ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
SCHOOL OF ECONOMIC SCIENCES
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
DEPARTMENT OF ECONOMICS

Department of Economics

Athens University of Economics and Business

WORKING PAPER no. 01-2024

Market Timing & Predictive Complexity

Stelios Arvanitis, Foteini Kyriazi, Dimitrios Thomakos

February 2024

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5

Market Timing & Predictive Complexity*

Stelios Arvanitis[†], *Foteini Kyriazi*[‡], *Dimitrios Thomakos*[§]

Abstract

We propose a new method for selecting the local estimation window for forecasting and trading financial returns. It is built around one particular definition of predictive complexity and we apply it in the simplest of predictors, the sample mean. We derive the exact conditions for the process of optimally selecting the local estimation window among a, theoretically found, grid of potential values of it. We use different loss functions, statistical and financial, which are considered individually and then pooled under two selection concepts, stochastic dominance and minimum description length, and find exact expressions as to how the associated complexities and their combinations can be derived and applied. Our results are based on a set of probabilistic assumptions for the time series under study and, based on those, we offer an inferential procedure for testing the presence of excess trading returns. Our empirical illustration on a set of diverse exchange trade funds (ETFs) across different asset classes suggests that the method works very well in practice and that it can generate both statistical and financial performance enhancements. A number of extensions to different predictors and different underlying assumptions is discussed.

JEL Codes: C53, C58, G17. *MSC2020*: 60G25, 68Q30, 90.

Keywords: approximate bounds, asset returns, complexity, ETF, forecasting, market timing, minimum description length, prediction, predictive complexity, quantitative trading, stochastic dominance.

---

*All computations performed by the authors in Python.

[†]Corresponding author. Department of Economics, Athens University of Economics and Business, Greece. Email: stelios@aueb.gr

[‡]Department of Agribusiness and Supply Chain Management, Agricultural University of Athens, Greece. Email: fkyriazi@aua.gr

[§]Department of Business Administration, National and Kapodistrian University of Athens, Greece. Email: dthomakos@ba.uoa.gr

1 Introduction

There is a vast literature, both among academics and finance professionals, about the ability to time the market and accurately forecast market swings. The claims lay at either one of the extremes, some saying that market timing is impossible and a futile exercise and others claiming that the opposite is true and forecasting market returns is indeed a viable and profitable occupation. In this paper we take the path of the mean between these two extremes and suggest that, when it is possible to time the market and predict and trade successfully future returns, it can be done by a coherent and consistent procedure that minimizes the minimum predictive complexity associated with a set of loss function or alternatively comes as close as possible to be the infeasible unanimous choice of predictive complexities associated with individual losses in a stochastic dominance framework.

Our suggested methodology rests on two pillars, that cover at the same time simplicity and complexity. Our first point of departure relates to the space of loss functions: successful trading requires accurate *sign* forecasts and not necessarily accurate *magnitude* forecasts of future returns. Thus, any attempt of forecasting future returns should be a blend of statistical and financial loss functions. To achieve this blending one must resort to some form of "model selection criteria" and we show how two such approaches could be used that are based on the notion of predictive density and predictive complexity. Our second point of departure relates to the method of generating forecasts of future returns: one does not necessarily need a complex method or complex model to generate successful *sign* forecasts, the sample mean (or a weighted sample mean) is more than enough for that, if one can select successfully the local window of observations over which that sample mean is computed. That this is so can be easily understood via a simple example. Suppose that your last 5, say, returns of an asset are these: -0.02, -0.01, 0.00, 0.03, -0.01. If you use all five observations your average would be -0.002, and if you use the last 2, 3 and 4 observations your average would be 0.01, 0.0067 and 0.0025. Now, if the next period's return is positive you have 3 windows that will give you the right sign while if the next period's return is negative you only have one window that will give you the right sign; the point is that there always going to be a local window of observations whose sample mean will give you the correct sign for next period's return. So, the method for getting a sign forecast can be as trivial as the sample mean, but the choice of the local window is far from a trivial problem to solve. We thus combine the two points of departure and provide a suggested solution for the "optimal choice" of a local, rolling, window based on the notion of predictive complexity.

Our work relates to four lines of research: prior work on predictive complexity, prior work on market timing, prior work on forecasting asset returns, and prior work on approximate bounds and stochastic dominance. Below we briefly review some of the literature on the above mentioned lines of research, mainly indicating that the work on market timing and, even more, on forecasting market research is actively pursued.

An important early work on the concept of predictive complexity, in the context of universal prediction and prediction with experts is that of Vovk and Watkins (1998); Vovk has made very important contributions to game-theoretic finance, prediction with experts and the latter years he is the one who introduced conformal prediction.[1] This work of Vovk and Vatkins was followed with a stream of other papers on the concept of predictive complexity from which we sample Kalnishkan and Vyugin (2002) [38] and Kalnishkan et al. (2004) on the foundations of predictive complexity [39],

[1] For a book length treatment of the concept of conformal prediction see Vovk, Gammernam and Shafer *Algorithmic Learning in a Random World*, 2nd ed. (2022) [78].

Vyugin and Vyugin (2005) [80] on the relationship between predictive complexity and information, and Chernov et al. (2010) [12] on the relationship of supermartingales in prediction with expert advice. Kalnishkan (2022) [40] has an excellent review, and practical introduction paper, for prediction with experts and one can consult this for additional references.[2]

Turning next to the concept of market timing we go quite a bit back and the literature below is entirely representative of earlier foundational papers and some recent advances. Sharpe (1975) [72], Henriksoon and Merton (1981) [31] and Henriksson (1984) [32] have initiated the post-1970s discussion on the concept of market timing and how to test for it. Following their work, and sampling across different time periods, we can find works like those of Jeffrey (1984) [34] calling market timing a "folly", Cumby and Modest (1987) [15] that provided another test for market timing in the context of forecast evaluation, Droms (1989) [20] who discussed market timing as an investment policy, Busse (1991) [9] who is an early reference on volatility timing and Shilling (1992) [73] that supported market timing as better than the passive buy& hold. Then, Pesaran and Timmermann (1994) [60] offered a generalization of the market timing test of Henriksson and Merton, Larsen and Wozniak (1995) [44] who also support market timing as a real-world, profitable strategy and Beebower and Varikooty (1999) [5] who analyzed market timing strategies. Getting into the 2000's we have Bollen and Busse (2001) [7] on the market timing ability of fund managers, Li and Lam (2002) [50] who fully explored in a mathematical context the optimal market timing strategies under transaction costs, Jiang (2003) [36] who suggested a non-parametric test for market timing and Lam and Li (2004) [43] who re-evaluated the previous work of Li and Lam on how good is the optimal market timing strategy.

[2] Additional material on the above concepts and the foundations of game-theoretic finance see Shafer and Vovk *Game-Theoretic Foundations for Probability and Finance* (2019) [71].

Christoffersen and Diebold (2006) [13] provided a seminal analysis on the interplay between direction-of-change (sign) forecasting and the impact of volatility dynamics – with additional work on the topic by Thomakos and Wang (2010) [75]; Then we have Jiang et al. (2007) [35] again on the ability of mutual funds to time the market, Thomakos et al. (2007) [74] on asset rotation for market timing – with follow-up work on asset rotation by Schizas and Thomakos (2015) [70]. Finally, we note the works of Elton et al. (2012) [22] once again for the ability of mutual funds to time the market, Matallin-Saez (2015) [53] on performance measures on market timing and Ferson and Mo (2016) [24] on a similar issue of performance measurement and finally Ding et al. (2023) [19] for a very recent non-parametric test of market timing.

Our next piece of literature relates to the forecastability of asset returns. One migth start with the Cowles (1934) [14] paper on whether stock market forecasters can actually successfully forecast the market – but the forecasting literature on the topic is older than that.[3]. We then "jump" to papers on or after 2020's just to illustrate the continued strong research interest on this topic. For example, Dai et al. (2020) [17] who consider technical indicators for forecasting stock returns, Dichtl (2020) [18] on forecasting excess returns of the gold market, Kyriazi and Thomakos (2020) [42] on the existence and interpretation of predictability in currencies, Liang et al. (2022) [48] on the use of dimensionality reduction methods for predicting market returns, Li et al. (2022) [49] on forecasting US stock market returns and Lv and Qi (2022) [52] that examine return predictability from a combination forecasting perspective. A stream of very recent papers includes Brennan and Taylor (2023) [8] on the interplay of expected returns and risk in the stock market, Casta (2023) [10] on the relationship between inflation, interest rates and the predictability of asset returns, Chen et al.

[3] See the review paper of Thomakos and Xidonas (2023) [76] on the foundations of forecasting pre-1940s.

(2023) [11] on use of economic policy uncertainty indices as predictors of market returns – with additional work on the topic by Huang et al. (2023) [33], Cotter et al. (2023) [16] on commodity futures return, Guerard et al. (2023) [29] on the long-term historical predictability of the S&P500 and Dow Jones indices, Haase and Neuenkirch (2023) [30] on the predictability of market returns during different regimes, Li and Sun (2023) [50] on predicting market returns using aggregate credit risk, Qiu et al. (2023) [67] on the use of long-term volatility for forecasting market returns, Saenz et al. (2023) [68] on the use of clustering and finally Zhang and Wang (2023) [81] on forecasting oil futures market returns.

Stochastic dominance relations are pre-orders on sets of probability measures on the real line (see Fishburn (1976) [25] for definitions). heir study has gained importance in the fields of economics, finance and statistics/econometrics (see among others Kroll and Levy (1980) [41], McFadden (1989) [54], Levy (1992) [45], Levy (2015) [46], Mosler and Scarsini (1993) [57], Gayant and Le Pape (2017) [26]), since it among others enables the investigation of issues of conservative choice under uncertainty, over large classes of preferences. This is due to that stochastic dominance relations usually have characterizations in terms of classes of utility functions (see Fishburn (1976) [25], Levy (1992) [45], Levy (2015) [46], Levy and Levy (2002) [47], whereas dominance w.r.t. such a relation is equivalent to preference by every utility in the class. Thus, those relations are among others useful for the characterization of robust choices in the case of ambiguity about the exact preferences of the decision maker. Since in the context of the expected utility paradigm, a loss function is dual to a utility function, the stochastic dominance maxim is applicable in forecasting whenever some form of robustness w.r.t. the choice of a loss function is desired; see for example Jin et al. (2017) [37], and Arvanitis et al. (2021) [3] in relation to the concept of superior predictive ability, or Post et al. (2019) [63] for the concept of robust optimization of forecast combinations.

From this very representative literature review above, we conjecture that we are not amiss in pursuing the theme of the current paper, as the recent literature shows continued and growing attention to the issue of market return predictability. Thus, synthesizing the above, and for a class of loss functions related to both statistical and financial considerations, we utilize the associated predictive complexity functionals as indicated in the works of Rissanen (1996) [64] and Rissanen (1997) [65]. Those are essentially the empirical risks of the loss functions involved, appropriately penalized so as to reflect the notion of statistical complexity as defined in the above-mentioned papers. They are functions of the predictive pairs consisting of the latent-at the time of prediction-random return and the past means, which are themselves functions of the estimating window. Then our methodology defines the optimal windows as either a minimizer of the minimal predictive complexity of the loss function class, or alternatively, as an approximate simultaneous minimizer of every complexity in the class in the context of stochastic dominance. We derive the limit theories of the extrema of the empirical processes associated with the predictive complexities involved. This enables the design and the derivation of the asymptotic properties of a statistical test of whether the population optimal window emerging from either the minimal predictive complexity part, or the stochastic dominance part of our methodology is also a population minimizer of a non-penalized risk over a loss function of interest. We apply our theory on a set of diverse exchange trade funds (ETFs) across different asset classes; the results suggest that the method works very well in practice and that it can generate both statistical and financial performance enhancements.

The rest of the paper is structured as follows: section 2 has the theoretical framework previously described, section 3 has the data, empirical implementation and discussion of results and section 4 offers some concluding remarks and conclusions for future research. The appendix holds all figures and tables that are appear in the main body of the paper.

2 Theoretical framework

2.1 Foundations & Definitions

$(X_t)_{t\in\mathbb{Z}}$ is an observable stochastic process of asset returns. At an arbitrary $t \in \mathbb{N}^\star$, X_t is unknown, the sample realization $(X_i)_{i=0,\ldots,t-1}$ is available to the researcher, and X_t is predicted by the α-weighted mean $y_t(\delta,\alpha) := (\sum_{i=t-\delta-1}^{t-1} i^\alpha)^{-1} \sum_{i=t-\delta-1}^{t-1} i^\alpha x_i$. Here, α is a non-negative fixed parameter that is pre-defined by the researcher. When $\alpha = 0$ the usual sample mean over the window δ is obtained. The averaging window δ is a strictly positive integer chosen by the researcher from a non-empty parameter set $\Delta \subset \mathbb{N}^\star$. We are interested in the optimal choice of the averaging window; the optimality considerations are directly related to a non empty set of loss functions, say $\mathcal{L}$, that is available to the researcher, which represent potential preferences towards the risk of the forecast pair $(X_t, y_t(\delta,\alpha))$. Given a total sample realization $(X_t)_{t=1,\ldots,T}$, the realized forecast pair time series available to the researcher is then $(X_t, y_t(\delta,\alpha))_{t=1,\ldots,T}$; this is usable for inference on the optimal choice of the window.

Following Rissanen (1996) [64] and Rissanen (1997) [65], for $\ell \in \mathcal{L}$, the predictive density associated with the loss ℓ, $f_{\ell,t}(x) = 2^{-\ell(x,y_t(\delta,\alpha))}(\int_{\min_{0\le i\le t-1} X_i}^{\max_{0\le i\le t-1} X_i} 2^{-\ell(z,y_t(\delta,\alpha))}dz)^{-1}$ is considered, provided that $-\ell$ is exponentially integrable. The predictive complexity of $(X_t)_{t=1,\ldots,T}$ associated with the loss ℓ, and the estimation window δ, is then defined by the functional $Q(\ell,\delta) := \mathbb{E}_{\mathbb{P}_T}[Q_t(\ell,\delta,\min_{0\le i\le t-1} X_i, \max_{0\le i\le t-1} X_i)]$, with

$Q_t(\ell, \delta, y_\star, y^\star) := \ell(X_t, y_t(\delta, \alpha)) + \log_2(P_t(\ell, \delta, y_\star, y^\star))$, and where for the penalization term we have that $P_t(\ell, \delta, y_\star, y^\star) := \int_{y_\star}^{y^\star} 2^{-\ell(x, y_t(\delta,\alpha))} dx$; $\mathbb{P}_T$ is the empirical distribution of the random element $(X_t, y_t(\delta, \alpha), \min_{0 \leq i \leq t-1} X_i, \max_{0 \leq i \leq t-1} X_i)_{t=1,\dots,T, \delta \in \Delta}$.

The minimization of $Q(\ell, \delta)$ over $\mathcal{L} \times \Delta$ produces the MDL optimal averaging window, associated with Rissanen (1996) [64] and Rissanen (1997) [65]; the finiteness of the parameter space $\mathcal{L} \times \Delta$ ensures its existence and measurability:

Definition 1. The MDL-optimal averaging window, δ_{MDL}, satisfies:

$$\delta_{\mathrm{MDL}} \in \arg \min_{\delta \in \Delta, \ell \in \mathcal{L}} Q(\ell, \delta). \tag{1}$$

The predictive complexities over the loss functions' set can be further used in order to construct a stochastic dominance relation on Δ. An averaging window δ dominates another $\delta^\star$, iff the first attains lower predictive complexity for any loss function in the class. Formally:

$$\delta \succeq \delta^\star \text{ iff } Q(\ell, \delta)) \leq Q(\ell, \delta^\star), \ \forall \ell \in \mathcal{L}.$$

Notice that $\delta_{\mathrm{MDL}}$is an efficient element of the order; it is not dominated by any other estimation window.

Given the stochastic dominance ordering, it is possible to choose the window optimally by the approximate super-efficiency criterion of Arvanitis et al. 2021 [2]:

Definition 2. The approximate bound of the SD relation, say δ_{SD} satisfies:

$$\delta_{\mathrm{SD}} \in \arg \min_{\delta \in \Delta} \sup_{\ell \in \mathcal{L}, \delta^\star \in \Delta} (Q(\ell, \delta) - Q(\ell, \delta^\star)). \tag{2}$$

The issue of existence and measurability for the approximate bound is also trivial. δ_{SD} "is as close as possible" to minimizing the predictive complexity regarding every loss function in the class. If such a "super-efficient" element exists, then $\delta_{\text{SD}} = \delta_{\text{MDL}}$. In general, whenever for example $\mathcal{L}$ is compact in some topology (e.g. of uniform convergence), then,

$$\sup_{\ell \in \mathcal{L}} (Q(\ell, \delta_{\text{SD}}) - Q(\ell, \delta_{\text{MDL}})) \geq 0.$$

Additionally, δ_{SD} is (Pareto) efficient (non-dominated); if it were dominated, then it would not optimize the criterion. Furthermore, δ_{SD} is generally characterized as an approximate bound. It can be justified as a conservative optimal selection under uncertainty concerning risk preferences. A sufficient condition for δ_{MDL} to be an approximate bound is that it is a bound for a subset of $\mathcal{L}$-this is trivial for at least a singleton set-and that it achieves almost optimality for the remaining loss functions.

2.2 Loss functions

Presently we work with a set of four loss functions; two statistical, namely the squared error and the absolute error, and two financial, namely the negative of the log cumulative wealth and the proportion of of incorrect sign predictions. Our motivation for window selection using multiple criteria that focus on both the statistical properties of the forecast errors $X_t - y_t(\delta, \alpha)$, $t = 1, \ldots, T$, as well as blending financial along with statistical information in a framework of that is suitable for testing the market timing ability of our methodology (via the active trading from the signs of our forecasts). Thus $\mathcal{L}$ is finite and at most equal to $\{\ell_{\text{SE}}, \ell_{\text{AE}}, \ell_{\text{W}}, \ell_{\text{sgn}}\}$, where $\ell_{\text{SE}}(x, y_t(\delta, \alpha)) := (x - y_t(\delta, \alpha))^2$, $\ell_{\text{AE}}(x, y_t(\delta, \alpha)) := |x - y_t(\delta, \alpha)|$, $\ell_{\text{W}}(x, y_t(\delta, \alpha)) := -\ln(x\text{sgn}(y_t(\delta, \alpha)) + 1)$,[4]

[4] This is a well defined real function as long as x assumes its values in the interior of $[-1, 1]$. Since our applications involve daily net financial returns, such a restriction is satisfied in the

and $\ell_{\text{sgn}}(x, y_t(\delta,\alpha)) = \mathbb{I}[\text{sgn}(x) \neq \text{sgn}(y_t(\delta,\alpha))]$.

For the above it is easily obtained that:

$$Q(\ell_{\text{SE}},\delta) := \frac{1}{T}\sum_{t=1}^{T}[(X_t - y_t(\delta,\alpha))^2 + \log_2(P_t(\ell_{\text{SE}},\delta, \min_{0\leq i\leq t-1} X_i, \max_{0\leq i\leq t-1} X_i))], \quad (3)$$

with

$$P_t(\ell_{\text{SE}},\delta, y_\star, y^\star) := \Phi(y^\star; y_t(\delta,\alpha); \ln^{-1/2}(4)) - \Phi(y_\star; y_t(\delta,\alpha); \ln^{-1/2}(4)),$$

and $\Phi(\cdot;\mu;v)$ denotes the cdf of $N(\mu, v)$, and $y_\star < y^\star$. Furthermore,

$$Q(\ell_{\text{AE}},\delta) := \frac{1}{T}\sum_{t=1}^{T}[|X_t - y_t(\delta,\alpha)| + \log_2(P_t(\ell_{\text{AE}},\delta, \min_{0\leq i\leq t-1} X_i, \max_{0\leq i\leq t-1} X_i)], \quad (4)$$

with

$$P_t(\ell_{\text{AE}},\delta, y_\star, y^\star) = \frac{2 - 2^{-(y^\star - y_t(\delta,\alpha))} - 2^{-(y_t(\delta,\alpha)-y_\star)}}{\ln(2)},$$

$$Q(\ell_{\text{W}},\delta) := -\frac{1}{T}\sum_{t=1}^{T}[\ln(X_t \text{sgn}(y_t(\delta,\alpha)) + 1) - \log_2(P_t(\ell_{\text{W}},\delta, \min_{0\leq i\leq t-1} X_i, \max_{0\leq i\leq t-1} X_i))], \quad (5)$$

where

$$P_t(\ell_{\text{W}},\delta, y_\star, y^\star) = \frac{(y^\star \text{sgn}(y_t(\delta,\alpha)) + 1)^{\ln(2)+1}}{(\ln(2)+1)\text{sgn}(y_t(\delta,\alpha))} - \frac{(y_\star \text{sgn}(y_t(\delta,\alpha)) + 1)^{\ln(2)+1}}{(\ln(2)+1)\text{sgn}(y_t(\delta,\alpha))},$$

---

totality of our sample. In cases where the supremum of the support of the distribution of X is greater than or equal to one, ℓ_{W}, can be modified to $-\ln(x\text{sgn}(y_t(\delta,\alpha)) + C)$, where $C > \max(1, \max_{0\leq t\leq T} X_t)$. The penalization term that is presented below would then be modified to $P_t(\ell_{\text{W}},\delta, y_\star, y^\star) = \frac{(y^\star \text{sgn}(y_t(\delta,\alpha))+C)^{\ln(2)+1}}{(\ln(2)+1)\text{sgn}(y_t(\delta,\alpha))} - \frac{(y_\star \text{sgn}(y_t(\delta,\alpha))+C)^{\ln(2)+1}}{(\ln(2)+1)\text{sgn}(y_t(\delta,\alpha))}$.

and,

$$Q(\ell_{\mathrm{sgn}}, \delta) := \frac{1}{T}\sum_{t=1}^{T}[\mathbb{I}[\mathrm{sgn}(X_t) \neq \mathrm{sgn}(y_t(\delta,\alpha))]+\log_2(P_t(\ell_{\mathrm{sgn}}, \delta, \min_{0\leq i\leq t-1} X_i, \max_{0\leq i\leq t-1} X_i))], \quad (6)$$

where,

$$P_t(\ell_{\mathrm{sgn}}, \delta, y_\star, y^\star) = \begin{cases} \frac{y^\star}{2}, & y_t(\delta,\alpha) < 0, y^\star > 0, y_\star \leq 0 \\ -\frac{y_\star}{2}, & y_t(\delta,\alpha) \geq 0, y^\star > 0, y_\star \leq 0 \\ \frac{y^\star - y_\star}{2}, & y_t(\delta,\alpha) < 0, y_\star > 0 \\ \frac{y^\star - y_\star}{2}, & y_t(\delta,\alpha) \geq 0, y^\star < 0. \end{cases}$$

Finiteness of the combined parameter space $\mathcal{L} \times \Delta$ implies that the optimizations involved for the derivations of the optimal windows via both the MDL and the SD principles are trivial. This also facilitates the extraction of the limit theories of the extrema of the empirical processes implied by the definitions above. Those limit theories are investigated in the following paragraph in order to construct statistical testing procedures about the relation of the population MDL or SD optimal windows with the optimization of the un-penalized financial loses.

2.3 Limit theory and inference

The interest in this section lies in the asymptotic behavior, as $T \to \infty$, of the following stochastic extremes:

$$M_1(\mathcal{L}_\star, \Delta, \mathbb{P}_T) := \inf_{\ell\in\mathcal{L}_\star} \inf_{\delta\in\Delta} \frac{1}{T}\sum_{t=1}^{T} \ell(X_t - y_t(\delta,\alpha)),$$

$$M_2(\mathcal{L}_\star, \Delta, \mathbb{P}_T) := \inf_{\ell \in \mathcal{L}_\star} \inf_{\delta \in \Delta} \frac{1}{T} \sum_{t=1}^{T} Q_t(\ell, \delta, \min_{0 \leq i \leq t-1} X_i, \max_{0 \leq i \leq t-1} X_i),$$

and

$$M_3(\mathcal{L}_\star, \Delta^2, \mathbb{P}_T) := \inf_{\delta \in \Delta} \sup_{\ell \in \mathcal{L}_\star} (\frac{1}{T} \sum_{t=1}^{T} Q_t(\ell, \delta, \min_{0 \leq i \leq t-1} X_i, \max_{0 \leq i \leq t-1} X_i) - \inf_{\delta \in \Delta} \frac{1}{T} \sum_{t=1}^{T} Q_t(\ell, \delta, \min_{0 \leq i \leq t-1} X_i, \max_{0 \leq i \leq t-1} X_i)).$$

$\mathcal{L}_\star \subseteq \{\ell_{\text{SE}}, \ell_{\text{AE}}, \ell_{\text{W}}, \ell_{\text{sgn}}\}$ is non-empty, and Δ is a non empty subset of $\mathbb{N}^\star$, with the obvious restriction that $\max \Delta \leq T$.

The derivation of the limit theories for the empirical extremals above is considered with a view towards statistical inference. The establishment of rates and limiting distributions can be used in order to construct testing procedures about the population counterparts of methodologies that derive optimal windows, via inversions of confidence regions for the extremes.

Whenever $\mathcal{L}_\star$ is a singleton, M_1 denotes the optimal empirical value associated with the window choice from minimization of the empirical risk associated with the loss function at hand. The optimal value can be used in order to construct confidence regions that could be helpful in testing whether windows obtained from other methodologies do not deviate significantly from being optimal w.r.t. the population risk associated with the particular loss. When $\mathcal{L}_\star$ has more than one elements, M_1 is associated with optimal choice based on the minimal empirical risk across the set of loss functions. This is not generally the same with the MDL extreme, as the penalization constants associated with normalization do not appear there. When the penalization terms are included, then M_2 is obtained; Whenever $\mathcal{L}_\star = \mathcal{L}$, the optimal value obtained from the MDL methodology is recovered. Whenever this equality does not hold, M_2 is interpretable as the optimal value of a restricted (on $\mathcal{L}_\star$) MDL criterion. In an analogous sense, M_3 is associated with the optimal value obtained from a restricted (on $\mathcal{L}_\star$) stochastic dominance criterion, that coincides with the optimal value of the criterion that results to the empirical approximate bound whenever $\mathcal{L}_\star$ is full, i.e. no dominance relations associated with losses that lie in $\mathcal{L} - \mathcal{L}_\star$ are discarded.

The following assumption framework facilitates the derivation of the asymptotic behavior sought:

A.1 The returns' process $(X_t)_{t\in\mathbb{Z}}$ is stationary and absolutely regular, with mixing coefficients $(\beta_k)_{k\in\mathbb{N}}$ that satisfy $\beta_k = O\left(k^{-r}\right)$ for $r > 1$.

A.2 The supremum of the support of the distribution of X_0 is less than one. The cdf of the distribution is continuous at -1, and the distribution has a second logarithmic and a second inverse moment.

A.3 $\#\Delta = O(\sqrt{T})$.

A.4 There exist sequences $a_T, a^\star_T, b_T, b^\star_T$, for which $\frac{\min_{t\leq T} X_t - b_T}{a_T}$, and $\frac{\max_{t\leq T} X_t - b^\star_T}{a^\star_T}$ weakly converge to non degenerate random variables.

The mixing condition in **A.1** holds for strictly stationary versions of ARMA, GARCH-type, and stochastic volatility processes (see for example Basrak et al. (2002) [4]). **A.2** implies first the well-definiteness as a real function of the ℓ_{W} loss $\mathbb{P}$ almost everywhere and its inclusion at the weighted Sobolev spaces used for the derivation of tightness for the empirical processes involved. The cdf continuity at -1 implies that the probability of a zero price is zero, and the moment existence restrictions imply restrictions on the rates of decay of the left tail of the distribution. Given that our application involves daily financial returns the support restriction in **A.2** is considered plausible. It can be extended to a more general compact support assumption, with a potentially greater than or equal to one supremum, if the modified wealth loss $-\ln(x\text{sgn}(y_t(\delta,\alpha))+C)$, for $C>\max(1,\max_{0\leq t\leq T}X_t)$ is considered. If the wealth loss is totally excluded from the analysis then the cdf continuity and the moment existence conditions can also be dropped. In any case, either the present form of the assumption or its aforementioned modification, ensure the $\sqrt{T}$-tightness for the penalization terms that appear in the definitions of M_2 and M_3. The upper bounded support condition-either in its current form, or modified as mentioned above-can be weakened if **A.4** is appropriately strengthened, at the cost of asymptotic negligibility of the non-penalization terms that appear in the aforementioned extremes and potentially unknown rates that depend on the $a_T, a_T^\star, b_T, b_T^\star$ rates that appear in **A.4**. **A.3** controls the rate of growth of Δ. The upper bound support restriction ensures that the parameter space's growth is tame enough so that a maximal inequality on the empirical processes associated with the extremes above is applicable. **A.4** can be established, using Theorem 6.4.15 from Meerschaert and Scheffler (2001) [55], and the proof of Lemma 2.4 of Meerschaert et al. (2013) [56], if the occurrence of an extreme value for a base asset return at some time instance has asymptotically negligible effects on the occurrence of extremes at sufficiently distant epochs. Condition (2.16) in Basrak et al. (2002) [4] verifies this property for processes that satisfy general stochastic recurrence equations, including GARCH-type ones. The strictly stationary versions of the ARMA and stochastic volatility models are also qualified, if the relevant errors are serially iid.

The required limit theory is obtained in the following result; there $\rightsquigarrow$ denotes convergence in distribution, and $\overset{\text{rank}}{\rightsquigarrow}$ denotes convergence in distribution conditionally on the filtration $(\mathcal{F}_T)_{T\in\mathbb{N}^\star}$ where $\mathcal{F}_T$ is generated by the rank statistics $\{\min_{t\leq T} X_t, \max_{t\leq T} X_t\}$:

Theorem 1. *Under **A.1**, **A.2**, **A.3**, and **A.4**, as $T\to\infty$,*

$$\sqrt{T}(M_1(\mathcal{L}_\star,\Delta,\mathbb{P}_T) - M_1(\mathcal{L}_\star,\Delta,\mathbb{P})) \rightsquigarrow \inf_{(\ell,\delta)\in\Gamma_{M_1}} \mathcal{G}_{M_1}(\ell,\delta),$$

$$\sqrt{T}(M_2(\mathcal{L}_\star,\Delta,\mathbb{P}_T) - M_2(\mathcal{L}_\star,\Delta,\mathbb{P})) \overset{\text{rank}}{\rightsquigarrow} \inf_{(\ell,\delta)\in\Gamma_{M_2}} \mathcal{G}_{M_2}(\ell,\delta),$$

$$\sqrt{T}(M_3(\mathcal{L}_\star,\Delta^2,\mathbb{P}_T) - M_3(\mathcal{L}_\star,\Delta^2,\mathbb{P})) \overset{\text{rank}}{\rightsquigarrow} \inf\sup\nolimits_{(\delta,\ell,\delta^\star)\in\Gamma_{M_3}} \mathcal{G}_{M_3}(\ell,\delta,\delta^\star),$$

where, $\mathcal{G}_{M_1}$, $\mathcal{G}_{M_2}$, and $\mathcal{G}_{M_3}$ are zero mean Gaussian processes with covariance kernels:

$$\mathrm{Var}_{\mathcal{G}_{M_1}}((\ell_1,\delta_1),(\ell_2,\delta_2)) := \sum_{t\in\mathbb{Z}} \mathrm{Cov}(\ell_1(X_0 - y_{0,\delta_1}), \ell_2(X_t - y_{t,\delta_2})),$$

$$\mathrm{Var}_{\mathcal{G}_{M_2}}((\ell_1,\delta_1),(\ell_2,\delta_2)) := \sum_{t\in\mathbb{Z}} \mathrm{Cov}(Q_0(\ell_1,\delta_1,\underline{X},\overline{X}), Q_t(\ell_2,\delta_2,\underline{X},\overline{X})),$$

and,

$$\mathrm{Var}_{\mathcal{G}_{M_3}}((\ell_1,\delta_1,\delta_1^\star),(\ell_2,\delta_2,\delta_2^\star)) := \sum_{t\in\mathbb{Z}} \mathrm{Cov}(Q_0^\star(\ell_1,\delta_1,\delta_1^\star,\underline{X},\overline{X}), Q_t^\star(\ell_2,\delta_2,\delta_2^\star,\underline{X},\overline{X})).$$

There, $\underline{X},\overline{X}$ respectively denote the infimum and the supremum of the support of the distribution of X_0, $Q_t^\star(\ell,\delta,\delta^\star,y_\star,y^\star) := Q_t(\ell,\delta,y_\star,y^\star) - Q_t(\ell,\delta^\star,y_\star,y^\star)$, and $\Gamma_{M^\star}\subseteq \mathcal{L}_\star\times\Delta$, $\Gamma_M\subseteq\mathcal{L}_\star\times\Delta$, $\Gamma_N\subseteq\mathcal{L}_\star\times\Delta\times\Delta$ are the sets of optimizers that appear in the definitions of $M^\star$, M and N respectively.

Proof. For each $\delta\in\Delta$, $(y_t(\delta,\alpha))_t$ adheres to **A.1** with the super-harmonic mixing coefficients, as well as to **A.2**. Also, the functions sets involved are bounded subsets of the weighted Sobolev space $H^1_l\left(\mathbb{R}^2, \langle x\rangle^{2+\delta}\right)$, i.e. the semi-normed space $\left\{ f : \mathbb{R}^2 \to \mathbb{R}, \begin{array}{c} \|f\|_{l,2+\delta,\mu} := \\ \left(\int_{\mathbb{R}^2}\left[\left|\frac{f(x)}{(1+\|x\|)^{2+\delta}}\right|^l + \left|D\frac{f(x)}{(1+\|x\|)^{2+\delta}}\right|^l\right]d\mu\right)^{1/l} < +\infty \end{array} \right\}$, where D denotes partial derivation in the sense of distributions, and μ denotes the Lebesgue measure on $\mathbb{R}^2$-see 3.3.2 of Nickl and Potcher (2007) [58], due to **A.2**. In the notation of the aforementioned paper, choosing $l = 2$, $r = 2 + \varepsilon$ and $\gamma = 3 + \delta$, $\beta = 2 + \delta$, $\mathfrak{M}$ the singleton set consisted of the population measure, we have that, due to Corollary 4.2 of Nickl and Potcher (2007) [58], and for large enough T, the bracketing entropy of the functions involved in the definitions of $\mathcal{M}_i$, $i = 1, 2, 3$, for each $\delta \in \Delta$ as a function of $\epsilon > 0$, is universally bounded from above by $c\epsilon^{-1}$ for some universal constant $c > 0$.

Then, from (i) above, (ii) the fact that $\beta_k \sim b^k$, and (iii) the fact that the class has an $L^{2+\varepsilon}(\mathbb{P})$-integrable envelope due to **A.2**, and due to **A.3** and the bounding from above of a finite max by a sum, we get that Theorems 1 and 2 of Doukhan, Massart, and Rio (1995) [21] are applicable. Then the results are obtained from Theorem 2.1 and Lemma B.1 of Fang and Santos (2014) [23]-the lemma can be extended to hold for totally bounded metric spaces, where the metric is not necessarily norm-induced- and then Δ is considered as a closed subset of $\mathbb{N}^\star \cup \omega$ endowed with the metric $d(m, n) = |\frac{1}{m} - \frac{1}{n}|$, where $\frac{1}{\omega} := 0$, and an extension of Lemma 21.19 of van der Vaart (2000) [77] to stationary and strongly mixing sequences, where in the original proof the Lindeberg-Feller CLT for strongly mixing sequences (see for example Theorem 3 in Phillip (1969) [61]) substitutes the classical Lindeberg-Feller CLT. The extension of Lemma 21.19 of van der Vaart (2000) [77] is applicable due to **A.1**, **A.2**, and **A.4**. □

The results are based on limit theorems for empirical processes, as well as the generalized delta method of Fang and Santos (2014) [23], applicable due to the

Hadamard differentiability of the optimization operators involved. For the second and the third cases the convergence is conditionally on the order statistics that appear in the penalization terms; the asymptotic independence between this conditioning information and the random elements obtained via summation that are present in the definitions of the stochastic extremes, along with a Cramer-Wald argument, imply that the results above can be easily re-expressed as a conditional joint convergence for the three extremal processes considered. The limiting distributions are latent; the covariance kernels of the Gaussian processes involved are unknown since they among others depend on latent temporal dependence characteristics of the processes involved. The sets of optimizers involved in the definitions of the limits are also latent, since they among others depend on the unknown population distribution.

Despite the aforementioned latency, Theorem 1 allows for the construction of a feasible inferential procedure based on subsampling in the spirit of Linton et al. (2014) (see also Linton et al. (2005)) that approximates the asymptotic quantiles of the limits that appear in the result. Let $0 < b_T \leq T$, and consider the subsamples from the original observations $(X_j)_{j=t,\ldots t+b_T-1}$ for all $t = 1, 2, \ldots, T - b_T + 1$. For $\alpha \in (0, 1)$, and for any $i = 1, 2, 3$, denote with $q_{i,B_T}(1-\alpha)$ the $1-\alpha$ quantile of the subsample empirical distribution of $\left(\sqrt{b_T}\left(M_i\left(\mathcal{L}_\star, \Delta_\star, \mathbb{P}_{t,b_T}\right) - M_i\left(\mathcal{L}_\star, \Delta_\star, \mathbb{P}_T\right)\right)\right)_{t=1,\ldots,T-b_T+1}$, where $\mathbb{P}_{t,b_T}$ denotes the empirical distribution of $(X_j)_{j=t,\ldots t+b_T-1}$, and $\Delta^\star$ equals Δ when $i = 1, 2$, while it equals Δ^2 when $i = 3$.

Our second result depends on a condition on the elements of Γ_{M_i} that avoids limiting degeneracies and an immediate adaptation of **A.3** to the subsampling rate:

A.3' $\#\Delta = O(\sqrt{b_T})$ while $\max \Delta \leq b_T$.

A.5 For all $i = 1, \ldots, 3$, the variance of $\mathcal{G}_{M_i}$ evaluated at any element of Γ_{M_i} is strictly positive.

The assumption **A.5**, due to a uniform integrability argument enabled by **A.2**, holds whenever $\liminf_{T\to\infty} \inf_{\delta\in\Delta} \mathbb{E}[(\ell(X_0 - y_{0,\delta}))^2] > 0$ for all $\ell \in \mathcal{L}_\star$, which in the present context of stationarity simply holds if (a) X_0 is not degenerate, and (b) the support of $X_0 - X_{-\kappa}$ contains an interval centered at zero for every $\kappa > 0$; when (a) holds, **A.1** implies (b) for large enough κ.

The following result is then obtained:

Theorem 2. *Suppose that **A.1**, **A.2**, **A.3'**, **A.4**, and **A.5** hold, and $b_T \to \infty$, while $\frac{b_T}{T} \to 0$. Then, for all $i = 1, 2, 3$, and as $T \to \infty$,*

$$\lim_{T\to\infty} \mathbb{P}\left[M_i(\mathcal{L}_\star, \Delta_\star, \mathbb{P}) \in \left[M_i\left(\mathcal{L}_\star, \Delta_\star, \mathbb{P}_T\right) \mp \frac{q_{i,B_T}\left(1-\alpha\right)}{\sqrt{T}}\right]\right] = 1 - \alpha. \tag{7}$$

Proof. The result follows from Theorem 1 Proposition 7.3.1 of Politis, Romano and Wolf (1999) [62], as long as the cdf of the distribution of $Q_i := \begin{cases} \inf_{\Gamma_{M_i}} \mathcal{G}_i, i = 1, 2 \\ \inf \sup_{\Gamma_{M_3}} \mathcal{G}_3, i = 3, \end{cases}$ is continuous at its $1-\alpha$ quantile. To this end, and from Lemma 18.15 of van der Vaart (2000) [77], we have that for $\mu, v \in \Gamma_i$ and $\mathcal{G}_{i,\mu}, \mathcal{G}_{i,v}$ the Gaussian process $\mathcal{G}_i$ evaluated there,

$$0 \leq \sigma^2 := \sup_{\Gamma_\mathrm{i}} \mathbb{E}\left[\mathcal{G}_{i,\mu}^2\right] \leq \sup_{\mu,v\in\Gamma_i} \mathbb{E}\left[\left(\mathcal{G}_{i,\mu} - \mathcal{G}_{i,v}\right]^2\right) < +\infty.$$

Hence due to the zero mean function of $\mathcal{G}_{i,\mu}$, and Furnique's inequality (see Relation (1,1) in Samorodnitsky (1991) [69]), we have that for $0 < \varepsilon < 1$, there exists a $\kappa\left(\varepsilon\right)$, such that

$$\mathbb{E}\left[\sup_{\Gamma_\mathrm{i}} \mathcal{G}_{i,\mu}^2\right] = \int_0^{+\infty} \mathbb{P}\left(\sup_{\Gamma_i} |\mathcal{G}_{i,\mu}| > \sqrt{y}\right) dy \leq 2\kappa\left(\varepsilon\right) \int_0^{+\infty} \exp\left(\tfrac{-(1-\varepsilon)}{2\sigma^2} y\right) dy < +\infty.$$

Then, Ch. 2 of Nualart (2006) [59], (see the remark after the proof of Proposition 2.1.11 (p. 109)) implies the existence of the square integrable Malliavin derivative for $\mathcal{G}_{i,\mu}$. Nualart (2006) [59] implies then that the Malliavin derivative of $\mathcal{G}_{\mu}$ equals zero only at trivial triples. The previous along with **A.5** imply the validity of Assumption 1 of Arvanitis, Scaillet and Topaloglou (2019) [2] for $\mathcal{T} = \emptyset$ in their notation. Then, Theorem 1 of Arvanitis, Scaillet and Topaloglou (2019) [2], implies (7). □

The result in (7) allows for asymptotically exact inference for several hypotheses structures; for example consider the optimal δ arising from the MDL criterion, say δ_{MDL} (respectively let δ_{SD} denote the optimal window arising as an approximate bound) obtained via the optimization defined in the definition of $M_2(\mathcal{L}_{\star}, \Delta, \mathbb{P}_T)$ (respectively $M_3(\mathcal{L}_{\star}, \Delta, \mathbb{P}_T)$) above, for $\mathcal{L}_{\star} = \{\ell_{\mathrm{SE}}, \ell_{\mathrm{AE}}, \ell_{\mathrm{W}}, \ell_{\mathrm{sgn}}\}$.

Suppose that the researcher is interested in testing the null hypothesis that every weak accumulation point of the optimal MDL (respectively SD) window is also optimal for the population risk formed by the wealth loss function. This could be motivated by the question of whether the window chosen by minimizing an algorithmic complexity (respectively the stochastic dominance) criterion is "close" to the one obtained by focusing only on financial wealth maximization.

Given $M_1(\{\ell_{\mathrm{W}}\}, \Delta, \mathbb{P}_T)$, a significance level $\alpha \in (0,1)$, and for $i =$ MDL, SD, consider the decision rule that the null is not rejected iff the empirical inclusion $\frac{1}{T}\sum_{t=1}^{T} \ell_{\mathrm{W}}(X_t, y_{t,\delta_i}) \in \left[M_1\left(\{\ell_{\mathrm{W}}\}, \Delta, \mathbb{P}_T\right) \mp \frac{q_{1,B_T}(1-\alpha)}{\sqrt{T}}\right]$ holds. Our final result derives limiting properties of this testing procedure:

Theorem 3. *For* $i =$ MDL, SD*, consider the premises of Theorem 2. Then, the testing procedure that rejects the composite hypothesis that* $\mathbb{E}(\ell_{\mathrm{W}}(X_0, y_{0,\delta_i^{\star}})) = M_1(\{\ell_{\mathrm{W}}\}, \Delta, \mathbb{P})$*, for* $\delta_i^{\star}$ *assuming every value that attain the accumulation points of*

δ_i, iff $\frac{1}{T}\sum_{t=1}^{T}\ell_{\mathrm{W}}(X_t, y_{t,\delta_i}) \notin \left[M_1\left(\{\ell_{\mathrm{W}}\},\Delta,\mathbb{P}_T\right) \mp \frac{q_{1,B_T}(1-\alpha)}{\sqrt{T}}\right]$ is asymptotically exact. If $\delta_i^{\star}$ is unique, then it is also consistent.

Proof. Theorem 1 along with Lusin's Theorem and the results of Georganopoulos (1967) [27] regarding the approximability of continuous real functions on compact metric spaces by Lipschitz continuous functions, imply the asymptotic independence between the weak accumulation points of δ_i and $M_1\left(\{\ell_{\mathrm{W}}\},\Delta,\mathbb{P}_T\right)$. Then the first result follows directly from Theorem 2. When $\delta_i^{\star}$ is unique, then, since due to the proof of Theorem 1, $\frac{1}{T}\sum_{t=1}^{T}\ell_{\mathrm{W}}(X_t, y_t(\delta,\alpha))$ converges uniformly in probability to $\mathbb{E}(\ell_{\mathrm{W}}(X_0, y_{0,\delta}))$, δ_i converges in probability to $\delta_i^{\star}$. If $\mathbb{E}(\ell_{\mathrm{W}}(X_0, y_{0,\delta_i^{\star}})) \neq M_1(\{\ell_{\mathrm{W}}\},\Delta,\mathbb{P})$, then the fact that q_{1,B_T} converges in probability to the $1-\alpha$ quantile of the distribution of $\inf_{(\ell,\delta)\in\Gamma_{M_1}}\mathcal{G}_{M_1}(\ell,\delta)$, implies that with probability converging to one, the null hypothesis gets rejected. □

The testing procedures thus defined are asymptotically exact due to Theorem 1 and an asymptotic independence argument between the aforementioned (potentially stochastic) accumulation points and $M_1\left(\{\ell_{\mathrm{W}}\},\Delta,\mathbb{P}_T\right)$. They are consistent when the δ_{MDL} (respectively the δ_{SD}) has a well defined non stochastic probability limit; uniqueness is not generally expected to hold due to non-convexities arising in the population risks associated with the financial loss functions.

2.4 Discussion

In the above the space of criteria $\mathcal{L}$ and the set of potential forecasts are finite. The parameter α is considered fixed. The analysis is extendable beyond those considerations.

The set $\mathcal{L}$ can be allowed to contain large classes of loss functions. It can for example be augmented with the class of convex loss functions; those are the ones that are uniquely minimized at zero, exhibit convexity and are continuous on the boundary of the support. Given Assumption **A.2** it can be proven that any member of the particular class can be expressed as a sum of convex mixtures of Russell-Seo ramp functions (see Russell and Seo (1989) [66]), one on the positive and one on the negative part of the support, i.e. $\ell(x, y_t(\delta, \alpha)) = \int \mathbb{I}[z < 0] w_-(z - (x - y_t(\delta, \alpha)))_+ dz + \int \mathbb{I}[z \geq 0] w_+((x - y_t(\delta, \alpha)) - z)_+ dz$, with $w_-, w_+ \geq 0$, and Lebesgue integrable to one (see Arvanitis et al. (2021) [2]). This representation can be used in order to facilitate the optimization over $\mathcal{L}$ that lies inside the definitions of δ_{MDL} and δ_{SD}, as well as the derivation of the limit theory of testing procedures like the ones defined in the previous section.

The optimal selection of α from a finite set of potential values is readily feasible. Optimization can be extended on (larger parts of) the convex hull of the finite set of forecast means obtained from each choice of a pair (δ, α), whereas $y_t(\delta, \alpha)$ can be replaced by $\sum_{\delta,\alpha} w_{\delta \in \Delta, \alpha \in A} y_t(\delta, \alpha)$, where A is the set of admissible values for α, and the weights satisfy $w_{\delta,\alpha} \geq 0$, $\sum_{\delta,\alpha} w_{\delta \in \Delta, \alpha \in A}$, i.e. they are elements of the standard simplex of dimension $\#(\Delta \times A)$-1. Thus the optimization w.r.t. δ in the procedures above can be replaced by optimization w.r.t. (a non-empty subset of) $\Delta \times A$, so that optimal convex forecast combinations can be obtained. The present analysis is essentially a restriction of this more general framework to the case where A is singleton and optimization is performed on the set of extremal points of $\Delta \times A$.

The numerical aspects related to the more general optimization procedures implied by the aforementioned extensions can be involved and are left for future research. A taste of the performance of some forecast combinations is obtained in the empirical analysis.

3 An empirical application

We illustrate the theoretical concepts presented above by an empirical application on financial time series of asset returns. We offer a rather complete view on the way the theoretical arguments can be put into practice and showcase the efficacy of the method as an efficacious approach to market timing.

3.1 Data

We choose to work with publicly available data on Exchange Traded Funds (ETFs) that cover a broad spectrum of asset classes. ETFs are highly liquid with minimal transaction costs and available to anyone with access to a brokerage account. We opt for the use of daily data covering two years, 2022 and 2023, that correspond to the downmarket of 2022 and the upmarket of 2023 – both being years of high volatility and sudden direction changes. Figure 1 has the daily price evolution of the S&P500-equivalent ETF, SPY, which shows the main characteristics of these two years, and for comparison we have Figure 2 with the daily price evolution of the Bitcoin.

Besides the SPY ETF we also consider the following asset classes:

1. For emerging markets the EEM ETF,
2. For oil services the OIH ETF,
3. For gold the GLD ETF,

4. For aggregate commodities the DBC ETF,

5. For agricultural commodities the DBA ETF,

6. For small capitalization the leveraged TNA ETF,

7. For cryptocurrencies the price of Bitcoin, BTC.[5]

All data are open-sourced from the Yahoo! Finance website. The descriptive statistics of the daily percent returns are presented in Table 1 and their associated correlations in Table 2.

We can observe that we have a variety of volatility levels and considerable fluctuations in the average and total returns across the ETFs under examination. For example, the leveraged ETF, TNA, has annualized volatility of 74.2%, an almost identical maximum drawdown of -74.9% and a total loss of -52.6%; in comparison the market ETF, SPY, has an annualized volatility of 19.9%, a maximum drawdown of -24.5% and a total gain of 3.24%, while the oil services ETF, OIH, has an annualized volatility of 41.7%, a maximum drawdown of -36.4% and a total gain of 71.5%. So, there is plenty of variation across asset classes and descriptive metrics to provide us with different information as input to our empirical analysis.

The correlations among the ETFs also illustrate the different informational content of their returns. We can see that the Bitcoin, for example, has minimal correlation with the agricultural and commodity index ETFs, DBA and DBC, and also with the gold ETF, GLD, and has maximal correlation of 44.5% with the market ETF, SPY. The latter in turn has minimal correlation with the agricultural ETF, DBA, and

[5] We expect that the limit theories and the inferential procedures investigated in Section 2.3 are not directly applicable to BTC as its price dynamics show strong indications of local mild explocivity and thereby of implausibility for **A.1**-see the analysis in Anyfantaki et al. 2021 [1].

with the gold ETF, GLD, and maximal correlation wit of 69.8% with the emerging markets ETF, EEM. Similar observations can be made and for the other correlations among these ETFs, for we cannot observe correlations of more than 50% but only in four pairs (SPY, EEM), (SPY, TNA), (EEM, TNA) and (DBC, OIH) - these are quite high but this is a fairly anticipated result.

3.2 Implementation and discussion of results

The empirical implementation of our suggested methodology needs the following required inputs: the number of required windows $\delta \in \Delta$, the exponent parameter $\alpha > 0$ and the number of observations $n \leq T$ over which one evaluates the loss and predictive complexity functions. The choice of the rolling window δ clearly does not coincide with the observations n used for this evaluation, and we explain more of this below. We proceed as follows:

- Out of the total $T = 501$ observations (the maximum available over the sample period 2022-2023 for the SPY ETF) we construct our Δ space as in $\delta \in \Delta = \left\{2, 3, \ldots, \sqrt{T}\right\}$ so that the maximum δ is $\sqrt{T}$ and the $\#\Delta = \sqrt{T} - 1$.
- Given a parameter value of $\alpha \in A = \{0, 0.5, 2, 5, 7\}$, we then compute all required quantities for all the loss functions and predictive complexities, over all rolling windows in Δ, for all observations $t = \sqrt{T} + 1, \sqrt{T} + 2, \ldots, T$, and where the evaluation of the loss functions and predictive complexities is done either recursively (as T increases) or over a local window of $n < T$ observations; we used $n = \{2, 5, 7\}$ local observations.
- In our empirical analysis we include one more, complexity-based, order selection criterion that is simpler and based on past work by Rissanen (1987), denoted

by ℓ_{MDL} in the tables and termed the "plain" MDL.

- Having computed the loss functions and the predictive complexities, we then proceed to compute, for all periods t, the optimal rolling windows for: (a) each of the 5 individual loss functions and predictive complexities, denoted by $(\ell_{\mathrm{SE}}, \ell_{\mathrm{AE}}, \ell_{\mathrm{W}}, \ell_{\mathrm{sgn}}, \ell_{\mathrm{MDL}})$ in the tables (b) for the combination of the two statistical loss functions in the context of the SD approach, denoted by $\mathcal{L}_{\mathrm{SE,AE}}$ in the tables (c) for the combination of the two financial loss functions in the context of the SD approach, denoted by $\mathcal{L}_{\mathrm{W,sgn}}$ in the tables (d) for the combination of all loss functions in the context of the SD approach, denoted by $\mathcal{L}_{\mathrm{A,SD}}$ in the tables; (e) for the combination of all loss functions in the context of the MDL approach, denoted by $\mathcal{L}_{\mathrm{A,MDL}}$ in the tables; and, finally, (f) from the average window across all individual loss functions, denoted by $\bar{\delta}_A$ in the tables.

- Using the optimal rolling windows from the previous step (computed using information only up to period t) we then select the optimal forecasts (for the next period $t+1$, so no hindsight is involved), according to each loss function or combination thereof, trade the sign of the optimal forecast and store the corresponding optimal forecast error.

- Finally, we evaluate the statistical performance of our optimally selected forecasts and the financial performance of the resulting trading returns; for the latter we also report the significance of our findings based on the results of Theorem 3.

- We tabulate and report the best performing loss function and predictive complexity, or combinations of them, for each possible combination of (α, n), for each ETF, and then proceed to aggregate the distribution of the loss functions according to each performance attribution criterion.

Based on this set-up we will provide some tentative answers to the following questions:

Q1. Does the suggested methodology "work", in the sense of providing us with financial performance over and above the passive benchmark? And are the potential excess returns significant, in the sense of Theorem 3?

Q2. Which loss function or combination thereof works better most of the time?

Q3. Is the "optimal" loss function dependent on the performance attribution measure?

Q4. Should we be using the suggested methodology for any performance attribution measure?

Q5. Do the results relate to the properties of the data on which the methodology was applied?

Our results are summarized in 15 Tables in total, which are grouped according to the attribution measure used to evaluate performance. We thus start with Tables 3, 4 and 5 that refer to results based on total excess returns. The excess returns refer to the difference between the active trading returns based on the forecasts of our methodology minus the passive returns of the buy & hold strategy of the corresponding ETF. A star, *, indicates that we reject the corresponding hypothesis of Theorem 3, thus indicating significance of the corresponding entry.[6] Table 3 has half

[6] A significant excess return implies that the method selected optimal local windows that out-

our ETFs under examination and Table 4 the other half. The tables present the best performing loss function and next the total excess return. We start with Table 3 which has the results for BTC, DBA, DBC and EEM. As a reminder, see Table 1, these ETFs have the following characteristics: BTC has maximum drawdown of -66.8%, annualized volatility of 57.2% (second largest below TNA) and negative total return, DBA has maximum drawdown of -15.9%, annualized volatility of 13.6% (the smallest across ETFs) and positive total return, DBC has maximumd drawdown of -27.3%, annualized voaltility of 22.1% and positive total return (equivalent to that of DBA) and, finally, EEM has maximum drawdown of -32.7%, annualized volatility of 20.3% and negative total return. A casual look at Table 3 reveals that the results indicate that in all cases considered (across combinations of (α, n)), except in just 4 (equivalent to only 4% in the table), we have positive excess returns. Different assets have different distribution of good performance across the (α, n) combinations: for example, looking at BTC we find that the top performance with an excess return of 181.8% is given by the stochastic dominance methodology with all loss functions being included in the selection of the optimal window and a recursive evaluation of the loss function with the plain sample mean as the forecasting function. However, this is not a unique occurrence for we can find similar performance across different combinations like the use of the stochastic dominance methodology using the statistical loss functions and $n = 7$-period evaluation with a total excess return of 170/9%, and the sample mean as the forecasting function. Or, you can see the use of the wealth loss function with an $\alpha = 7$ weighted mean as the forecasting function, an $n = 5$-period evaluation and a total excess return of 169.3%.

performed the ex-post optimal individual window; a non-significant excess return implies that the method has not performed *statistically* better than the ex-post optimal local window but it might have performed better it financially.

This is illustrative of the argument that we made in the introduction of the paper, on the simplicity of sign forecasting of returns, but also on the importance of having a method that selects the local window successfully. For the other assets in the table what we can observe is that their top performing loss functions are concentrated for larger values of the α-parameter and smaller values of the n-parameter: for DBA the top excess return performance is 31.8% with a combination of $\alpha, n) = (7,7)$ and the optimal local window being obtained from the use of all loss functions and the MDL criterion and a close second with a total excess return of 30.0% with a combination of $(7,5)$ and the use of the stochastic dominance criterion using the financial loss functions. The magnitude of excess returns does not necessarily matter for the interpretation of our results, not so much as the parametrization being used and the loss functions that perform at the top. For the DBC and the EEM ETFs the top performance is obtained by statistical loss functions (SE loss at the top position, AE loss at the second top position) for DBC and by stochastic dominance loss functions (financial loss at the top position, statistical loss at the second top position) for EEM. From the results of this Table what we can say conclusively is that selecting the local window via the suggested methodology works almost all the time but we cannot conclude as to a "globaly optimal" loss function across all ETFs; it does show, however, that there is *at least one* loss functions from each of the considered groups that will provide positive out-performance.

The results are similar if we look into Table 4, with some notable exceptions. First, we see that now we have more instances where none of the methods will provide positive excess returns, in particular for the oil services ETF, OIH, and mostly for the market ETF, SPY. For this latter ETF we have that in 15 out of 24 combinations the excess return is negative, but still we can get positive excess returns of up to 38.9%

and 22.0%, as we can see at the lower panel of the table. Second, it appears that volatility is helping the method perform quite well. If one looks at the OIH and TNA ETFs (which have the third and first largest annualized volatilities of 41.7% and 74.2%) will easily see that there are notable excess returns to be made by different combinations of the (α, n) parameters. What makes these two ETFs important for our analysis, besides their large volatility, is that they have the largest and smallest passive total return among all ETFs being examined (with 71.5% for OIH and -52.63% for TNA respectively). Therefore, to achieve overperformance on both instances using a optimally-selected local window is of considerable theoretical and practical interest. We note that for the OIH ETF the top two performing parameter combinations are as follows: for OIH we have $(\alpha, n) = (7, 7)$ using stochastic dominance with the two financial loss functions and a total excess return of 52.8%, with the second best performance of 47.3% being obtained with the same loss functions and a similar combination of parameters at $(7, 5)$; for TNA we have $(\alpha, n) = (0, 2)$ and a total excess return of 221.9% and the use of stochastic dominance with the two statistical loss functions, with the second best performance being given by the $(7, 5)$ parameter combination, again with the same loss functions, and a total excess return of 181.1%.

In actual practice getting the best or the second best performing loss function is not an event of very high probability. Thus, it is imperative to examine further the average, across assets and loss functions, performance of our suggested methodology and for this we turn now to Table 5. There we have tabulated the distribution of all loss functions across all ETFs and some interesting, and practical results, can be obtained from it. First, it is very clear that the most frequently top performing loss function is the most standard one, the square loss; it appears in Tables 3 and 4 with a frequency of 31%, one third thus of the time one can obtain top performance using this loss function. The next top performing loss function is the "plain" MDL (which in essence is also a square loss with a different penalty function) with a participation of 17.7% in the two tables, and the third best performing loss is the stochastic dominance combination of the two statistical loss functions with a participation of 9.90%. These three loss functions account for almost 60% of the top performing loss functions in Tables 3 and 4. If we next find the average excess return for each of these loss functions (across all 8 $\times$24 instances of combinations) we find them to be 5.46% for the square loss, 3.74% for the "plain" MDL loss and 3.79% for the stochastic dominace with the statistical loss functions. It is furthermore interesting to note that the next best average excess return is given by the stochastic dominance with all loss functions at 3.07%, although this loss function accounts for only 5.7% in the total participation of loss functions. If we concentrate only on the instances in which these loss functions appeared in the tables to compute their average excess return (i.e., not in all 8$\times$24 combinations) then a different result emerges, which is of equal practical significance: the order of average performance reverses. Now, the average excess return for the stochastic dominance with all loss functions is 53.6%, for stochastic dominance with the statistical loss functions is 38.3%, for the "plain" MDL loss function is 21.1% and for the square loss is 17.8%. Thus, there is a trade-off between performance magnitude and performance occurrence, with the less frequently appearing, as top performer, stochastic dominance approach to provide higher excess returns.[7]

We conclude our discussion of Tables 3 to 5 by addressing the issue of significance, per the notion of the test of Theorem 3. As previously explained, the theorem

[7] Note, however, that in this statement we are not taking into account how many other times the stochastic dominance approach has appeared in the totality of results, available on request. This implies that the actual average return of the stochastic dominance approach will be of necessity smaller than the 53.6% appearing here.

provides for a way of assessing whether the method-based choice of the optimal local window outperforms *statistically* the optimal ex-post individual window. Thus, the advantages of having such a significance test are twofold: first, we can locate (and possibly devise some cross-validation procedure) the parametrizations that lead to consistent statistical significance (that has to be combined with financial excess returns); second, we can relate the comparisons of significant parametrizations to non-significant ones (but with financial over-performance in terms of total excess returns) to the characteristics of the asset under study. Scanning again Tables 3 and 4 we first note that in all ETFs examined either all top 3 performing loss functions or at least one of them give rise to total excess returns that are statistically significant: that implies that our suggested method did outperform the ex-post optimal individual window. We note in particular that the two most volatile assets, the Bitcoin and TNA, are the ones that have instances of parametrizations that provide considerable total excess returns but these are not statistically significant, in the sense of Theorem 3. If we next concentrate on where the bulk of significant results lies, in terms of parametrizations, and we count the significant entries per α-block in the two tables we find that the extreme values of α i.e., 0 for the plain sample mean and 7 for the weighted sample mean, we have the highest concentration of significant total excess returns and, in fact, for $\alpha = 7$ we have 24 out of 32 of n combinations to provide us with significant total excess returns, followed by $\alpha = 0$ where we have 18 out of 32 such combinations. This result is of potential theoretical and practical usefulness: one the one hand we can further explore the impact of joint optimizations of the (α, δ) pair, and on the other hand one can device additional benchmarking options, where the case of the plain sample mean with $\alpha = 0$ serves as a yardstick for performance comparisons.

We next turn to Tables 6 through 11 that contain the evaluation results for the statistical measures, the MAE and the MSE. We first note that there is no real benchmark for comparison here, as we are using the simplest possible method for forecasting, the sample mean. Direct comparisons are thus possible between the plain sample mean, $\alpha = 0$, and the weighted sample mean, $\alpha > 0$, from the tables. Starting off with Tables 6 and 7 we immediately note that the MAE does not vary considerable within each ETF but only across ETFs: we have significant performance equivalence based on this attribution measure. However, what is also immediately eminent is that the plain sample mean leads to best MAE-based statistical performance overall and this is true whether we look into Table 6 or Table 7. This is a significant finding for practical purposes because it indicates that top statistical performance and top financial performance are not to be obtained via the same "model" or method. It also implies that the objective loss function for active trading and the the objective loss function for performance evaluation need not be the same - remember that we found in Table 5 how the statistical loss functions can bring performance enhancements in terms of financial performance. If we look into Table 8 we find another critical result: almost 66% of the top performing loss function is the "plain" MDL one, with the stochastic dominance with statistical loss functions being top performing with a 17.2% participation rate and the average local window across all individual loss functions covering the rest 15.1% across all loss functions and ETFs. What we can meaningfully get from this table is that the stochastic dominance approach with statistical loss functions has a participation of 10% or more in both the financial evaluation criterion of total excess returns and in the statistical evaluation criterion of the MAE.

In Tables 9 and 10, where the evaluation criterion is the MSE, the results are almost completely in line with Tables 6 and 7 - a non-surprising result. We find that, again, the plain sample mean, $\alpha = 0$, gives overall the best performance - although we note that the differences with the weighted sample mean, $\alpha > 0$, are now much less pronounced that they were before with the MAE. This result can be interpreted in two different ways: one would be that the MSE is not sensitive enough to differentiate model performance (the well-known scaling issue, taking the square root of the MSE will change the scaling of the results), and the other way would be that the MSE is a "robust" evaluation method when one sees it along with the results of Tables 3 to 5 (i.e., that the MSE gives top financial performance most of the time). The final interpretation rests with the researcher and the objective one has in mind. Finally, the results of Table 11 are also in line with those in Table 8, with the "plain" MDL being top performing loss function 48.4% of the time, followed by the average local window across individual loss functions with a participation of 16.7% and the stochastic dominance with statistical loss funcitons with a participation of 12.5% - here it is interest to note that non-zero participation rates as top performing loss functions have other ones, like the MDL with all loss functions with 7.3%, stochastic dominance with financial loss functions with 5.2% and stochastic dominance with all loss functions with 4.2%. In our view these latter results support those found in Tables 3 through 5 and make for a practical recommendation on the use of those loss functions as discussed in Table 5, for they practically coincide with those presented in Table 8.[8]

We finally turn to our last performance attribution measure, the sign success ratio which counts the percentage of times that the forecasts captured accurately the sign of next period's returns. This is an important performance measure because getting

[8] It might be claimed that in this respect the squared error is robust both as a loss function and as an evaluation measure.

the sign of next period right is important for active trading. The results in Tables 12 through 14 show higher variation compared to the statistical evaluation measures but lower compared to the excess returns evaluation measure. What is important to point out in this context is that the results are now more “scattered” across different values of α and one can find the same or very similar SSRs within each ETF in both Table 12 and Table 13. Furthermore, it is also evident that now we have different loss functions as top performers. Starting with the comparison on the differentiation of the SSRs we see that there is a tendency to have top performance for both $\alpha = 0$ and for $\alpha = 7$, with the in-between results being in general less good compared to these extremes. Second, it is also evident that now the stochastic dominance and MDL loss functions appear much more frequently in the tables than they were before in the statistical evaluation measures - and this recommends a further comparison of the results in Tables 3 and 4 with the current Tables 12 and 13. Third, we see that the SSRs do not deviate significantly from the 50% benchmark but nevertheless the financial performance is very good; this suggests that the method is able to potentially capture the signs of critical periods (either in upmarket or downmarket conditions) and thus all that is required is that one gets just a numerically different from 50% SSR. Turning next to Table 14 we find the same variety of top performing loss functions as we did find in Table 5. Here the results are as follows: the “plain” MDL is the top participating loss function with a rate of 25.0% closely followed by the square loss function with a participation rate of 19.3% and then by the stochastic dominance with statistical loss functions with 12.5% participation rate. If note is that the full MDL also participates as a top loss function with a rate of 7.8% followed by the average local window across individual loss function with a rate of 7.3%, then followed by stochastic dominance with financial loss functions with 6.8% and stochastic dominance with all loss functions with a participation rate of 5.7%. Again we see the same loss functions as we did in Table 3, making an even stronger case for the use of financial loss functions for performance attribution with a variety of the more advanced loss functions for local window selection and forecasting.

The combined results of all tables prompt us to make one final comparison using the results of Tables 3, 4, 12 and 14, so as to better understand which loss functions jointly appear the most under the two financial performance attribution measures. We do this in Table 15. In the table we first list the top 3 performing loss functions based on total excess return attribution and then we list the corresponding loss functions for the same parametrizations but that appeared in the sign success ratio attribution. The results of the table are not entirely conclusive (for they are clearly dependent on the particular dataset that we have used) but they are highly suggestive: for each ETF examined there are at least two (one for SPY) loss functions that are top performers which are the same in the total excess returns attribution and sign success ratio attribution. Among the loss functions appearing in the table we have that the square loss function ℓ_{SE} appears 4 times and the stochastic dominance with the financial loss functions $\mathcal{L}_{\mathrm{W,sgn}}$ also 4 times, with the stochastic dominance with the statistical loss functions appearing 3 times. Thus, from the 24 top 3 performing loss functions we have that 15 of them match i.e., 62.5%, and we might make the modest claim that our methodology works fairly robust across the two financial performance attribution measures. Furthermore, and this of practical importance, if one were to look at the top 3 performing loss functions based on the sign success ratio attribution would *not* obtain best total excess return performance: it is not just *how many* correct signs you forecast that matters that will lead to financial performance, it is the combination of a loss function that combines timing (of market turning points) and sign predictability at the same time.

We end our discussion by reviewing and summarizing our answers on the questions posed at the beginning of this section - with the understanding that our results are representative and suggestive for a broader class of asset returns. Thus, our suggested methodology readily works on providing a theoretically grounded, coherent way of selecting the optimal local window to use in generating forecasts. It works by providing a very agreeable blend of financial performance across a broad range of asset classes, as represented by the corresponding ETFs we used, and that financial performance can be obtained by a number of different loss functions. Our results allow us to conjecture that the standard square loss function will work most of the time but it will not necessarily give optimal financial performance; the latter is obtained by the more "advanced" combinations of loss functions either by the stochastic dominance principle or by the MDL principle. We find that the combination of the two statistical loss functions and the combination of the two financial loss functions, and also all four of them, provides a very solid approach for local window selection that leads to very good financial performance. This performance is, in about 40% of our total parametrizations, statistically significant, thus providing positive total returns that outperform the ex-post optimal local window. We further illustrated that the two financial performance attribution measures, the total excess returns and the sign success ratio, give consistent selection of the loss function compared to the statistical performance attribution measures; for the latter we find that not only the optimal selections are for different loss functions, but also that they do not allow for sufficient separation across parametrizations. Finally, our results are fairly consistent across all ETFs we have examined. We propose some further suggestions for future research in the concluding remarks that follow.

4 Concluding remarks

We propose a new methodology for selecting the optimal local window of estimation for forecasting asset returns. At this stage of our research our results are confined to the simplest of predictors, the (weighted) sample mean, and we fully develop the theory and implementation of the method. Our approach is based on a notion of predictive complexity, whose minimization results in the optimal selection of the local window. We work within the confines of five individual loss functions and their combinations based on the stochastic dominance and the minimum description length principles and derive explicit formulas that are of immediate applicability and easy to use. We implement the derived theoretical results into an empirical application using several exchange trade funds across different asset classes. Our results indicate that the method is commendable for practical work as it provides accurate estimation of the local window, as judged by the derived financial performance: we can generate excess returns (compared to the passive benchmark) that are found to be (in most tested instances) “significant” according to the inference theory developed for the method.

The implications of our results are multi-fold. First, the current research illustrates that market timing is possible even with the simplest of predictors, if the information window is chosen appropriately and “objectively”. Second, it provides for expanding our analysis to more complex predictors, that can be built gradually over the simple predictor considered here, and thus examine the need and requirements for additional complexity in forecasting asset returns. Third, it opens-up comparisons with the first part of the literature in the introduction i.e., prediction with expert advice; this is of particular practical usefulness as both our method and the prediction with expert advice work on the same foundational premise: that there is a range of possible combinations of information/models/parameters each of which might be appropriate to use in prediction. Fourth, we can further examine the potential of our method to an expanded asset universe and data frequencies. We are pursuing some of the above in our continued research on this topic.

The deeper understanding of the possible conflations between the notions of statistical/predictive complexity of Rissanen (see for example [64]-[65]) with the stochastic dominance paradigm provides an interesting path of research that may be proven fruitful for some of the extensions above; predictive complexity interprets the exponentially integrable signed loss functions as densities, thus providing with parametric statistical models and penalized likelihoods that approximate the algorithmic complexity of the data. Stochastic dominance interprets the loss functions as preferences towards risk of erroneous predictions (or of economic loss due to erroneous predictions). The first methodology seeks to minimize complexity over the class of statistical models involved; the second seeks to do so by also taking into account the risk of misspecified preferences in a conservative manner. The derivation of theoretical results investigating the relations between the optimal choices of each approach as functions of the loss function classes involved could thereby be another interesting path of further research.

References

[1] Anyfantaki, S., Arvanitis, S. and Topaloglou, N., 2021. Diversification benefits in the cryptocurrency market under mild explosivity. European Journal of Operational Research, 295(1), pp.378-393.

[2] Arvanitis, S., Post, T., & Topaloglou, N., 2021. Stochastic Bounds for reference sets in portfolio analysis. Management Science, 67(12), 7737-7754.

[3] Arvanitis, S., Post, T., Potì, V., & Karabati, S., 2021. Nonparametric tests for optimal predictive ability. International Journal of Forecasting, 37(2), 881-898.

[4] Basrak B, Davis RA, Mikosch T, 2002. Regular variation of GARCH processes. Stochastic Processes Their Appl. 99(1):95–115.

[5] Beebower, G. L., & Varikooty, A. P., 1991. Measuring Market Timing Strategies. Financial Analysts Journal, 47(6), 78-84 & 92.

[6] Bialek, W., Nemenman, I. and Tishby, N., 2001. Predictability, complexity, and learning. Neural computation, 13(11), pp.2409-2463.

[7] Bollen, N., & Busse, J., 2001. On the Timing Ability of Mutual Fund Managers. Journal of Finance, 56, 1075- 1094.

[8] Brennan, M.J. and Taylor, A.P., 2023. Expected returns and risk in the stock market. Journal of Empirical Finance, 72, pp.276-300.

[9] Busse, J., 1999. Volatility Timing in Mutual Funds: Evidence from Daily Returns. Review of Financial Studies, 12, 1009-1041.

[10] Časta, M., 2023. Inflation, interest rates and the predictability of stock returns. Finance Research Letters, 58, p.104380.

[11] Chen, J., Ma, F., Qiu, X. and Li, T., 2023. The role of categorical EPU indices in predicting stock-market returns. International Review of Economics & Finance, 87, pp.365-378.

[12] Chernov, A., Kalnishkan, Y., Zhdanov, F. and Vovk, V., 2010. Supermartingales in prediction with expert advice. Theoretical Computer Science, 411(29-30), pp.2647-2669.

[13] Christoffersen, P.F. and Diebold, F.X., 2006. Financial asset returns, direction-of-change forecasting, and volatility dynamics. Management Science, 52(8), pp.1273-1287.

[14] Cowles, A., 1934. Can stock market forecasters forecast?. Econometrica, 1, 309-24.

[15] Cumby, R., & Modest, D., 1987. Testing for market timing ability: A framework for forecast evaluation. Journal of Financial Economics, 19(1), 169-189.

[16] Cotter, J., Eyiah-Donkor, E. and Potì, V., 2023. Commodity futures return predictability and intertemporal asset pricing. Journal of Commodity Markets, 31, p.100289.

[17] Dai, Z., Dong, X., Kang, J. and Hong, L., 2020. Forecasting stock market returns: New technical indicators and two-step economic constraint method. The North American Journal of Economics and Finance, 53, p.101216.

[18] Dichtl, H., 2020. Forecasting excess returns of the gold market: Can we learn from stock market predictions?. Journal of Commodity Markets, 19, p.100106.

[19] Ding, J., Jiang, L., Liu, X., & Peng, L., 2023. Non parametric tests for market timing ability using daily mutual fund returns. Journal of Economic Dynamics and Control, 150, 104635.

[20] Droms, W., 1989. Market Timing as an Investment Policy. Financial Analysts Journal, 45, 73-77.

[21] Doukhan, P., Massart, P., and Rio, E., 1995. Invariance Principles for Absolutely Regular Empirical Processes. Annales de l'IHP Probabilités et Statistiques 31(2), 393-427.

[22] Elton, E. J., Gruber, M. J., & Blake, C. R., 2012. An examination of mutual fund timing ability using monthly holdings data. Review of Finance, 16(3), 619-645.

[23] Fang, Z., and Santos, A., 2014. Inference on Directionally Differentiable Functions. arXiv preprint arXiv:1404.3763.

[24] Ferson, W., & Mo, H., 2016. Performance measurement with selectivity, market and volatility timing. Journal of Financial Economics, 121(1), 93-110.

[25] Fishburn, P. C., 1976. Continua of stochastic dominance relations for bounded probability distributions, Journal of Mathematical Economics, 3(3), 295-311.

[26] Gayant, J. P., & Le Pape, N., 2017. Increasing degree inequality, Journal of Mathematical Economics, 70, 185-189.

[27] Georganopoulos. G, 1967. Sur l' approximation des fonctions continues par des fonctions Lipschitziennes. Comptes Rendus Hebdomadiaires des Seances de l Academie des Sciences Serie A, 264(7), p.319.

[28] Ghosh, M. and Nandakumar, S., 2012. Predictive complexity and generalized entropy rate of stationary ergodic processes. In International Conference on Algorithmic Learning Theory (pp. 365-379). Berlin, Heidelberg: Springer Berlin Heidelberg Vancouver.

[29] Guerard, J., Thomakos, D., Kyriazi, F., & Mamais, K. 2023. On the Predictability of the DJIA and S&P500 Indices. forthcoming at Wilmott Magazine, Ltd.

[30] Haase, F. and Neuenkirch, M., 2023. Predictability of bull and bear markets: A new look at forecasting stock market regimes (and returns) in the US. International Journal of Forecasting, 39(2), pp.587-605.

[31] Henriksson R. D., & Merton, R. C., 1981. On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills. The Journal of Business. 54(4), 513-533.

[32] Henriksson R. D., 1984. Market Timing and Mutual Fund Performance: An Empirical Investigation. The Journal of Business. 57(1), 73-96.

[33] Huang, Y., Ma, F., Bouri, E. and Huang, D., 2023. A comprehensive investigation on the predictive power of economic policy uncertainty from non-US countries for US stock market returns. International Review of Financial Analysis, 87, p.102656.

[34] Jeffrey R.H., 1984. The Folly of Stock Market Timing. Harvard Business Review, 62, 102-110.

[35] Jiang, G., Yao, T., & Yu, T., 2007. Do Mutual Funds Time the Market? Evidence from Portfolio Holdings. Journal of Financial Economics, 86(3), 724-758.

[36] Jiang, W., 2003. A nonparametric test of market timing. Journal of Empirical Finance, 10, 399-425.

[37] Jin, S., Corradi, V. and Swanson, N.R., 2017. Robust forecast comparison. Econometric Theory, 33(6), pp.1306-1351.

[38] Kalnishkan, Y. and Vyugin, M.V., 2002. On the absence of predictive complexity for some games. In Algorithmic Learning Theory: 13th International Conference, ALT 2002 Lübeck, Germany, November 24–26, 2002 Proceedings 13 (pp. 164-172). Springer Berlin Heidelberg.

[39] Kalnishkan, Y., Vovk, V. and Vyugin, M.V., 2004. A criterion for the existence of predictive complexity for binary games. In Algorithmic Learning Theory: 15th International Conference, ALT 2004, Padova, Italy, October 2-5, 2004. Proceedings 15 (pp. 249-263). Springer Berlin Heidelberg.

[40] Kalnishkan, Y., 2022. Prediction with expert advice for a finite number of experts: A practical introduction. Pattern Recognition, 126, p.108557.

[41] Kroll, Y., & Levy, H., 1980. Stochastic Dominance Criteria: A Review and Some New Evidence, Research in Finance, Vol. II, Greenwich: JAI Press, pp. 263–227

[42] Kyriazi, F., & Thomakos, D., 2020. Foreign Exchange Rate Predictability: Seek and Ye Shall Find It. Handbook of Applied Investment Research, World Scientific Publishers, Chapter 19, 511-556.

[43] Lam, K. and Li, W., 2004. Is the 'perfect'timing strategy truly perfect?. Review of Quantitative Finance and Accounting, 22, pp.39-51.

[44] Larsen, G., A., & Wozniak, G., D., 1995. Market timing can work in the real world. The Journal of Portfolio Management, 21(3), 74-81.

[45] Levy, H., 1992. Stochastic Dominance and Expected Utility: Survey and Analysis, Management Science, 38, 555–593.

[46] Levy, H., 2015. Stochastic dominance: Investment decision making under uncertainty, Springer.

[47] M. Levy & Levy, H., 2002. Prospect Theory: Much Ado about Nothing?, Management Science 48, 1334-1349.

[48] Li, Y., Liang, C. and Huynh, T.L.D., 2022. Forecasting US stock market returns by the aggressive stock-selection opportunity. Finance Research Letters, 50, p.103323.

[49] Liang, C., Xu, Y., Wang, J. and Yang, M., 2022. Whether dimensionality reduction techniques can improve the ability of sentiment proxies to predict stock market returns. International Review of Financial Analysis, 82, p.102169.

[50] Li, W. and Lam, K., 2002. Optimal market timing strategies under transaction costs. Omega, 30(2), pp.97-108.

[51] Li, T. and Sun, X., 2023. Predicting stock market returns using aggregate credit risk. International Review of Economics & Finance, 88, pp.1087-1103.

[52] Lv, W. and Qi, J., 2022. Stock market return predictability: A combination forecast perspective. International Review of Financial Analysis, 84, p.102376.

[53] Matallín-Sáez, J.C., 2015. A note on market timing: Interim trading and the performance of holdings-based and return-based measures. International Review of Economics & Finance, 35, pp.90-99.

[54] McFadden, D., 1989. Testing for Stochastic Dominance. in Studies in the Economics of Uncertainty, eds. T. Fomby and T. Seo, New York: Springer- Verlag, pp. 113–134.

[55] Meerschaert MM, Scheffler HP, 2001. Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice (Wiley, New York).

[56] Meerschaert MM, Scheffler HP, Stoev SA. 2013. Extreme value theory with operator norming. Extremes 16(4):407–428.

[57] Mosler, K., & Scarsini, M., 1993. Stochastic Orders and Applications, a Classified Bibliography, Berlin: Springer-Verlag.

[58] Nickl, R., and Pötscher, B. M., 2007. Bracketing Metric Entropy Rates and Empirical Central Limit Theorems for Function Classes of Besov-and Sobolev-type. Journal of Theoretical Probability, 20(2), 177-199.

[59] Nualart, D., 2006. The Malliavin calculus and related topics (Vol. 1995, p. 317). Berlin: Springer.

[60] Pesaran, H., & Timmermann, A., (1994). A generalization of the non-parametric Henriksson-Merton test of market timing. *Economics Letters*, *44(1-2)*, 1-7.

[61] Philipp, W., 1969. The central limit problem for mixing sequences of random variables. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 12(2), pp.155-171.

[62] Politis, D. N., Romano, J. P., and Wolf, M., 1999. Subsampling. Springer Science and Business Media.

[63] Post, T., Karabatı, S. and Arvanitis, S., 2019. Robust optimization of forecast combinations. International Journal of Forecasting, 35(3), pp.910-926.

[64] Rissanen, J., 1986. Stochastic complexity and modeling. The annals of statistics, pp.1080-1100.

[65] Rissanen, J., 1997. Stochastic complexity in learning. Journal of computer and system sciences, 55(1), pp.89-95.

[66] Russell, W.R. and T.K. Seo, 1989. Representative sets for stochastic dominance rules, in: Studies in the economics of uncertainty: In honor of Josef Hadar, Springer-Verlag.

[67] Qiu, R., Liu, J. and Li, Y., 2023. Long-term adjusted volatility: Powerful capability in forecasting stock market returns. International Review of Financial Analysis, 86, p.102530.

[68] Sáenz, J.V., Quiroga, F.M. and Bariviera, A.F., 2023. Data vs. information: Using clustering techniques to enhance stock returns forecasting. International Review of Financial Analysis, 88, p.102657.

[69] Samorodnitsky, G., 1991. Probability Tails of Gaussian Extrema. Stochastic Processes and their Applications 38(1), 55-84.

[70] Schizas, P., & Thomakos, D., 2015. Market timing and trading strategies using asset rotation: non-neutral market positioning for exploiting arbitrage opportunities. Quantitative Finance, 15(2), 285-298.

[71] Shafer, G. and Vovk, V., 2019. Game-theoretic foundations for probability and finance (Vol. 455). John Wiley & Sons.

[72] Sharpe W.F., 1975. Likely Gains from Market Timing. Financial Analysts Journal, 31, 60-69.

[73] Shilling, A. G., 1992. Market timing: better than a buy-and-hold strategy. Financial Analysts Journal, 48(2), 46-50.

[74] Thomakos, D., Wang, T., & Wu, J., 2007. Market Timing and Cap Rotation. Mathematical and Computer Modelling, 46, 278-291,

[75] Thomakos, D.D. and Wang, T., 2010. 'Optimal'probabilistic and directional predictions of financial returns. Journal of Empirical Finance, 17(1), pp.102-119.

[76] Thomakos, D. and Xidonas, P., 2023. The origins of forward-looking decision making: Cybernetics, operational research, and the foundations of forecasting. Decision Analytics Journal, 8, p.100284.

[77] van der Vaart, A.W., 2000. Asymptotic Statistics. Cambridge University Press.

[78] Vovk, V., Gammerman, A. and Shafer, G., 2005. Algorithmic learning in a random world (Vol. 29). New York: Springer.

[79] Vovk, V. and Watkins, C., 1998. Universal portfolio selection. In Proceedings of the eleventh annual conference on Computational learning theory (pp. 12-23).

[80] Vyugin, M.V. and Vyugin, V.V., 2005. Predictive complexity and information. Journal of Computer and System Sciences, 70(4), pp.539-554.

[81] Zhang, Y. and Wang, Y., 2023. Forecasting crude oil futures market returns: A principal component analysis combination approach. International Journal of Forecasting, 39(2), pp.659-673.

Appendix: figures and tables

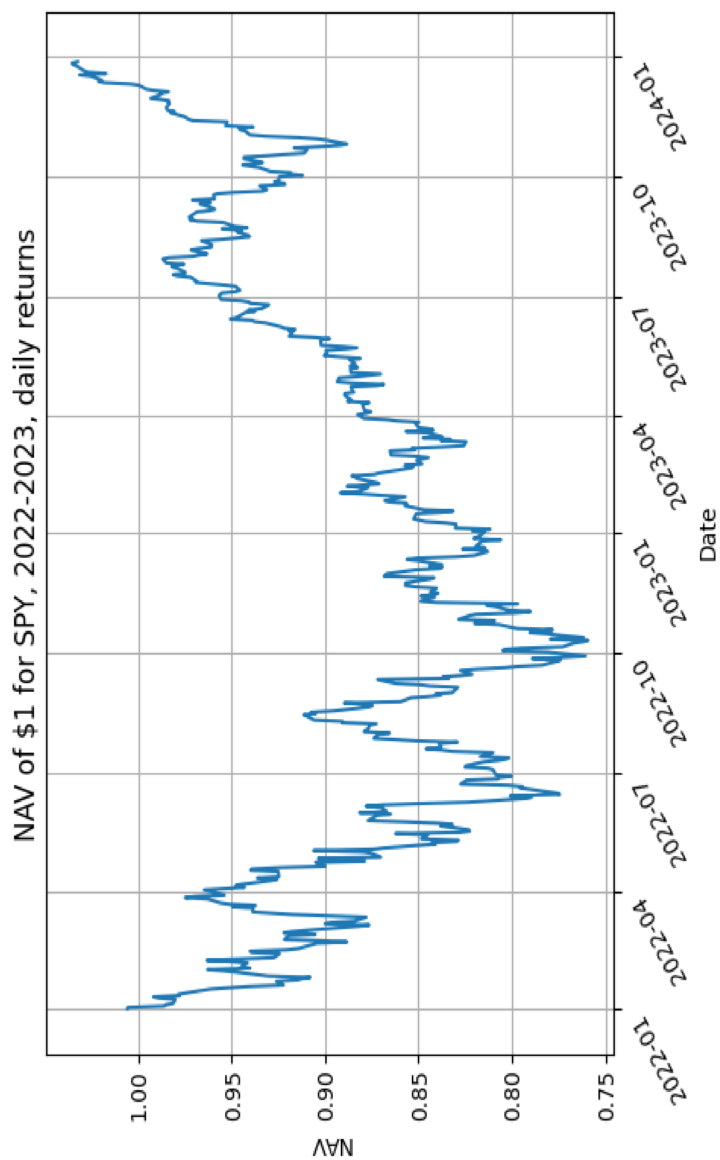

Fig. 1: Net asset value of $1 invested in SPY

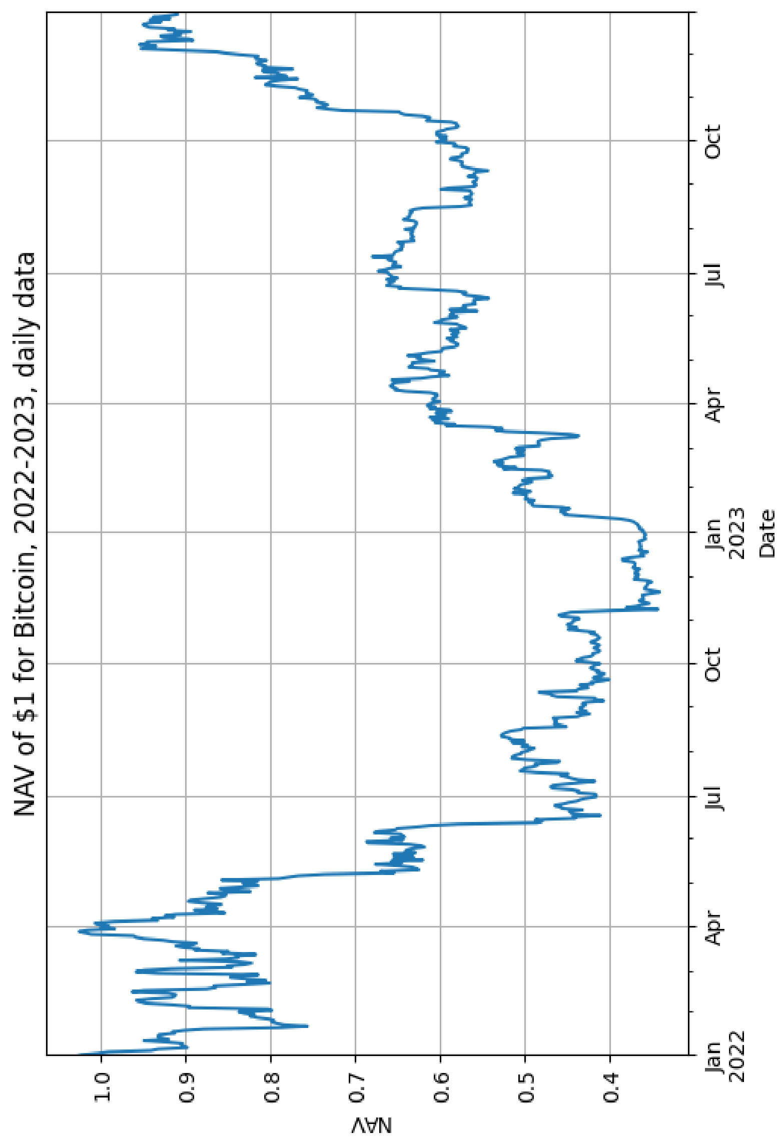

Fig. 2: Net asset value of $1 invested in Bitcoin

Tab. 1: Descriptive statistics of ETFs, 2022-2023, daily data, annualized

	Mean	**SD**	**Sharpe**	**Sortino**	**TR**	**MaxDD**
BTC	11.59%	57.21%	0.203	4.604	-9.09%	-66.74%
DBA	6.04%	13.62%	0.443	10.219	10.35%	-15.94%
DBC	8.31%	22.12%	0.376	8.270	11.95%	-27.34%
EEM	-5.45%	20.31%	-0.268	-6.338	-13.46%	-32.71%
GLD	6.86%	14.59%	0.470	11.280	11.82%	-21.03%
OIH	36.70%	41.68%	0.881	20.646	71.50%	-36.42%
SPY	3.62%	19.82%	0.183	4.173	3.24%	-24.50%
TNA	-11.29%	74.17%	-0.152	-3.487	-52.63%	-74.85%

(a) Mean is the annualized average return, SD is the annualized standard deviation, Sharpe is the annualized Sharpe ratio, Sortino is the annualized Sortino ratio, TR is the total return and MaxDD is the maximum drawdown.

Tab. 2: Correlations of ETFs, 2022-2023, daily data

	BTC	**DBA**	**DBC**	**EEM**	**GLD**	**OIH**	**SPY**
DBA	9.03%						
DBC	9.41%	59.08%					
EEM	39.84%	14.45%	21.57%				
GLD	12.77%	22.56%	39.89%	27.64%			
OIH	18.79%	37.97%	65.74%	36.27%	23.14%		
SPY	44.51%	13.64%	18.56%	69.79%	13.08%	41.38%	
TNA	42.78%	12.90%	21.06%	69.20%	12.25%	49.73%	88.56%

Tab. 3: Top performing loss functions, attribution by total excess return, Part I

α	n	**BTC**	TR_e^{BTC}	**DBA**	TR_e^{DBA}	**DBC**	TR_e^{DBC}	**EEM**	TR_e^{EEM}
0	T	$\mathcal{L}_{\text{A,SD}}$	181.8%*	ℓ_{SE}	10.4%*	ℓ_{SE}	31.1%*	$\mathcal{L}_{\text{SE,AE}}$	19.2%*
	2	$\mathcal{L}_{\text{A,SD}}$	115.9%	ℓ_{AE}	16.1%*	ℓ_{SE}	26.3%	$\bar{\delta}_A$	10.7%
	5	ℓ_{SE}	29.2%	ℓ_{SE}	10.5%*	ℓ_{SE}	32.7%*	$\mathcal{L}_{\text{SE,AE}}$	8.1%*
	7	$\mathcal{L}_{\text{SE,AE}}$	170.9%*	ℓ_{W}	19.4%*	ℓ_{SE}	36.2%*	$\bar{\delta}_A$	17.9%*
0.5	T	$\mathcal{L}_{\text{W,sgn}}$	88.2%	ℓ_{SE}	-4.3%	ℓ_{MDL}	10.6%*	ℓ_{MDL}	16.3%*
	2	$\mathcal{L}_{\text{A,SD}}$	37.1%	ℓ_{W}	14.6%	$\mathcal{L}_{\text{A,MDL}}$	36.0%	ℓ_{MDL}	17.1%
	5	$\mathcal{L}_{\text{A,SD}}$	20.5%	ℓ_{W}	4.2%*	$\mathcal{L}_{\text{W,sgn}}$	35.7%	ℓ_{MDL}	19.4%
	7	ℓ_{SE}	40.8%	ℓ_{W}	9.2%*	$\mathcal{L}_{\text{W,sgn}}$	27.5%	ℓ_{MDL}	19.0%
1	T	ℓ_{SE}	3.5%*	ℓ_{SE}	5.6%*	ℓ_{SE}	17.5%*	ℓ_{MDL}	7.4%*
	2	ℓ_{SE}	4.2%	ℓ_{SE}	7.0%	$\mathcal{L}_{\text{A,MDL}}$	33.5%	ℓ_{W}	19.3%
	5	ℓ_{SE}	4.2%	ℓ_{SE}	5.9%*	$\mathcal{L}_{\text{A,SD}}$	57.3%	$\bar{\delta}_A$	15.0%
	7	ℓ_{W}	42.5%*	ℓ_{SE}	5.0%	ℓ_{SE}	26.2%*	ℓ_{W}	25.2%
2	T	ℓ_{SE}	-17.5%	ℓ_{SE}	25.9%	ℓ_{SE}	59.1%	ℓ_{MDL}	5.0%
	2	ℓ_{sgn}	-17.9%	ℓ_{SE}	25.5%*	ℓ_{SE}	61.4%*	ℓ_{W}	16.5%*
	5	ℓ_{sgn}	-10.6%	ℓ_{SE}	26.5%*	ℓ_{SE}	65.8%*	ℓ_{W}	13.7%
	7	$\mathcal{L}_{\text{A,MDL}}$	25.1%	ℓ_{SE}	26.4%*	ℓ_{SE}	69.7%*	$\mathcal{L}_{\text{A,SD}}$	21.1%
5	T	ℓ_{sgn}	34.9%*	ℓ_{SE}	27.1%*	ℓ_{SE}	41.9%*	ℓ_{SE}	6.5%*
	2	ℓ_{sgn}	32.5%	ℓ_{SE}	27.8%	ℓ_{SE}	44.2%	$\mathcal{L}_{\text{A,MDL}}$	13.5%
	5	$\mathcal{L}_{\text{A,MDL}}$	67.3%	ℓ_{SE}	27.7%*	ℓ_{SE}	48.0%	$\mathcal{L}_{\text{A,SD}}$	9.7%
	7	$\mathcal{L}_{\text{A,MDL}}$	81.4%	$\mathcal{L}_{\text{A,MDL}}$	27.7%*	ℓ_{SE}	51.7%	$\mathcal{L}_{\text{A,SD}}$	21.4%
7	T	$\mathcal{L}_{\text{W,sgn}}$	51.0%*	ℓ_{sgn}	23.4%*	δ_A	44.1%*	ℓ_{MDL}	15.3%*
	2	ℓ_{sgn}	91.9%**	ℓ_{AE}	22.5%*	ℓ_{AE}	66.1%*	$\mathcal{L}_{\text{W,sgn}}$	56.4%*
	5	ℓ_{W}	169.3%*	$\mathcal{L}_{\text{W,sgn}}$	30.0%*	$\mathcal{L}_{\text{A,MDL}}$	47.5%*	$\mathcal{L}_{\text{SE,AE}}$	33.9%*
	7	ℓ_{W}	147.3%*	$\mathcal{L}_{\text{A,MDL}}$	31.8%*	ℓ_{SE}	49.1%*	ℓ_{MDL}	15.3%*

(a) When multiple loss functions give the same result then only one is listed based on its simplicity as follows: if statistical only then list square loss; if financial only then list wealth loss, if statistical and financial then list square loss, if complexity-based and individual loss then list the individual loss, if complexity-based only then list the "simpler" one.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).
(c) TR_e^{XXX} denotes the total excess return of ticker XXX.
(d) A star * indicates rejection of the null hypothesis of Theorem 3.

Tab. 4: Top performing loss functions, attribution by total excess return, Part II

α	n	**GLD**	TR_e^{GLD}	**OIH**	TR_e^{OIH}	**SPY**	TR_e^{SPY}	**TNA**	TR_e^{TNA}
0	T	$\bar{\delta}_A$	30.3%*	$\mathcal{L}_{\text{A,MDL}}$	-55.9%	ℓ_{MDL}	1.1%	$\mathcal{L}_{\text{SE,AE}}$	46.3%*
	2	$\bar{\delta}_A$	28.3%	ℓ_{sgn}	-18.7%	$\mathcal{L}_{\text{SE,AE}}$	5.8%	$\mathcal{L}_{\text{SE,AE}}$	221.9%
	5	$\bar{\delta}_A$	9.5%*	$\mathcal{L}_{\text{A,SD}}$	-5.2%	ℓ_{MDL}	-0.2%	$\mathcal{L}_{\text{W,sgn}}$	127.4%*
	7	$\bar{\delta}_A$	10.3%*	$\mathcal{L}_{\text{W,sgn}}$	-27.3%	ℓ_{MDL}	-1.3%	$\mathcal{L}_{\text{A,SD}}$	99.0% *
0.5	T	ℓ_{sgn}	15.9%*	$\mathcal{L}_{\text{A,MDL}}$	-4.0%	ℓ_{MDL}	13.2%*	ℓ_{MDL}	54.0%
	2	$\mathcal{L}_{\text{W,sgn}}$	12.5%	ℓ_{SE}	2.1%	$\mathcal{L}_{\text{SE,AE}}$	18.6%	ℓ_{MDL}	64.4%
	5	$\bar{\delta}_A$	-0.2%	ℓ_{SE}	-8.8%	ℓ_{MDL}	16.2%	ℓ_{MDL}	65.8%
	7	ℓ_{W}	11.4%*	ℓ_{SE}	-5.0%	ℓ_{MDL}	15.2%	ℓ_{MDL}	73.1%
1	T	$\mathcal{L}_{\text{W,sgn}}$	16.6%*	ℓ_{SE}	8.8%*	$\mathcal{L}_{\text{SE,AE}}$	-2.7%	$\mathcal{L}_{\text{A,SD}}$	30.7%
	2	$\bar{\delta}_A$	11.0%	ℓ_{SE}	23.9%	ℓ_{MDL}	-4.7%	ℓ_{MDL}	24.3%
	5	ℓ_{W}	-0.3%	$\mathcal{L}_{\text{A,MDL}}$	17.1%	ℓ_{MDL}	-2.7%	ℓ_{MDL}	41.7%
	7	ℓ_{W}	11.8%*	ℓ_{SE}	18.3%	ℓ_{MDL}	-3.9%	ℓ_{MDL}	47.1%
2	T	ℓ_{MDL}	5.7%*	ℓ_{SE}	21.9%*	$\mathcal{L}_{\text{SE,AE}}$	-3.0%	ℓ_{SE}	-24.0%
	2	$\mathcal{L}_{\text{SE,AE}}$	11.2%	ℓ_{SE}	26.8%	ℓ_{MDL}	-4.4%	$\mathcal{L}_{\text{A,MDL}}$	39.6%
	5	ℓ_{MDL}	5.3%	ℓ_{SE}	22.8%	ℓ_{MDL}	-1.4%	$\mathcal{L}_{\text{A,MDL}}$	-15.4%
	7	ℓ_{W}	9.9%	ℓ_{SE}	31.8%	ℓ_{MDL}	-2.5%	$\mathcal{L}_{\text{A,MDL}}$	-20.5%
5	T	$\mathcal{L}_{\text{SE,AE}}$	0.6%	ℓ_{SE}	18.5%*	ℓ_{SE}	-22.7%	$\bar{\delta}_A$	-10.6%
	2	$\mathcal{L}_{\text{SE,AE}}$	2.5%	ℓ_{SE}	29.9%	ℓ_{SE}	-23.9%	ℓ_{SE}	-13.2%
	5	ℓ_{MDL}	-3.6%	ℓ_{SE}	33.9%	ℓ_{SE}	-22.5%	ℓ_{SE}	-12.4%
	7	$\mathcal{L}_{\text{SE,AE}}$	5.6%	ℓ_{SE}	43.3%	ℓ_{SE}	-23.9%	ℓ_{SE}	-11.2%
7	T	$\mathcal{L}_{\text{SE,AE}}$	8.8%*	ℓ_{W}	-1.0%	ℓ_{SE}	-24.8%	$\mathcal{L}_{\text{W,sgn}}$	28.7%*
	2	$\mathcal{L}_{\text{SE,AE}}$	-19.2%	ℓ_{sgn}	36.9%*	ℓ_{W}	38.9%*	ℓ_{MDL}	86.5%
	5	$\mathcal{L}_{\text{SE,AE}}$	-2.9%	$\mathcal{L}_{\text{W,sgn}}$	47.3%*	ℓ_{MDL}	6.6%*	$\mathcal{L}_{\text{SE,AE}}$	181.1%
	7	ℓ_{sgn}	-3.7%	$\mathcal{L}_{\text{W,sgn}}$	52.8%*	$\mathcal{L}_{\text{SE,AE}}$	22.0%*	ℓ_{MDL}	96.7%

(a) When multiple loss functions give the same result then only one is listed based on its simplicity as follows: if statistical only then list square loss; if financial only then list wealth loss, if statistical and financial then list square loss, if complexity-based and individual loss then list the individual loss, if complexity-based only then list the "simpler" one.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).
(c) TR_e^{XXX} denotes the total excess return of ticker XXX.
(d) A star * indicates rejection of the null hypothesis of Theorem 3.

Tab. 5: Distribution of top performing loss functions by total excess returns

Loss	**BTC**	**DBA**	**DBC**	**EEM**	**GLD**	**OIH**	**SPY**	**TNA**	**Total**	**%**
ℓ_{SE}	6	14	15	1		14	5	4	59	30.73%
ℓ_{AE}		2	1						3	1.56%
ℓ_{W}	3	4		4	4	1	1		17	8.85%
ℓ_{sgn}	5	1			2	2			10	5.21%
ℓ_{MDL}			1	8	3		13	9	34	17.71%
$\mathcal{L}_{\text{W,sgn}}$	2	1	2	1	2	3		2	13	6.77%
$\mathcal{L}_{\text{SE,AE}}$	1			3	7		5	3	19	9.90%
$\mathcal{L}_{\text{A,SD}}$	4		1	3		1		2	11	5.73%
$\mathcal{L}_{\text{A,MDL}}$	3	2	3	1		3		3	15	7.81%
$\bar{\delta}_A$			1	3	6			1	11	5.73%

(a) The table counts and numbers and percentages, over the total, of the distribution of loss functions appearing in Tables 3 and 4 for performance attribution by total excess returns.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).

Tab. 6: Top performing loss functions, attribution by MAE, Part I

α	n	**BTC**	MAE^{BTC}	**DBA**	MAE^{DBA}	**DBC**	MAE^{DBC}	**EEM**	MAE^{EEM}
0	T	ℓ_{MDL}	1.97%	$\mathcal{L}_{\text{SE,AE}}$	0.68%	$\mathcal{L}_{\text{SE,AE}}$	1.05%	$\mathcal{L}_{\text{SE,AE}}$	0.98%
	2	ℓ_{MDL}	1.97%	$\mathcal{L}_{\text{SE,AE}}$	0.68%	ℓ_{MDL}	1.06%	ℓ_{MDL}	0.99%
	5	ℓ_{MDL}	1.97%	ℓ_{MDL}	0.68%	ℓ_{MDL}	1.06%	ℓ_{MDL}	0.98%
	7	ℓ_{MDL}	1.96%	ℓ_{MDL}	0.68%	ℓ_{MDL}	1.06%	$\mathcal{L}_{\text{SE,AE}}$	0.98%
0.5	T	ℓ_{MDL}	1.98%	$\mathcal{L}_{\text{SE,AE}}$	0.68%	ℓ_{MDL}	1.05%	ℓ_{MDL}	0.98%
	2	ℓ_{MDL}	1.98%	ℓ_{MDL}	0.68%	ℓ_{MDL}	1.06%	$\mathcal{L}_{\text{SE,AE}}$	0.99%
	5	ℓ_{MDL}	1.98%	ℓ_{MDL}	0.68%	ℓ_{MDL}	1.06%	ℓ_{MDL}	0.98%
	7	ℓ_{MDL}	1.97%	ℓ_{MDL}	0.68%	ℓ_{MDL}	1.06%	ℓ_{MDL}	0.98%
1	T	ℓ_{MDL}	1.99%	ℓ_{MDL}	0.68%	ℓ_{MDL}	1.06%	ℓ_{MDL}	0.99%
	2	ℓ_{MDL}	2.00%	ℓ_{MDL}	0.68%	$\mathcal{L}_{\text{SE,AE}}$	1.06%	ℓ_{MDL}	0.99%
	5	ℓ_{MDL}	2.00%	ℓ_{MDL}	0.68%	ℓ_{MDL}	1.06%	ℓ_{MDL}	0.99%
	7	ℓ_{MDL}	1.98%	ℓ_{MDL}	0.68%	ℓ_{MDL}	1.06%	ℓ_{MDL}	0.99%
2	T	$\mathcal{L}_{\text{W,sgn}}$	2.02%	ℓ_{MDL}	0.69%	ℓ_{MDL}	1.06%	ℓ_{MDL}	1.00%
	2	ℓ_{MDL}	2.03%	ℓ_{MDL}	0.69%	$\mathcal{L}_{\text{SE,AE}}$	1.06%	ℓ_{MDL}	1.00%
	5	ℓ_{MDL}	2.03%	ℓ_{MDL}	0.69%	$\mathcal{L}_{\text{SE,AE}}$	1.06%	ℓ_{MDL}	1.00%
	7	ℓ_{MDL}	2.02%	ℓ_{MDL}	0.69%	ℓ_{MDL}	1.06%	ℓ_{MDL}	1.00%
5	T	$\mathcal{L}_{\text{SE,AE}}$	2.10%	ℓ_{MDL}	0.71%	ℓ_{MDL}	1.09%	ℓ_{MDL}	1.03%
	2	ℓ_{MDL}	2.12%	ℓ_{MDL}	0.71%	ℓ_{MDL}	1.09%	ℓ_{MDL}	1.04%
	5	ℓ_{MDL}	2.12%	ℓ_{MDL}	0.71%	ℓ_{MDL}	1.09%	ℓ_{MDL}	1.03%
	7	$\mathcal{L}_{\text{SE,AE}}$	2.10%	ℓ_{MDL}	0.71%	ℓ_{MDL}	1.09%	ℓ_{MDL}	1.03%
7	T	δ_A	2.56%	δ_A	0.83%	ℓ_{sgn}	1.24%	δ_A	1.24%
	2	$\bar{\delta}_A$	2.43%	$\bar{\delta}_A$	0.81%	$\bar{\delta}_A$	1.24%	$\bar{\delta}_A$	1.19%
	5	$\bar{\delta}_A$	2.45%	$\bar{\delta}_A$	0.81%	$\bar{\delta}_A$	1.25%	$\bar{\delta}_A$	1.20%
	7	$\bar{\delta}_A$	2.45%	$\bar{\delta}_A$	0.82%	$\bar{\delta}_A$	1.25%	$\bar{\delta}_A$	1.20%

(a) When multiple loss functions give the same result then only one is listed based on its simplicity as follows: if statistical only then list square loss; if financial only then list wealth loss, if statistical and financial then list square loss, if complexity-based and individual loss then list the individual loss, if complexity-based only then list the "simpler" one.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).
(c) MAE^{XXX} denotes the mean-absolute error of the forecasts for ticker XXX.

Tab. 7: Top performing loss functions, attribution by MAE, Part II

α	n	**GLD**	MAE^{GLD}	**OIH**	MAE^{OIH}	**SPY**	MAE^{SPY}	**TNA**	MAE^{TNA}
0	T	$\mathcal{L}_{\text{A,SD}}$	0.71%	$\mathcal{L}_{\text{SE,AE}}$	2.04%	$\mathcal{L}_{\text{SE,AE}}$	0.94%	$\mathcal{L}_{\text{SE,AE}}$	3.65%
	2	ℓ_{MDL}	0.71%	$\mathcal{L}_{\text{SE,AE}}$	2.04%	ℓ_{MDL}	0.94%	$\mathcal{L}_{\text{SE,AE}}$	3.65%
	5	ℓ_{MDL}	0.71%	ℓ_{MDL}	2.05%	ℓ_{MDL}	0.94%	$\mathcal{L}_{\text{SE,AE}}$	3.65%
	7	ℓ_{MDL}	0.71%	$\mathcal{L}_{\text{SE,AE}}$	2.04%	ℓ_{MDL}	0.94%	ℓ_{MDL}	3.65%
0.5	T	ℓ_{MDL}	0.71%	$\mathcal{L}_{\text{SE,AE}}$	2.04%	ℓ_{MDL}	0.95%	ℓ_{MDL}	3.67%
	2	ℓ_{MDL}	0.71%	ℓ_{MDL}	2.05%	ℓ_{MDL}	0.95%	ℓ_{MDL}	3.67%
	5	ℓ_{MDL}	0.71%	ℓ_{MDL}	2.05%	ℓ_{MDL}	0.94%	ℓ_{MDL}	3.66%
	7	ℓ_{MDL}	0.71%	ℓ_{MDL}	2.05%	ℓ_{MDL}	0.94%	ℓ_{MDL}	3.66%
1	T	ℓ_{MDL}	0.72%	ℓ_{MDL}	2.05%	ℓ_{MDL}	0.95%	ℓ_{MDL}	3.70%
	2	ℓ_{MDL}	0.72%	ℓ_{MDL}	2.05%	ℓ_{MDL}	0.96%	ℓ_{MDL}	3.70%
	5	ℓ_{MDL}	0.72%	$\mathcal{L}_{\text{SE,AE}}$	2.05%	ℓ_{MDL}	0.95%	$\mathcal{L}_{\text{A,SD}}$	3.96%
	7	ℓ_{MDL}	0.72%	ℓ_{MDL}	2.05%	ℓ_{MDL}	0.95%	ℓ_{MDL}	3.69%
2	T	ℓ_{MDL}	0.73%	ℓ_{MDL}	2.07%	ℓ_{MDL}	0.97%	$\mathcal{L}_{\text{SE,AE}}$	3.76%
	2	ℓ_{MDL}	0.73%	ℓ_{MDL}	2.08%	ℓ_{MDL}	0.97%	ℓ_{MDL}	3.76%
	5	ℓ_{MDL}	0.73%	$\mathcal{L}_{\text{SE,AE}}$	2.08%	ℓ_{MDL}	0.96%	ℓ_{MDL}	3.75%
	7	ℓ_{MDL}	0.73%	ℓ_{MDL}	2.08%	ℓ_{MDL}	0.96%	ℓ_{MDL}	3.75%
5	T	$\mathcal{L}_{\text{SE,AE}}$	0.76%	$\mathcal{L}_{\text{SE,AE}}$	2.14%	ℓ_{MDL}	1.00%	$\mathcal{L}_{\text{SE,AE}}$	3.90%
	2	$\mathcal{L}_{\text{SE,AE}}$	0.77%	ℓ_{MDL}	2.15%	ℓ_{MDL}	1.00%	ℓ_{MDL}	3.91%
	5	$\mathcal{L}_{\text{SE,AE}}$	0.77%	$\mathcal{L}_{\text{SE,AE}}$	2.14%	ℓ_{MDL}	1.00%	ℓ_{MDL}	3.90%
	7	$\mathcal{L}_{\text{SE,AE}}$	0.77%	$\mathcal{L}_{\text{SE,AE}}$	2.14%	ℓ_{MDL}	1.00%	ℓ_{MDL}	3.90%
7	T	δ_A	0.91%	$\mathcal{L}_{\text{SE,AE}}$	2.52%	δ_A	1.19%	$\mathcal{L}_{\text{SE,AE}}$	4.41%
	2	$\bar{\delta}_A$	0.88%	$\bar{\delta}_A$	2.47%	$\bar{\delta}_A$	1.14%	$\bar{\delta}_A$	4.39%
	5	$\bar{\delta}_A$	0.87%	$\bar{\delta}_A$	2.47%	$\bar{\delta}_A$	1.15%	$\bar{\delta}_A$	4.44%
	7	$\bar{\delta}_A$	0.88%	$\bar{\delta}_A$	2.47%	$\bar{\delta}_A$	1.15%	$\bar{\delta}_A$	4.44%

(a) When multiple loss functions give the same result then only one is listed based on its simplicity as follows: if statistical only then list square loss; if financial only then list wealth loss, if statistical and financial then list square loss, if complexity-based and individual loss then list the individual loss, if complexity-based only then list the "simpler" one.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).
(c) MAE^{XXX} denotes the mean-absolute error of the forecasts for ticker XXX.

Tab. 8: Distribution of top performing loss functions by MAE

Loss	**BTC**	**DBA**	**DBC**	**EEM**	**GLD**	**OIH**	**SPY**	**TNA**	**Total**	%
ℓ_{SE}									0	0.00%
ℓ_{AE}									0	0.00%
ℓ_{W}									0	0.00%
ℓ_{sgn}			1						1	0.52%
ℓ_{MDL}	17	17	16	17	15	11	19	14	126	65.63%
$\mathcal{L}_{\mathrm{W,sgn}}$	1								1	0.52%
$\mathcal{L}_{\mathrm{SE,AE}}$	2	3	4	3	4	10	1	6	33	17.19%
$\mathcal{L}_{\mathrm{A,SD}}$					1			1	2	1.04%
$\mathcal{L}_{\mathrm{A,MDL}}$									0	0.00%
$\bar{\delta}_A$	4	4	3	4	4	3	4	3	29	15.10%

(a) The table counts and numbers and percentages, over the total, of the distribution of loss functions appearing in Tables 5 and 6 for performance attribution by MAE.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\mathrm{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\mathrm{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).

Tab. 9: Top performing loss functions, attribution by MSE, Part I

α	n	**BTC**	MSE^{BTC}	**DBA**	MSE^{DBA}	**DBC**	MSE^{DBC}	**EEM**	MSE^{EEM}
0	T	$\mathcal{L}_{\mathrm{A,SD}}$	0.085%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.011%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.020%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.017%
	2	ℓ_{MDL}	0.085%	ℓ_{sgn}	0.010%	ℓ_{MDL}	0.020%	ℓ_{MDL}	0.017%
	5	ℓ_{MDL}	0.086%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.011%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
	7	ℓ_{MDL}	0.084%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.010%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
0.5	T	$\mathcal{L}_{\mathrm{W,sgn}}$	0.086%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.011%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
	2	ℓ_{MDL}	0.086%	ℓ_{SE}	0.011%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
	5	ℓ_{MDL}	0.086%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.010%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
	7	ℓ_{MDL}	0.084%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.010%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
1	T	$\mathcal{L}_{\mathrm{A,SD}}$	0.086%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.011%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
	2	ℓ_{MDL}	0.087%	ℓ_{sgn}	0.010%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.021%	ℓ_{MDL}	0.017%
	5	ℓ_{MDL}	0.087%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.010%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
	7	ℓ_{MDL}	0.085%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.011%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
2	T	$\mathcal{L}_{\mathrm{W,sgn}}$	0.088%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.012%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
	2	ℓ_{MDL}	0.088%	ℓ_{sgn}	0.010%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.021%	ℓ_{MDL}	0.017%
	5	$\mathcal{L}_{\mathrm{SE,AE}}$	0.088%	ℓ_{sgn}	0.010%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.021%	ℓ_{MDL}	0.017%
	7	$\mathcal{L}_{\mathrm{SE,AE}}$	0.086%	$\mathcal{L}_{\mathrm{A,MDL}}$	0.011%	ℓ_{MDL}	0.021%	ℓ_{MDL}	0.017%
5	T	$\mathcal{L}_{\mathrm{SE,AE}}$	0.092%	$\mathcal{L}_{\mathrm{A,SD}}$	0.010%	ℓ_{MDL}	0.022%	ℓ_{MDL}	0.018%
	2	$\mathcal{L}_{\mathrm{SE,AE}}$	0.094%	$\mathcal{L}_{\mathrm{W,sgn}}$	0.010%	ℓ_{MDL}	0.022%	ℓ_{MDL}	0.018%
	5	$\mathcal{L}_{\mathrm{SE,AE}}$	0.092%	$\mathcal{L}_{\mathrm{W,sgn}}$	0.011%	ℓ_{MDL}	0.022%	ℓ_{MDL}	0.018%
	7	$\mathcal{L}_{\mathrm{SE,AE}}$	0.090%	$\mathcal{L}_{\mathrm{A,SD}}$	0.011%	ℓ_{MDL}	0.023%	ℓ_{MDL}	0.018%
7	T	δ_A	0.132%	δ_A	0.011%	ℓ_{sgn}	0.028%	δ_A	0.026%
	2	$\bar{\delta}_A$	0.119%	$\bar{\delta}_A$	0.010%	$\bar{\delta}_A$	0.027%	$\bar{\delta}_A$	0.024%
	5	$\bar{\delta}_A$	0.121%	$\bar{\delta}_A$	0.011%	$\bar{\delta}_A$	0.028%	$\bar{\delta}_A$	0.024%
	7	$\bar{\delta}_A$	0.121%	$\bar{\delta}_A$	0.011%	$\bar{\delta}_A$	0.028%	$\bar{\delta}_A$	0.025%

(a) When multiple loss functions give the same result then only one is listed based on its simplicity as follows: if statistical only then list square loss; if financial only then list wealth loss, if statistical and financial then list square loss, if complexity-based and individual loss then list the individual loss, if complexity-based only then list the "simpler" one.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\mathrm{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\mathrm{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).
(c) MSE^{XXX} denotes the mean-squared error of the forecasts for ticker XXX.

Tab. 10: Top performing loss functions, attribution by MSE, Part II

α	n	**GLD**	MSE^{GLD}	**OIH**	MSE^{OIH}	**SPY**	MSE^{SPY}	**TNA**	MSE^{TNA}
0	T	$\mathcal{L}_{\mathrm{A,MDL}}$	0.013%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.070%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.016%	ℓ_{MDL}	0.217%
	2	$\mathcal{L}_{\mathrm{A,SD}}$	0.010%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.217%
	5	$\mathcal{L}_{\mathrm{A,SD}}$	0.010%	ℓ_{MDL}	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.216%
	7	ℓ_{sgn}	0.010%	ℓ_{MDL}	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.216%
0.5	T	$\mathcal{L}_{\mathrm{A,MDL}}$	0.013%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.219%
	2	$\mathcal{L}_{\mathrm{A,SD}}$	0.010%	ℓ_{MDL}	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.219%
	5	$\mathcal{L}_{\mathrm{W,sgn}}$	0.010%	ℓ_{MDL}	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.219%
	7	$\mathcal{L}_{\mathrm{W,sgn}}$	0.010%	ℓ_{MDL}	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.218%
1	T	$\mathcal{L}_{\mathrm{A,MDL}}$	0.013%	ℓ_{MDL}	0.070%	ℓ_{MDL}	0.016%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.222%
	2	$\mathcal{L}_{\mathrm{A,SD}}$	0.011%	ℓ_{MDL}	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.223%
	5	ℓ_{W}	0.010%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.222%
	7	$\mathcal{L}_{\mathrm{W,sgn}}$	0.010%	ℓ_{MDL}	0.070%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.222%
2	T	$\mathcal{L}_{\mathrm{W,sgn}}$	0.010%	ℓ_{MDL}	0.072%	ℓ_{MDL}	0.016%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.228%
	2	$\mathcal{L}_{\mathrm{W,sgn}}$	0.010%	ℓ_{MDL}	0.072%	ℓ_{MDL}	0.017%	ℓ_{MDL}	0.228%
	5	$\bar{\delta}_A$	0.010%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.072%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.228%
	7	ℓ_{W}	0.010%	ℓ_{MDL}	0.072%	ℓ_{MDL}	0.016%	ℓ_{MDL}	0.228%
5	T	$\bar{\delta}_A$	0.013%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.076%	ℓ_{MDL}	0.018%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.244%
	2	$\mathcal{L}_{\mathrm{W,sgn}}$	0.012%	ℓ_{MDL}	0.077%	ℓ_{MDL}	0.018%	ℓ_{MDL}	0.246%
	5	ℓ_{W}	0.012%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.076%	ℓ_{MDL}	0.018%	ℓ_{MDL}	0.246%
	7	ℓ_{W}	0.012%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.076%	ℓ_{MDL}	0.017%	ℓ_{MDL}	0.245%
7	T	δ_A	0.014%	δ_A	0.103%	δ_A	0.024%	$\mathcal{L}_{\mathrm{SE,AE}}$	0.320%
	2	$\bar{\delta}_A$	0.013%	$\bar{\delta}_A$	0.098%	$\bar{\delta}_A$	0.023%	$\bar{\delta}_A$	0.312%
	5	$\bar{\delta}_A$	0.013%	$\bar{\delta}_A$	0.098%	$\bar{\delta}_A$	0.022%	$\bar{\delta}_A$	0.317%
	7	$\bar{\delta}_A$	0.013%	$\bar{\delta}_A$	0.098%	$\bar{\delta}_A$	0.023%	$\bar{\delta}_A$	0.319%

(a) When multiple loss functions give the same result then only one is listed based on its simplicity as follows: if statistical only then list square loss; if financial only then list wealth loss, if statistical and financial then list square loss, if complexity-based and individual loss then list the individual loss, if complexity-based only then list the "simpler" one.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\mathrm{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\mathrm{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).
(c) MSE^{XXX} denotes the mean-squared error of the forecasts for ticker XXX.

Tab. 11: Distribution of top performing loss functions by MSE

Loss	**BTC**	**DBA**	**DBC**	**EEM**	**GLD**	**OIH**	**SPY**	**TNA**	**Total**	%
ℓ_{SE}		1							1	0.52%
ℓ_{AE}									0	0.00%
ℓ_{W}					4				4	2.08%
ℓ_{sgn}		4	1		1				6	3.13%
ℓ_{MDL}	10		16	19		12	19	17	93	48.44%
$\mathcal{L}_{\text{W,sgn}}$	2	2			6				10	5.21%
$\mathcal{L}_{\text{SE,AE}}$	6		4	1		8	1	4	24	12.50%
$\mathcal{L}_{\text{A,SD}}$	2	2			4				8	4.17%
$\mathcal{L}_{\text{A,MDL}}$		11			3				14	7.29%
$\bar{\delta}_A$	4	4	3	4	6	4	4	3	32	16.67%

[(a)] The table counts and numbers and percentages, over the total, of the distribution of loss functions appearing in Tables 7 and 8 for performance attribution by MSE.
[(b)] The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).

Tab. 12: Top performing loss functions, attribution by SSR, Part I

α	n	**BTC**	SSR^{BTC}	**DBA**	SSR^{DBA}	**DBC**	SSR^{DBC}	**EEM**	SSR^{EEM}
0	T	$\mathcal{L}_{\text{SE,AE}}$	50.14%	ℓ_{SE}	51.57%	$\bar{\delta}_A$	53.25%	$\mathcal{L}_{\text{SE,AE}}$	50.52%
	2	ℓ_{MDL}	49.14%	ℓ_{AE}	52.21%	$\mathcal{L}_{\text{W,sgn}}$	52.00%	$\bar{\delta}_A$	50.11%
	5	$\mathcal{L}_{\text{SE,AE}}$	49.93%	ℓ_{SE}	51.27%	ℓ_{SE}	51.91%	ℓ_{W}	50.21%
	7	$\mathcal{L}_{\text{SE,AE}}$	50.07%	$\bar{\delta}_A$	51.49%	$\mathcal{L}_{\text{A,MDL}}$	52.55%	$\bar{\delta}_A$	50.43%
0.5	T	ℓ_{MDL}	49.57%	ℓ_{SE}	49.06%	$\bar{\delta}_A$	51.36%	ℓ_{MDL}	49.69%
	2	ℓ_{MDL}	49.57%	ℓ_{W}	51.16%	$\mathcal{L}_{\text{A,MDL}}$	52.84%	ℓ_{MDL}	49.68%
	5	$\mathcal{L}_{\text{A,SD}}$	49.93%	$\bar{\delta}_A$	50.64%	$\mathcal{L}_{\text{W,sgn}}$	52.33%	ℓ_{W}	50.64%
	7	ℓ_{MDL}	49.78%	ℓ_{W}	51.06%	$\mathcal{L}_{\text{A,MDL}}$	52.98%	ℓ_{MDL}	49.79%
1	T	ℓ_{MDL}	48.29%	ℓ_{SE}	49.27%	ℓ_{SE}	51.99%	$\bar{\delta}_A$	49.27%
	2	ℓ_{MDL}	48.28%	ℓ_{SE}	49.26%	$\mathcal{L}_{\text{A,MDL}}$	53.05%	ℓ_{W}	51.58%
	5	ℓ_{MDL}	48.35%	ℓ_{SE}	48.94%	$\mathcal{L}_{\text{A,MDL}}$	53.39%	$\bar{\delta}_A$	49.79%
	7	ℓ_{MDL}	48.48%	ℓ_{SE}	48.72%	$\mathcal{L}_{\text{A,MDL}}$	53.83%	ℓ_{W}	50.85%
2	T	$\mathcal{L}_{\text{W,sgn}}$	47.71%	ℓ_{SE}	50.31%	ℓ_{SE}	53.46%	$\mathcal{L}_{\text{A,SD}}$	49.48%
	2	ℓ_{sgn}	47.71%	ℓ_{sgn}	51.37%	ℓ_{SE}	53.68%	ℓ_{W}	50.95%
	5	$\mathcal{L}_{\text{SE,AE}}$	48.06%	ℓ_{SE}	50.21%	ℓ_{SE}	53.81%	ℓ_{W}	49.58%
	7	$\mathcal{L}_{\text{SE,AE}}$	48.05%	ℓ_{SE}	50.21%	ℓ_{SE}	53.83%	$\mathcal{L}_{\text{A,SD}}$	50.64%
5	T	$\mathcal{L}_{\text{A,MDL}}$	47.57%	$\mathcal{L}_{\text{A,MDL}}$	50.52%	ℓ_{SE}	52.83%	$\mathcal{L}_{\text{A,SD}}$	49.48%
	2	ℓ_{W}	47.85%	ℓ_{SE}	50.74%	ℓ_{SE}	53.05%	ℓ_{W}	49.47%
	5	$\mathcal{L}_{\text{A,MDL}}$	48.06%	ℓ_{SE}	50.42%	ℓ_{SE}	53.18%	$\mathcal{L}_{\text{A,SD}}$	49.36%
	7	$\mathcal{L}_{\text{A,MDL}}$	48.77%	$\mathcal{L}_{\text{A,MDL}}$	50.85%	ℓ_{SE}	53.19%	$\mathcal{L}_{\text{A,SD}}$	50.85%
7	T	$\mathcal{L}_{\text{W,sgn}}$	51.71%	ℓ_{MDL}	50.73%	$\bar{\delta}_A$	53.46%	ℓ_{MDL}	51.36%
	2	$\mathcal{L}_{\text{A,SD}}$	51.43%	ℓ_{W}	52.21%	ℓ_{W}	53.89%	$\mathcal{L}_{\text{W,sgn}}$	53.68%
	5	ℓ_{W}	52.95%	$\mathcal{L}_{\text{SE,AE}}$	53.18%	ℓ_{W}	54.24%	ℓ_{MDL}	51.27%
	7	ℓ_{W}	50.65%	$\mathcal{L}_{\text{W,sgn}}$	51.49%	$\mathcal{L}_{\text{W,sgn}}$	54.47%	ℓ_{MDL}	51.49%

(a) When multiple loss functions give the same result then only one is listed based on its simplicity as follows: if statistical only then list square loss; if financial only then list wealth loss, if statistical and financial then list square loss, if complexity-based and individual loss then list the individual loss, if complexity-based only then list the "simpler" one.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).
(c) SSR^{XXX} denotes the sign success ratio of the forecasts for ticker XXX.

Tab. 13: Top performing loss functions, attribution by SSR, Part I

α	n	**GLD**	SSR^{GLD}	**OIH**	SSR^{OIH}	**SPY**	SSR^{SPY}	**TNA**	SSR^{TNA}
0	T	ℓ_{sgn}	52.41%	$\mathcal{L}_{\text{W,sgn}}$	48.64%	ℓ_{MDL}	51.78%	$\mathcal{L}_{\text{SE,AE}}$	52.62%
	2	$\bar{\delta}_A$	52.00%	ℓ_{sgn}	49.68%	$\mathcal{L}_{\text{SE,AE}}$	52.63%	$\mathcal{L}_{\text{SE,AE}}$	53.68%
	5	ℓ_{MDL}	51.69%	$\mathcal{L}_{\text{A,SD}}$	50.42%	$\bar{\delta}_A$	52.33%	$\mathcal{L}_{\text{SE,AE}}$	53.39%
	7	ℓ_{MDL}	51.70%	ℓ_{sgn}	49.57%	$\bar{\delta}_A$	52.13%	$\mathcal{L}_{\text{SE,AE}}$	52.77%
0.5	T	$\bar{\delta}_A$	50.52%	ℓ_{MDL}	49.06%	$\mathcal{L}_{\text{SE,AE}}$	53.25%	$\mathcal{L}_{\text{SE,AE}}$	53.04%
	2	$\mathcal{L}_{\text{SE,AE}}$	50.11%	ℓ_{MDL}	49.05%	$\mathcal{L}_{\text{SE,AE}}$	53.26%	ℓ_{MDL}	52.63%
	5	ℓ_{W}	52.97%	$\bar{\delta}_A$	49.15%	ℓ_{MDL}	53.18%	ℓ_{MDL}	52.54%
	7	ℓ_{W}	51.91%	ℓ_{W}	50.21%	ℓ_{MDL}	53.19%	ℓ_{MDL}	52.55%
1	T	$\mathcal{L}_{\text{W,sgn}}$	51.99%	$\mathcal{L}_{\text{SE,AE}}$	49.48%	ℓ_{MDL}	51.78%	$\mathcal{L}_{\text{A,SD}}$	51.57%
	2	$\mathcal{L}_{\text{SE,AE}}$	50.32%	ℓ_{SE}	49.26%	ℓ_{MDL}	51.79%	ℓ_{MDL}	51.58%
	5	ℓ_{W}	50.64%	$\mathcal{L}_{\text{A,MDL}}$	50.21%	ℓ_{MDL}	51.91%	ℓ_{MDL}	51.91%
	7	ℓ_{W}	50.85%	ℓ_{W}	51.28%	ℓ_{MDL}	51.91%	ℓ_{MDL}	51.91%
2	T	ℓ_{MDL}	50.10%	ℓ_{SE}	50.52%	$\mathcal{L}_{\text{SE,AE}}$	54.09%	ℓ_{SE}	49.27%
	2	ℓ_{MDL}	50.11%	ℓ_{SE}	50.32%	ℓ_{MDL}	53.68%	$\mathcal{L}_{\text{A,MDL}}$	50.53%
	5	ℓ_{MDL}	50.00%	ℓ_{SE}	50.64%	ℓ_{MDL}	53.81%	$\mathcal{L}_{\text{A,MDL}}$	49.58%
	7	ℓ_{MDL}	50.00%	ℓ_{SE}	50.85%	ℓ_{MDL}	53.83%	$\mathcal{L}_{\text{A,MDL}}$	49.57%
5	T	$\mathcal{L}_{\text{W,sgn}}$	49.27%	ℓ_{SE}	52.62%	ℓ_{MDL}	50.52%	ℓ_{SE}	49.48%
	2	$\mathcal{L}_{\text{SE,AE}}$	48.84%	ℓ_{SE}	52.84%	ℓ_{MDL}	50.53%	ℓ_{SE}	49.47%
	5	$\mathcal{L}_{\text{SE,AE}}$	48.52%	ℓ_{SE}	52.97%	ℓ_{MDL}	50.42%	ℓ_{SE}	49.15%
	7	$\mathcal{L}_{\text{A,SD}}$	49.79%	ℓ_{SE}	53.19%	ℓ_{MDL}	50.43%	ℓ_{SE}	49.15%
7	T	$\mathcal{L}_{\text{A,SD}}$	50.31%	$\mathcal{L}_{\text{W,sgn}}$	53.88%	ℓ_{MDL}	50.10%	ℓ_{MDL}	51.78%
	2	ℓ_{sgn}	48.63%	ℓ_{W}	53.47%	ℓ_{W}	52.21%	ℓ_{MDL}	52.21%
	5	$\mathcal{L}_{\text{SE,AE}}$	51.06%	$\mathcal{L}_{\text{W,sgn}}$	53.60%	ℓ_{MDL}	50.85%	$\mathcal{L}_{\text{SE,AE}}$	54.24%
	7	$\mathcal{L}_{\text{SE,AE}}$	49.57%	$\mathcal{L}_{\text{W,sgn}}$	54.45%	$\mathcal{L}_{\text{SE,AE}}$	51.91%	ℓ_{MDL}	52.34%

(a) When multiple loss functions give the same result then only one is listed based on its simplicity as follows: if statistical only then list square loss; if financial only then list wealth loss, if statistical and financial then list square loss, if complexity-based and individual loss then list the individual loss, if complexity-based only then list the "simpler" one.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).
(c) SSR^{XXX} denotes the sign success ratio of the forecasts for ticker XXX.

Tab. 14: Distribution of top performing loss functions by SSR

Loss	**BTC**	**DBA**	**DBC**	**EEM**	**GLD**	**OIH**	**SPY**	**TNA**	**Total**	**%**
ℓ_{SE}		13	10			9		5	37	19.27%
ℓ_{AE}		1							1	0.52%
ℓ_{W}	3	3	2	7	4	3	1		23	11.98%
ℓ_{sgn}	1	1			2	2			6	3.13%
ℓ_{MDL}	8	1		6	6	2	16	9	48	25.00%
$\mathcal{L}_{\text{W,sgn}}$	2	1	3	1	2	4			13	6.77%
$\mathcal{L}_{\text{SE,AE}}$	5			1	6	1	5	6	24	12.50%
$\mathcal{L}_{\text{A,SD}}$	2			5	2	1		1	11	5.73%
$\mathcal{L}_{\text{A,MDL}}$	3	2	6			1		3	15	7.81%
$\bar{\delta}_A$		2	3	4	2	1	2		14	7.29%

(a) The table counts and numbers and percentages, over the total, of the distribution of loss functions appearing in Tables 9 and 10 for performance attribution by SSR.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).

Tab. 15: Comparison of loss functions on total excess returns and sign success ratio

	Total Excess Returns Ranking			Corresponding Sign Success Ratios		
	1st	2nd	3rd	of 1st	of 2nd	of 3rd
BTC	181.80%	170.90%	169.30%	50.14%	50.07%	52.95%
	$\mathcal{L}_{\text{A,SD}}$	$\mathcal{L}_{\text{SE,AE}}$	ℓ_{W}	$\mathcal{L}_{\text{SE,AE}}$	$\mathcal{L}_{\text{SE,AE}}$	ℓ_{W}
DBA	31.80%	30%	27.80%	51.49%	53.18%	50.74%
	$\mathcal{L}_{\text{A,MDL}}$	$\mathcal{L}_{\text{W,sgn}}$	ℓ_{SE}	$\mathcal{L}_{\text{W,sgn}}$	$\mathcal{L}_{\text{SE,AE}}$	ℓ_{SE}
DBC	69.70%	66.10%	65.80%	53.83%	53.89%	53.81%
	ℓ_{SE}	ℓ_{AE}	ℓ_{SE}	ℓ_{SE}	ℓ_{W}	ℓ_{SE}
EEM	56.40%	33.90%	25.20%	53.68%	51.27%	50.85%
	$\mathcal{L}_{\text{W,sgn}}$	$\mathcal{L}_{\text{SE,AE}}$	ℓ_{W}	$\mathcal{L}_{\text{W,sgn}}$	ℓ_{MDL}	ℓ_{W}
GLD	30.30%	28.30%	16.60%	0.52	52%	51.99%
	$\bar{\delta}_A$	$\bar{\delta}_A$	$\mathcal{L}_{\text{W,sgn}}$	ℓ_{sgn}	$\bar{\delta}_A$	$\mathcal{L}_{\text{W,sgn}}$
OIH	52.80%	47.30%	43.30%	54.45%	53.60%	53.19%
	$\mathcal{L}_{\text{W,sgn}}$	$\mathcal{L}_{\text{W,sgn}}$	ℓ_{SE}	$\mathcal{L}_{\text{W,sgn}}$	$\mathcal{L}_{\text{W,sgn}}$	ℓ_{SE}
SPY	38.90%	22%	18.60%	52.21%	50.85%	50.43%
	ℓ_{W}	$\mathcal{L}_{\text{SE,AE}}$	$\mathcal{L}_{\text{SE,AE}}$	ℓ_{W}	ℓ_{MDL}	ℓ_{MDL}
TNA	221.90%	181.10%	127.40%	53.68%	54.24%	49.15%
	$\mathcal{L}_{\text{SE,AE}}$	$\mathcal{L}_{\text{SE,AE}}$	$\mathcal{L}_{\text{W,sgn}}$	$\mathcal{L}_{\text{SE,AE}}$	$\mathcal{L}_{\text{SE,AE}}$	ℓ_{SE}

(a) The table lists the top 3 performing loss functions for total excess returns and then lists the corresponding sign success ratio loss functions that appeared for the same parametrizations as in the total excess returns attribution.
(b) The loss function nomenclature is the following: $\mathcal{L}_{A,SD}$ is the collection of all 4 loss functions using SD; $\mathcal{L}_{A,MDL}$ is the same but using MDL; $\mathcal{L}_{\text{SE,AE}}$ is the collection of the two statistical loss functions using SD; $\mathcal{L}_{\text{W,sgn}}$ is the collection of the two financial loss functions using SD; $\bar{\delta}_A$ is the average optimal δ from all individual loss functions (including ℓ_{MDL}).

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
SCHOOL OF ECONOMIC SCIENCES
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
DEPARTMENT OF ECONOMICS

Department of Economics
Athens University of Economics and Business

List of Recent Working Papers

2022

01-22 Is Ireland the most intangible intensive economy in Europe? A growth accounting perspective, Ilias Kostarakos, KieranMcQuinn and Petros Varthalitis
02-22 Common bank supervision and profitability convergence in the EU, Ioanna Avgeri, Yiannis Dendramis and Helen Louri
03-22 Missing Values in Panel Data Unit Root Tests, Yiannis Karavias, Elias Tzavalis and Haotian Zhang
04-22 Ordering Arbitrage Portfolios and Finding Arbitrage Opportunities, Stelios Arvanitis and Thierry Post
05-22 Concentration Inequalities for Kernel Density Estimators under Uniform Mixing, Stelios Arvanitis
06-22 Public Sector Corruption and the Valuation of Systemically Important Banks, Georgios Bertsatos, Spyros Pagratis, Plutarchos Sakellaris
07-22 Finance or Demand: What drives the Responses of Young and Small Firms to Financial Crises? Stelios Giannoulakis and Plutarchos Sakellaris
08-22 Production function estimation controlling for endogenous productivity disruptions, Plutarchos Sakellaris and Dimitris Zaverdas
09-22 A panel bounds testing procedure, Georgios Bertsatos, Plutarchos Sakellaris, Mike G. Tsionas
10-22 Social policy gone bad educationally: Unintended peer effects from transferred students, Christos Genakos and Eleni Kyrkopoulou
11-22 Inconsistency for the Gaussian QMLE in GARCH-type models with infinite variance, Stelios Arvanitis and Alexandros Louka
12-22 Time to question the wisdom of active monetary policies, George C. Bitros
13-22 Investors' Behavior in Cryptocurrency Market, Stelios Arvanitis, Nikolas Topaloglou and Georgios Tsomidis
14-22 On the asking price for selling Chelsea FC, Georgios Bertsatos and Gerassimos Sapountzoglou
15-22 Hysteresis, Financial Frictions and Monetary Policy, Konstantinos Giakas
16-22 Delay in Childbearing and the Evolution of Fertility Rates, Evangelos Dioikitopoulos and Dimitrios Varvarigos
17-22 Human capital threshold effects in economic development: A panel data approach with endogenous threshold, Dimitris Christopoulos, Dimitris Smyrnakis and Elias Tzavalis
18-22 Distributional aspects of rent seeking activities in a Real Business Cycle model, Tryfonas Christou, Apostolis Philippopoulos and Vanghelis Vassilatos

2023

- **01-23** **Real interest rate and monetary policy in the post Bretton Woods United States, George C. Bitros and Mara Vidali**
- **02-23** **Debt targets and fiscal consolidation in a two-country HANK model: the case of Euro Area, Xiaoshan Chen, Spyridon Lazarakis and Petros Varthalitis**
- **03-23** **Central bank digital currencies: Foundational issues and prospects looking forward, George C. Bitros and Anastasios G. Malliaris**
- **04-23** **The State and the Economy of Modern Greece. Key Drivers from 1821 to the Present, George Alogoskoufis**
- **05-23** **Sparse spanning portfolios and under-diversification with second-order stochastic dominance, Stelios Arvanitis, Olivier Scaillet, Nikolas Topaloglou**
- **06-23** **What makes for survival? Key characteristics of Greek incubated early-stage startup(per)s during the Crisis: a multivariate and machine learning approach, Ioannis Besis, Ioanna Sapfo Pepelasis and Spiros Paraskevas**
- **07-23** **The Twin Deficits, Monetary Instability and Debt Crises in the History of Modern Greece, George Alogoskoufis**
- **08-23** **Dealing with endogenous regressors using copulas; on the problem of near multicollinearity, Dimitris Christopoulos, Dimitris Smyrnakis and Elias Tzavalis**
- **09-23** **A machine learning approach to construct quarterly data on intangible investment for Eurozone, Angelos Alexopoulos and Petros Varthalitis**
- **10-23** **Asymmetries in Post-War Monetary Arrangements in Europe: From Bretton Woods to the Euro Area, George Alogoskoufis, Konstantinos Gravas and Laurent Jacque**
- **11-23** **Unanticipated Inflation, Unemployment Persistence and the New Keynesian Phillips Curve, George Alogoskoufis and Stelios Giannoulakis**
- **12-23** **Threshold Endogeneity in Threshold VARs: An Application to Monetary State Dependence, Dimitris Christopoulos, Peter McAdam and Elias Tzavalis**
- **13-23** **A DSGE Model for the European Unemployment Persistence, Konstantinos Giakas**
- **14-23** **Binary public decisions with a status quo: undominated mechanisms without coercion, Efthymios Athanasiou and Giacomo Valletta**
- **15-23** **Does Agents' learning explain deviations in the Euro Area between the Core and the Periphery? George Economides, Konstantinos Mavrigiannakis and Vanghelis Vassilatos**
- **16-23** **Mild Explocivity, Persistent Homology and Cryptocurrencies' Bubbles: An Empirical Exercise, Stelios Arvanitis and Michalis Detsis**
- **17-23** **A network and machine learning approach to detect Value Added Tax fraud, Angelos Alexopoulos, Petros Dellaportas, Stanley Gyoshev, Christos Kotsogiannis, Sofia C. Olhede, Trifon Pavkov**
- **18-23** **Time Varying Three Pass Regression Filter, Yiannis Dendramis, George Kapetanios, Massimiliano Marcellino**
- **19-23** **From debt arithmetic to fiscal sustainability and fiscal rules: Taking stock, George Economides, Natasha Miouli and Apostolis Philippopoulos**
- **20-23** **Stochastic Arbitrage Opportunities: Set Estimation and Statistical Testing, Stelios Arvanitis and Thierry Post**
- **21-23** **Behavioral Personae, Stochastic Dominance, and the Cryptocurrency Market, Stelios Arvanitis, Nikolas Topaloglou, and Georgios Tsomidis**
- **22-23** **Block Empirical Likelihood Inference for Stochastic Bounding: Large Deviations Asymptotics Under *m*-Dependence, Stelios Arvanitis and Nikolas Topaloglou**
- **23-23** **A Consolidation of the Neoclassical Macroeconomic Competitive General Equilibrium Theory via Keynesianism (Part 1 and Part 2), Angelos Angelopoulos**
- **24-23** **Limit Theory for Martingale Transforms with Heavy-Tailed Noise, Stelios Arvanitis and Alexandros Louka**

Department of Economics
Athens University of Economics and Business

The Department is the oldest Department of Economics in Greece with a pioneering role in organising postgraduate studies in Economics since 1978. Its priority has always been to bring together highly qualified academics and top quality students. Faculty members specialize in a wide range of topics in economics, with teaching and research experience in world-class universities and publications in top academic journals.

The Department constantly strives to maintain its high level of research and teaching standards. It covers a wide range of economic studies in micro-and macroeconomic analysis, banking and finance, public and monetary economics, international and rural economics, labour economics, industrial organization and strategy, economics of the environment and natural resources, economic history and relevant quantitative tools of mathematics, statistics and econometrics.

Its undergraduate program attracts high quality students who, after successful completion of their studies, have excellent prospects for employment in the private and public sector, including areas such as business, banking, finance and advisory services. Also, graduates of the program have solid foundations in economics and related tools and are regularly admitted to top graduate programs internationally. Three specializations are offered:1. Economic Theory and Policy, 2. Business Economics and Finance and 3. International and European Economics. The postgraduate programs of the Department (M.Sc and Ph.D) are highly regarded and attract a large number of quality candidates every year.

For more information:

https://www.dept.aueb.gr/en/econ/

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5