Assessing the Performance of AI-Labelled Portfolios

Mauricio Praxmarer and Iván Simon*†

This draft: November 2024

Abstract

We evaluate the performance of AI funds. Using a hand-collected dataset of US mutual funds, our analysis reveals that AI funds yield similar performance as human-managed counterparts with similar investment objectives and fund characteristics. Yet, they fail to outperform a market benchmark. AI tends to benefit funds in activities such as market screening. On the other hand, funds did not manage to employ AI to their advantage to optimize their portfolio allocation, probably due to high turnover. Compared to their peers, AI funds exhibit superior market timing and inferior stock-picking abilities.

Keywords: machine learning, artificial intelligence, natural language processing, portfolio management

(JEL G10, G11)

*Department of Banking and Finance, University of Innsbruck, Universitätstraße 15, A-6020 Innsbruck, Austria, corresponding emails: mauricio.praxmarer@uibk.ac.at and ivan.simon@student.uibk.ac.at

†We thank Moritz Dauber, Moreno Frigg, Dennis Umlandt, Patrick Weiss as well as participants of the Austrian Working Group on Banking and Finance 2024, Annual Meeting of the German Finance Association 2024, University of Innsbruck and University of Linz for valuable comments. The computational results presented here have been achieved (in part) using the LEO HPC infrastructure of the University of Innsbruck.

1 Introduction

In the past decade, AI has made major leaps forward and has found its way into the fund management industry. Especially during the past year, funds that apply certain AI-enhanced strategies have gained popularity among investors. The Amplified AI-Enhanced Equity ETF (or short by its ticker *AIEQ*) was launched in October 2017 and has contributed to that hype by making AI-enhanced strategies available to the general public through a retail investor-friendly ETF structure. In its inaugural month alone, the AIEQ collected 70 million U$D from investors. According to Google's global interest index, interest in "Artificial Intelligence" was rated 7 (out of 100) in November 2017 (the month after the launch of the AIEQ fund). As of February 2024, this index has reached its all-time high, defined as 100. This has led to the emergence of more AI funds. These funds claim to apply Artificial Intelligence at various stages of their portfolio management process to improve their overall performance. Yet, despite their increasing popularity, very little research has been conducted on whether these self-proclaimed AI funds deliver on their promise to yield better investment returns for investors. In this article, we provide new insights on whether such funds that market themselves as AI-enhanced provide their investors with any advantage.

While there is extensive literature proving that AI strategies lead to improved portfolio allocations (see, for example, Gu et al., 2020, Freyberger et al., 2020, Cong et al., 2021, Azevedo et al., 2023, Chen et al., 2024, Kelly et al., 2024), it is still questionable whether (1). funds can adopt such AI-enhanced strategies and, if so, as to (2). how much of the reported paper returns from these strategies can be converted into real performance (after accounting for fund fees, implicit costs from turnover, market impact, and costs of initializing AI strategies such as data gathering, training models, etc.). Chen and Velikov (2023) show that most anomalies disappear once one accounts for such external costs. Therefore, if such self-declared AI funds cannot outperform their benchmarks, it could mean that they are not (yet) able to apply AI to their advantage. Alternatively, it could also hint that these funds use the tag AI as a signal in the sense of "AI-washing", especially since our methodology of identifying AI funds relies on reportings from the mutual funds rather than from a third independent institution. Likewise, Heath et al. (2023) find that while SRI investing does select firms with higher ESG scores into their portfolio, they do not engage in urging those firms' behaviour into improving their ESG criteria even further. The authors refer to that behaviour as "impact washing".

We take an investor's viewpoint to identify AI funds. As such, we limit ourselves to publicly available information. In the 497K filings, funds must report (amongst other things) a description of their investment objective and methodology. We screen those filings for AI-related keywords, which gives us a sample of funds that actively claim to employ AI at some stage of their investment process. We then test whether AI's advantages over human fund managers translate into higher yields compared to conventional funds and the market benchmark. Our findings indicate that the AI-labelled funds investigated do not outperform a market portfolio. However, when compared to peer funds that are matched upon having similar fund characteristics, AI funds exhibit comparable performance. A portfolio comprising equal weights of AI funds yields a SR of 0.129 over the observation period. In comparison, a similar portfolio composed of conventional benchmark funds, which we match based on similar fund attributes, yields a SR of 0.133 during the same period, with the difference in SRs being statistically insignificant. We subsequently proceed to separate funds into sub-portfolios to investigate whether AI excels at certain tasks of setting up and managing portfolios. Our results reveal that mutual funds manage to incorporate AI more successfully for purposes that happen at an early stage of the portfolio creation process. Portfolios incorporating AI for screening undervalued assets based on predefined characteristics outperformed a subgroup of portfolios that utilized AI for optimizing allocations, such as estimating expected returns and variances. The former group of portfolios demonstrated larger returns compared to the latter.

In a second step, we investigate whether the observed returns were due to skill or luck and delve into the stock selection and timing abilities of AI funds. Leveraging AI technology allows funds and managers to process unprecedented amounts of data when making investment decisions. Kelly et al. (2024) theoretically demonstrate that "complex" models, with parameters exceeding the number of observations, should outperform simpler models in terms of return predictability. Therefore, one would expect higher skill levels for the presumed AI funds. We decompose a fund's excess return following Pástor et al. (2020) into two components: a fund-specific variable capturing the overall skill level of the fund and a second component, namely a function that measures how actively a given fund exploits this specific set of skills. Building on this decomposition, we test two additional hypotheses. First, we investigate if AI funds exhibit a considerable skill level, further distinguishing between market timing and stock picking as advocated in Kacperczyk et al. (2014). And consequently, we examine how actively an AI fund applies that skill. Our results in Section 3 reveal that, on average, AI funds display lower activity levels than their peers, especially those that apply AI at earlier stages of portfolio management. Regarding the skill itself, our results reveal that AI funds exhibit timing abilities. This ability is associated with an additional monthly excess return of 44 basis points (bps) over their rivals. However, this advantage is offset by an inferior stock-picking ability worth 57 bps in monthly return loss.

We subsequently investigate how money flows in and out of AI funds. Sirri and Tufano (1998) mention that investors buy funds that are easier to find. As such, AI might serve as a quality signal. Nevertheless, advertising a fund with AI-enhancing strategies alone does not lead to any significant future inflow of money. In line with numerous previous studies, we find that fund flow relies predominantly on past performance and its expense ratios instead (Gruber, 1996, Jain and Wu, 2000, Huang et al., 2007, to name a few).

We contribute to the literature in several ways. Previous studies proxy costs and limitations of AI strategies and, as a result, only report paper returns. Following a more hands-on approach, we can consider the outcomes of (the subset of funds that claim to use) AI strategies after considering all costs and expenses. Like Chen and Ren (2022) we find that AI funds do not outperform a market benchmark or a benchmark of alike non-AI rival funds. This aligns with Chen and Velikov (2023) who find that three effects, namely 1) transaction costs, 2) post-publication decay (McLean and Pontiff, 2016) and 3) liquidity implications of improved trading technology (Chordia et al., 2014), render most of the AI strategies they investigate unprofitable. We extend the findings of Chen and Ren (2022) to further distinguish the returns into subclasses of AI funds based on the stage at which AI is presumably applied. We find that there are certain tasks for which the considered AI funds already managed to exploit the possibilities of AI and such at which they still lag behind. These results also relate to the main claim of Edmans (2023). The author argues that considering ESG should be for investing as the ESG criteria are highly relevant to the firm's value. However, it is nothing special in that one should also consider other drivers of value equally. Likewise, our results might conclude that AI, rather than its labelling, might be viewed as adding value by screening the market for profitable investments but that one should not entirely rely on AI in the portfolio management context. By splitting up the AI funds into different sub-categories, we reveal that AI funds that apply AI to optimize their allocation trade considerably more and invest in less liquid securities than AI funds that only apply AI to screen the market. This is further evidence for the findings of Avramov et al. (2023), who argue that ML strategies characterize themselves by extremely high turnover or trading in high limits-to-arbitrage markets (such as markets with high volatility or low liquidity), which cuts down strongly on the potential real returns of these ML strategies. Concerning "AI washing", we demonstrate that the tag AI does not lead to additional inflows. Contrary to ESG, where AI leads to additional flows, inducing price pressures and inelasticity of demand (Beck, 2021). Additionally, we also provide an alternative benchmark by using quantitative funds, putting the results of Miguel and Chen (2021) into perspective. We further argue that comparing AI funds to quantitative funds poses a more ceteris-paribus interpretation. Besides that, we expand the observation period and the cross-section, which in the analysis of Chen and Ren (2022), for instance, only consists of 24 months of observations with 14 AI funds considered. Yet, our sample still faces challenges, such a a potential incubation bias which we are not able to resolve due to a still rather small sample size.

This study proceeds as follows: We provide a concise overview of our datasets and how we gather them in Section 2, with a more extensive explanation in Appendix A. Additionally, in Section 2, we outline how we match benchmark funds to each identified AI-labeled fund and compute the measures used to quantify the effects of AI in portfolio creation and optimization. The subsequent sections delve into our main analysis and present the results in Section 3 and Section 4. In Section 5, we revisit parts of the analysis and alter some of the specifications used in the two previous sections to test the robustness of our results. Finally, Section 6 summarizes our findings.

2 Data & Methodology

2.1 Data

For US mutual funds data, we rely on the CRSP Survivor-Bias-Free US Mutual Fund Database. Fund attributes for the characterization of quantitative funds are from the Lipper Hindsight Database, accessed via Refinitiv [1] Eikon. Stock prices, returns, and dividends are from the monthly CRSP files. Balance sheet and income statement data at the firm level are from COMPUSTAT. We

[1] now LSEG

obtain factor returns and the risk-free rate from Kenneth French's website[2]. Inflation is the change in the Consumer Price Index (All Urban Consumers)[3] from the US Bureau of Labor Statistics. Data on the US business cycle is from the National Bureau of Economic Research (NBER)[4].

2.2 Identifying AI funds

We rely on SEC filings of the funds at the CIK level, which we subsequently merge with the CRSP mutual fund database. We especially focus on SEC filing 497K. The Summary Prospectus 497K contains general information about the respective fund, such as the fee structure, assigned manager, a fund description, or an outline of its applied methodology. The latter one is the one we are interested in. We string-search a predefined library of keywords in the Summary Prospectus. For all matches, we read through the description of the fund and its methodology and decide, based on the provided information, whether the fund tries to enhance its performance with the help of AI. We only label a fund as an AI fund if we are certain that AI is employed[5]. In ambiguous cases, we abstain from adding the fund to our sample of AI funds. In total, we screen through 155,630 filings from 44,582 distinct mutual funds. Table 1 summarizes how we obtain our sample of AI funds. For a more detailed description of the filters we apply, the keywords we matched in our final sample and other additional information, we refer to Appendix A.

We further sub-classify AI funds into two sub-groups based on the stage of portfolio management at which the fund resorts to AI. The first group applies AI to screen the market for an investable universe or mispriced constituents. This can be, for example, discovering all stocks that operate in a given industry to create an enhanced industry ETF. Another example would be to use AI to uncover all stocks with a certain kind of mispricing that is not explicitly visible from balance sheet data. Zhu (2019), for instance, finds that funds that incorporate such alternative data, for example, satellite data or consumer transactions, invest and divest more efficiently. Similarly, Grennan and Michaely (2021) and Dugast and Foucault (2018) also suggest that alternative data collected

[2] `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

[3] Series Id: CUUR0000SA0

[4] `https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions`

[5] For instance, we found a fund that claims to employ "LASSO" in its portfolio management process. One would assume they apply a "least absolute shrinkage and selection operator" (LASSO) regression to identify suitable predictors of future premiums. However, upon further investigation, we found that this fund uses the acronym "LASSO" in the sense of "Long and Short Strategic Opportunities". We eliminate such "false positives" from our sample of AI funds.

Table 1: **Selecting Criteria to identify AI-managed funds:** We obtain Summary Prospectus 497K from EDGAR. Panel A shows the filters that we use to identify which of them only use AI keywords in their name or descriptions, which of them only invest in companies that operate predominantly in AI and which funds use AI somewhere along their investment process. Panel B reports the number of matches in the filings that we obtain with our keywords.

Panel A: Selecting Funds	**N**
Number of screened filings	155,630
Number of unique funds	44,582
Filtering Step 1:	
Fund Filing contains AI keywords	669
Filtering Step 2:	
Fund uses AI at some stage of their investment process	103
Filtering Step 3:	
Fund has a connected CRSP Portfolio Number	85
Filtering Step 4:	
Account for subclasses of funds	67

Panel B: Matched Keywords after Step 2	**N**
Artificial Intelligence / AI	70
Machine Learning	36
Natural Language Processing / NLP	35
Deep Learning	5
Ai-Enhanced	4
AI Powered / AI-Powered	4
Neural Networks	3
Cluster Analysis	1
Data Mining	1
Natural Language	1
Predictive Analytics	1

by fin-techs or hedge funds may enhance price informativeness. This group is labelled *Screening* in our results. The second sub-group includes AI in a later stage of portfolio construction. An example would be to use AI to enhance predictions about future returns: Chinco et al. (2019), for instance, find that their LASSO regressions select predominantly stocks that contain news about fundamentals, leading to improved out-of-sample return predictions and subsequently to higher Sharpe Ratios. Our results refer to this kind of fund as *Optimization*.

The methodology above allows us to identify AI funds based on their reporting. This, however, also implies that We may only find those funds that actively mention that they employ AI to enhance their portfolio. This might leave us with a sample that inherits a selection bias, which may be coined "AI-enhanced-*labelled*". Funds that claim to apply "statistical methods" or "quantitative analysis" could potentially also utilize AI, ML or DL in their investment process. However, as long as they do not explicitly mention this within their SEC filings, our methodology is not capable of distinguishing between funds that apply "statistical methods" with a simple OLS and those that, for instance, compute multiple-layered neural networks to update any priors on expected returns. Since this constitutes an ambiguous case, we drop this fund from the list of identified AI funds. As a result, we identify those funds that explicitly claim that AI is part of their portfolio management process. Thus, our results should be interpreted as how the label "AI" influences returns or fund flows rather than whether AI affects portfolio returns. To achieve the latter, one must proceed with additional data about the funds and dive deeper into their investment strategies. We refrain from this in this analysis and leave such a task to future research.

2.3 Identifying benchmark funds

Besides the best possible identification of AI-labelled funds, the second crucial aspect of investigating realized AI-related returns is the selection of an appropriate counterfactual. In that regard, we adapt the methodology of Hoberg et al. (2018). The underlying idea is to define a set of fund characteristics and subsequently find "rival" funds that exhibit similar characteristics. The fund characteristics are derived from the individual characteristics of their respective holdings. Based on (orthogonalized) ranks and z-scores, they identify rival funds as funds with the smallest Euclidean distance in an n-dimensional space, where n is the number of fund characteristics used. We alter this methodology into a two-step approach. First, we filter all funds that operate within the same fund style provided by CRSP[6]. We ensure we do not compare funds that follow different objectives like this. We do not consider fund style in the matching vector since the fund style is measured by a categorical variable, meaning that one cannot interpret distances between the categories. However, the methodology relies on those *metric* distances. In the second step, we search for similar funds based on fund properties.

For the main analysis, we compute z-scores based on the value-weighted stock characteristics in each fund. We winsorize all matching variables at the 1% and the 99% levels. We then orthogonalize the respective z-scores. This gives us the matching vector $C_{f,t} = \begin{pmatrix} z_{lme} & z_{lmom} & z_{ldy} \end{pmatrix}$, where $lme \equiv log(ME)$ is the logarithm of the market equity of a fund, $lmom \equiv log(1 + r_{t-2,t-12})$ and ldy is the log dividend yield. We then adapt the cutoff distance $d^* = 8.858\%$ from Hoberg et al. (2018). However, according to them, the results are robust to the choice of d^*. We, furthermore, impose a minimum of ten rival funds to achieve a broader sample of comparable fund performance. This process identifies several benchmark funds for the previously identified AI-labelled funds. We then aggregate these rival funds to create a portfolio of rival funds that better mimic the characteristics

[6]The CRSP Style Code consists of up to four letters, each representing a classification dimension. Level 1, the first letter, differentiates, e.g. between equity, fixed income, mixed and other funds. With each extra level, the investment objective becomes more granular. We use the full four-level classification to match rival funds. In the robustness checks, we use quantitative funds as benchmark funds. Due to the low number of identified quantitative funds, we filter at the third level. Otherwise, only 47 of 67 AI-labelled portfolios are matched to any quantitative funds.

of a specific AI fund. We employ two aggregations that differ only in the weights they assign to the rival funds, equal weights and distance weights. The latter uses the inverse Euclidean distance to the respective AI fund as weight. In such a distance-weighted portfolio, the rival funds, which are more similar to the respective AI fund, receive a higher weight and vice versa.

2.4 Defining Measures of Skill and Activeness

In the model motivated by Pástor et al. (2020), the expected adjusted gross return is defined as the product of a fund-specific constant representing skill and a function that quantifies how actively that skill is measured. To measure the respective skill of a self-proclaimed AI fund or a rival fund, we adapt the timing and picking skill defined by Kacperczyk et al. (2014) as

$$Timing_t^j = \sum_{i=1}^{N^j} \left(w_{i,t}^j - w_{i_t}^m \right) \left(\beta_{i,t} R_{t+1}^m \right) \tag{1}$$

and

$$Picking_t^j = \sum_{i=1}^{N^j} \left(w_{i,t}^j - w_{i_t}^m \right) \left(R_{t+1}^i - \beta_{i,t} R_{t+1}^m \right), \tag{2}$$

where N is the number of stocks in a fund's portfolio, j. $\left(\beta_{i,t} R_{t+1}^m\right)$ in the *Timing*-equation is the systematic part of the future returns. $\left(R_{t+1}^i - \beta_{i,t} R_{t+1}^m\right)$ in the *Picking*-equation, on the other hand, measures the idiosyncratic part of the future return of asset i within the portfolio. $\left(w_{i,t}^j - w_{i_t}^m \right)$ captures the over- or under-weighting that the portfolio has in a given stock compared with the market benchmark, meaning how much the portfolio exposes itself to the systematic or idiosyncratic part of the returns. This means that a fund with high "timing" ability over- (under-) weights high (low) beta assets before market upswings and, vice versa, over- (under-) weights low (high) beta assets before market downturns. Similarly, a fund with a high "Picking" ability would build up positions in stocks with high idiosyncratic returns in the advent of market increases and reduce them in sight of market downswings.

For the activeness measure, we stick to the measure proposed by Pástor et al. (2020)[7]

$$Activeness = TL^{-1/2}, \tag{3}$$

[7]Equation 37 of Pástor et al. (2020)

where T is the turnover scaled by the NAV. $L = (\Sigma_{i=1}^{N} \frac{w_i^2}{m_i})^{-1}$ is portfolio liquidity where N is the number of stocks in a given portfolio, w_i and m_i are the respective weights of constituent i in the fund's portfolio and the market portfolio. Activeness is supposed to be positively related to turnover – to generate excess returns, an active fund has to move away from the (market) benchmark – and negatively with L, as mispriced stocks tend to be less liquid and smaller.

3 Performance

As mentioned in the previous section, we identify 67 funds that apply artificial intelligence for screening the investable universe to identify suitable fund constituents or, in the subsequent stage, for integrating selected stocks, bonds, etc., into a portfolio.

Figure 1 gives a first overview of the identified AI-labelled funds. The left plot illustrates the distribution of these funds' size (measured by Assets under Management (AuM)). At the end of 2017, there were only two self-proclaimed AI funds, AIEQ and a second one, which did not advertise their adoption of AI as aggressively as AIEQ. Subsequently, AIEQ experienced growth, coinciding with the emergence of numerous new funds over time, as depicted in the right plot. The solid grey line represents the total number of unique funds utilizing AI, while the dashed black line adjusts for share classes and funds sharing the same portfolio. Those new funds started with modest AuM[8]. Relative to these smaller funds, AIEQ's magnitude became significant, resulting in a notable spread between the 5th and 95th percentiles, particularly from 2020 to 2022. The issue with smaller funds is that overhead costs can be a substantial factor in performance after fees, which we want to investigate. Additionally, we posit that funds may not accord adequate attention to portfolios below a certain threshold, potentially hindering the ability to generate outperformance. Evans (2010) refer to this as the incubation bias and point out that one might eliminate the effect of this bias by removing the first three years of return observations[9]. Nevertheless, this is impossible for our sample since it would leave us hardly any fund-month return observations. Furthermore, it is also common to include a size filter of 15 million U$D as the reported returns might be upward biased (see, for example, Elton et al. (2001), Chen et al. (2004), Pástor et al. (2015)). Since this would again eliminate half of our observations, we exclude all AI-labeled funds below a minimum

[8]The smallest AuM reported by CRSP is 0.1, indicating a fund size up to 100,000 USD.
[9]see Evans (2010), p. 1584

size of five million. However, as a robustness check, we revisit the subsequent analysis, including smaller funds, in Section 5.

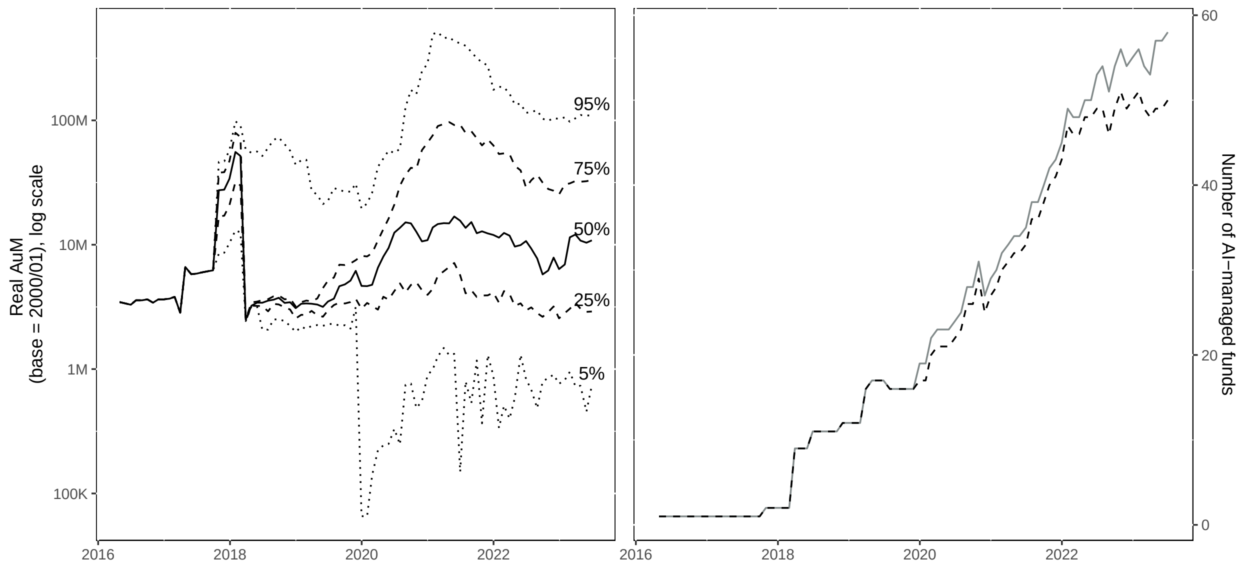

Figure 1: **Distribution AuM & Number of AI-managed funds/portfolios:** The left plot depicts the distribution of the Assets under Management (AuM) across all sub-classes of the AI-managed funds. The right plot shows the number of AI-managed funds (grey solid line) and unique portfolios (i.e. funds adjusted for different share classes of the same portfolio; black dashed line) over time. We further consider portfolios with different portfolio numbers identical if their time-series correlation of daily returns is above 99.9 % and if the same company offers them. For more extensive information in that regard, we refer to Appendix A.

Our primary inquiry centres on whether AI usage furnishes fund managers with an advantage over counterparts eschewing artificial assistance. As an initial step, we examine the cumulative returns of such funds and compare them with their respective peers. Figure 2 displays those returns. We aggregate AI-labelled funds into portfolios and compute the respective returns. The left plot delineates the aggregated returns over time (solid line) for an equally weighted portfolio (after deducting management fees and 12b fees), while the right plot does so in an AuM-weighted manner. With the dot-dashed line, we compare the returns of the AI portfolios with their matched non-AI peers. Similarly, the dotted line represents the market benchmark, defined as the equal- and value-weighed CRSP stock market portfolio. Examination of the size-weighted outcomes reveals that between 2020 and 2022, AI funds potentially outperformed their fund and market benchmarks. However, cumulative returns diminish over time, with performance falling below that of the stock market portfolio. Yet, the value-weighted portfolio of the AI-tagged funds still keeps up with the accumulated performance of the non-AI equivalents. The equal-weighted portfolios exhibit high relevance since the AIEQ makes up to 30% of the AuM of AI-managed portfolios at certain points. Again, the AI funds perform similarly to their conventional analogues. Additionally, the small AI-managed funds seem to outperform the equal-weighted stock market.

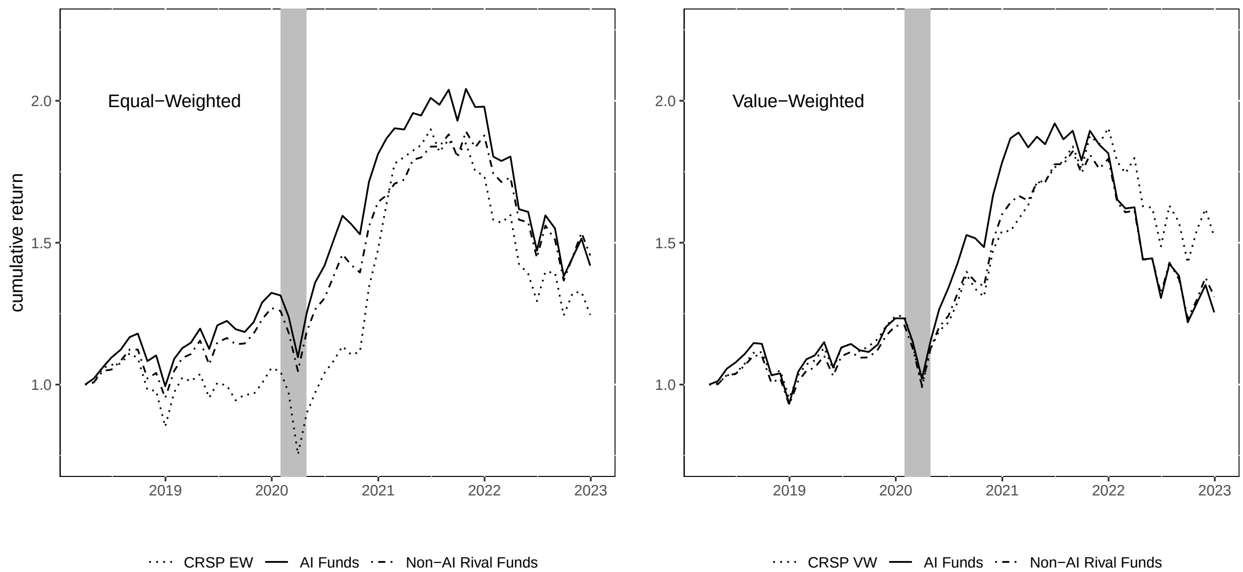

Figure 2: **Return over time:** We aggregate the AI-managed and the non-AI-managed funds into equal- and value-weighted portfolios and compute the respective cumulative returns. The left panel represents the equal-weighted aggregation, and the right is value-weighted. The rival funds were identified using orthogonal z-scores based on Hoberg et al. (2018). The dotted line represents the CRSP Equal-/Value-Weighted Stock Market. We require AI-labelled funds to have a minimum fund size of five million. NBER recessions are shaded in grey.

Table 2 delves deeper into these returns. Panels A and B report the average excess returns for equal and value-weighted AI and rival funds. Since we obtain several potential rivals for each AI fund, we aggregate them with equal weights (ew) and based on their relevance with the inverse of the Euclidean distance (dw) to the respective AI fund. These portfolios could be considered fictitious, AI-fund-style, mimicking non-AI counterfactual benchmark funds. We then aggregate these rival fund portfolios, representing one fund for each AI fund, into equal-weighted portfolios (Panel A) and value-weighted portfolios (Panel B). The average monthly excess returns over the risk-free rate of the equal-weighted AI-labelled funds in Panel A amount to 71 basis points (bps). Depending on the aggregation method, AI funds, on average, slightly under- or overperform their peers regarding average excess returns. Equal-weighted rival funds yield 0.3 bps less per month, while distance-weighted rival fund portfolios generate an average monthly excess performance over AI funds of 0.8 bps. A similar pattern is observed in Panel B for the value-weighted aggregation of AI and rival funds. AI funds generate an aggregate average monthly excess return of 52 bps.

Their rivals yield 55 and 53 bps for equal-/distance-weighted alternatives. The resulting returns differences are low, and a t-test indicates no significance for any of them. Panel C further investigates the Sharpe Ratios (SR). AI-labelled funds exhibit smaller risk-adjusted returns compared to conventional funds. The SR of equally-weighted AI funds stands at 0.129 since October 2017, while non-AI rivals exhibit SRs of 0.136 for equal-weighted rivals and 0.133 for distance-weighted ones. A significance test of the differences in the average risk-adjusted returns, following Ledoit and Wolf (2008) using the Parzen kernel fails to reject the hypotheses that the SRs differ for all combinations of equal- and value-weighted AI funds and equal- and distance-weighted connected peer funds. Our results align with the previous literature: Zhang et al. (2023) find similar results for the Chinese Mutual Fund market. Their studied "Big Data" funds have average Sharpe Ratios of 0.123, far below the 0.208 of the conventional funds benchmark. The average monthly return among those big data funds is 0.70 %, whereas traditional funds average 1.28 % over the same period.

In Table 3, we explore potential alphas and the loadings of AI funds on factors, interpreted as fund styles, via several factor regressions based on monthly returns:

$$r_{j,t}^{AI} - r_t^f = \alpha_j + \sum_{k=1}^{K} \beta_{j,k} f_{k,t} + \varepsilon_{j,t}, \tag{4}$$

where $r_{j,t}^{AI}$ is the return of AI-labelled portfolio j in month t, r_t^f is the risk-free rate and r_t^m is the value-weighted market return of CRSP firms. $f_{k,t}$ is the return of zero-investment factor-mimicking portfolios constructed based on size, book-to-market, operating profitability, investment and past returns, respectively. We vary K from 1 to 6 and denote these variations as CAPM, FFC4 and FFC6, when $f_{k,t}$ is $\left\{\left(r_t^m - r_t^f\right)\right\}$, $\left\{\left(r_t^m - r_t^f\right) \quad SMB_t \quad HML_t \quad MOM_t\right\}$, and $\left\{\left(r_t^m - r_t^f\right) \quad SMB_t \quad HML_t \quad RMW_t \quad CMA_t \quad MOM_t\right\}$ respectively. Columns (3)-(5) and (8)-(10) give the results of the regressions for equal and value-weighted portfolios of AI-labelled fund returns. Most importantly, none of the alphas are significant across all columns for equal-weighted or value-weighted AI portfolios. This result blends in with the results of Zhang et al. (2023), who study big data funds in the Chinese mutual fund market. Their identified big data funds do not produce any improvements in alphas compared to traditional funds. The reported average alpha is significantly smaller than those of conventional funds. Columns (4), (5), (9) and (10) reveal that AI-branded funds tilt themselves towards small-sized firms. Furthermore, the results show that

Table 2: **Performance metrics:** We evaluate the performance of AI-enhanced funds compared to matched rival funds. Suitable rival funds are matched based on the methodology used in Hoberg et al. (2018). For each AI fund, we construct a portfolio comprising similar benchmark funds. We select the potential benchmark funds from two different samples: 1. all available funds in the CRSP universe that operate at the same time as the AI funds (which we denote "Rival Funds") and 2. a sub-selection of funds that claim to apply rule-based mathematical models in their portfolio creation process (which we call "Quant Funds"). Subsequently, these benchmark funds are aggregated into two distinct portfolios: one employing equal-weighted (*ew*) methodology, and the other utilizing distance-weighted (*dw*) methodology. In the *dw* approach, weights are determined based on the inverse of Euclidean distance, assigning higher weights to closer funds and lower weights to less similar ones within the benchmark portfolio. "Market" denotes the results from the CRSP market portfolio with the corresponding aggregation that is applied in each Panel. Panel A and B present the average *excess* returns and return differences for both AI and rival portfolios. We calculate these metrics using both equal-weighted and value-weighted portfolios of AI funds and their respective rival portfolios. We multiply the returns by 100. Returns, and consequently, the return differences in columns (4)-(6) are reported in per cent. N reports the number of months. Panel C displays the Sharpe Ratios (SR) and their differences. Columns (1)-(4) present SR for portfolios with equal weights, while columns (5)-(8) consider weights based on the total net assets (TNA) of the constituting funds. Standard Errors for SR differences are computed following Ledoit and Wolf (2008), based on the prewhitened Parzen kernel. Throughout Panel A-C, we compute p-values based on a t-distribution.

Panel A: Equal Weighted Returns

				$r^x_{AI} - r^x_{Rv}$			
	$\overline{r^x}$	se	p-val	Δr^x	se	p-val	N
AI funds	0.714	0.701	0.308				62
Rival funds (ew)	0.717	0.670	0.284	−0.003	0.969	0.997	62
Rival funds (dw)	0.707	0.673	0.294	0.008	0.972	0.994	62
Quant funds (ew)	0.408	0.612	0.505	0.306	0.931	0.743	62
Quant funds (dw)	0.410	0.612	0.504	0.305	0.931	0.744	62
Market	0.588	0.912	0.519	0.126	1.150	0.913	62

Panel B: Value Weighted Returns

				$r^x_{AI} - r^x_{Rv}$			
	$\overline{r^x}$	se	p-val	Δr^x	se	p-val	N
AI funds	0.525	0.755	0.487				62
Rival funds (ew)	0.553	0.680	0.416	−0.028	1.016	0.978	62
Rival funds (dw)	0.536	0.684	0.433	−0.010	1.019	0.992	62
Quant funds (ew)	0.487	0.635	0.443	0.038	0.987	0.969	62
Quant funds (dw)	0.476	0.635	0.454	0.050	0.987	0.960	62
Market	0.796	0.701	0.256	−0.270	1.030	0.793	62

Table continues on next page

Table continued from previous page

Panel C: Sharpe Ratios

	Equal Weighted Portfolios					Value Weighted Portfolios				
	SR	SR_{ann}	ΔSR	se	pval	SR	SR_{ann}	ΔSR	se	pval
AI	0.129	0.448				0.088	0.306			
Rivals (ew)	0.136	0.471	−0.007	0.027	0.807	0.103	0.358	−0.015	0.035	0.671
Rivals (dw)	0.133	0.462	−0.004	0.025	0.879	0.099	0.345	−0.011	0.032	0.727
Quant (ew)	0.085	0.293	0.045	0.069	0.519	0.097	0.338	−0.009	0.078	0.907
Quant (dw)	0.085	0.294	0.044	0.069	0.524	0.095	0.329	−0.007	0.076	0.930
Market	0.082	0.284	0.048	0.065	0.465	0.144	0.500	−0.056	0.070	0.428

these funds tilt towards growth stocks and weak stocks with low operating profitability.

Additionally, we regress excess fund returns on excess rival returns, *ew* and *dw*,

$$r_{j,t}^{AI} - r_t^f = \alpha + \beta_j^{MKT}\left(r_{j,t}^{Rv} - r_t^f\right) + \varepsilon_t, \tag{5}$$

where $r_{j,t}^{Rv}$ is the aggregated return, *ew* or *dw*, of the rival funds of AI fund j at month t. The results are reported in Table 3, columns (1) & (2) and (6) & (7), respectively. The idea is to see whether AI funds can generate some alpha beyond conventional funds, in the sense that Jensen's alpha investigates an outperformance over the market return. Once again, the constants are insignificant, indicating that AI funds fail to generate alpha beyond their matched rival funds. This finding mirrors those from Table 2.

Given the absence of substantial return differences between AI-labelled funds and the stock market or other funds and the lack of significant alphas, we delve deeper into whether AI may be more beneficial at the first or second stage of portfolio management. Figure 3, thus, disentangles the returns reported in Figure 2 into funds that advertise themselves to exert the power of AI in iden-

Table 3: **Fama/French/Carhart Factor Regressions:** Columns (1)&(2) and (5)&(6) report the coefficients obtained from regressing the excess returns of AI funds onto the excess returns of equal-weighted and value-weighted portfolios of Rival funds. Rival funds are identified using the methodology outlined in Hoberg et al., 2018. For each AI fund, we select a minimum of ten similar funds and aggregate them into a portfolio designed to replicate a counterpart fund with comparable characteristics to the AI funds. This aggregation employs both equal weights (ew) and distance weights (dw), where we compute weights for dw with the inverse of the Euclidean distance of each rival fund. In columns (3)-(5) and (8)-(10), we follow the conventional approach of factor regressions, reporting the coefficients obtained from regressing the equal-weighted and value-weighted AI returns onto the factor returns. Columns (1)-(5) employ equal weights on AI fund returns, while columns (6)-(10) use fund size (TNA) as weights. P-values are computed based on a t-distribution with $n-k-1$ degrees of freedom.

	Dependent variable:									
	AI fund returns (ew)					AI fund returns (vw)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	−0.025	−0.011	−0.103	−0.084	0.007	−0.077	−0.055	−0.296	−0.273	−0.186
	(0.146)	(0.140)	(0.238)	(0.148)	(0.161)	(0.196)	(0.185)	(0.303)	(0.170)	(0.170)
Rivals (ew)	1.031***					1.089***				
	(0.019)					(0.022)				
Rivals (vw)		1.027***					1.083***			
		(0.017)					(0.022)			
Mkt-Rf			0.980***	0.934***	0.971***			1.036***	0.982***	1.025***
			(0.030)	(0.030)	(0.026)			(0.044)	(0.035)	(0.026)
SMB				0.288***	0.173**				0.414***	0.294***
				(0.059)	(0.066)				(0.089)	(0.070)
HML				−0.184***	−0.163***				−0.240***	−0.233***
				(0.046)	(0.055)				(0.063)	(0.078)
RMW					−0.260***					−0.284***
					(0.057)					(0.060)
CMA					0.062					0.102
					(0.079)					(0.091)
MOM				0.007	−0.020				0.041	0.007
				(0.066)	(0.065)				(0.080)	(0.076)
Observations	62	62	68	68	68	62	62	68	68	68
R^2	0.970	0.973	0.917	0.948	0.956	0.959	0.960	0.886	0.936	0.946
Adjusted R^2	0.970	0.972	0.916	0.944	0.952	0.959	0.959	0.884	0.932	0.940

Note: Newey and West (1994) Standard Errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01

tifying the relevant constituents to invest in, and those that employ AI to arrive at a weighting for the previously selected stocks, bonds, etc. We tag the first group "*Screening*" while we refer to the latter as "*Optimization*". Again, the two solid lines represent the performance of the respective AI fund portfolios in the *Screening* and the *Optimization* group. The dot-dashed lines represent the matched rival fund returns. The dotted lines again compare the funds to the aggregated stock market. The left plot illustrates returns when all funds have identical weights in their respective portfolios, whereas the right plot weights funds based on their AuM. Only the *Screening* sub-portfolio with equal weights manages to outpace the market benchmark. The *Optimization* AI funds perform even worse than the sub-portfolio of their rivals. In the value-weighted case, both AI subgroups of funds do not manage to surpass the stock market. Apart from that underperformance relative to the aggregate stock market, the results hint that AI might seem to enable its fund managers to find better constituents for their portfolios: The AI-*Screening* portfolios yield higher cumulated returns than the conventional benchmark funds, especially in the equal-weighted case that gives less weight to the AIEQ. On the other hand, the AI funds may not yet have managed to exploit the full potential of AI regarding portfolio optimization: These *Allocation* funds do not exhibit considerable excess returns compared to their peers and perform much worse than the market.

Figure 3 suggests that AI might play a beneficial role in pinpointing potentially profitable assets for portfolio construction. In contrast, AI funds seem to fail to use the potential of AI regarding allocating weights to a predefined universe. To further explore this, we study whether those AI funds demonstrate skill and effectively utilise it. Pástor et al. (2020) posit that a fund's excess return, $a = \mu g (T, L)$, comprises a fund-specific constant (μ) representing its skill level, and a measure (g) indicating how actively the fund applies its skill. Activeness is gauged by turnover (T) and a liquidity/diversification metric (L). Activeness increases in turnover T. The rationale is that an active fund has to trade away from the benchmark to generate returns over its benchmark. Activeness is negatively related to liquidity L, because mispricing is greater among less-liquid and smaller stocks. We recompute the fund activeness measure for the funds and then regress fund activeness on the Dummy variable for AI funds, including time-fixed effects. The results are in columns (1) and (2) of Table 4. AI-labelled funds tend to be less active than their peers, meaning that they trade less or select more liquid stocks than their peers. This is somewhat surprising, as previous studies find the opposite (see, for example, Avramov et al., 2023). However, column (2) results reveal that the sub-category of Screening funds drove this result. AI funds that claim

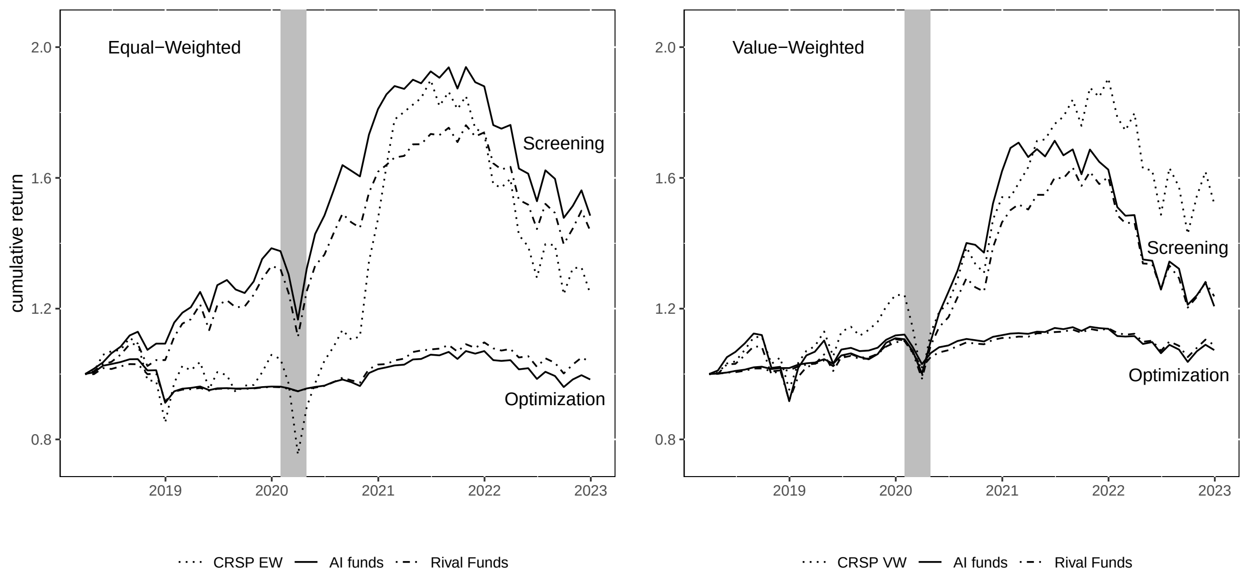

Figure 3: **AI to screen the universe and allocate portfolios:** We sort AI-labelled funds into two categories: The first uses AI to screen the market and establish an investment universe of potentially profitable constituents. We label this group *Selection*. The second group of funds employs AI to arrive at desired portfolio weights. This group has the tag *Allocation*. The solid lines are the cumulative returns of an equal- and value-weighted portfolio of those groups of AI funds. The dashed lines are the cumulative returns of the portfolio of benchmark funds, which we again match using orthogonal z-scores based on Hoberg et al. (2018). The left plot depicts the returns if all funds have the same weights in their respective portfolios. The right plot weights the funds based on their AuM. Further information on how we select funds and find appropriate benchmark funds is given in Appendix A. US recessions, as dated by the NBER, are highlighted in grey.

optimal portfolio weighting with AI are more active, which aligns with previous findings. In Table 3, we find that AI funds tend to hold smaller stocks, which also tend to be less liquid. This finding aligns with our results since the activeness measure also comprises holdings in less liquid assets.

Furthermore, we assess the general skill level of the AI and Non-AI funds using the two measures motivated in Kacperczyk et al. (2014), *Timing* and *Picking*. We explain the measures in Section 2. Timing evaluates how well a portfolio times its holdings (with over-/underweighting) to the systematic component of returns. Picking captures the idiosyncratic part of the portfolio's holdings. We compute the betas of the assets using rolling windows of the previous 12 months. We regress the returns resulting from market timing and stock picking onto an AI dummy, that is one, if the respective fund is flagged as an AI fund. The results are in columns (3) and (4) for Timing and (9) and (10) of Table 4 for Picking. In the remaining columns, we added further control variables as covariates. Contrary to other parts of the previous analysis, we do not use *ew* or *dw* portfolios of the set of rival portfolios but consider them as they are. Across all specifications, AI funds exhibit significantly better timing abilities than their conventional rival funds. Their over/underweighting attributes to monthly returns between 44 bps and 24 bps compared to their non-AI peers, depending on the control variables used. This outperformance is offset by AI funds significantly underperforming their rivals in stock picking. Their inferior picking ability creates negative returns of 57 bps for overall AI-labelled funds. Separating AI-labelled funds into subgroups based on the presumed stage of AI application yields similar results. The AI portfolios have relatively higher timing ability worth 47 and 37 bps in extra return units. This timing return is, nevertheless, offset by the comparably worse picking ability of approximately the same size, producing returns of -58 bps and -53 bps on average. These results continue to hold if we add further explanatory variables again. Timing is positive for the subgroups of AI portfolios with 27 and 46 bps of return, while picking cancels these returns out with -44 bps relative to the matched rivals. These results are opposite to those reported in Chen and Ren (2022), which find relative picking ability and no significant difference in timing ability. However, the significance of the picking ability is low, indicating that extending the time series and cross-section with new AI-managed funds may improve the power of tests, providing a clearer picture of AI-labelled fund skill levels.

4 Persistence and Fund Flows

We follow the classical approach put forward by Carhart (1997) to asses the persistence of AI returns. Given the limited number of AI funds, we categorise them into terciles based on the previous month's returns of all funds (including those that are neither AI funds nor rival funds) instead of deciles. We then hold the sorts constant for the next period (month, quarter or semester[10]) and compute the returns of equally weighted portfolios of each tercile. This process is repeated, rolling forward until the beginning of the next holding period, yielding a time series of non-overlapping returns for tercile portfolios. Additionally, we evaluate the performance of a zero-investment portfolio that takes long positions in the top-performing tercile and short positions in the bottom-performing tercile. We assess the persistence of the equally weighted tercile portfolios by regressing their excess returns onto K factors:

[10]We refrain from investigating persistence over the next year or successive years due to the constraints of our sample, as outlined in Section 2. Our sample is restricted by the inception of the first AI fund in 2017, resulting in a limited number of observations. In the case of FFC6, this would entail $K > T$.

Table 4: **Activeness and Skill:** We run Fixed Effects Regressions with fixed time effects to examine the impact of AI funds on activeness and skill. We employ a dummy variable indicating the use of AI at any stage of portfolio management as the explanatory variable. Additionally, we disaggregate this AI dummy into two distinct stages: AI employed in the stock selection process and AI utilized in determining the weighting scheme. The *Activeness* measure, as proposed in Pástor et al. (2020), is defined as $Activeness = TL^{-1/2}$, where T represents fund turnover and L denotes a measure of portfolio liquidity. Skill is assessed through *Timing* and *Picking* metrics as outlined in Kacperczyk et al. (2014): $Timing_t^j = \sum_{i=1}^{N^j} \left(w_{i,t}^j - w_{i_t}^m \right) \left(\beta_{i,t} R_{t+1}^m \right)$ and $Picking_t^j = \sum_{i=1}^{N^j} \left(w_{i,t}^j - w_{i_t}^m \right) \left(R_{t+1}^i - \beta_{i,t} R_{t+1}^m \right)$. For the abbreviated models, we include only a dummy variable for AI as the explanatory factor. We incorporate two sets of additional covariates: one where matching variables used to identify rival funds are excluded, and another where these variables are included. Standard Errors are clustered following Arellano (1987). We compute p-values based on a t-distribution.

	Dependent variable:													
	Fund Activeness		Timing						Picking					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
D(AI Fund)	−5.4153***		0.0044***		0.0024**		0.0032***		−0.0057***		−0.0020		−0.0027*	
	(2.0725)		(0.0007)		(0.0011)		(0.0011)		(0.0015)		(0.0017)		(0.0015)	
D(Screening)		−9.1351***		0.0047***		0.0020		0.0027**		−0.0058***		−0.0016		−0.0022
		(1.6799)		(0.0010)		(0.0014)		(0.0013)		(0.0020)		(0.0021)		(0.0019)
D(Optimization)		7.5145**		0.0037***		0.0036**		0.0046***		−0.0053***		−0.0033		−0.0044**
		(3.8266)		(0.0010)		(0.0017)		(0.0016)		(0.0015)		(0.0022)		(0.0022)
log(ME)							−0.0024***	−0.0024***					0.0020***	0.0020***
							(0.0001)	(0.0001)					(0.0001)	(0.0001)
log(BM)					−0.0034***	−0.0034***	−0.0062***	−0.0062***			0.0004**	0.0004**	0.0024***	0.0024***
					(0.0003)	(0.0003)	(0.0003)	(0.0003)			(0.0002)	(0.0002)	(0.0002)	(0.0002)
log($r_{2,12}$)							−0.0013	−0.0013					0.0026***	0.0026***
							(0.0009)	(0.0009)					(0.0007)	(0.0007)
log(DY)							0.0015***	0.0015***					−0.0006***	−0.0006***
							(0.0003)	(0.0003)					(0.0002)	(0.0002)
log(Fund Age)					0.0006***	0.0006***	0.0007***	0.0007***			−0.00002	−0.00002	−0.0001*	−0.0001*
					(0.0001)	(0.0001)	(0.0001)	(0.0001)			(0.0001)	(0.0001)	(0.0001)	(0.0001)
Expense Ratio					0.0124	0.0123	−0.0031	−0.0032			−0.0389**	−0.0388**	−0.0142	−0.0140
					(0.0267)	(0.0267)	(0.0251)	(0.0252)			(0.0197)	(0.0197)	(0.0184)	(0.0184)
Turnover					−0.0001	−0.0001	−0.0001	−0.0001			−0.0001*	−0.0001*	−0.0001*	−0.0001*
					(0.0001)	(0.0001)	(0.0001)	(0.0001)			(0.0001)	(0.0001)	(0.0001)	(0.0001)
Groups	6,520	6,520	13,578	13,578	6,658	6,658	6,537	6,537	13,578	13,578	6,658	6,658	6,537	6,537
Observations	151,456	151,456	485,137	485,137	202,005	202,005	197,246	197,246	485,137	485,137	202,005	202,005	197,246	197,246
R^2	0.00005	0.0001	0.0001	0.0001	0.0029	0.0029	0.0161	0.0161	0.0002	0.0002	0.0002	0.0002	0.0097	0.0097
Adjusted R^2	−0.0002	−0.0002	−0.0001	−0.0001	0.0026	0.0026	0.0157	0.0157	0.0001	0.0001	−0.0002	−0.0002	0.0094	0.0094

Note: *p<0.1; **p<0.05; ***p<0.01

$$r_{tc,t} - r_t^f = \alpha_{tc} + \sum_{k=1}^{K} \beta_{tc,k} f_{k,t} + \varepsilon_{tc,t}, \tag{6}$$

where $r_{tc,t}$ is the return of the tercile portfolio or a zero-investment portfolio constructed solely with AI funds or rival funds, respectively, and r_t^f is the risk-free rate. Like Equation 4, we vary the number of factor-mimicking portfolios, K, in $f_{t,k}$.

Busse et al. (2010) conduct similar factor regressions for all US funds and identify a reversal in the performance of poorly-performing funds and momentum in well-performing funds. Panel A, B and C in Table 5 report the intercepts from the corresponding CAPM, four-factor (FFC4) and six-factor (FFC6) regressions. The first three columns are the alphas from sub-sampling solely on AI funds, and the last three columns filter on rival funds. Throughout both groups, AI and rival funds, and for all variations of considered factors, we observe that the bottom and mid-performing funds yield subsequent significant negative alphas. This finding also holds for all investigated holding periods. This means that, contrary to the results in the overall fund sample of Busse et al. (2010), bad and mediocre performance on average persists for at least the following six months.

Concerning the top tercile, previous findings of Carhart (1997) and Busse et al. (2010) suggest positive momentum in fund returns. However, we only observe significant positive future semester alphas within the FFC4 regressions for AI funds. The remaining intercepts of AI-labelled funds are not significant. For the rival funds, we even observe the opposite. All but one of the significant intercepts, all at 1Q and 1S horizons, exhibit a negative sign, indicating short-term reversal rather than momentum.

In the model of Farboodi and Veldkamp (2020), higher computational power enables funds to process more information simultaneously, enhancing price informativeness. However, this can lead to diminishing returns as all market participants react to the same information, resulting in a scenario where assets are bought uniformly. As a result, increasing inflows towards these assets will create upward price pressures towards them. Similarly, Gabaix and Koijen (2022) find that price elasticities in the equity market are small, indicating that flows in and out of the market exert significant price pressure on stocks experiencing substantial flows. The magnitude of this effect is noteworthy, with a \$ 1 inflow generating a five-fold impact on the aggregate market. Additionally,

Table 5: **Persistence of alphas**: We asses the persistence of returns following Carhart (1997): At the beginning of each month, quarter or semester, we sort each fund portfolio into terciles based on their past-month return. We then aggregate the returns to equally-weighted tercile portfolios. Additionally, we compute returns for a portfolio with a long position in the top tercile and a short position in the bottom portfolio. We re-balance the portfolios at the start of each holding period, meaning monthly, quarterly or half-annually. The table reports the alphas from factor regressions on these portfolio returns. 1M - next month, 1Q - next quarter, 1S - next semester. Newey and West (1994) Standard Errors are in parentheses. We compute the p-values based on a t-distribution with $n-k-1$ degrees of freedom.

Panel A: CAPM

	AI			Rivals		
	1M	1Q	1S	1M	1Q	1S
1	-0.011***	-0.0131	-0.0303	-0.0027	-0.0027	-0.0261**
	(0.0027)	(0.008)	(0.0273)	(0.0022)	(0.0064)	(0.0093)
2	-0.0026	-0.0197***	-0.0123	-0.0017*	-0.0052**	-0.0128***
	(0.0028)	(0.0063)	(0.0184)	(0.001)	(0.002)	(0.0018)
3	0.003	-0.011	0.0089	-0.0006	-0.0068*	-0.0048
	(0.0035)	(0.0082)	(0.0198)	(0.0024)	(0.0038)	(0.0074)
3-1	0.013***	-0.0012	0.0412	0.001	-0.0076	0.0222
	(0.0038)	(0.0079)	(0.0301)	(0.0039)	(0.0091)	(0.0135)

Panel B: FFC4

	AI			Rivals		
	1M	1Q	1S	1M	1Q	1S
1	-0.0104***	-0.0084	-0.0145	-0.0026	-7,00E-04	-0.0457***
	(0.0019)	(0.0057)	(0.0205)	(0.0022)	(0.0072)	(0.0042)
2	-0.0029	-0.0215***	-0.0101	-0.0015*	-0.0061	-0.0166***
	(0.0027)	(0.007)	(0.0151)	(0.0009)	(0.0288)	(0.0019)
3	0.0034	-0.0105	0.0572**	-0.0006	-0.0086*	0.0104**
	(0.0032)	(0.0084)	(0.0215)	(0.0028)	(0.004)	(0.0026)
3-1	0.0128***	-0.0048	0.0774	0.001	-0.0112*	0.0635***
	(0.0044)	(0.0102)	(0.0408)	(0.0046)	(0.0057)	(0.0153)

Panel C: FFC6

	AI			Rivals		
	1M	1Q	1S	1M	1Q	1S
1	-0.0091***	-0.0102	0.021	-0.0021	-0.0046	-0.0154
	(0.002)	(0.0064)	(0.0189)	(0.0021)	(0.0046)	(0.0217)
2	-0.0022	-0.0236***	0.0692***	-0.0019*	-0.0083***	-0.0164**
	(0.0035)	(0.0054)	(0.011)	(0.001)	(0.0016)	(0.004)
3	0.0039	-0.01	-0.0042	-0.0014	-0.007**	-0.0442**
	(0.003)	(0.0071)	(0.0067)	(0.0027)	(0.0031)	(0.008)
3-1	0.0119***	-0.0027	-0.0244	-0.0002	-0.0056	-0.0202
	(0.0039)	(0.0112)	(0.1303)	(0.0042)	(0.0055)	(0.0752)

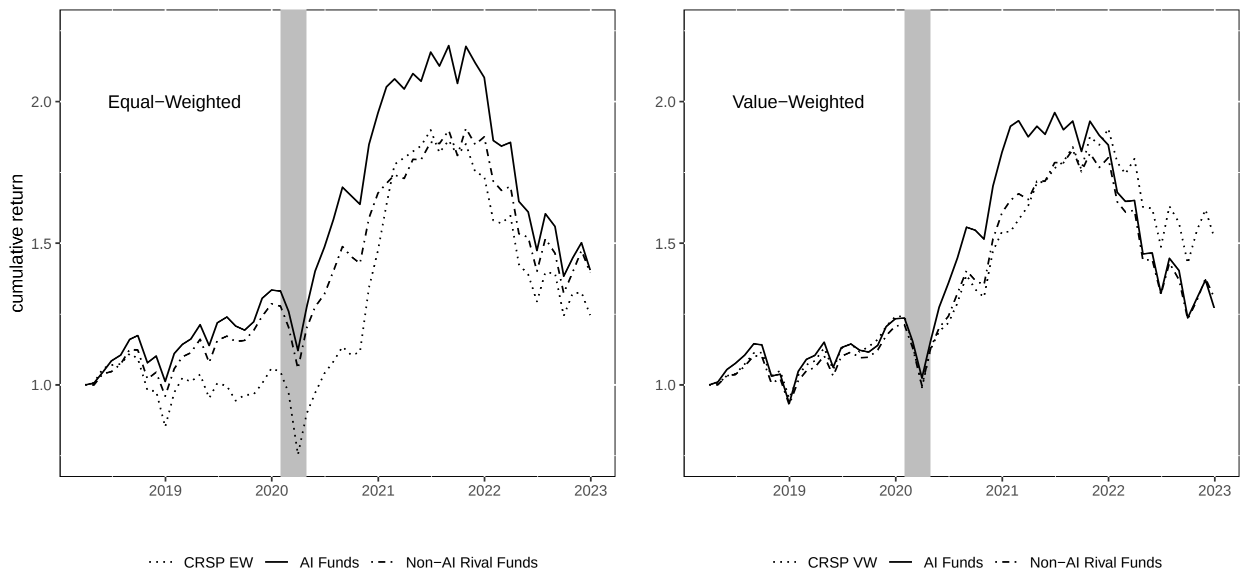

Figure 4: **Top 50 % (TNA) AI funds, cumulative returns:** We compute the cumulative returns for AI-labelled funds, same as for Figure 2 for the funds with the largest 50 % (measured in TNA) AI-labelled funds. The solid line depicts the returns of an equal and value-weighted portfolio of AI funds, respectively. The dash-dotted line is the portfolio of rival funds, matched adapting the methodology in Hoberg et al. (2018). A detailed description is in Section 2. The dotted line represents the cumulative returns of CRSP equity market portfolios. The grey shaded areas are recessions as reported by NBER.

findings by Pástor et al. (2021) and Pástor et al. (2022) argue that although green firms typically have low expected returns, they tend to outperform following positive demand for the ESG factor. Moreover, Beck (2021) examines ESG returns at the fund level. Even though ESG funds should not outperform the market, the flow-driven return of such funds is 2.07%. Therefore, we hypothesise that the inelasticity of demand could - analogous to ESG funds - be viable for AI-labelled funds.

To explore this further, we examine the impact of inflows and outflows on the identified AI funds. We suspect that this effect is particularly relevant for more considerable funds, given their greater visibility to public investors compared to smaller funds. Figure 4 plots the cumulative returns of the top 50 % of AI-labelled funds. The first prominent observation is that the aggregate returns are larger for the big funds, compared to the cumulative returns in Figure 2, especially throughout the years post-COVID. Notably, the equal-weighted portfolio (shown in the left plot with a solid line) yields higher returns than the equal-weighted market portfolio in the first half of the observation period. We suspect these excess returns over the benchmark could be attributable to the hype of such AI funds and the AI labelling of such funds.

The availability of data on fund flows is scarce in the CRSP dataset. In addition, due to the change in the filing system, the numbers are not comparable. To obtain more reliable data, we proxy fund flows via the change in TNA, corrected for internal portfolio growth and dividends:

$$flow^{abs}_{j,t} = NAV_{j,t} - NAV_{j,t-1} * (1 + r_{j,t}) \tag{7a}$$

$$flow^{rel}_{j,t} = \frac{NAV_{j,t} - NAV_{j,t-1} * (1 + r_{j,t})}{NAV_{j,t-1}}, \tag{7b}$$

where $NAV_{j,t}$ is the aggregated net asset value of portfolio j at time t and $(1 + r_{j,t})$ denotes the investment-weighted portfolio returns of portfolio j from time $t-1$ to t. Berk and Tonks (2007) argue that Equation 7b does not correctly capture the relative in- or outflow towards funds. The authors argue that this computation does not perfectly control the change of flows for internal growth. Alternatively, they propose the following measure for relative fund flows

$$flow^{rel,BT07}_{j,t} = \frac{NAV_{j,t} - NAV_{j,t-1} * (1 + r_{j,t})}{NAV_{j,t-1} * (1 + r_{j,t})} \tag{7c}$$

$$\tag{7d}$$

To analyse the implications of fund flows into AI and rival funds, we conduct fixed effects regressions. The results are in Table 6. We test whether AI funds alone have a direct connection. AI funds alone do not attract any significant extra flows compared to similar rival funds. As such ”AI washing”, contrary to ”impact washing” as in Heath et al. (2023), is not effective and does not significantly attract investors. We then include past performance to evaluate whether AI funds, in combination with past performance, have an influence. We proxy perceived past performance to investors via dummies indicating performance in the top/bottom tercile of the previous month's return. Additionally, we include the past risk-adjusted performance from rolling four-factor regressions. Contrary to Miguel and Chen (2021), we do not find robust evidence for a strong sensitivity of relative fund flows to past returns, as columns (5) and (8) do not show any significance. Column (2) reveals a significant positive relation regarding past risk-adjusted performance for absolute flows. Regarding absolute flows, we also observe that past positive performance significantly leads to future inflows. On the contrary, we do not see significant outflows following past bad performance. This finding would be in line with Christoffersen and Musto (2002). They argue that funds with good past performance attract more performance-sensitive investors, which leaves funds with a history of inferior past performance with investors who are less reactive to performance. Therefore, good performance will lead to money inflows from performance-sensitive investors, whereas low performance will not trigger outflows of inelastic investors.

We further proceed to control for further potential influences on flows. Regarding the choice of control variables, we rely on findings from previous studies and add some of our own: Size seems to be negatively related to future flows. Also, younger funds tend to receive more inflows (see e.g. Chevalier and Ellison, 1997, Sirri and Tufano, 1998, Jain and Wu, 2000, Barber et al., 2005, Nanda et al., 2009), which seems reasonable as younger funds tend to be smaller in size. We further include expense ratios and turnover ratios to account for fee structures, as lower participation costs or higher marketing expenses seem to attract more investors (c.f. Sirri and Tufano, 1998, Barber et al., 2005, Nanda et al., 2009, Huang et al., 2007). Additionally, we rely again on our previously motivated measures of skill, fund activeness, timing and picking to control for the skill of the funds in addition to its past performance, which plays a role throughout most of the previously cited papers and is also significant in our results. We further add rolling standard deviations of flows for the steadiness of flows, book-to-market, momentum and dividend yield to account for the influence of the matching criteria. Regarding the initial question, we find no evidence that the label AI alone generates future inflows as the label ESG does in previous studies. If anything, column (3) reveals that, including several controls, the label AI will have negative absolute flows. Nevertheless, past risk-adjusted returns become more important. In line with the explanation of past performance, skill measures, timing, and picking also yield significant positive coefficients throughout all versions of fund flows. Therefore, the label AI plays a minor role, while past performance mainly influences flows. In line with previous findings, costs are negatively related to absolute flows and younger funds tend to receive significantly more relative flows.

5 Further Analyses and Robustness

So far, we have restricted our analysis by several assumptions. However, these design choices might be influential for the subsequent results. Below, we investigate the robustness of our results concerning some alternations in the analysis.

Table 6: **Fund flows**: We run Fixed Effects Regressions with fixed time effects to investigate flows towards funds. In the short model, we only include a dummy for AI funds. In the second specification, we further the variables used in Miguel and Chen (2021): That includes two further dummy variables that are one if the past performance was in the top or bottom tercile of all funds respectively. Furthermore, we include past performance, past alphas from a four-factor regression and the prevalent standard deviation of the fund-specific flow computed with rolling regressions. The last specification adds additional fund-specific control variables. In the table, we abbreviate Berk and Tonks (2007) with BT07. Standard Errors are clustered following Arellano (1987). We compute the p-values based on a t-distribution.

	Dependent variable:								
	fund flows (abs)			fund flows (rel)			fund flows (rel) BT07		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
D(AI Fund)	0.427	0.373	−19.594***	0.007	−0.005	0.013	0.006	−0.006	0.009
	(0.800)	(0.898)	(4.521)	(0.012)	(0.010)	(0.016)	(0.012)	(0.010)	(0.015)
D(Bottom_{t-1})		4.726***	−5.247		0.041	−0.001		0.041	−0.001
		(1.705)	(5.627)		(0.036)	(0.005)		(0.036)	(0.005)
D(Top_{t-1})		−0.234	2.296		−0.022	0.003		−0.021	0.002
		(1.438)	(5.193)		(0.031)	(0.005)		(0.031)	(0.005)
α_{t-1}		162.757***	780.329***		0.711	0.507***		0.715	0.462**
		(61.384)	(177.070)		(0.805)	(0.190)		(0.792)	(0.193)
σ^{abs}_{t-1}			−0.102***						
			(0.018)						
σ^{rel}_{t-1}						0.001			
						(0.002)			
σ^{BT07}_{t-1}									0.001
									(0.002)
Fund Activeness_{t-1}			0.076**			0.00003			0.00003
			(0.030)			(0.00003)			(0.00003)
Timing_{t-1}			143.234**			0.153**			0.147**
			(62.604)			(0.062)			(0.061)
Picking_{t-1}			223.295***			0.381***			0.357***
			(41.896)			(0.075)			(0.073)
Fund Age_{t-1}			−4.533			−0.014***			−0.013***
			(3.361)			(0.002)			(0.002)
Expense Ratio_{t-1}			−3,168.832***			−0.264			−0.287
			(997.442)			(0.364)			(0.359)
Turnover_{t-1}			0.235			0.004			0.004
			(0.323)			(0.002)			(0.002)
log(ME_{t-1})			1.702			−0.001			−0.0003
			(1.468)			(0.002)			(0.002)
log(BM_{t-1})			2.948			0.005			0.005
			(2.924)			(0.005)			(0.005)
log($r_{2,12}$)			4.621			0.044**			0.044**
			(6.784)			(0.018)			(0.018)
log(DY_{t-1})			2.035			−0.004			−0.004
			(2.145)			(0.003)			(0.003)
Groups	13,953	13,408	5,930	13,953	13,408	5,930	13,953	13,408	5,930
Observations	732,794	655,206	119,399	732,794	655,206	119,399	732,794	655,206	119,399
R^2	0.000	0.00001	0.018	0.000	0.00000	0.001	0.000	0.00000	0.001
Adjusted R^2	−0.0001	−0.0001	0.017	−0.0001	−0.0001	0.0004	−0.0001	−0.0001	0.0004
Note:									*p<0.1; **p<0.05; ***p<0.01

5.1 Keep tiny funds

For the main analysis, we removed all funds with a fund size below five million. Appendix D repeats the analysis, including those tiny funds, and presents the results. We find that the overall results are similar to the main analysis. When separating funds based on the stage at which AI is used, we see that the tiny funds perform worse for screening the market and better for optimizing the allocation.

5.2 Incubation Bias

Evans (2010) shows that it is common for funds to launch new funds privately and only subsequently open them for investors after surviving a certain test period. This strategy is referred to as incubation. The issue with incubation is that incubated funds perform better than non-incubated ones, leading to bias. The proposed strategy to eliminate the bias is to consider only return observations with a valid Ticker or to remove the first three years, as the incubation period is generally shorter than this threshold period. Previous literature additionally removes all observations with a fund size smaller than 15 million U$D. In our main analysis, we refrained from these two filters. As pointed out in Evans (2010), a size filter of 25 million (10 million more than common) only eliminates 47 % of all incubated funds while removing 24 % of non-incubated funds from the sample. Given that we already have a relatively small sample of AI funds, both in the cross-section and the time series, we do not have much margin to eliminate valid observation. A filter of 5 million, as in the main analysis, already eliminates roughly a quarter of the cross-section, while a filter of 15 million would more than halve the sample. For robustness and completeness, though, we repeat the analysis with the established 15 million-dollar filter and require a valid ticker symbol for each observation. We refrain from the age filter, as it would drop observations of non-incubated funds and remove valid observations where the incubation has already been completed. Instead, we opt for the filter that addresses the bias from where it originates. We only consider funds with a valid ticker, meaning where incubation is already over. This filter was effective in the original paper and does not remove valid observations.

Figure 5 reports the results from Figure 3, adjusted by the two new filters. Notice that the sample starts in June 2019, more than one and a half later than in the main analysis. The reason lies in the availability of funds after the applied filter, specifically, the funds that declare AI components

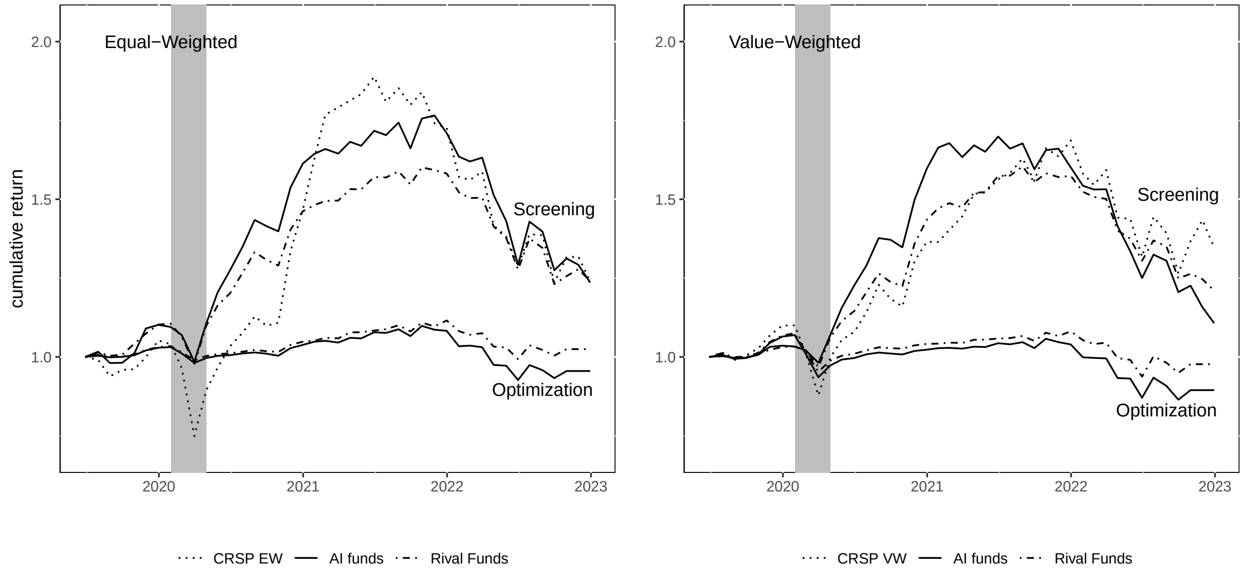

Figure 5: Following Evans (2010), we only consider fund observations with a valid ticker symbol. Additionally, we exclude funds with a size below 15 million. *Screening* combines all funds into a portfolio that uses AI at an early stage of portfolio creation, while *Optimization* tries to optimize the allocation in a later stage. Recession periods are grey-shaded.

in stock screening. Compared to the results from Figure 3, the AI funds perform slightly worse, and returns deteriorate towards the end of the observation periods. The best return periods of the AI funds in Figure 3 lie in the first months. Therefore, we conclude that a slight bias might be inherent. Pástor et al. (2015) provide an alternative explanation: As more funds emerge, the industry becomes more competitive, and more funds compete for the same market. As a result, due to industry growth, returns deteriorate over time. Nevertheless, in line with previous results, AI funds perform comparably to their matched rival funds. The question is how interpretable these results become with an even shorter time series and smaller cross-section. Furthermore, the first months lie in the COVID era. Future research will be able to disentangle these issues with more funds and time observations.

5.3 Drop orthogonalization and z-scores

The analysis' results rely heavily on the choice of the rival funds. Especially the orthogonalization of the z-scores puts a lot of emphasis on the order of the variables in the characteristics vector $C_{f,t}$. Matching variables placed in the vector's first elements receive much higher importance than those in the latter elements of the matching vector. We, therefore, re-match the rival funds based on z-score without orthogonalization. We abbreviate these results with *zs*. Alternatively, instead of using z-scores, we use the ranks in the $C_{f,t}$ vector. Following the first matching methodology of Hoberg et al. (2018), we construct ranks as the percentile of a stock's characteristic in the distribution of all NYSE stocks with share code 11 or 12. We then aggregate the ranks of all stock holdings at the portfolio level using the relative portfolio weights of the constituents. *pcrk* flag all results based on these percentile ranks throughout this paper.

Figure 6 plots the analogous returns to Figure 2 but with alternative rival funds, matched without orthogonalizing either ranks or z-scores. Panel A shows results computed based on ranking the variables. Panel B depicts the respective plot based on z-scores instead. The solid lines represent the funds relying on AI, whereas the dash-dotted lines depict the non-AI peers. A grey line indicates an aggregation where each portfolio receives an equal weight; a black line, on the other hand, aggregates relative to the value managed within the respective portfolio. Comparing the results between ranks and z-scores, we see no substantial differences. The graphs paint a very similar image, meaning that our results are robust to the choice of methods in the matching process. Similar to the rival funds from the main analysis, the AI-managed funds perform better than their peers initially. However, depending on the set of matching variables, the performance of the AI-labelled funds deteriorates in the second half of the observation period. We further observe that the result varies across the variables used to match the funds. Therefore, we conclude that z-scores or ranks do not greatly affect the set of rivals. Further, the orthogonalization of variables in the matching vector does not play a major role.

5.4 Alternative variables

Regarding the variables to match the rival funds, we stay close to the selection of Hoberg et al. (2018). However, one could, for instance, argue that momentum as a matching variable may lead to a self-fulfilling prophecy of similar subsequent fund returns. We, therefore, alter the variables in the $C_{f,t}$ vector. *HKP18* adopts the variables used in the 4D network of Hoberg et al. (2018), namely log size, log book-to-market ratio, log past returns and log dividend yield. *alt1* picks up the previous argument and replaces the momentum characteristic from the main analysis with fund activeness. *alt2* tries to find alternatives for all three variables from the main analysis. Fund age replaces size, as longer existing funds should also have a higher TNA. Log turnover poses the alternative

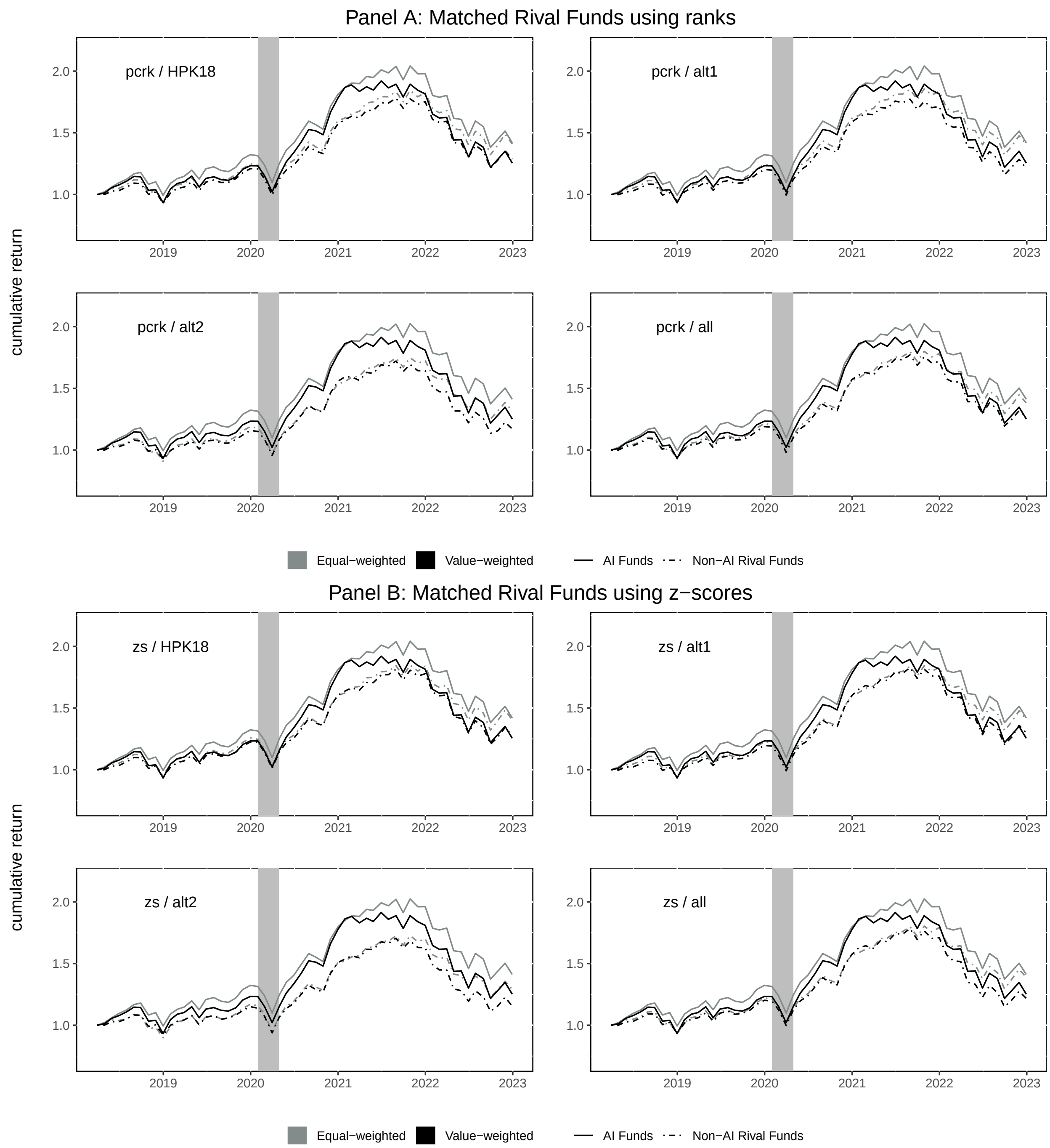

Figure 6: **Return over time using alternative variables to match rival funds:** We use the same methodology as previously to match (portfolios of) rival funds. However, this time we drop orthogonalization and use both ranks ($pcrk$) and z-scores (zs). Furthermore, we use alternative variables in $C_{f,t}$. For *HKP18*, we adopt $C^{HPK18}_{f,t} = \big(lme \quad lbm \quad lmom \quad ldy\big)$, for *alt1*, $C^{alt1}_{f,t} = \big(lme \quad fa \quad lbm\big)$, *alt2* uses $C^{alt2}_{f,t} = \big(lage \quad ltnr \quad lep\big)$, and *all* combines all previous variables and adds one extra $C^{all}_{f,t} = \big(lme \quad lbm \quad lmom \quad ldy \quad lage \quad fa \quad lep \quad ltnr \quad lexr\big)$, where lme is the log size, lbm is log book-to-market, $lmom$ is log momentum, ldy is log dividend yield, $lage$ is the log fund age in years, fa represents a measure of fund activeness, lep is log earings/price, $ltnr$ is the log turnover ratio, and $lexr$ is the log of reported expense ratio. Grey lines indicate equal weights in portfolios; black lines use market weights. Solid lines represent AI funds, and dot-dashed lines depict rival funds. Periods of recessions (as reported by NBER) are shaded in grey.

to momentum, following the idea of fund activeness in Pástor et al. (2020) that one has to trade away from a given benchmark to create outperformance, leading to positive momentum. The log earnings-to-price ratio is used as an alternative multiple for valuation. Furthermore, earnings should transform into cash dividends sooner or later, leading us back to the dividend yield. Finally, *all* combines all previously mentioned variables into the spacial basis. The graphs in Figure 6 show that the cumulative returns for *HPK18* are similar to our baseline model in the main analysis. The cumulative returns of the other alternatives do not yield substantially different results either. Thus, we conclude that the choice of the matching variable does not play an influential role in our analysis.

5.5 Quantitative Funds

To evaluate the robustness of our rival funds and compare the performance of the identified AI funds to an alternative sample, we use quantitative funds. Quantitative funds are similar to AI funds in using mathematical models to produce a security selection. Likewise, both tend to follow a rule-based approach in allocating weights to the constituents. However, a human fund manager develops and applies statistical analysis. In contrast, the research and investment process is (claimed to be) enhanced through AI for the latter group of funds. We excluded AI funds from our quantitative fund sample to prevent any overlap. A detailed outline of how we obtained our sample of quantitative funds is in Table A1 in Appendix B. Thus, comparing quantitative funds may provide further insights in two ways. First, it constitutes an alternative sample of rival funds, providing further insights regarding the robustness of our results. Second, and more importantly, it may serve as a ceteris paribus analysis. Both types of funds apply rule-based methods and, due to the matching procedure, also resemble each other in their investment objective and fund characteristics. Thus, the quant funds are even closer to the AI funds, and all else equal, they may only differ on the aspect of applying AI. The results may, therefore, evaluate the incremental effect of AI.

Row 4 and 5 of Panel A-C in Table 2 report the analogous results for the respective quantitative funds. Panels A and B report the equal and value-weighted averages of the fund returns. The equally aggregated quantitative funds yield, on average, 0.408 % (ew) or 0.41 % (dw) per month (over the risk-free rate), which is 30 bps less than the AI funds. Nevertheless, this return difference is not statistically different from zero. Using market weights for the aggregation (Panel B), average quant fund returns are closer to those of AI funds. The return difference is 3.8 bps (ew) or five bps (dw). As such, AI funds do not significantly outperform quantitative funds. The results from Panel C reflect the same conclusion. AI funds produce Sharpe ratios that are higher than those of quantitative funds. Nevertheless, the differences in Sharpe ratios are not significant. As such, AI funds do not produce strategic advantages over quantitative funds either. On the other hand, AI-labelled funds do not underperform compared to their rule-based quantitative alternatives.

Figure 7 and 8, however provide further insights. Identical to Figure 2, we aggregate the returns of the quantitative fund using equal and market weights and subsequently compute the cumulative portfolio returns. AI-labelled funds recovered well from the COVID-19 recession, especially the smaller AI funds. However, this return advantage was lost gradually in the subsequent months. In the value-weighted case, the AI funds are already on par with the quantitative funds. A closer look at Figure 8 reveals that this return pattern was mainly driven again by the sub-group *Screening*. In equal and value-weighted cases, the funds with market screening AI outperformed in the first half of the observation period. In the second half, the quantitative peers performed better and caught up. However, similar to the results from Table 2 and to the results from Figure 2, Figure 3 and Figure 6, sub-groups of AI funds do not manage to outperform their peers significantly.

6 Conclusion

We evaluate funds that claim to enhance their portfolio management by implementing AI and related methods. In summary, funds that advertise themselves as AI-enhanced perform comparably to non-AI funds, showing neither significantly superior nor inferior performance. We do not find that using the label AI would lead to future inflow either. However, AI funds underperform in comparison to a value-weighted market portfolio. Compared to traditional funds, AI-driven funds exhibit lower activity levels and greater timing ability. However, their stock selection skills fall short when compared to their counterparts. Additionally, our findings reveal persistent negative performance in AI and rival funds, with weak evidence towards momentum in well-performing AI funds and reversal for rival funds. We arrive at similar conclusions when adopting rule-based quantitative funds without reported AI components as benchmarks.

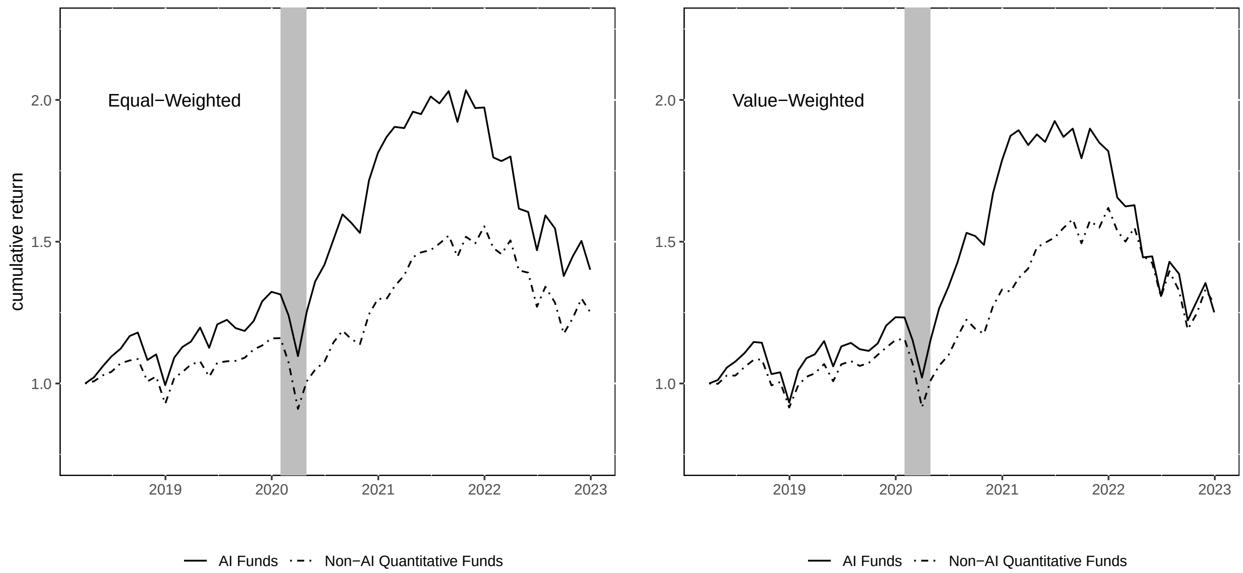

Figure 7: The figure illustrates the cumulative returns of AI funds (solid line) and matched quantitative benchmark funds (dot-dashed line). The left plot aggregates the monthly returns of AI-labelled and quant funds with equal weights, while the right plot considers the size of the funds in the aggregation. Contrary to the matching in Figure 2, we match quantitative funds based on the first three levels of investment objectives provided by CRSP to ensure that all AI funds are paired with enough rival funds. We exclude funds with a TNA below 5 million. The grey-shaded area is in a recession period, as NBER indicates.

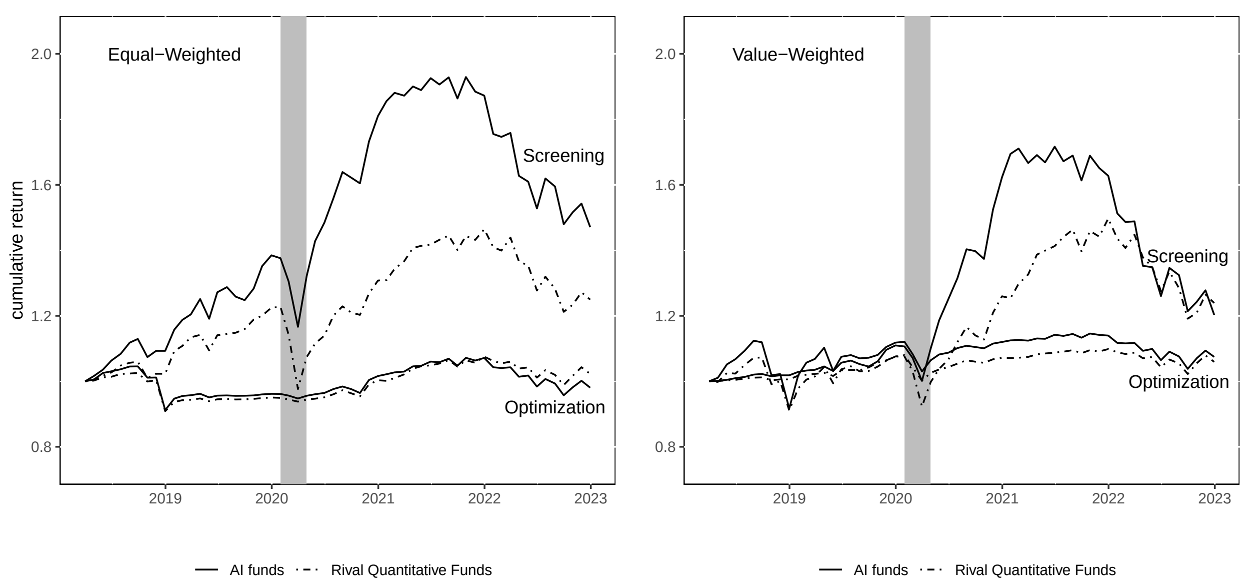

Figure 8: The figure separates the cumulative returns from Figure 7 into those of the sub-groups *Selection* and *Optimization*. The first (second) group contains AI funds - solid lines - or their rivals - dot-dashed lines - that declare to use AI to filter profitable investments into the investment universe (to deviate from a classic weighting scheme and improve the allocation). Analogous to Figure 7, we match rivals only based on the first three layers of fund investment objective classification. Furthermore, this analysis excludes funds with a TNA smaller than 5 million USD.

Although our sample contains more data than the analysis conducted by Chen and Ren (2022), which closely aligns with ours, our data availability remains limited. The integration of AI into portfolio management processes began in 2017 and is, therefore, relatively recent. More conclusive findings will likely be drawn in the coming years, with more return observations and AI funds in the cross-section. A non-resolved question regarding a potential incubation bias can then be revisited. Future research could also add the CSMAR and RESSET databases to the analysis and broaden the scope of analysis from US to Chinese mutual funds.

References

Arellano, M. (1987). "Computing Robust Standard Errors for Within-Groups Estimators". In: *Oxford Bulletin of Economics and Statistics* 49.4, pp. 431–434.

Avramov, D., S. Cheng, and L. Metzker (2023). "Machine Learning vs. Economic Restrictions: Evidence from Stock Return Predictability". In: *Management Science* 69.5, pp. 2587–2619.

Azevedo, V., C. Hoegner, and M. Velikov (2023). "The Expected Returns on Machine-Learning Strategies". In: *Available at SSRN 4702406*.

Barber, B. M., T. Odean, and L. Zheng (2005). "Out of sight, out of mind: The effects of expenses on mutual fund flows". In: *Journal of Business* 78.6, pp. 2095–2120.

Beck, P. van der (2021). "Flow-driven ESG returns". In: *Swiss Finance Institute Research Paper* 21–71. Available at SSRN: https://ssrn.com/abstract=3929359.

Berk, J. B. and I. Tonks (2007). "Return Persistence and Fund Flows in the Worst Performing Mutual Funds". NBER Working Paper 13042.

Busse, J. A., A. Goyal, and S. Wahal (2010). "Performance and Persistence in Institutional Investment Management". In: *Journal of Finance* 65.2, pp. 765–790.

Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance". In: *Journal of Finance* 52.1, pp. 57–82.

Chen, A. Y. and M. Velikov (2023). "Zeroing in on the Expected Returns of Anomalies". In: *Journal of Financial and Quantitative Analysis* 58.3, pp. 968–1004.

Chen, J., H. Hong, M. Huang, and J. D. Kubik (2004). "Does fund size erode mutual fund performance? The role of liquidity and organization". In: *American Economic Review* 94.5, pp. 1276–1302.

Chen, L., M. Pelger, and J. Zhu (2024). "Deep Learning in Asset Pricing". In: *Management Science* 70.2, pp. 714–750.

Chen, R. and J. Ren (2022). "Do AI-powered Mutual Funds Perform Better?" In: *Finance Research Letters* 47, p. 102616.

Chevalier, J. and G. Ellison (1997). "Risk taking by mutual funds as a response to incentives". In: *Journal of Political Economy* 105.6, pp. 1167–1200.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019). "Sparse Signals in the Cross-Section of Returns". In: *Journal of Finance* 74.1, pp. 449–492.

Chordia, T., A. Subrahmanyam, and Q. Tong (2014). "Have Capital Market Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity?" In: *Journal of Accounting and Economics* 58.1, pp. 41–58.

Christoffersen, S. E. and D. K. Musto (2002). "Demand curves and the pricing of money management". In: *Review of Financial Studies* 15.5, pp. 1499–1524.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2021). "AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI". In: *Available at SSRN 3554486*.

Dugast, J. and T. Foucault (2018). "Data Abundance and Asset Price Informativeness". In: *Journal of Financial Economics* 130.2, pp. 367–391.

Edmans, A. (2023). "The end of ESG". In: *Financial Management* 52.1, pp. 3–17.

Elton, E. J., M. J. Gruber, and C. R. Blake (2001). "A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases". In: *Journal of Finance* 56.6, pp. 2415–2430.

Evans, R. B. (2010). "Mutual fund incubation". In: *Journal of Finance* 65.4, pp. 1581–1611.

Farboodi, M. and L. Veldkamp (2020). "Long-Run Growth of Financial Data Technology". In: *American Economic Review* 110.8, pp. 2485–2523.

Freyberger, J., A. Neuhierl, and M. Weber (2020). "Dissecting Characteristics Nonparametrically". In: *Review of Financial Studies* 33.5, pp. 2326–2377.

Gabaix, X. and R. S. J. Koijen (2022). "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis". Swiss Finance Institute Research Paper No. 20-91, Available at SSRN: https://ssrn.com/abstract=3686935.

Grennan, J. and R. Michaely (2021). "Fintechs and the Market for Financial Analysis". In: *Journal of Financial and Quantitative Analysis* 56.6, pp. 1877–1907.

Gruber, M. J. (1996). "Another puzzle: the growth in actively managed mutual funds". In: *Journal of Finance* 52, pp. 783–810.

Gu, S., B. Kelly, and D. Xiu (2020). "Empirical Asset Pricing via Machine Learning". In: *Review of Financial Studies* 33.5, pp. 2223–2273.

Heath, D., D. Macciocchi, R. Michaely, and M. C. Ringgenberg (2023). "Does socially responsible investing change firm behavior?" In: *Review of Finance* 27.6, pp. 2057–2083.

Hoberg, G., N. Kumar, and N. Prabhala (2018). "Mutual Fund Competition, Managerial Skill, and Alpha Persistence". In: *Review of Financial Studies* 31.5, pp. 1896–1929.

Huang, J., K. D. Wei, and H. Yan (2007). "Participation costs and the sensitivity of fund flows to past performance". In: *Journal of Finance* 62.3, pp. 1273–1311.

Jain, P. C. and J. S. Wu (2000). "Truth in mutual fund advertising: Evidence on future performance and fund flows". In: *Journal of Finance* 55.2, pp. 937–958.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp (2014). "Time-varying Fund Manager Skill". In: *Journal of Finance* 69.4, pp. 1455–1484.

Kelly, B., S. Malamud, and K. Zhou (2024). "The virtue of Complexity in Return Prediction". In: *Journal of Finance* 79.1, pp. 459–503.

Ledoit, O. and M. Wolf (2008). "Robust Performance Hypothesis Testing with the Sharpe Ratio". In: *Journal of Empirical Finance* 15.5, pp. 850–859.

McLean, R. D. and J. Pontiff (2016). "Does Academic Research Destroy Stock Return Predictability?" In: *Journal of Finance* 71.1, pp. 5–32.

Miguel, A. F. and Y. Chen (2021). "Do Machines Beat Humans? Evidence from Mutual Fund Performance Persistence". In: *International Review of Financial Analysis* 78, p. 101913.

Nanda, V. K., Z. J. Wang, and L. Zheng (2009). "The ABCs of mutual funds: On the introduction of multiple share classes". In: *Journal of Financial Intermediation* 18.3, pp. 329–361.

Newey, W. K. and K. D. West (1994). "Automatic Lag Selection in Covariance Matrix Estimation". In: *Review of Economic Studies* 61.4, pp. 631–653.

Pástor, L., R. F. Stambaugh, and L. A. Taylor (2015). "Scale and Skill in Active Management". In: *Journal of Financial Economics* 116.1, pp. 23–45.

— (2020). "Fund Tradeoffs". In: *Journal of Financial Economics* 138.3, pp. 614–634.

— (2021). "Sustainable Investing in Equilibrium". In: *Journal of Financial Economics* 142.2, pp. 550–571.

— (2022). "Dissecting Green Returns". In: *Journal of Financial Economics* 146.2, pp. 403–424.

Sirri, E. R. and P. Tufano (1998). "Costly search and mutual fund flows". In: *Journal of Finance* 53.5, pp. 1589–1622.

Zhang, J., Z. Peng, Y. Zeng, and H. Yang (2023). "Do Big Data Mutual Funds Outperform?" In: *Journal of International Financial Markets, Institutions and Money* 88, p. 101842.

Zhu, C. (2019). "Big Data as a Governance Mechanism". In: *Review of Financial Studies* 32.5, pp. 2021–2061.

A Identifying AI funds

At the heart of our study lies the identification of funds that feature some artificial intelligence (AI) in the portfolio construction process. We rely on the Summary Prospectus Type 497K obtained from the Securities and Exchange Commission's (SEC's) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) filing system to achieve this classification. Chen and Ren (2022) report that the first AI-managed fund was the *ETF Managers Trust: AI Powered Equity ETF* (AIEQ)[11]. It was issued on 19/10/2017. They find 15 AI-managed funds until the end of 2019[12]. We download the Summary Prospectus Type 497K for all CIK starting from 2019. We download 155,630 filings for 44,582 funds from the beginning of 2017 until the first quarter of 2024. We subsequently apply several filters. Table 1 outlines this process of arriving from all fund filings to the final identified AI-labelled funds. We download 155,630 497-K filings for 44,582 unique CIK (i.e. unique funds) from EDGAR. In the first step, we search for AI-related keywords in the filings. Of the 44,582 funds, 669 contain any keywords in their filings. However, most of these "hits" are misleading. While we are interested in Artificial Intelligence, *AI* also serves as an abbreviation for "American Independence", an arbitrary name of a screened fund. The Summary Prospectus also contains information, for instance, about the fund manager. If they have a master's degree in "Machine Learning", this has nothing to do with the fund or how it is being managed. To give another example of such non-intended hits, many are AI-sector funds and somehow contain variations of our search words in their title or fund description. Nevertheless, we are not interested in funds that invest in companies that predominantly operate within or prominently focus their R&D on AI. We are keen on finding those funds with AI solely in their portfolio creation process. Therefore, we review the descriptions of the remaining filings manually and decide whether or not - given the information in the Summary Prospectus - the fund uses AI in some stage of the portfolio construction. In this step, we also classify whether a fund uses AI in selecting the investable universe or whether AI is used in computing a suitable weighting. This leaves us with 103 funds that use AI at some stage of the asset management process to develop a portfolio. Given that the company's CIK identifies the filings, we merge the list of identified funds with the CRSP Survivor-Bias-Free U.S. Mutual Fund Database. However, we encounter 18 funds without matching CIKs in the CRSP

[11] CUSIP: 26924G813

[12] In an initial version, we adopted their 15 funds and searched from 2019 onwards. However, we also found one fund before AIEQ. As a result, we identified all funds in our sample based on our methodology and updated our sample.

database, resulting in a final sample of 85 funds.

Pástor et al. (2015) note that many mutual funds hold the same portfolio. These funds represent different share classes with different fee structures. We aggregate sub-classes of the same fund in two ways. First, we connect funds if they share the same portfolio number. Pástor et al. (2015) already mention that many mutual funds offer multiple share classes, representing only a claim onto the same underlying portfolio. These funds then carry different fund identifiers but the same portfolio identifier. Where possible, we, therefore, run our analysis based on portfolio level rather than fund level. A mutual fund that offers the same portfolio with multiple fee structures would otherwise lead to multiple versions of the same portfolio, resulting in a higher weight for that specific portfolio in the aggregate analysis. The main analysis considers returns after expenses. As a result, the fee structure does not play a role. On the other hand, a fund with multiple classes would lead to an overrepresentation of that fund in our sample compared to a fund with only one fee structure. Thus, we combine those funds using TNA weights. This step already finds most of these connected funds. However, there are still some unmatched funds left. For example, CRSP Portno 1023744 (*Voya Small Company Fund*) and 1024790 (*Voya Small Company Portfolio*) hold a close to similar portfolio. We link these remaining funds based on correlation. If the time-series correlation of daily returns between two funds is above 99.9 % and if the same company issues them, we consider them the same portfolio. We require at least 252 days of pairwise complete observations for the correlations. In this fourth filtering step, we reduce our sample to 67 unique portfolios managed entirely or partially with artificial intelligence.

Panel B in Table 1 provides insight into the keywords identified during the AI fund screening process. Approximately half of the screened AI funds report using Natural Language Processing (NLP), primarily in funds that employ NLP to refine their investment universe. Consequently, most of these funds fall into our first AI fund subgroup, *Screening*. The remaining keyword matches primarily describe specific Machine Learning (ML) techniques. In most cases, funds using these methods belong to the second subgroup, which focuses on portfolio optimization. It is important to note that the total number of keyword matches exceeds the number of funds identified in Panel A, as some AI funds match multiple keywords.

Using this methodology, one could interpret the main analysis results in terms of AI-*managed* rather than *labelled* funds. However, identifying AI-managed funds would require additional methodology or analysis, as many funds may be reluctant to disclose their portfolio management processes or strategies. As a result, funds that apply AI, ML, or DL may not mention these technologies in their filings or may refer to them ambiguously. For example, we encountered thousands of matches for terms like “statistical analysis” or “quantitative methods.” While the set of AI funds likely overlaps significantly with funds using quantitative analysis in portfolio management, without more specific information, it would be premature to classify such funds as “AI funds.” Consequently, we can only reliably identify funds that explicitly report the application of AI in their filings. While both interpretations are valid to some degree, to maintain clarity and rigour in our conclusions, we restrict our analysis to AI-*labelled* funds, reserving the identification of a broader set of AI-*managed* funds for future research.

For the benchmark funds, we then match funds in the CRSP Survivor-Bias-Free U.S. Mutual Fund Database based on their fund characteristics. Contrary to the datasets of previous research (c.f., e.g. Berk and Van Binsbergen, 2015 or Pástor et al., 2015), we keep ETFs, sector and index funds in the dataset. Most AI-labelled funds are offered through the fund construct of ETFs or use AI methods, such as NLP, to identify the investment universe for an improved version of a predefined sector fund - filtering them would eliminate potential (suitable) benchmark funds.

B Identifying Quantitative funds

We outlined the process of constructing our sample of quantitative funds in Table A1. We rely on the classification of funds from Lipper Hindsight via Refinitiv Eikon (now LSEG). Lipper screens the reportings of mutual funds and flags funds as ”quantitative” if they report relying on a rule-based mathematical model for their investment decisions. In that sense, their methodology resembles ours to identify funds with AI components. Miguel and Chen (2021) also source this database to evaluate the persistence of quantitative funds over human funds. For a detailed description of the peculiarities of the Lipper Hindsight database, we refer to Section 2.2 of their paper. The sample of the paper mentioned above contains 1,592 funds until 2019. In total, we identify 2,441 funds until 2024/11. We match these funds to our CRSP Mutual Fund database by 1. their CUSIP and

Table A1: **Selection Criteria to obtain our sample of Quantitative Funds:** We download funds with the "quantitative fund" attribute from the Lipper Hindsight database via Refinitiv Eikon. Quantitative funds are defined as those that adopt a rule-based mathematical model for investment decisions. This table reports the number of funds (portfolios) after each (the last) filtering step.

Selecting Quant Funds		**N**
Quantitative Funds from LSEG		2,441
Filtering Step 1: Matching with CRSP		
matched with CUSIP	2,176	
(remaining) matched with NASDAQ Ticker	113	
unmatched	152	2,289
Filtering Step 2:		
Fund is active after 2017/10		1,545
Filtering Step 3:		
Fund is not AI Fund		1,540
Filtering Step 4:		
Fund has a connected CRSP Portfolio Number		1,533
Filtering Step 5:		
Account for sub-clases		488

then the remaining, still unmatched funds 2. through the (Nasdaq) Ticker symbol. We can connect 2,176 funds with the CRSP data through the first matching loop. Of the remaining unmatched funds, we combine 113 funds from LSEG with CRSP by their Ticker symbol, leaving us with 2,289 matched funds. To match our sample with that of AI funds, we subsequently follow the steps outlined in A. Given that the *AIEQ* was launched in October 2017, we limit our sample of quant funds to the same starting point. This eliminates 744 funds, leaving us with 1,545 funds. Like quantitative funds, AI funds may rely on rule-based investment strategies. Alternatively, it may be possible that quant funds use AI at some stage of their market screening or portfolio optimization process. Thus, an overlap of these funds is evident. We eliminate all AI funds from the quantitative fund sample to prevent comparing AI funds with AI funds again. However, this step only removes five funds from our sample, which would further hint towards the previously mentioned thought that the identified AI funds may largely consist of those those that use AI proprietary as a label. We then proceed to account for different share classes. In step 1, we matched based on the fund. However, as mentioned in Section A, multiple funds may be managed within the same portfolio since they only constitute a different fee structure of the same underlying portfolio. For a discussion on the necessity of grouping those funds into one portfolio for our analysis, we refer to Section A. We lose 7 funds due to missing CRSP portfolio numbers. Of the remaining 1,533 funds, we group them into 488 effective quantitative portfolios. Again, we group based on the same CRSP portfolio identifying number and based on return correlation among these funds.

C Descriptive Statistics

C.1 Variables

We compute momentum as a stock's returns over the past 12 months, excluding the most recent month. Since delisting returns induce substantial biases (c.f. Shumway, 1997, Shumway and Warther, 1999), we correct stock returns by a delisting return as in Chen and Zimmermann (2022), following suggestions from Johnson and Zhao (2012). Market equity is the number of shares times the last price observed in a current month. Similar to Vuolteenaho (2002), we replace missing observations in month t with the market equity in month $t-1$, compounded with the monthly return (excluding dividends). We construct book equity as in Fama and French (1992)[13] As mentioned in the original paper, book equity is often unavailable. To increase the number of available book equities in our cross-section of stocks, we resort to the corrections from Vuolteenaho (2002): We proxy missing book equity by last year's book equity, plus earnings, minus dividends paid in the current year. If book equity is still unavailable, we assume that book equity has grown at the same rate as market equity. We treat negative or zero book equity as missing observations. The Book-to-market constitutes the ratio between the book equity reported at the end of June, computed as indicated above, and the most recent December market equity (c.f. Fama and French, 1992). Stock size on the other hand uses the prevailing market equity. Dividends and earnings are the sums of the dividends paid or the reported earnings throughout the past 12 months (see e.g. Welch and Goyal, 2008). To obtain the dividend yield (dy) we divide our aggregate of dividends by the most recent end-of-year price. Similarly, we compute the price-earnings ratio using the most recent December price and the accumulated earnings.

To reduce the effect of outliers and potential wrong variables, we winsorize all variables each month along the cross-section of stocks at the 1 % and 99 % levels. We additionally perform log transformations.

C.2 Funds

We aggregate the stock-level variable at the portfolio level using their respective weight in the respective portfolio. Table A2 summarizes the previously discussed variables at the portfolio level,

[13] c.f. Chen and Zimmermann (2022) for the full code, Signal "*BMdec*"

split into groups for further insights. The numbers are averages of the fund characteristics averaged over the different groups of portfolios. The size of the compared rival funds appears much higher than those of the AI-labelled funds. Nevertheless, this is due to some funds that may resemble all the other matching characteristics but already have more assets under management. This explanation would also align with rival funds' higher average fund age. AI funds are overall more active than rival funds. This means they trade more and/or in markets with lower liquidity. Especially those funds that claim to rely on AI for an optimized allocation shift their portfolio a lot. On average, these funds report a turnover of 356.98 %, which compares high to the 87 % and 93 % of the rival funds. While this excess of trading activity results in average positive market timing returns, it does not lead to improved stock-picking returns. Detailed results are reported in Table 4 and discussed in Section 2 and 3.

Table A2: **Summary Statistics - Fund Characteristics:** The table reports the mean fund characteristics for the variables used throughout the analysis. *Screening* combines funds that engage in AI to improve their procedure to screen the market for a profitable investment universe (or their respective rivals), whereas *Optimization* comprises all (benchmark funds for) funds that rely on AI for a more effective weighting scheme within their portfolio. *All* combines both subsets. Rival funds are aggregated based on two schemes: ew - equally weighted and dw - distance weighted. Fund activeness = $TL^{-1/2}$ is computed following Pástor et al. (2020), where T is the reported turnover and L is a liquidity measure. Market timing and Stock Picking are motivated by Kacperczyk et al. (2014) and represent measurements of skill inherent in a fund. Market timing measures the return generated from over- (under-)weighting constituents with high (low) future systematic returns. Stock picking captures the same return attribution as market timing but with abnormal returns.

	AI Funds			Rival Funds (ew)			Rival Funds (dw)		
	All	Screening	Optimization	All	Screening	Optimization	All	Screening	Optimization
TNA (mln U$D)	70.768	79.201	55.931	1,953.515	1,808.199	2,241.138	1,948.207	1,856.753	2,129.221
ret (%)	0.363	0.375	0.342	0.382	0.308	0.529	0.357	0.270	0.528
fund age	1.415	1.491	1.281	11.764	11.866	11.561	11.518	11.544	11.468
expense ratio (%)	0.555	0.418	0.822	0.761	0.740	0.802	0.747	0.720	0.801
turnover (%)	137.662	26.170	356.977	64.283	52.471	87.638	64.892	50.808	92.740
fund activeness	10.513	6.819	22.043	7.759	6.774	9.757	7.935	6.846	10.141
market timing (%)	0.748	0.762	0.720	0.753	0.805	0.648	0.755	0.809	0.648
stock picking (%)	-0.640	-0.662	-0.598	-0.483	-0.579	-0.291	-0.490	-0.587	-0.298

C.3 Competition

We evaluate the environment an AI fund operates in based on the two measures used in Hoberg et al. (2018): One is the number of matched rival funds, N_{peers}, with a small enough Euclidean distance, $d_{i,j,t}$. A higher number indicates a crowded market with more funds competing for investors. The second measure, $tsim_t = \sum_{j=1}^{N} s_{i,j,t} + d_t^*$, where $s_{i,j} = -d_{i,j,t}$ and d_t^* is the cut-off value with the highest still acceptable distance, tries to capture how tightly these funds are clustered. A $tsim$ of zero would mean that all rival funds are completely similar and competition is high. The higher $tsim$, the less related the group of rival funds around the investigated AI fund. These measures not only describe the competition in the field but may also serve as quality measures of the matching itself. A high number of rival funds with low $tsim$ would be desirable as this would mean that we would compare an AI fund with a large portfolio of very close funds, leading to a better counterfactual property for these (portfolios of) rival funds. Table A3 reports the average distances and competition measures. On average, we match 33 rival funds to an AI fund. AI funds are thus benchmarked against a portfolio of rival funds. This should ensure that outliers in fund characteristics of some rival funds, such as the aforementioned higher TNA, are counterbalanced by other funds. Remember that the number of funds is artificially slightly higher than expected because we require a minimum of 10 rival funds for each evaluated AI fund. The matching algorithm finds very close substitutes for the funds that report levering their portfolio selection through AI in the screening process. Even though the average number of rival funds is lower for this subset of AU funds, the similarity measure, $tsim$, is relatively lower compared to the funds that claim to use AI in portfolio optimization. Thus, these funds are benchmarked with a small but very dense rival group, whereas the subgroup of funds that employ AI to improve their weighting of securities is benchmarked against a larger but less similar group of rivals.

Table A3: **Summary Statistics - Competition:** The table reports the averages of the measures of how competitive the area is that AI funds engage in. $d_{i,j,t}$ denotes the Euclidean Distance of AI fund i with (potential) rival fund j at time t. N_{Peers} tells the average number of rival funds matched to a single AI fund. A high number indicates that an AI fund is benchmarked against a diverse portfolio of many rival funds. $tsim_t = \sum_{j=1}^{N} s_{i,j,t} + d_t^*$, where $s_{i,j} = -d_{i,j,t}$ and d_t^* is the cut-off value with the highest still acceptable distance, measures how close these funds are related. A lower value indicates similar funds, whereas a high value hints towards little competition for investors. *Screening* includes all funds that use AI early in the portfolio creation process to screen the market for profitable investments. *Optimization* represents the subgroup of funds that claim to use AI to optimize their weighting scheme.

	All	Screening	Optimization
$d_{i,j,t}$	0.616	0.641	0.566
N_{Peers}	33.059	15.109	69.109
$tsim_t$	4.748	1.688	10.893

D Repeat analysis including tiny funds

We re-conduct the analysis from Section 3 without removing funds with TNA below 5 million from our sample. Figure A1 draws the cumulative returns, as we do in Figure 2 in the main analysis. The results are very similar to the ones where we eliminate the tiny funds. Equally-weighted AI funds and their conventional fund benchmark outpace the market benchmark during the first months. Towards the end, they perform relatively worse than the market portfolio. At the end of the observation period, the cumulative are very close already. Value-weighted, the graph shows close to no alteration to the original plot, as the tiny funds receive close to no weight anyway. Nevertheless, the tiny funds alter the results in Figure A2. While the *Selection* sub-portfolio exhibited returns beyond that of the equally-weighted market portfolio, the tiny funds that use AI to screen for investment opportunities seem to deteriorate the performance. On the other hand, the tiny funds with AI components in their portfolio optimization perform relatively better. Towards the end of our sample, all five portfolios - screening and optimizing AI funds and their rivals and the market- yield similar cumulative returns. Regardless of the small weights in the value-weighted aggregation, the tiny AI funds improve the performance of the optimization sub-group funds and lower those of the screening sub-group.

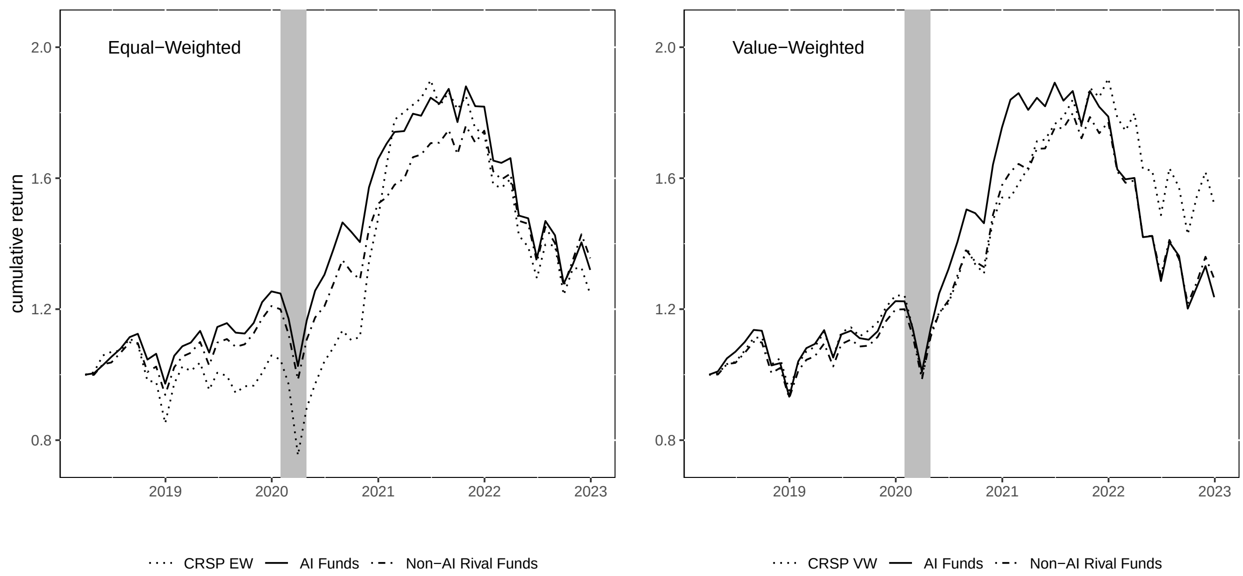

Figure A1: **Return over time incl. tiny funds:** Analogous to Figure 2, we aggregate the returns of AI-labelled funds and their peers. However, we now include funds with TNA below five million, which we eliminated in the main analysis. The rivals are matched based on z-scores as outlined in Section 2, based on Hoberg et al. (2018). The solid line represents AI funds, dash-doted plots the cumulative returns of the peers, and the dotted line is the equal and value-weighted return of the CRSP market. The grey-shaded area is a recession period, defined by NBER.

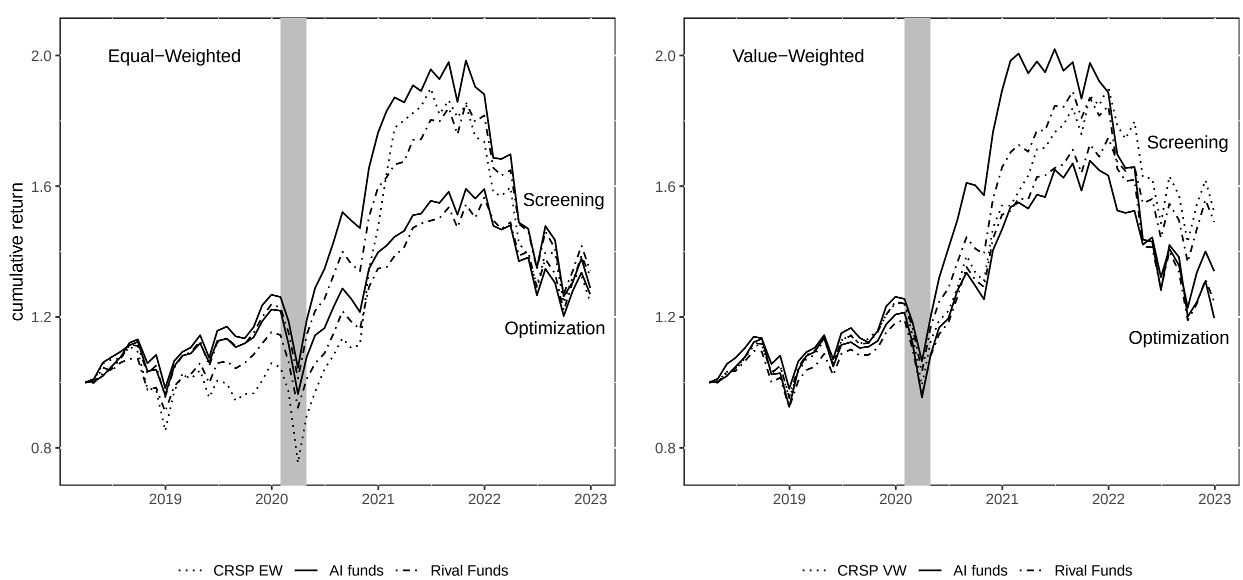

Figure A2: **Subgroups of AI funds, incl. tiny funds:** Again, we sort the funds from Figure A1 into the two categories of Figure 3: Selection an Allocation. The latter employs AI to weigh assets in the portfolio. Selection uses AI at an earlier stage of portfolio management to skim the investment universe for potentially profitable constituents. The solid line depicts the AI-labelled funds combined into an AI fund portfolio with equal weights in the left panel and TNA weights in the right plot. Solid lines represent sub-portfolios of AI funds. Dash-dotted lines are respective sub-portfolios of rival funds, and the dotted lines are the equal and value-weighted market portfolios of CRSP.

References

Berk, J. B. and J. H. Van Binsbergen (2015). "Measuring Skill in the Mutual Fund Industry". In: *Journal of Financial Economics* 118.1, pp. 1–20.

Chen, A. Y. and T. Zimmermann (2022). "Open Source Cross-Sectional Asset Pricing". In: *Critical Finance Review* 27.2, pp. 207–264.

Chen, R. and J. Ren (2022). "Do AI-powered Mutual Funds Perform Better?" In: *Finance Research Letters* 47, p. 102616.

Fama, E. F. and K. R. French (1992). "The cross-section of expected stock returns". In: *Journal of Finance* 47.2, pp. 427–465.

Hoberg, G., N. Kumar, and N. Prabhala (2018). "Mutual Fund Competition, Managerial Skill, and Alpha Persistence". In: *Review of Financial Studies* 31.5, pp. 1896–1929.

Johnson, W. B. and R. Zhao (2012). "Contrarian share price reactions to earnings surprises". In: *Journal of Accounting, Auditing & Finance* 27.2, pp. 236–266.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp (2014). "Time-varying Fund Manager Skill". In: *Journal of Finance* 69.4, pp. 1455–1484.

Miguel, A. F. and Y. Chen (2021). "Do Machines Beat Humans? Evidence from Mutual Fund Performance Persistence". In: *International Review of Financial Analysis* 78, p. 101913.

Pástor, L., R. F. Stambaugh, and L. A. Taylor (2015). "Scale and Skill in Active Management". In: *Journal of Financial Economics* 116.1, pp. 23–45.

— (2020). "Fund Tradeoffs". In: *Journal of Financial Economics* 138.3, pp. 614–634.

Shumway, T. (1997). "The delisting bias in CRSP data". In: *Journal of Finance* 52.1, pp. 327–340.

Shumway, T. and V. A. Warther (1999). "The delisting bias in CRSP's Nasdaq data and its implications for the size effect". In: *Journal of Finance* 54.6, pp. 2361–2379.

Vuolteenaho, T. (2002). "What drives firm-level stock returns?" In: *Journal of Finance* 57.1, pp. 233–264.

Welch, I. and A. Goyal (2008). "A comprehensive look at the empirical performance of equity premium prediction". In: *Review of Financial Studies* 21.4, pp. 1455–1508.