	62.56%	2.365	-0.046	-0.65	0.52	0.91	0.36
DBE ARIMAX Stck. B. 1	-12.02%	28.69%	-0.419	61.99%	3.496	-0.081	-0.88	0.38	0.61	0.54
DBE ARIMAX Stck. B. 2	-11.75%	28.69%	-0.410	57.44%	3.623	-0.084	-0.92	0.36	0.67	0.5
DBE ARIMAX Stck. B. 3	-10.33%	28.69%	-0.360	48.45%	2.944	-0.077	-0.76	0.45	0.83	0.41
DBE ARIMAX Cmb. B. 1	-9.72%	28.69%	-0.339	64.07%	2.381	-0.051	-0.73	0.47	0.96	0.33
DBE ARIMAX Cmb. B. 2	-4.47%	28.69%	-0.156	56.16%	3.718	-0.012	-0.17	0.86	1.65	0.1
DBE ARIMAX Cmb. B. 3	-7.24%	28.69%	-0.252	62.95%	3.115	-0.029	-0.44	0.66	1.19	0.23

Note: Table with performance-metrics results and the t-test for dependent groups, verifying the first and third hypotheses of ARIMAX models investing in the basket of energy commodities from 02.01.2008 to 01.12.2022. The best and only positive result was achieved by the ARIMA-GARCH model (IR*=0.160). As many as 5 out of 11 models showed a double-digit negative return.

Table 11: The signals of the ARIMAX model for energy commodities

Instrument	long_signals	short_signals	neutral_signals
DBE ARIMA	1882	1877	0
DBE ARIMA_GARCH	**2159**	**1600**	**0**
DBE ARIMAX Commodities Basket 1	1863	1896	0
DBE ARIMAX Commodities Basket 2	1805	1954	0
DBE ARIMAX Commodities Basket 3	1851	1908	0
DBE ARIMAX Stock Basket 1	1793	1966	0
DBE ARIMAX Stock Basket 2	1861	1898	0
DBE ARIMAX Stock Basket 3	1873	1886	0
DBE ARIMAX Combined Basket 1	1885	1874	0
DBE ARIMAX Combined Basket 2	1866	1893	0
DBE ARIMAX Combined Basket 3	1810	1949	0

Note: Tables with performance-metrics and positions of models investing in baskets of energy commodities. The largest difference between long and short positions in favor of long positions occurred in the case of the model with the only positive rate of return.

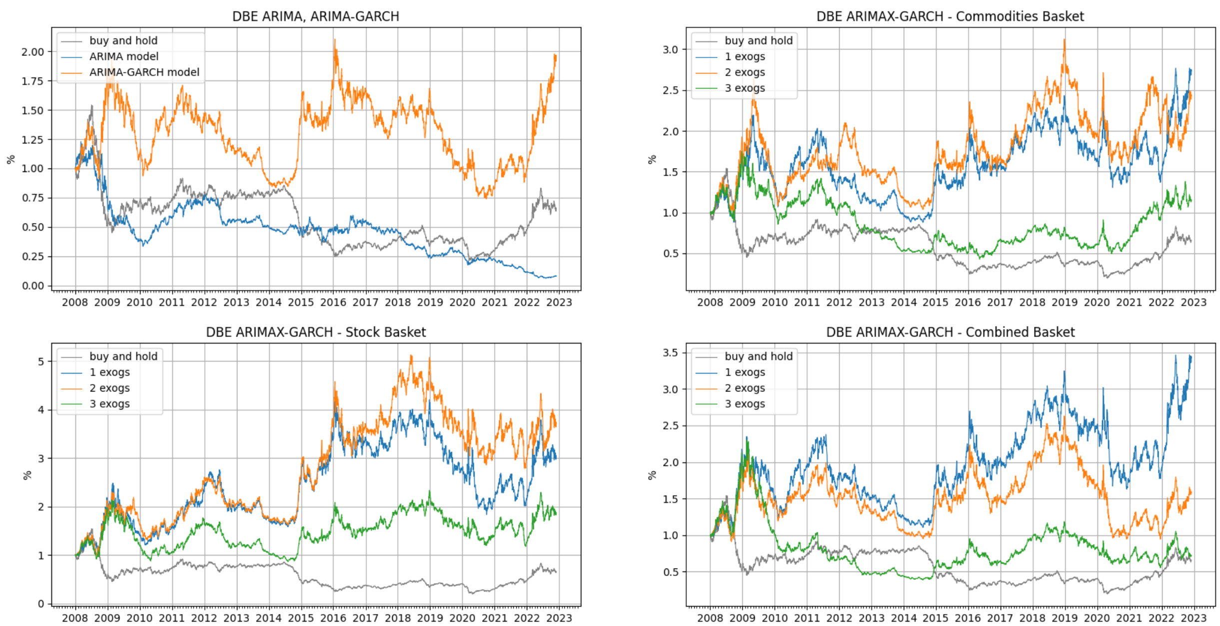

Figure 8: Results of ARIMAX-GARCH models investing in a basket of energy commodities

Note: Chart panel showing ARIMAX-GARCH models investing in a basket of energy commodities from 02.01.2008 to 01.12.2022. The main upward impulses of the models in this set came from sharp movements of the underlying instrument.

Table 12: Results of ARIMAX-GARCH models investing in a basket of energy commodities

Instruments	ARC	aSD	IR*	MD	MLD	IR**	t(H1)	p(H1)	t(H3)	p(H3)
DBE	-2.78%	28.69%	-0.097	70.79%	2.794	-0.004	-	-	-	-
DBE ARIMA	-15.6%	28.68%	-0.544	64.99%	6.310	-0.131	-1.33	0.18	-	-
DBE ARIMA_GARCH	4.59%	28.69%	0.160	51.83%	3.504	0.014	0.67	0.51	2.49	0.01
DBE ARIMAX_GARCH Cmdt. B. 1	6.98%	28.68%	0.243	51.3%	2.317	0.033	0.87	0.39	2.69	0.01
DBE ARIMAX_GARCH Cmdt. B. 2	6.14%	28.68%	0.214	59.04%	2.722	0.022	0.8	0.42	2.57	0.01
DBE ARIMAX_GARCH Cmdt. B. 3	0.94%	28.69%	0.033	49.49%	3.464	0.001	0.35	0.73	1.95	0.05
DBE ARIMAX_GARCH Stck. B. 1	7.72%	28.68%	0.269	50.69%	3.179	0.041	0.92	0.36	2.7	0.01
DBE ARIMAX_GARCH Stck. B. 2	**9.2%**	**28.68%**	**0.321**	**41.1%**	**3.933**	**0.072**	**1.05**	**0.29**	**2.93**	**0**
DBE ARIMAX_GARCH Stck. B. 3	4.26%	28.69%	0.148	53.38%	2.202	0.012	0.63	0.53	2.35	0.02
DBE ARIMAX_GARCH Cmb. B. 1	8.61%	28.68%	0.300	48.75%	3.302	0.053	1	0.32	2.87	0
DBE ARIMAX_GARCH Cmb. B. 2	3.14%	28.69%	0.109	53.58%	3.282	0.006	0.54	0.59	2.2	0.03
DBE ARIMAX_GARCH Cmb. B. 3	-2.16%	28.69%	-0.075	64.92%	3.194	-0.003	0.06	0.96	1.59	0.11

Note: Tables with the performance-metrics results and the t-test for dependent groups, verifying the first and third hypotheses of ARIMAX-GARCH models investing in an energy commodity basket from 02.01.2008 to 01.12.2022. The best model was DBE-

ARIMAX-GARCH with an IR* of 0.321, which clearly outperformed DBE-ARIMA (0.160). Only one model achieved a negative annual return.

Table 13: The signals of the ARIMAX-GARCH model for energy commodities.

Instrument	long_signals	short_signals	neutral_signals
DBE ARIMA	1882	1877	0
DBE ARIMA_GARCH	2159	1600	0
DBE ARIMAX_GARCH Commodities Basket 1	2123	1636	0
DBE ARIMAX_GARCH Commodities Basket 2	2060	1699	0
DBE ARIMAX_GARCH Commodities Basket 3	2058	1701	0
DBE ARIMAX_GARCH Stock Basket 1	2153	1606	0
DBE ARIMAX_GARCH Stock Basket 2	**2100**	**1659**	**0**
DBE ARIMAX_GARCH Stock Basket 3	2108	1651	0
DBE ARIMAX_GARCH Combined Basket 1	2157	1602	0
DBE ARIMAX_GARCH Combined Basket 2	2113	1646	0
DBE ARIMAX_GARCH Combined Basket 3	2052	1707	0

Note: The table presents buy and sell signals generated by individual models. All models except DBE-ARIMA had a significant long position advantage, even if the underlying instrument was characterized by a downward trend, allowing for a much higher return than buy and hold.

4.4 Precious metals

The only model among all investing in precious metals that outperforms the Buy & Hold strategy is the DBP-ARIMA with IR** = 0.012. All other models achieve a similar or worse result. The DBP-ARIMAX models with one and two commodity variables + DJIA, similarly to the DBP-ARIMA, have a positive return rate since mid-2013. The best result was characterized by the largest number of short positions. The worst result in terms of ARC = -7.33% was characterized by the DBP-ARIMAX-GARCH with one DJIA variable.

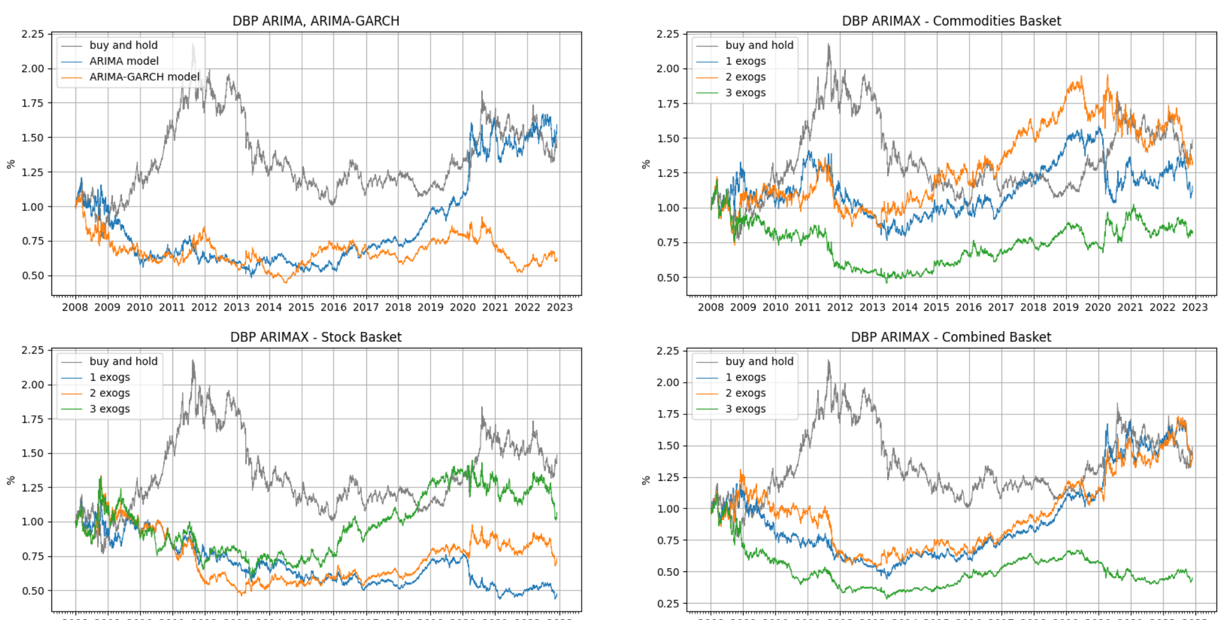

Figure 9: Results of ARIMAX models investing in a basket of precious metals

Note: Panel with charts of ARIMAX models investing in a basket of precious metals from 02.01.2008 to 01.12.2022. The models in this group achieved results similar to or worse than the Buy&Hold strategy. The DBP-ARIMAX models with two and three commodity variables + DJIA were characterized by a stable upward trend.

Table 14: Results of ARIMAX models investing in a basket of precious metals

Instruments	ARC	aSD	IR*	MD	MLD	IR**	t(H1)	p(H1)	t(H3)	p(H3)
DBP	2.69%	20.18%	0.133	37.35%	4.671	0.010	-	-	-	-
DBP ARIMA	**3.17%**	**20.18%**	**0.157**	**43.07%**	**2.655**	**0.012**	**0.06**	**0.95**	**-**	**-**
DBP ARIMA_GARCH	-3.11%	20.19%	-0.154	40.79%	3.405	-0.012	-0.78	0.43	-1.11	0.27
DBP ARIMAX Cmdt. B. 1	0.97%	20.19%	0.048	34.87%	3.218	0.001	-0.22	0.83	-0.46	0.65
DBP ARIMAX Cmdt. B. 2	1.94%	20.19%	0.096	40.55%	1.869	0.005	-0.1	0.92	-0.22	0.82
DBP ARIMAX Cmdt. B. 3	-1.25%	20.19%	-0.062	37.63%	3.421	-0.002	-0.53	0.6	-0.8	0.42
DBP ARIMAX Stck. B. 1	-4.86%	20.19%	-0.241	39.63%	5.448	-0.030	-0.98	0.33	-2.39	0.02
DBP ARIMAX Stck. B. 2	-2.07%	20.19%	-0.103	40.28%	5.262	-0.005	-0.64	0.52	-1.07	0.29
DBP ARIMAX Stck. B. 3	0.44%	20.19%	0.022	35.13%	2.675	0.000	-0.3	0.76	-0.52	0.61
DBP ARIMAX Cmb. B. 1	2.5%	20.18%	0.124	33.77%	4.948	0.009	-0.03	0.98	-0.18	0.86
DBP ARIMAX Cmb. B. 2	2.49%	20.18%	0.123	45.98%	4.393	0.007	-0.03	0.98	-0.12	0.9
DBP ARIMAX Cmb. B. 3	-5.21%	20.19%	-0.258	47.59%	3.302	-0.028	-1.08	0.28	-1.5	0.13

Note: The tables show the performance metrics and the t-test for dependent groups, verifying the first and third hypotheses of ARIMAX models investing in the basket of energy commodities from 02.01.2008 to 01.12.2022. The best result was achieved by DBP-ARIMA with IR* = 0.157, compared to Buy&Hold (0.133). DBP-ARIMA was the only model that outperformed the benchmark.

Table 15: The ARIMAX model signals for precious metals.

Instrument	long_signals	short_signals	neutral_signals
DBP ARIMA	**1906**	**1853**	**0**
DBP ARIMA_GARCH	2050	1709	0
DBP ARIMAX Commodities Basket 1	1921	1838	0
DBP ARIMAX Commodities Basket 2	1904	1855	0
DBP ARIMAX Commodities Basket 3	1895	1864	0
DBP ARIMAX Stock Basket 1	1851	1908	0
DBP ARIMAX Stock Basket 2	1875	1884	0
DBP ARIMAX Stock Basket 3	1897	1862	0
DBP ARIMAX Combined Basket 1	1871	1888	0
DBP ARIMAX Combined Basket 2	1881	1878	0
DBP ARIMAX Combined Basket 3	1884	1875	0

Note: The table presents buy and sell signals generated by individual models.

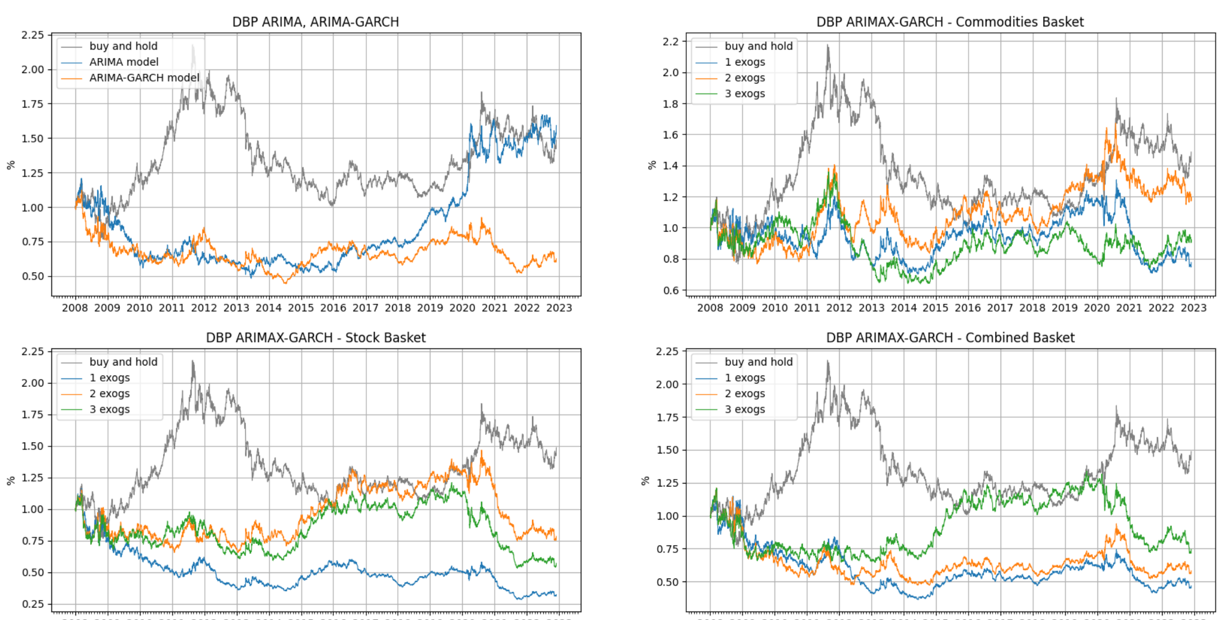

Figure 10: Results of ARIMAX-GARCH models investing in a basket of precious metals.

Note: Panel with charts of ARIMAX-GARCH models investing in an energy commodities basket from 02.01.2008 to 01.12.2022. Throughout the vast majority of the testing period, the Buy&Hold strategy performed the best.

Table 16: Results of ARIMAX-GARCH models investing in a basket of precious metals

Instruments	ARC	aSD	IR*	MD	MLD	IR**	t(H1)	p(H1)	t(H3)	p(H3)
DBP	2.69%	20.18%	0.133	37.35%	4.671	0.010	-	-	-	-
DBP ARIMA	**3.17%**	**20.18%**	**0.157**	**43.07%**	**2.655**	**0.012**	**0.06**	**0.95**	**-**	**-**
DBP ARIMA_GARCH	-3.11%	20.19%	-0.154	40.79%	3.405	-0.012	-0.78	0.43	-1.11	0.27
DBP ARIMAX_GARCH Cmdt. B. 1	-1.69%	20.19%	-0.084	42.88%	2.032	-0.003	-0.6	0.55	-0.79	0.43
DBP ARIMAX_GARCH Cmdt. B. 2	1.2%	20.19%	0.059	35.96%	2.032	0.002	-0.2	0.84	-0.31	0.76
DBP ARIMAX_GARCH Cmdt. B. 3	-0.55%	20.19%	-0.027	40.61%	3.230	0.000	-0.43	0.67	-0.61	0.54
DBP ARIMAX_GARCH Stck. B. 1	-7.33%	20.18%	-0.363	47.7%	3.452	-0.056	-1.39	0.16	-1.86	0.06
DBP ARIMAX_GARCH Stck. B. 2	-1.63%	20.19%	-0.081	45.46%	3.123	-0.003	-0.58	0.56	-0.83	0.41
DBP ARIMAX_GARCH Stck. B. 3	-3.72%	20.19%	-0.184	46.31%	3.254	-0.015	-0.86	0.39	-1.15	0.25
DBP ARIMAX_GARCH Cmb. B. 1	-4.98%	20.19%	-0.247	47.36%	3.052	-0.026	-1.05	0.29	-1.4	0.16
DBP ARIMAX_GARCH Cmb. B. 2	-3.57%	20.19%	-0.177	44.5%	3.123	-0.014	-0.85	0.4	-1.14	0.25
DBP ARIMAX_GARCH Cmb. B. 3	-1.97%	20.19%	-0.098	44.06%	3.329	-0.004	-0.62	0.54	-0.84	0.4

Note: The table shows the performance metrics and the t-test for dependent groups, verifying the first and third hypotheses of ARIMAX-GARCH models investing in a basket of energy goods from 02.01.2008 to 01.12.2022. The best result was achieved by the DBP-ARIMA model with an IR* of 0.157. Only one of the remaining models (DBP-ARIMAX-GARCH with two commodity variables) achieved a positive return (ARC = 1.2%).

Table 17: The ARIMAX-GARCH model signals for precious metals

Instrument	long_signals	short_signals	neutral_signals
DBP ARIMA	**1906**	**1853**	**0**
DBP ARIMA_GARCH	2050	1709	0
DBP ARIMAX_GARCH Commodities Basket 1	2045	1714	0
DBP ARIMAX_GARCH Commodities Basket 2	2026	1733	0
DBP ARIMAX_GARCH Commodities Basket 3	2010	1749	0
DBP ARIMAX_GARCH Stock Basket 1	2002	1757	0
DBP ARIMAX_GARCH Stock Basket 2	2028	1731	0
DBP ARIMAX_GARCH Stock Basket 3	2050	1709	0
DBP ARIMAX_GARCH Combined Basket 1	1997	1762	0
DBP ARIMAX_GARCH Combined Basket 2	2017	1742	0
DBP ARIMAX_GARCH Combined Basket 3	2025	1734	0

Note: The table shows the buy and sell signals generated by each model. The best performance was characterized by the largest number of short positions.

4.5 Classification of the model based on the number of variables

In Table 18, we placed a classification ranking to examine which quantity of variables in a particular model was the most effective. In each of the baskets, we analyzed IR** values for one, two, or three exogenous variables. The strategy with the highest IR** in each of the three baskets (commodity basket, DJIA basket, commodity basket + DJIA basket) achieved 3 points. The strategy with IR** in second place achieved 2 points, and on third place 1 point. The total points for each of the baskets for all underlying instruments reflect the ranking result.

Table 19 evaluates the performance of baskets of exogenous variables. The ranking of points awarded works similarly to Table 18, meaning that for each number of variables, we analyze the IR** of the model using one of the three baskets of variables. The classification conducted indicates the superiority of the Commodities Basket. As a result, the basket of stock variables achieved the worst results.

Table 18: Results of model classification based on the number of variables for ARIMAX models

No. of variables	DBE	DBA	DBP	DBB	Result
1	6	5	6	7	24
2	5	8	7	3	23
3	**7**	**5**	**5**	**8**	**25**

Note: Small differences indicate that the results are not stable and repeatable for each group, which has made it difficult to objectively evaluate and answer the question of which number of variables performed best.

Table 19: Results of model classification based on the basket of exogenous data for ARIMAX models

Basket type	DBE	DBA	DBP	DBB	Result
Commodities Basket	**6**	**7**	**6**	**9**	**28**
Stock Basket	4	6	5	3	18
Combined Basket	8	5	7	6	26

Note: A significant difference between the baskets indicated that the best exogenous variables for forecasting the movements of commodity baskets were the prices of individual commodities.

4.6 Ensemble models

In this chapter, the results of the equity line and performance metrics of the ensemble models are presented. The first method of combining models allows us to evaluate which amount of exogenous variables for all 4 ETFs achieved the best results and which group of exogenous variables achieved the best results in terms of the risk-adjusted return rate (IR**).

The second method was to combine all individual models into one representing the entire group (ARIMA, ARIMA-GARCH, ARIMAX, ARIMAX-GARCH). For example, the ARIMA ensemble consisted of 4 ARIMA models (4 instruments), and the ARIMAX ensemble consisted of 36 single models (4 instruments, 3 baskets, 3 options of variables). The combination of models was carried out analogously for all sets.

We applied the approach of combining the equity lines of different combinations of models because we assumed that diversification would reduce investment risk without significantly affecting return rates. In short, we wanted to increase the informational value of our strategies as measured by the Information Ratio**. Investment capital was divided equally between each individual strategy at the beginning of the investment period.

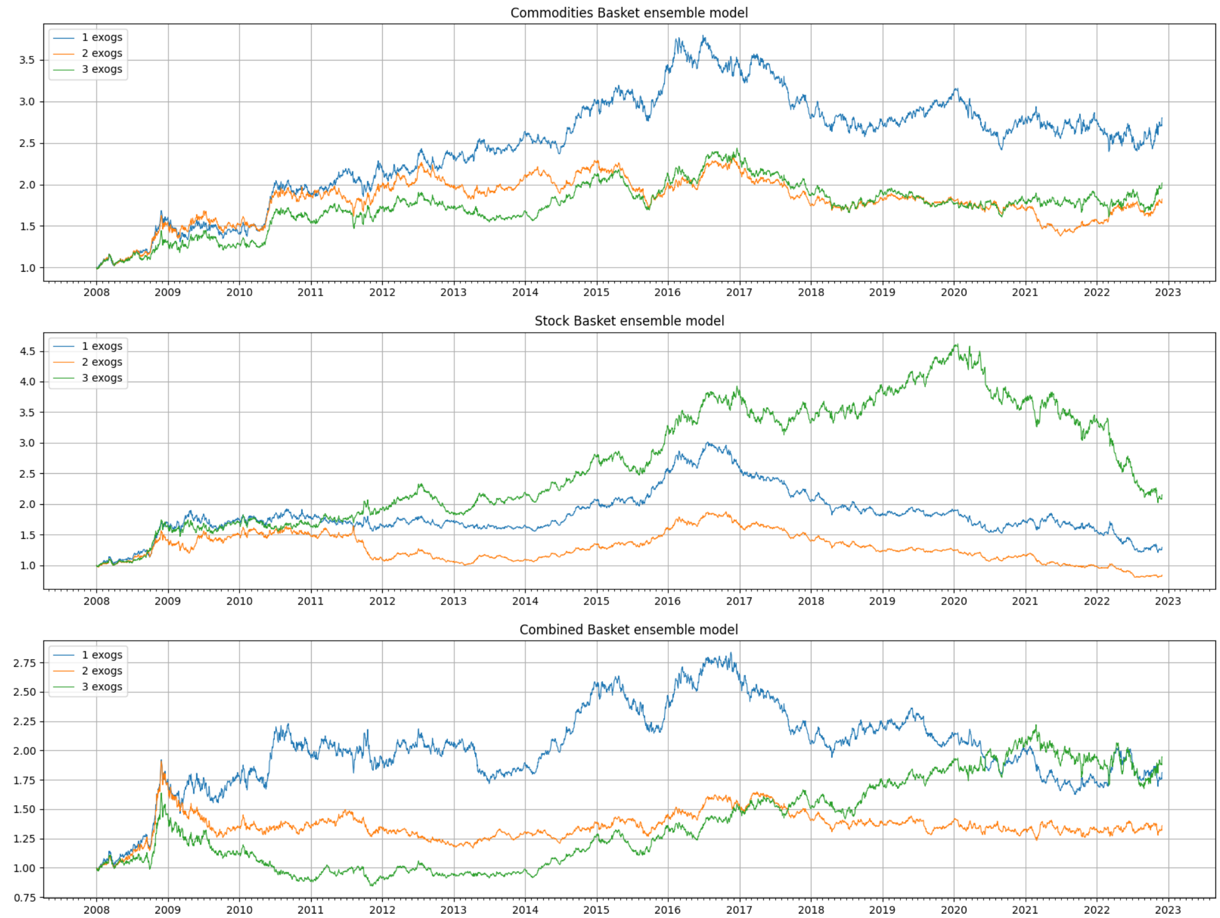

Figure 11: Analysis of ensemble models based on the number of variables

Note: The chart illustrates ensemble models based on the number of variables used to predict movements of each of the 4 ETFs. The chart shows that the best combination of exogenous variables until 2020 were three equity variables.

Table 20: Table with results of ensemble models based on the number of variables

Instrument	No. of models	ARC	aSD	IR*	MD	MLD	IR**
ARIMAX Commodities Basket 1	**4**	**7.15%**	**15.69%**	**0.456**	**25.13%**	**2.873**	**0.130**
ARIMAX Commodities Basket 2	4	4.11%	13.7%	0.300	24.85%	3.044	0.050
ARIMAX Commodities Basket 3	4	4.82%	14.91%	0.323	22.76%	2.147	0.068
ARIMAX Stock Basket 1	4	1.73%	14.75%	0.117	29.58%	4.738	0.007
ARIMAX Stock Basket 2	4	-1.17%	13.44%	-0.087	35.24%	6.107	-0.003
ARIMAX Stock Basket 3	4	5.27%	15.74%	0.335	41.52%	2.873	0.042
ARIMAX Combined Basket 1	4	4.07%	15.04%	0.270	24.75%	2.770	0.044
ARIMAX Combined Basket 2	4	2.08%	12.46%	0.167	34.11%	2.159	0.010
ARIMAX Combined Basket 3	4	4.56%	15.32%	0.298	34.08%	3.337	0.040

Note: The table shows the performance-metrics of ensemble models based on the number of variables used to forecast the movements of each of the 4 ETFs. The best result is achieved by the ARIMAX model with one commodity variable. The second-best result belongs to the ARIMAX model with three stock variables. It can be noted that, on average, commodity variables achieved better results than the other two groups.

For the first method of combining, both among models in the commodity group and the two others, the ARIMAX ensemble with one commodity exogenous variable achieved the best result (IR** = 0.130). The second result (IR** = 0.068) was given to the ARIMAX ensemble with three DJIA variables. Summing up the Information Ratio** results for the individual baskets, the best was the basket of commodity variables, then commodities + DJIA, and the worst was DJIA.

Table 21: Verification of the H(6) hypothesis

Ensemble	IR**	Individual	t(H6)	p(H6)	IR**
ARIMAX Commodities Basket 1	0.130	ARIMAX-DBB-Commodities Basket-1	-1.22	0.22	0.311
		ARIMAX-DBA-Commodities Basket-1	0.87	0.38	0.003
		ARIMAX-DBP-Commodities Basket-1	0.78	0.44	0.001
		ARIMAX-DBE-Commodities Basket-1	1.25	0.21	-0.024
ARIMAX Commodities Basket 2	0.050	ARIMAX-DBB-Commodities Basket-2	-0.99	0.32	0.099
		ARIMAX-DBA-Commodities Basket-2	-0.16	0.87	0.035
		ARIMAX-DBP-Commodities Basket-2	0.16	0.87	0.005
		ARIMAX-DBE-Commodities Basket-2	2.08	0.04	-0.123
ARIMAX Commodities Basket 3	0.068	ARIMAX-DBB-Commodities Basket-3	-1.19	0.24	0.215
		ARIMAX-DBA-Commodities Basket-3	0.4	0.69	0.007
		ARIMAX-DBP-Commodities Basket-3	0.78	0.44	-0.002
		ARIMAX-DBE-Commodities Basket-3	1.34	0.18	-0.046
ARIMAX Stock Basket 1	0.007	ARIMAX-DBB-Stock Basket-1	-1.11	0.27	0.074
		ARIMAX-DBA-Stock Basket-1	-0.01	0.99	0.003
		ARIMAX-DBP-Stock Basket-1	0.89	0.37	-0.030
		ARIMAX-DBE-Stock Basket-1	1.38	0.17	-0.081
ARIMAX Stock Basket 2	-0.003	ARIMAX-DBB-Stock Basket-2	-0.39	0.7	0.000
		ARIMAX-DBA-Stock Basket-2	-0.8	0.42	0.011
		ARIMAX-DBP-Stock Basket-2	-0.03	0.97	-0.005
		ARIMAX-DBE-Stock Basket-2	0.99	0.32	-0.084
ARIMAX Stock Basket 3	0.042	ARIMAX-DBB-Stock Basket-3	-1.16	0.25	0.144
		ARIMAX-DBA-Stock Basket-3	0.57	0.57	0.003
		ARIMAX-DBP-Stock Basket-3	0.59	0.56	0.000
		ARIMAX-DBE-Stock Basket-3	1.55	0.12	-0.077
ARIMAX Combined Basket 1	0.044	ARIMAX-DBB-Combined Basket-1	-0.94	0.35	0.110
		ARIMAX-DBA-Combined Basket-1	0.03	0.97	0.019
		ARIMAX-DBP-Combined Basket-1	0.09	0.92	0.009
		ARIMAX-DBE-Combined Basket-1	1.34	0.18	-0.051
ARIMAX Combined Basket 2	**0.010**	**ARIMAX-DBB-Combined Basket-2**	**-0.8**	**0.42**	**0.030**
		ARIMAX-DBA-Combined Basket-2	0.05	0.96	0.002
		ARIMAX-DBP-Combined Basket-2	-0.27	0.79	0.007
		ARIMAX-DBE-Combined Basket-2	0.41	0.68	-0.012
ARIMAX Combined Basket 3	0.040	ARIMAX-DBB-Combined Basket-3	-1.28	0.2	0.166
		ARIMAX-DBA-Combined Basket-3	0.84	0.4	-0.001
		ARIMAX-DBP-Combined Basket-3	1.36	0.17	-0.028
		ARIMAX-DBE-Combined Basket-3	1.07	0.28	-0.029

Note: The above table presents the results of the verification of hypothesis H(6). The hypothesis assumed that creating a diversified strategy composed of four individual models by investing in four instruments simultaneously would improve the results in terms of risk-adjusted return (IR**). To test this, we conducted a test for equality of means for independent samples. The results indicate that only one model's results are statistically significant.

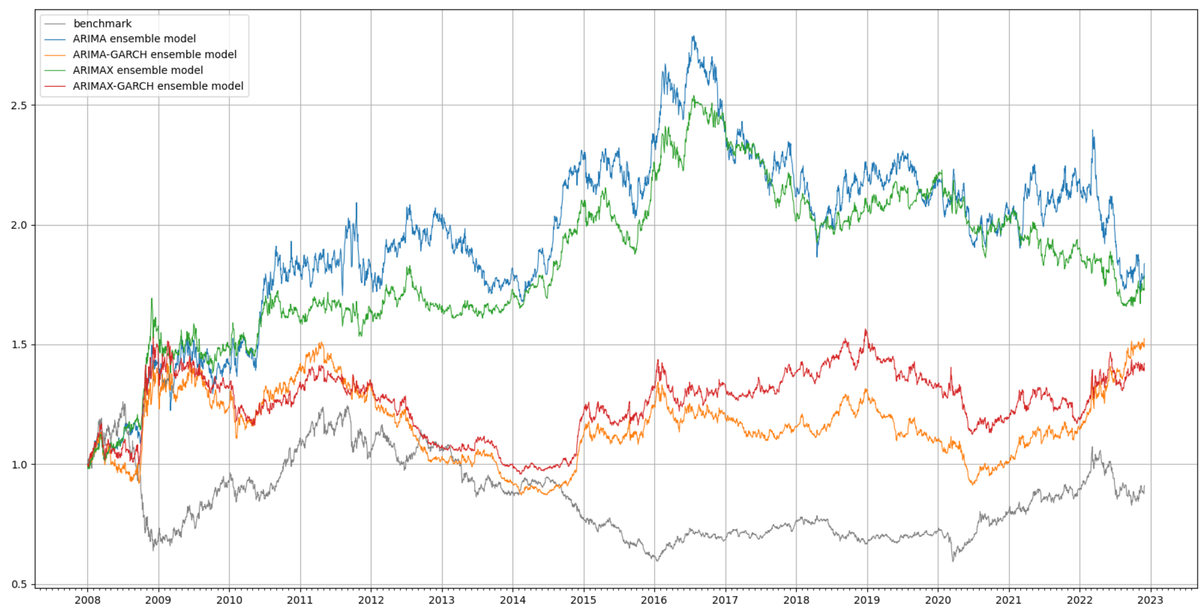

Figure 12: Ensemble models

Note: The chart shows ensemble models from the beginning of 2008 to December 2, 2022. For the majority of the test period, the models used in the study achieved better results than the benchmark, which is a combination of four commodity baskets (DBA, DBB, DBE, DBP) with equal weight at the beginning of the period. ARIMA and ARIMAX are highly correlated, as are the ARIMA-GARCH and ARIMAX-GARCH models. It can be concluded that adding the GARCH model results in a much stronger signal creation than adding exogenous variables.

Table 22: Table with results of ensemble models

Instrument	No. of models	ARC	aSD	IR*	MD	MLD	IR**	t(H2)	p(H2)
benchmark	4	-0.63%	15.85%	-0.040	49.44%	4.813	-0.001	NA	NA
ARIMA-100-20	4	4.16%	15.32%	0.272	28.77%	2.603	0.039	0.81	0.42
ARIMA_GARCH-100-20	4	2.85%	14.18%	0.201	25.48%	3.524	0.022	0.57	0.57
ARIMAX-100-20	**36**	**3.95%**	**11.54%**	**0.342**	**18.47%**	**2.873**	**0.073**	**0.77**	**0.44**
ARIMAX_GARCH-100-20	36	2.37%	13.36%	0.178	22.17%	3.532	0.019	0.47	0.64

Note: The table presents the performance metrics results of ensemble models. The best performing model was an autoregressive model with a moving average and exogenous variables (ARIMAX) with an IR* score of 0.342. Although the autoregressive model without exogenous variables (ARIMA) has a higher return rate (ARC = 4.16%), it carries more risk, as can be seen from the All Risk or Average Maximum Drawdown metrics.

The second way of combining showed similarities between ARIMA and ARIMAX models and between ARIMA-GARCH and ARIMAX-GARCH. The addition of the GARCH model had a much greater impact on generating forecasts than adding another exogenous variable. The highest result was achieved by the ARIMAX ensemble with IR** = 0.073. Although the ARIMA ensemble showed a higher annual return rate (ARC = 4.16% vs ARC = 3.95%), it was subject to higher risk. All final models were able to beat the Buy & Hold, which was an index composed of 4 ETFs.

4.7 Discussion of the empirical results

Our results indicated that it is difficult to build a model that, after appropriate parameterization, achieves stable, positive returns. Our investment strategies investing in four different commodity groups often led to annual losses. In total, 51 out of 80 models created by us achieved a positive annual return. A little more than half (44 models) were able to beat the benchmark, which was the Buy & Hold strategy. The models achieved results ranging from -15.6% (DBE-ARIMA) to +15.7% (DBB-ARIMAX with one commodity variable). The only group that showed solid results was the set of models investing in a basket of industrial metals.

Model classification did not provide clear evidence that any number of variables (1, 2 or 3) was the most effective. On the other hand, the classification based on the basket of exogenous variables showed the advantage of the commodity basket over the other two.

By considering the average Information Ratio** of the ensemble models, we confirmed the conclusions from the basket variable classification. Additionally, we noticed that the order for classification by number of variables is 3, 1 and 2, respectively.

The second stage of composition visualized the relationships between the estimated models. By summing up the results of individual strategies, we were able to create an average result. This allowed us to notice that ARIMA models are significantly more sensitive to the addition of the GARCH model than to the addition of a new variable in the form of a return from a selected basket. This also gave us the possibility to formulate final conclusions about which models performed best and in which group their best combinations should be sought. The most effective final model turned out to be ARIMAX with an Information Ratio** = 0.073. This is the 9th best result in the entire study. The combination of models definitely protected us from capital loss and allowed us to achieve above-average profits (Buy & Hold ARC = -0.63%).

5 Sensitivity analysis

In this part of the work, we presented a sensitivity analysis carried out for all final models. The initial assumption used for analyzing the models in this study were the lengths: In-Sample = 100 and Out-Of-Sample = 20. In the following analysis, we will consider combinations of In-Sample with 50, 100, and 200 periods and Out-Of-Sample with 10, 20, and 40 periods. Figure 13 indicates that the parameters chosen by us were not the best combination. In table 22, we see that the highest IR*=0.384 was achieved by the ARIMAX combination with IS = 100 and OOS = 40.

Changing the initial assumptions of the ARIMA and ARIMAX models improved the result only in one case (changing OOS from 20 to 40). The ARIMA-GARCH and ARIMAX-GARCH models are more sensitive to changes in parameters. Their result can be improved in as many as 5 out of 8 possible combinations of parameters. In all four final models, extending IS to 200 periods worsened the results.

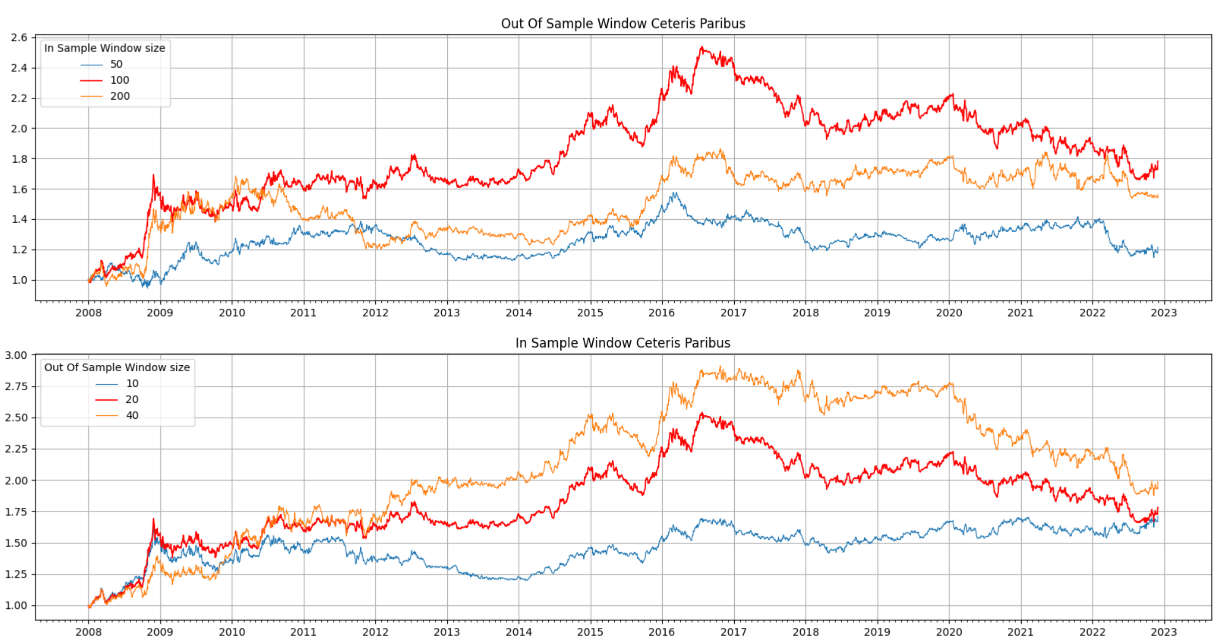

Figure 13: Sensitivity analysis for the ensmemble ARIMAX model

Note: The chart shows equity lines after conducting a sensitivity analysis for the final ARIMAX model. The sensitivity analysis in this case involved changing two variables - the length of the In Sample and Out of Sample windows. The sensitivity analysis is performed by changing only one of the two parameters at a time.

Table 23: Results of sensitivity analysis based on Information Ratio**

Instrument	ARC	aSD	IR*	MD	MLD	IR**
benchmark	-0.63%	15.85%	-0.040	49.44%	4.813	-0.001
ARIMA-50-10	0.81%	14.78%	0.055	34.2%	4.036	0.001
ARIMA_GARCH-50-10	7.07%	16.78%	0.421	26.3%	3.341	0.113
ARIMAX-50-10	0.57%	9.85%	0.058	18.59%	4.627	0.002
ARIMAX_GARCH-50-10	3.69%	12.53%	0.294	20.11%	3.385	0.054
ARIMA-50-20	-0.52%	14.27%	-0.036	25.77%	4.056	-0.001
ARIMA_GARCH-50-20	5.0%	14.47%	0.345	30.14%	2.849	0.057
ARIMAX-50-20	1.31%	9.94%	0.132	18.68%	2.774	0.009
ARIMAX_GARCH-50-20	4.48%	11.25%	0.398	18.47%	3.536	0.097
ARIMA-50-40	2.65%	14.09%	0.188	28.33%	2.706	0.018
ARIMA_GARCH-50-40	**7.96%**	**16.33%**	**0.487**	**30.34%**	**1.762**	**0.128**
ARIMAX-50-40	2.25%	9.83%	0.229	16.37%	2.560	0.031
ARIMAX_GARCH-50-40	**6.49%**	**11.59%**	**0.560**	**17.72%**	**2.238**	**0.205**
ARIMA-100-10	0.99%	14.54%	0.068	26.72%	2.893	0.003
ARIMA_GARCH-100-10	3.47%	16.28%	0.213	34.56%	2.690	0.021
ARIMAX-100-10	3.66%	10.73%	0.342	22.11%	3.758	0.057
ARIMAX_GARCH-100-10	3.95%	13.09%	0.302	21.16%	3.528	0.056
ARIMA-100-20	4.16%	15.32%	0.272	28.77%	2.603	0.039
ARIMA_GARCH-100-20	2.85%	14.18%	0.201	25.48%	3.524	0.022
ARIMAX-100-20	3.95%	11.54%	0.342	18.47%	2.873	0.073
ARIMAX_GARCH-100-20	2.37%	13.36%	0.178	22.17%	3.532	0.019
ARIMA-100-40	**5.33%**	**14.29%**	**0.373**	**34.48%**	**3.353**	**0.058**
ARIMA_GARCH-100-40	4.5%	14.11%	0.319	25.83%	3.516	0.056
ARIMAX-100-40	**4.68%**	**12.2%**	**0.384**	**19.4%**	**3.353**	**0.093**
ARIMAX_GARCH-100-40	4.54%	12.63%	0.359	17.32%	3.655	0.094
ARIMA-200-10	2.03%	14.62%	0.139	25.87%	2.937	0.011
ARIMA_GARCH-200-10	1.09%	14.08%	0.078	30.43%	5.107	0.003
ARIMAX-200-10	3.03%	12.37%	0.245	20.06%	4.353	0.037
ARIMAX_GARCH-200-10	0.84%	13.01%	0.064	24.67%	5.933	0.002
ARIMA-200-20	2.79%	14.32%	0.195	26.8%	3.103	0.020
ARIMA_GARCH-200-20	0.63%	13.84%	0.046	26.09%	5.246	0.001
ARIMAX-200-20	3.0%	12.12%	0.248	19.33%	4.516	0.038
ARIMAX_GARCH-200-20	-0.07%	12.74%	-0.006	23.84%	5.770	0.000
ARIMA-200-40	2.33%	16.01%	0.146	31.09%	2.940	0.011
ARIMA_GARCH-200-40	0.55%	13.87%	0.040	32.18%	3.337	0.001
ARIMAX-200-40	2.72%	12.98%	0.210	28.05%	2.452	0.020
ARIMAX_GARCH-200-40	0.05%	13.03%	0.003	27.86%	5.909	0.000

Note: Translation: The table shows the sensitivity analysis of the final models. Changing the parameters of the ARIMA model only improved the final result in one case (changing OOS from 20 to 40). Similarly, in the ARIMAX model, changing the OOS window from 20 to 10 gave the same result, while changing OOS from 20 to 40 improved the outcome. For the ARIMA-GARCH and ARIMAX-GARCH models, changing the parameters improved the results in 5 to 8 cases. All models achieved a worse IR* result when we changed the initial assumptions to IS higher than 200.

The sensitivity analysis cast doubt on the robustness of the ARIMA and ARIMAX models. The change in parameters had a significant negative impact, indicating that the chosen lengths of the training and testing windows based on intuition may not have been among the best solutions. This could also suggest that examining the results for other window lengths to increase the value of the risk-adjusted return rate would not produce the desired effects.

The ARIMA-GARCH and ARIMAX-GARCH models showed better results with shorter IS and OOS windows. It is possible that applying different initial assumptions would bring better results for single models.

6 Conclusions

The goal of this work was to create an algorithmic investment strategy that generates above average returns. In our research, we used four time series corresponding to open investment funds (ETFs) from the Invesco group. The ETFs we selected are mainly used as an alternative to investing through futures contracts on commodity markets. We used four models (ARIMA, ARIMAX, ARIMA-GARCH, ARIMAX-GARCH) to create buy and sell signals. We compared the results of all models with the Buy&Hold strategy based on performance metrics and equity lines.

In this study, we were guided by the following research hypotheses:

($H1$) Based on predictions from individual ARIMA/ARIMA-Garch and ARIMAX/ARIMAX-Garch models, we are unable to generate signals for algorithmic investment strategies that outperform the market (i. e. characterized by a higher IR** than Buy&Hold).

- Out of the 80 individual models created in this study, 44 of them had a higher IR** than the Buy&Hold strategy. Due to the significant differences in the results of the models investing in different ETFs, and the fact that none of the 80 models showed statistical significance, we have no grounds to reject this hypothesis.

($H2$) Based on predictions from ensemble ARIMA/ARIMA-Garch and ARIMAX/ARIMAX-Garch models, we are unable to generate signals for or algorithmic investment strategies that outperform the market.

- All four final strategies achieved risk-adjusted returns better than Buy&Hold. The final strategies exhibited much lower risk than the benchmark, which would allow for leverage to be used to increase returns. However, the significance test indicated that the final strategies are not fully reliable. This contradiction makes it difficult for us to verify the hypothesis, and we decide to leave it for personal verification.

($H3$) Predictions from the ARIMA-Garch/ARIMAX/ARIMAX-Garch model are more effective in algorithmic investment strategies than predictions from the ARIMA model. Both exogenous data and volatility forecasting factor contribute to improving the informational value of the ARIMA model in the context of IR**.

- Among the individual ARIMA models for all four ETFs, model frequently exhibited the best or one of the better risk-adjusted returns. The t-statistic analysis indicated that 35 out of 76 individual model observations had an advantage over ARIMA (in terms of return direction). On the other hand, the analysis of Table 12 indicates a significant impact of the volatility factor and exogenous variables on improving the overall strategy's performance in terms of IR** and statistical significance. Similar to (H3), we are able to partially reject this hypothesis.

($H4$) The number of exogenous variables affects the results of algorithmic investment strategies based on the ARIMAX model.

- The classification results in table 18 did not provide clear evidence, but the results of the ensemble models indicated that the amount of variables influenced the informational value of the model. We had no ground to reject.

($H5$) The length of the training and test window affects the results of the tested investment strategies.

- The results of the sensitivity analysis did not provide us with grounds to reject.

($H6$) The final strategy diversified through simultaneous investment in four ETFs (DBA, DBE, DBB, DBP) improves results in terms of risk-adjusted return (IR**) compared to individual models.

- 22 out of 36 individual models from Table 21 exhibited a positive t-statistic. This indicates that portfolio diversification through simultaneous investment in four ETFs has a positive impact on the final strategy's performance. However, only one strategy among the individual models turned out to be statistically significant. The hypothesis has been partially rejected.

The main aspect that we focused on while analyzing the results is that they are dependent on the choice of the underlying instrument. Changing the time series from DBB to DBE, despite the same parameters and exogenous variables, had a drastic impact on the deterioration of the results. This means that the models used by us are not an appropriate tool for making investment decisions.

A precise classification of variables and their amounts allowed us to understand the essence of selecting exogenous variables. Analyzing the composition of models led us to the conclusion that the choice of the appropriate basket of variables is more important than the number of variables added to the model.

Finally, the sensitivity analysis suggested to us that shortening the In-Sample and Out-Of-Sample windows could have resulted in better results for the ARIMA-GARCH and ARIMAX-GARCH models. The deterioration of the results of ARIMA and ARIMAX demonstrated their lack of robustness.

We believe that our research thoroughly addressed the essence of properly conducting parameterization and analysis of models in terms of their combination and the impact of individual parameters. The application of the proper testing methodology along with metrics allowed for a comparison of many strategies and verification of hypotheses. We hope that the results of this work will be useful to people analyzing similar issues.

References

Abda, H. T., Essaa, A. K., & Jassima, F. M. (2021). Analyzing the relationship between the dow jones index and oil prices using the ARIMAX model. *International Journal on Advanced Science, Engineering and Information Technology*, *11*(2), 465–473.

Adcock, R., & Gradojevic, N. (2019). Non-fundamental, non-parametric bitcoin forecasting. *Physica A: Statistical Mechanics and Its Applications*, *531*, 121727.

Alexander, B., Aaron, K. Y. S., Svetlozar, R., Michael, F., & Frank, F. (2013). Empirical analysis of ARMA-GARCH models in market risk estimation on high-frequency US data. *Studies in Nonlinear Dynamics & Econometrics*, *17*(2), 167–177. https://doi.org/10.1515/snde-2012-0033

Atsalakis, G. S., & Valavanis, K. P. (2009). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert Systems with Applications*, *36*(7), 10696–10707.

Bailey, D. H., Borwein, J., Lopez de Prado, M., Salehipour, A., & Zhu, Q. J. (2016). Backtest overfitting in financial markets. *Automated Trader*.

Bailey, D. H., Borwein, J., Lopez de Prado, M., & Zhu, Q. J. (2016). The probability of backtest overfitting. *Journal of Computational Finance, Forthcoming*.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance*, *32*(3), 663–682.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, *52*(1-2), 5–59.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

Bui, Q., & Ślepaczuk, R. (2022). Applying hurst exponent in pair trading strategies on nasdaq 100 index. *Physica A: Statistical Mechanics and Its Applications*, *592*, 126784. https://doi.org/10.1016/j.physa.2021.126784

Castellano Gomez, S., & Ślepaczuk, R. (2021). Robust optimisation in algorithmic investment strategies. *Working Papers of Faculty of Economic Sciences, University of Warsaw*, (WP 27/2021 (375)).

Chen, H., & Maher, D. (2013). On the predictive role of large futures trades for s&P500 index returns: An analysis of COT data as an informative trading signal. *Journal of International Financial Markets, Institutions and Money*, *27*, 177–201.

Chumacero, R. A. (2001). Estimating ARMA Models Efficiently. *Studies in Nonlinear Dynamics & Econometrics*, *5*(2), 1–14. https://doi.org/10.2202/1558-3708.1074

Cowles 3rd, A. (1933). Can stock market forecasters forecast? *Econometrica: Journal of the Econometric Society*, 309–324.

Di Persio, L., & Honchar, O. (2016). Artificial neural networks architectures for stock price prediction: Comparisons and applications. *International Journal of Circuits, Systems and Signal Processing*, *10*(2016), 403–413.

Dimson, E., & Mussavian, M. (1998). A brief history of market efficiency. *European Financial Management*, *4*(1), 91–103.

Elena, G., & Hiroki, T. (2005). Bayesian Analysis of a Doubly Truncated ARMA-GARCH Model. *Studies in Nonlinear Dynamics & Econometrics*, *9*(2), 1–38. https://doi.org/10.2202/1558-3708.1166

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

Georgios, B., & Theodore, P. (2015). On the relationship between oil and gold before and after financial crisis: linear, nonlinear and time-varying causality testing. *Studies in Nonlinear Dynamics & Econometrics*, *19*(5), 657–668. https://doi.org/10.1515/snde-2014-0060

Gökbulut, R., & Pekkaya, M. (2014). Estimating and forecasting volatility of financial markets using asymmetric GARCH models: An application on turkish financial markets. *International Journal of Economics and Finance*, *6*(4), 23–35.

Granger, C. (1969). Econometrica: Journal of the econometric society. *JSTOR*, *37*, 424.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424–438.

Granger, C. W. (1992). Forecasting stock market prices: Lessons for forecasters. *International Journal of Forecasting*, *8*(1), 3–13.

John, M. (2005). Can GARCH Models Capture Long-Range Dependence? *Studies in Nonlinear Dynamics & Econometrics*, *9*(4), 1–43. https://doi.org/10.2202/1558-3708.1269

Julián, A.-F., Fernando, F.-R., Maria-Dolores, G.-A., & Simon, S.-R. (2003). An Empirical Evaluation of Non-Linear Trading Rules. *Studies in Nonlinear Dynamics & Econometrics*, *7*(3), 1–32. https://doi.org/10.2202/1558-3708.1160

Kijewski, M., & Ślepaczuk, R. (2020). Predicting prices of s&P500 index using classical methods and recurrent neural networks. *Working Papers of Faculty of Economic Sciences, University of Warsaw*, (WP 27/2020 (333)).

Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, *10*(2), 169–181.

Kriechbaumer, T., Angus, A., Parsons, D., & Casado, M. R. (2014). An improved wavelet–ARIMA approach for forecasting metal prices. *Resources Policy*, *39*, 32–41.

Latoszek, M., & Ślepaczuk, R. (2020). Does the inclusion of exposure to volatility into diversified portfolio improve the investment results? Portfolio construction from the perspective of a polish investor. *Economics and Business Review*, *6*(1), 46–81. https://doi.org/10.18559/ebr.2020.1.3

Li, X., Ma, J., Wang, S., & Zhang, X. (2015). How does google search affect trader positions and crude oil prices? *Economic Modelling*, *49*, 162–171.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, *20*(2), 130–141.

Ma, C. K., Mercer, J. M., & Walker, M. A. (1992). Rolling over futures contracts: A note. *The Journal of Futures Markets (1986-1998)*, *43*(1), 203.

Markowitz, H. M. (1968). Portfolio selection. In *Portfolio selection*. Yale university press.

Michańków, J., Sakowski, P., & Ślepaczuk, R. (2022). LSTM in algorithmic investment strategies on BTC and s&P500 index. *Sensors*, *22*(3), 917. https://doi.org/10.3390/s22030917

Mondal, P., Shit, L., & Goswami, S. (2014). Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *International Journal of Computer Science, Engineering and Applications*, *4*(2), 13.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, *42*(4), 2162–2172.

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

Roondiwala, M., Patel, H., & Varma, S. (2017). Predicting stock prices using LSTM. *International Journal of Science and Research (IJSR)*, *6*(4), 1754–1756.

Ryś, P., & Ślepaczuk, R. (2019). Machine learning in algorithmic trading strategy optimization-implementation and efficiency. *Central European Economic Journal*, *5*(1), 206–229. https://doi.org/10.1515/ceej-2018-0021

Sanders, D. R., Irwin, S. H., & Merrin, R. P. (2009). Smart money: The forecasting ability of CFTC large traders in agricultural futures markets. *Journal of Agricultural and Resource Economics*, 276–296.

Shen, S., Jiang, H., & Zhang, T. (2012). Stock market forecasting using machine learning algorithms. *Department of Electrical Engineering, Stanford University, Stanford, CA*, 1–5.

Ślepaczuk, R. (2004). Efektywność rynku instrumentów pochodnych na GPW w warszawie. *Rynek, Gospodarka, Społeczeństwo*, *12*, 163–188.

Ślepaczuk, R. (2006). Anomalie rynku kapitałowego w świetle hipotezy efektywności rynku. *Financ. Internet Q.(Former. E-Finans.)*, (1), 1–12.

Ślepaczuk, R., Sakowski, P., & Zakrzewski, G. (2018). Investment strategies that beat the market. What can we squeeze from the market? *Financ. Internet Q.(Former. E-Finans.)*, *14*, 36–55. https://doi.org/10.2478/fiqf-2018-0026,

Vo, N., & Ślepaczuk, R. (2022). Applying hybrid ARIMA-SGARCH in algorithmic investment strategies on s&P500 index. *Entropy*, *24*(2), 158. https://doi.org/10.3390/e24020158

Wang, Y., Liu, L., Diao, X., & Wu, C. (2015). Forecasting the real prices of crude oil under economic and statistical constraints. *Energy Economics*, *51*, 599–608.

Wiecki, T., Campbell, A., Lent, J., & Stauth, J. (2016). All that glitters is not gold: Comparing backtest and out-of-sample performance on a large cohort of trading algorithms. *The Journal of Investing*, *25*(3), 69–80.

Wysocki, M., & Ślepaczuk, R. (2021). Artificial neural networks performance in WIG20 index options pricing. *Entropy*, *24*(1), 35. https://doi.org/10.3390/e24010035

Xiong, T., Li, C., Bao, Y., Hu, Z., & Zhang, L. (2015). A combination method for interval forecasting of agricultural commodity futures prices. *Knowledge-Based Systems*, *77*, 92–102.

Zhang, Y., Chu, G., & Shen, D. (2021). The role of investor attention in predicting stock prices: The long short-term memory networks perspective. *Finance Research Letters*, *38*, 101484.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl