AI Narrative and Stock Mispricing *

Arka P. Bandyopadhyay† Dat Mai‡ Kuntara Pukthuanthong§

May 24, 2023

Abstract

We apply advanced natural language processing to develop a dynamic dictionary of artificial intelligence (AI). Using this dictionary, we construct a real-time index of AI attention from more than 3,000,000 *New York Times* articles. Firms having high exposures to AI have higher returns one month ahead and lower returns five to seven months ahead, suggesting initial overreactions to AI news and subsequent corrections. The connection between AI exposures and future returns is concentrated among non-big stocks, indicating that small AI stocks are more difficult to value. A long-short AI exposure portfolio among non-big stocks generates significant annual alphas of 3% against benchmark multifactor models.

Keywords: news narrative, artificial intelligence, NLP, return prediction, mispricing

*We thank David Andolfatto, Gennaro Bernile, Paymon Khorrami, and Manuel Santos for helpful comments.
†University of Miami, Email: axb2914@miami.edu
‡University of Missouri-Columbia, Email: datmai@mail.missouri.edu
§University of Missouri-Columbia, Email: pukthuanthongk@missouri.edu

1 Introduction

The remarkable advancements in artificial intelligence (AI), epitomized by generative technologies such as ChatGPT, have profoundly reshaped the dynamics of society and businesses. The media consistently emphasizes AI-related developments, sparking investor interest in companies specializing in AI products and services. This increased focus has generated an urgent demand to understand the ramifications of AI on financial markets. Although prior research has delved into the implications of AI replacing human roles in areas like systematic risk (Zhang 2019), innovation (Babina et al. 2021), and capital structure (Bates, Du, and Wang 2020), the existing literature has yet to examine the correlation between the extent of AI relatedness at the firm level and its impact on stock returns. Our study pioneers this essential inquiry, bridging the gap in the current body of research and providing critical insights into the financial consequences of AI relatedness.

Our study addresses this critical gap in the literature by developing a novel measure of firm-level "AIness" that captures the extent of AI integration within a company. This measure is a valuable tool for investors and market participants to understand better the potential impact of AI adoption on a firm's performance and the broader financial landscape. Our research provides valuable insights for investment strategies, corporate decision-making, and risk management by demonstrating that our AIness metric can predict stock returns. Furthermore, our findings contribute to the ongoing discourse on the role of AI in shaping the future of industries and the global economy.

In our study, we extract AI narratives from news sources, building upon the ideas presented by Shiller (2017, 2020), which propose that certain narratives can spread, evolve, influence human behavior, and ultimately impact economic outcomes. We identify AI as a powerful narrative that has the potential to shape financial markets. Our measure is forward-looking as the news content captures the market's attention and the likelihood the market anticipates on particular subjects (Gentzkow and Shapiro 2010; Mullainathan and Shleifer 2005). Our AI metric is a proxy for the market's focus on AI-related developments. By examining AI narratives from as early as the 1970s, we establish that the attention paid to AI has had significant implications for financial markets.

To create a real-time index of public attention to AI, we analyze the frequency of AI-related keywords in 3,000,000 *New York Times (NYT)* articles from 1974 to 2020. We use the state-of-the-art word-embedding method, *word2vec* (Mikolov et al. 2013), to represent each word in the

NYT articles as a numeric vector. We then identify words with high similarity scores to *artificial intelligence*-related keywords, forming our AI dictionary.

We employ a rolling estimation scheme to ensure that our AI dictionary and index are available in real-time. For every month t, we use the last ten years of NYT articles (including the current month) to estimate the *word2vec* model and generate the AI dictionary for that specific month t. Subsequently, we compute the frequencies of terms in this AI dictionary to construct our AI index score using the NYT articles available in the following month $t + 1$. This approach ensures that our AI dictionary is updated at the end of each month, and our AI attention score is available in real time. We continue this rolling estimation scheme until the end of the sample period.

Our estimation strategy fulfills two essential objectives. Firstly, by computing our AI score from the NYT articles using only available real-time information, we effectively address the look-ahead bias, a critical issue in any test for return predictions. Secondly, our dynamic, monthly-updated AI dictionary accommodates the evolving semantics related to artificial intelligence over time. The AI-related concepts prevalent in the 1980s and 1990s differ considerably from contemporary discussions about AI. Consequently, utilizing a static AI dictionary may fail to accurately capture AI-related terms from earlier periods, leading to an underestimation of public attention to AI at the time. Our estimation strategy overcomes this challenge by employing real-time news data to construct a time-varying AI dictionary, ensuring the automatic discovery of relevant concepts and terminology.

After constructing the real-time daily AI attention score from the NYT articles, we estimate firm exposures to AI by regressing daily stock returns onto our daily AI score in a rolling window after controlling for common sources of stock return movements proxied by the six pricing factors from Fama and French (2018).[1] We refer to the stock beta concerning our AI score as the measure of "AIness." Our AIness measure reflects the public's perception of a firm's AI adoption: when investors consider a company an AI-driven firm, its stock price exhibits a high comovement with the AI index, indicating a high AIness score.

We document that AIness has been more volatile over the recent decade and is higher for firms in Business Equipment, Chemicals, Durables, Manufacturing, and Utilities than those in Energy, Health, Nondurables, and Shopping. The Health industry has the lowest AIness. We also find that

[1] As shown in Section 4, our AI beta maintains its predictive power of future returns even after we control for 101 stock predictors from Green, Hand, and Zhang (2017). Therefore, the choice of pricing factors to include when constructing our AI betas does not impact our main results.

AIness is higher for firms having bigger market equity, higher value ratios, higher market betas, higher profits, higher growth, and lower return volatility. Finally, AIness is not related to firm age.

Our approach to capturing firm-level AIness is distinct from measures constructed from firm fundamentals, as found in studies by Zolas et al. (2020), Babina et al. (2022), Lui, Lee, and Ngai (2022), Cao et al. (2020), Alsheibani, Cheung, and Messom (2018), Holmström (2022), Johnk, Weibert, and Wyrtki (2021), and Bäck et al. (2022). Our measure encapsulates the public's perception of the extent of AI usage within a company, including retail and institutional investors. This perception encompasses AI's role in product development pipelines, the automation of operations, and the communication of AI-related plans to investors.

Our primary empirical inquiry investigates the impact of firm-level AIness on future stock returns. We address this question by conducting standard return predictive regressions. Our findings reveal that firms with high AIness exhibit higher returns one month ahead. This result remains robust after controlling for 101 stock predictors from Green, Hand, and Zhang (2017). Firm characteristics are known to be highly correlated, as demonstrated by Rajan and Zingales (1995), Fazzari et al. (1988), Peters and Taylor (2017), and Alti (2003). Both Green, Hand, and Zhang (2017) and Chib et al. (2022) indicate that including more characteristics leads to a purer and more distinct characteristic representation. By incorporating the 101 characteristics examined in (Green, Hand, and Zhang 2017), we demonstrate that the predictive power of AIness persists. This outcome implies that AIness plays a unique role, not subsumed by other firm characteristics. Furthermore, our results remain consistent when using industry-adjusted returns, suggesting that specific industries do not drive the connection between firm AIness and future stock returns.

We further investigate the role of firm size in explaining the impact of AIness. We document that the predictive power of AIness is concentrated among stocks having a market value below the NYSE 50^{th} percentile break point, suggesting that AI news influences small firms more than large firms. Investigating the impact of AIness on long-horizon stock returns, we find that high AIness predicts higher returns one month ahead and lower returns five to seven months ahead. This suggests that investors initially overreact to AI news and subsequently correct their initial beliefs. Once again, this pattern of initial overreactions and subsequent corrections is concentrated among small stocks. Our findings indicate that investors find it more challenging to value small AI stocks, resulting in the return predictability pattern. Big established firms do not need to take costly actions to promote the use of AI in their firms, and the information is already priced in their stock prices. Smaller firms, however, take costly actions of using news and social media, and other outlets to credibly inform investors about the adoption of AI in their product pipeline and business strategy.

In addition to returns, AIness also positively predicts sales growth and returns on equity over the next 12 months. Again, this prediction is concentrated among small stocks. This indicates that investors overreact to AI exposures of small firms because they anticipate higher growth and stronger financial performance. This overreaction leads to higher stock returns one month ahead and lower returns five to seven months ahead when the market corrects.

There could be several reasons why AIness only predicts stock returns and financial performance for small firms but not large ones. One possible explanation is that small firms may have more flexibility and agility to adapt to and implement new technologies like AI compared to large firms, which may have established and rigid structures. As a result, small firms may leverage AI more efficiently and effectively to gain a competitive advantage, which could lead to higher stock returns. Additionally, small firms may have a higher potential for growth, which could be more easily realized through AI.

On the other hand, large firms may already have implemented AI to a greater extent and may have less room for further improvement or efficiency gains. Large firms may also face higher implementation costs and organizational challenges when adopting new technologies like AI, which could limit their ability to reap the benefits of AI adoption. Furthermore, larger firms may be subject to more scrutiny and regulation, which could limit their ability to exploit AI's potential fully.

In our last empirical test, we investigate whether AIness can be used to create a profitable trading strategy. We apply standard portfolio sorting exercises using firm-level AIness. A portfolio sorted on AIness among non-big stocks earns an annualized alpha of around 3% against all leading factors models.[2] This portfolio sorting result is consistent with the regression analysis, suggesting a profitable trading pattern among small AI stocks.

Our study offers several significant contributions to the literature. Firstly, we are pioneers in developing a measure of AIness at the firm level, enriching the growing research on AI replacement

[2] A portfolio sorted on AIness alone earns an annualized alpha of 2%.

and finance. Zhang (2019) presents a theoretical framework emphasizing firms' ability to employ automation to replace routine-task workers, thereby hedging against macroeconomic shocks and reducing systematic risk. Bates, Du, and Wang (2020) explores replacing non-routine tasks with AI-driven automation, shedding light on the potential influence of workplace automation on corporate financial policies in publicly traded US firms. Their empirical results suggest that firms capable of substituting routine and non-routine jobs with automation tend to adopt more aggressive financial strategies. These studies on workplace automation collectively highlight the benefits of advanced technologies in corporate finance.

Empirical findings from Babina et al. (2021) reveal that AI technology firms boast a high-quality workforce characterized by education, training, and marketable skills, leading to increased productivity and product innovation. Our research demonstrates that AIness can predict stock returns, particularly for smaller firms. In addition, our study contributes to the literature on mispricing and risk by showing that the predictive power of AIness reverts within seven months, suggesting that AIness represents mispricing rather than risk. Overall, our work significantly advances understanding in the field by providing novel insights into the relationship between AI and firm performance.

Our research holds considerable practical implications for investors, corporate executives, and policymakers. Investors can leverage the insights gleaned from our study to refine their investment strategies by incorporating AI-related information. Corporate executives may utilize our findings to devise more effective AI strategies that enhance their firm's stock returns. Policymakers can draw on the results to develop policies that promote ethical AI adoption while mitigating potential adverse outcomes. Our study offers valuable guidance for various stakeholders to make informed decisions regarding AI and its impact on financial markets.

2 Construction of News-Based AI Index

In this section, we discuss first our data source and then the construction of our AI dictionary and index.

2.1 Data

We construct the index of AI attention discussed in the NYT articles. We choose the NYT as it is one of the most prestigious and widely circulated worldwide and has been used in finance research such as Garcia (2013) and Hillert and Ungeheuer (2019). From the archive of all NYT articles from 1974 to 2020, we remove articles having fewer than 100 content words to reduce noises introduced by these short articles. Other than this restriction, we keep all articles in all columns and sections of the NYT since we want to study the proportion of NYT content related to AI. This serves as a good approximation of public attention to AI. Our final news sample consists of around 3 million news articles having an average of 400 words per article.

We start our sample in 1974 because the NYT began to talk about AI frequently over the ten years from 1974 to 1984. As discussed in the following subsection, we estimate our model using *word2vec* on a rolling 120-month basis to construct the time-varying dictionary of AI.[3] To prepare texts for *word2vec*, we remove one-letter words, tags, and special characters and perform lemmatization (i.e., remove word endings such as *es* and *ing* and revert words to roots such as *was* → *be*). We do not remove stop words because *word2vec* needs the context of a word to learn its numeric representation.

2.2 Method

We apply an advanced word embedding method called *word2vec* developed by Mikolov et al. (2013) to construct the real-time AI score. *word2vec* has seen huge popularity and adoption in various fields. In finance, *word2vec* has been used to construct dictionaries of corporate culture (Li et al. 2021) or macroeconomics (van Binsbergen et al. 2022). In essence, *word2vec* is an unsupervised word embedding model to construct vector representations for words in a document. It does so by training a neural network to predict the probability of a word given its neighbor words. Thus, *word2vec* is a contextualized NLP method as compared to the traditional bag-of-word approach that treats words in isolation.

We differ from the previous applications of *word2vec* in finance by applying it to develop a time-varying dictionary of AI. More specifically, as illustrated in Figure 1, every month t, we use

[3] Our first AI dictionary is constructed in July 1984, and the first news-based AI score is in August 1984.

the past 120 months (including the current month) of news data to train the *word2vec* model.[4]

The input of *word2vec* is the collection of all sentences belonging to the news articles in the 120-month training window. Each sentence comprises one-word terms (unigrams) and two-word terms (bigrams). Instead of arbitrarily combining individual words into bigrams, we adopt the automatic standard phrase detection method introduced by Mikolov et al. (2013). Specifically, this method combines two unigrams into one bigram if these words commonly occur together. In our training, we combine two words into a bigram if they occur at least five times in the training news collection. For example, the sentence "Artificial intelligence will be widely used." is converted into "artificial_intelligence will be widely used" where *artificial_intelligence* is treated as one word by *word2vec*. The advantage of combining words into bigrams is twofold. First, it helps the model construct vector representations for sensible words. Second, it reduces the number of words that the model is required to learn, thus speeding up the training process.

The output of the monthly estimation process is one numeric vector for each word in the news collection. Following standard practice, we represent each word by a 100×1 vector. Among these produced word vectors, we have a numeric vector to represent the term *artificial_intelligence*. Using these word vectors, we compute the cosine similarity between all other terms in the news collection with *artificial_intelligence*. We can then identify the 100 words having the highest cosine similarities with *artificial_intelligence*, i.e., 100 words most related to *artificial_intelligence*. We refer to these 100 AI-related words as the AI dictionary in month t.

After constructing the AI dictionary in month t, we compute the frequency of these AI words in news articles in the next month $t+1$. We sum up the total counts of these AI-related words and divide it by the total number of words across articles each day in month $t+1$ to compute the daily AI attention scores in month $t+1$.

We roll this estimation method one month forward by running the *word2vec* model in month $t+1$ using news data from month $t-119$ to month $t+1$ and compute the AI score in month $t+2$. Iterating this process forward, we can thus construct real-time daily news-based AI attention.

In Figure 2, we plot the word cloud of our time-varying AI dictionary, which aggregates our monthly 100-word dictionaries. In this word cloud, the bigger the terms, the higher the frequency

[4] We use the `gensism` package in `Python` to train the *word2vec* model. In training, as recommended by the package author, we use a window size of five words and keep words with at least five appearances in the corpus.

of the words. We can see that all frequently used words are strongly related to AI, such as *artificial_intelligence, robotics, computational, molecular_biology,* and *neural_network*.

In Figure 3, we plot the time series of our real-time AI score from 1985 to 2020. Panel A plots the daily index, and Panel B plots the aggregated monthly index. As mentioned above, the NYT began discussing AI in the mid-1970s. As illustrated, the AI index fluctuated during the 1980s and 1990s. After that, news media paid less attention to AI during the 2000s. AI then regained popularity during the 2010s with advances in data storage methods and processing powers, i.e., the availability of "big data." AI attention saw a fallback around 2019-2020. We conjecture that at the time of writing this paper in early 2023, our AI index would see a resurgence due to the ongoing news and social media frenzy over ChatGPT and other recently introduced AI-powered chatbots.

3 Construction of Firm AIness

In this section, we first describe the construction of our firm-level AIness (AI beta) and its characteristics.

3.1 Construction

To measure the degree of firm exposure to AI, we use the comovements between firm returns and our constructed AI index, which proxies investors' attention to AI news. Specifically, we use a rolling six-month window to regress excess daily stock returns onto daily AI and control for common risk factors in Fama and French (2018), including market (MKT), size (SMB), value (HML), momentum (MOM), profitability (RMW), and investment (CMA). Our regression specification is as follows:

$$\begin{aligned} r_{i\tau_{t-5\to t}} = \alpha_{it} &+ \beta_{it}^{AI} AI_{\tau_{t-5\leftarrow t}} + \beta_{it}^{MKT} MKT_{\tau_{t-5\to t}} + \beta_{it}^{SMB} SMB_{\tau_{t-5\to t}} + \beta_{it}^{HML} HML_{\tau_{t-5\to t}} \\ &+ \beta_{it}^{MOM} MOM_{\tau_{t-5\to t}} + \beta_{it}^{RMW} RMW_{\tau_{t-5\to t}} + \beta_{it}^{CMA} CMA_{\tau_{t-5\to t}} + \epsilon_{\tau_{t-5\to t}}, \end{aligned} \tag{1}$$

where $r_{i\tau_{t-5\to t}}$ is the excess return of stock i on day τ in a rolling six-month window from month $t-5$ to month t and β_{it}^{AI} (referred to as *AI beta*) is the measure of firm exposure to AI news after netting out common sources of stock return movements. In the main results, we use a rolling window of six months to construct AI betas since, as shown in later sections, the impact of AI betas on future stock returns is short-term. In Appendix A, we show that using a rolling window of one or three months produces consistent results.

Since we construct AI betas at the firm level, the resulting measures inevitably contain noises. To mitigate the impact of noises and outliers and to put all return predictors on the same scale in predictive regressions, we follow recent papers (Kelly, Pruitt, and Su 2019; Kozak, Nagel, and Santosh 2020) in using rank-transformed characteristics. Specifically, every month t, we compute the rank-transformed characteristics as follows:

$$rc_{it}^{j} = \frac{rank(c_{it}^{j})}{N_t^j + 1} - 0.5, \tag{2}$$

where c_{it}^{j} is characteristic j of firm i in month t, $rank(.)$ is the rank function that yields rank from 1 to N_t^j, and N_t^j is the number of stocks having data on characteristic j in month t. We subtract the rank-transformed characteristics by 0.5 so that the resulting ranked characteristic rc_{it}^{j} falls within the range from -0.5 to 0.5.[5]

3.2 Characteristics

We first examine the heterogeneity of AI betas across industries. Our stock universe is the merged dataset between CRSP and Computstat following Fama and French (1993) merging convention. Our sample is from 1985 to 2020, corresponding to the availability of the AI score. Figure 4 plots the time series of AI betas for each of the 11 Fama-French industries. To compute the industry-level AI betas, we value-weight AI betas of stocks belonging to each industry. The common theme from all industries is that industry-level AI betas become more volatile during the past decade from 2010 to 2020, the period when news media increases coverage of AI-related news as evident in Figure 3.

To quantify the differences in AI betas among industries, we run a panel regression of firm-level AI betas on industry dummies and year-month fixed effects and cluster standard errors by both firms and year-months. We choose Business Equipment as our baseline industry. Panel A of Table 1 shows that Energy, Health, Nondurables, and Others have significantly lower AI betas than Business Equipment. Health has the lowest AI betas among industries and no industry has statistically higher AI betas than Business Equipment.

In Panel B of Table 1, we replace industry dummies with firm characteristics to examine what characteristics explain AI betas. We also include firm and year-month fixed effects. We find that

[5] Dividing ranks by the number of stocks and scaling them to range between -0.5 and 0.5 are deterministic functions that do not change statistical significance in regressions.

AI betas are higher in firms with larger market equity (mve), higher value (bm), higher market beta (beta), higher return on equity (roeq), higher leverage (lev), higher R&D over sales, and higher sales growth (sgr). AI betas are also higher in firms with lower return volatility and are unrelated to firm age.

In Table A.1, we report the five firms having the highest average AI betas for each industry from 2011 to 2020. We include only firms having at least 60 months of AI betas in computing the time-series averages. Unsurprisingly, many of these are high-tech companies. Finally, AI betas are fairly persistent, with an average cross-sectional auto-correlation of 75%.

4 Prediction Results

4.1 Predicting Next One-Month Returns

Our main research question is whether firm exposures to AI news predict future returns after controlling for common stock characteristics in Green, Hand, and Zhang (2017).[6] We employ the standard Fama-MacBeth regression (Fama and MacBeth 1973) to investigate this research question. First, every month t, we run the cross-sectional regression:

$$r_{it+1} = a_t + \lambda_t \times AI\_beta_{it} + \gamma_t^{'} X_{it} + \epsilon_{it+1}, \tag{3}$$

where r_{it+1} is the excess return of stock i in month $t+1$, AI_beta_{it} is AI exposure of firm i in month t, and X_{it} is the control variables from Green, Hand, and Zhang (2017). Second, we compute the time series averages of λ_t and γ_t to estimate the risk premium of AI beta and other control variables. We report the premium estimates with their t-statistics computed with Newey and West (1987) standard errors.

Panel A of Table 2 reports the results for the universe of all stocks. In column (1), when we do not control for any common return predictors, AI beta's estimated annualized risk premium is 2.8%, significant at the 1% level. In economic terms, on average, a firm with the highest AI beta earns an annualized return of 2.8% higher than a firm with the lowest AI beta. When we control for seven common return predictors (size, book-market, market beta, momentum, investment, return on equity, and idiosyncratic volatility) in column (2), the risk premium of AI is still significant at

[6] We drop variable IPO because including this variable substantially reduces the sample size.

the 1% level. In column (3), we extend the list of control variables in column (2) to further include gross profitability, leverage, capital expenditure growth, research, and development expenditure as a percentage of sales, sale concentration, and sales growth, the estimated risk premium of AI beta is still significant at the 5% level.

In Panel B, we examine the universe of stocks having market values below the NYSE 50$^{\text{th}}$ percentile every month and find that the results remain consistent. In Panel C, we examine stocks having a market value above the NYSE 50$^{\text{th}}$ percentile. For this set of large stocks, the risk premium of AI beta is still positive but far from significant. We conclude that AI news only has pricing effects on small stocks.[7]

In Table 2, to mitigate the impacts of noises and estimation errors in AI betas, we rank-transform them before using them in the Fama-MacBeth regression. However, if estimation errors are large, AI betas' ranks are still unreliable. To further mitigate the impact of estimation errors, we transform AI betas into deciles, i.e., instead of ranking AI betas from 1 to N_t, we now rank them from 1 to 10. We again scale the decile-transformed AI betas to range between -0.5 to 0.5. In Table 3, we use decile-transformed AI betas and document that the results are consistent with rank-transformed betas in Table 2.

To investigate whether specific industries drive the predictive power of AI betas, we replace excess returns with industry-adjusted returns in equation (3). We compute industry-adjusted returns as the difference between stock i's return in month t and the value-weighted average return in month t of the industry that stock i belongs to. We use the Fama-French 49 industry classifications. As reported in Table 4, the results for industry-adjusted returns are still consistent with the excess returns discussed above. After netting out the industry return effects, firms with higher AI exposures earn higher returns the next month. The link between AI exposures and future returns is concentrated among small stocks, i.e., stocks having market values below the NYSE 50$^{\text{th}}$ percentile.

To further examine the robustness of the premium of AI betas, we include all 101 stock characteristics from Green, Hand, and Zhang (2017) into our predictive regressions. Chib et al. (2022) show that including additional characteristics reveals a clearer slope in their corresponding relationships not subsumed by other characteristics. Since many of these characteristics are very sparse,

[7] We also try partitioning the universe of stocks into young and old firms but do not find a difference in return predictions between the two groups.

including them in our regressions substantially reduces the number of stock-month observations. Thus, instead of running the Fama-MacBeth regression, which requires us to run the cross-sectional regressions monthly, we pull all stock-month observations into a panel regression. We continue to use rank-transformed characteristics in our panel regression to minimize the impact of outliers. We include stock and year-month fixed effects and cluster standard errors by stocks and year-months. Panel A of Table 5 reports the results for excess returns. Accordingly, the coefficient on AI beta is still significant at the 5% level for all stocks (column 1) and small stocks (column 2). In Panel B, with industry-adjusted returns, the coefficient on AI beta is significant at the 5% level for all stocks and 10% for both small and large stocks.

Overall, firms with higher AI news exposure earn higher returns the next month. The result is concentrated among small stocks and is robust to different controls and specifications.

4.2 Predicting Next Twelve-Month Returns

To investigate the long-term effect of AI news on stock returns, we rerun the Fama-MacBeth regression (3) but with returns from one to twelve months ahead on the left-hand side and control for seven common return predictors (size, book-market, market beta, momentum, investment, return on equity, and idiosyncratic volatility). Table 6 reports the results with each column corresponding to a return horizon from the next one to twelve months. For the case of all stocks in Panel A, AI beta has a significant positive risk premium for one month ahead and yields negative risk premiums over months three to nine. Among these months, the risk premiums in months five and seven are significant at 10%. The results for small stocks in Panel B are consistent with those for all stocks, while big stocks in Panel C do have any significant results.

In Figure 5, we visualize the patterns of AI risk premiums from Table 6. These results suggest that AI news causes initial overreactions in small stocks having high AI exposures, resulting in a positive risk premium the next month. These initial overreactions are corrected in subsequent months, leading to significant negative premiums in months 5 to 7. Compared to large stocks, it is more challenging to value small stocks with high AI exposures, so the influence of AI news is more substantial in these small stocks.

In Table 7, we replace excess returns with industry-adjusted returns and redo regression analysis in Table 6. We find that industry-adjusted returns display the same patterns of initial overreactions to AI news in small stocks and subsequent corrections. Yet the subsequent corrections are not as significant as with excess returns.

We conclude that AI news causes initial overreactions and subsequent corrections among small AI stocks, which are more challenging to value than large tech firms. Thus, AI betas are more associated with mispricing than a risk-based mechanism.

4.3 Predicting Financial Performance

We have shown that AI betas forecast higher returns one month ahead and lower returns five to seven months ahead among small firms. In this section, we provide a possible explanation for this pattern. Specifically, we run the following panel regression using AI betas to predict firm financial fundamentals:

$$y_{i,t\rightarrow t+12} = \alpha + \beta \times AI\_beta_{it} + \epsilon_{i,t\rightarrow t+12}, \tag{4}$$

where $y_{i,t\rightarrow t+12}$ is either sales growth (sgr), capital expenditure growth (grcapx), return on asset (roaq), or return on equity (roeq) of firm i over the next 12 months. We include firm and year-month fixed effects and cluster standard errors by firms and year-months.

We find that for the sample of all stocks in Panel A of Table 8, AI betas positively predict the return on equity (significant at the 5% level), return on assets (significantly at the 10% level), and sales growth (marginally significant at the 10% level) over the next 12 months. For small stocks in Panel B, higher AI betas are associated with higher sales growth and higher return on equity, both significant at the 5% level. In Panel C, the link between AI betas and future financial performance is almost non-existent for big stocks, except for return on equity (significant at the 10% level).

Our results here, combined with the return prediction pattern presented earlier, paint a clearer picture: In response to AI news, investors overreact to high AI beta stocks in anticipation of higher sales growth and stronger financial performance in the short run. This overreaction leads to higher returns one month ahead and lower returns five to seven months ahead due to subsequent corrections. The links between AI exposures and future financial performance and returns are concentrated in small stocks.

5 Portfolio Sorts

This section investigates whether we can exploit AI news to create profitable trading strategies by conducting standard portfolio sorts based on AI betas.

5.1 Portfolios Sorted on AI Beta

We first consider portfolios sorted on AI betas. Specifically, at every month's end, we sort stocks into three equal buckets based on their AI betas as constructed in Section 3 and rebalance the portfolios monthly.

The first two rows of Table 9 report the average rank-transformed AI betas and market values of the three resulting portfolios sorted on AI betas. Stocks in the lowest and highest terciles have smaller market values than those in the middle bucket, indicating that returns of small firms are more likely to comove (in both directions) with AI news than large firms.

Panel A of Table 9 reports results for equal-weighted portfolios. The first row shows that time-series average returns of portfolios sorted on AI betas increase monotonically from the lowest to the highest. The high-minus-low portfolio has an average annualized return of 1.9%, significant at the 1% level. This value is consistent with the risk premium estimated in the univariate Fama-MacBeth regression reported in the first column of Table 2. The remaining rows of Panel A report the time-series alphas of the zero-cost portfolio sorted on AI betas against four leading factor models, including the six-factor model (FF6) of Fama and French (2018), q-factor model (Q5) of Hou et al. (2021), mispricing model of Stambaugh and Yuan (2017), and behavioral model of Daniel, Hirshleifer, and Sun (2020). Accordingly, the equal-weight high-minus-low portfolio yields annualized alphas above 2% against all factor models, all significant at the 1% level.

Panel B shows the results for value-weighted portfolios sorted on AI betas. Since the impact of AI news is concentrated among small stocks, the value-weighted high-minus-low portfolio yields smaller mean returns and alphas than the equal-weighted portfolio. Annualized mean return and alphas are above 1%, but none are significant.

5.2 Portfolios Double-Sorted on Size and AI Beta

We next examine the interplay between size and AI betas. Every month end, we first sort stocks into two size buckets using the NYSE 50^{th} percentile break point. Based on their AI betas, we sort stocks into three equal buckets within each size bucket. We rebalance the portfolios monthly.

Panel A of Table 10 reports the results for equal-weighted portfolios. Within each size bucket, average returns increase monotonically from the lowest to highest AI beta tercile. However, only within small stocks, the high-minus-low AI beta portfolio yields a significant mean return and alphas above 2%, all significant at the 1% level.

Panel B reports the results for value-weighted portfolios. We see a pattern similar to the equal-weighted portfolios: mean returns increase monotonically from the lowest to highest AI beta tercile within each size bucket; yet only within the small size bucket does the top-minus-bottom AI beta portfolio have significant mean return and alphas. Its alphas are greater than those of the equal-weighted counterpart, averaging 3% against the four factors model, all significant at the 1% level.

Overall, we can create profitable portfolios out of AI news, making annualized alphas of about 2% unconditionally, or if we first condition on size, we can earn alphas of annualized 3% within stocks having market values below the NYSE 50^{th} percentile break point.

6 Discussion and Possible Consequences

This section summarizes the advent of large language models (LLMs), e.g., ChatGPT, and the AI narrative this may create for firms. This has also led to an investment frenzy on pre-IPO firms across venture capitalists and private equity investors. Several public companies are gearing up to acquire these niche AI startups. This necessitates ethical and regulatory supervision and has policy implications.

6.1 ChatGPT and Large Language Models

OpenAI's ChatGPT, the fastest-growing application in human history, is built on an LLM using generative/conversational AI. This chatbot enables users to have human-like conversations by recognizing patterns. This growth has outpaced TikTok, Instagram, etc. An LLM is a deep-learning algorithm that can recognize, summarize, translate, predict, and generate text and other content based on knowledge gained from massive data sets. LLMs can teach AIs human languages, understand various subject disciplines, write software code and create graphics based on text descriptions.

Lopez-Lira and Tang (2023) have documented the increase in accuracy of stock price prediction using ChatGPT. Hansen and Kazinnik (2023) show that LLMs like ChatGPT can decode Fedspeak (i.e., the language used by the Fed to communicate on monetary policy decisions). Cowen and Tabarrok (2023) and Korinek (2023) demonstrate that ChatGPT is helpful in teaching economics and conducting economic research. Noy and Zhang (2023) find that ChatGPT can enhance productivity in professional writing jobs. Also, Yang and Menczer (2023) demonstrate that ChatGPT successfully identifies credible news outlets.

Arguably, ChatGPT and GPT-4 will create new AI narratives for firms. It remains to be seen how firms use these LLMs to their advantage. Opportunities to improve productivity seem most apparent in the information technology, education, government, and business services industries. But despite being embraced by businesses and organizations worldwide, ChatGPT poses risks. Considering the potential impact on society, rules, and regulations might be needed for artificial intelligence (AI) development.

6.2 AI-frenzy in Pre-IPO private investments

Artificial Intelligence stocks have been surging on the hype surrounding ChatGPT, the viral AI chatbot that reached 100 million users in just two months from November 2022. Microsoft is making a significant investment in OpenAI, the creator of ChatGPT. Alphabet, which controls 93% of the search market, scrambled to roll out its AI chatbot BARD. Bloomberg announced they would release ChatGPT Bloomberg, Elon Musk has launched an AI company, and Amazon is launching its generative AI service through its cloud computing platform. The AI war among tech giants is heating up as generative technologies capture investors' attention. Mentions of AI, machine learning, and related terms have surged in the recent earnings calls of the biggest software and semiconductor companies.

6.3 Similarity to the 2001 Dotcom Boom

The proliferation of the World Wide Web (WWW) during the latter half of the 1990s ushered in a new era of internet-based companies, keen to capitalize on the wave of hype and anticipation the online world was expected to offer. Indeed, the internet is widely heralded as one of the biggest technological breakthroughs. The term "Dot Com" was magic. It promised vast riches from companies that leveraged the power of the internet.

As with the "Dot Com" boom, there is a great deal of hope, marketing, and empty promises swirling around the recent AI explosion. The challenge is to separate fact from fiction. While this boom and crash were unfortunate for investors, it produced some of the most innovative ideas ahead of their time. Concepts developed by many companies that went under, including VoIP, eCommerce, big data, and the web experience, still live on today, in many cases, as the fundamental concepts driving success in large corporations. But, amid the app-building VC frenzy, we must remain mindful of fundamental differences between today's tech environment and the dot-com era.

7 Conclusion

This paper investigates the asset pricing impacts of public attention to artificial intelligence (AI). To extract public attention to AI, we apply an advanced NLP method called *word2vec* to build a dynamic dictionary of AI terms from the New York Times (NYT) articles. We then count the frequency of this AI dictionary from 3,000,000 NYT articles from 1974 to 2020 to construct a real-time index of AI attention.

After controlling for benchmark pricing factors, we compute firm exposures to AI (AIness) by regressing daily stock returns onto our daily AI score. Our main research question is how AIness forecasts future stock returns. We document that high AIness forecasts higher returns one month ahead and lower returns five to seven months ahead. The predictive power is concentrated among non-big stocks. Our findings indicate that small AI firms are more difficult to value, resulting in initial market overreactions to AI news and subsequent market corrections.

We conduct portfolio sorts on AIness to examine whether AI news can be used to create profitable trading strategies. We find that a high-minus-low AI portfolio among non-big stocks generates annual alphas of 3% against benchmark leading factor models. Our paper reveals that the AI narrative has a sizable impact on stock prices and can be exploited to create trading profits.

References

Alsheibani, S. A., Cheung, Y. P., & Messom, C. H. (2018). Artificial intelligence adoption: AI-readiness at firm-Level. *Pacific Asia Conference on Information Systems.*

Alti, A. (2003). How sensitive is investment to cash flow when financing is frictionless? *Journal of Finance*, *58*(2), 707–722.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2021). Artificial intelligence, firm growth, and product innovation. *Firm Growth, and Product Innovation.*

Babina, T., Fedyk, A., He, A. X., & Hodson, J. (2022). Firm investments in artificial intelligence technologies and changes in workforce composition. In *Technology, productivity, and economic growth.* University of Chicago Press.

Bäck, A., Hajikhani, A., Jäger, A., Schubert, T., & Suominen, A. (2022). *Return of the Solow-paradox in AI? AI-adoption and firm productivity* (Papers in Innovation Studies No. 2022/1). Lund University, CIRCLE - Centre for Innovation Research.

Bates, T. W., Du, F., & Wang, J. J. (2020). Workplace automation and corporate financial policy. *Working paper.*

Cao, S., Jiang, W., Yang, B., & Zhang, A. L. (2020). *How to talk when a machine is listening?: Corporate disclosure in the age of AI* (Working Paper No. 27950). National Bureau of Economic Research.

Chib, S., Lin, Y. C., Pukthuanthong, K., & Zeng, X. (2022). Asset pricing with slope factors: Model and evidence of outperformance.

Cowen, T., & Tabarrok, A. T. (2023). How to learn and teach economics with large language models, including GPT. *Working Paper.*

Daniel, K., Hirshleifer, D., & Sun, L. (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies*, *33*(4), 1673–1736.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, *128*(2), 234–252.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636.

Fazzari, S. M., Hubbard, R. G., Petersen, B. C., Blinder, A. S., & James, M. (1988). Financing corporate constraints investment. *Brookings Papers on Economic Activity*, *1*(1), 141–206.

Garcia, D. (2013). Sentiment during recessions. *Journal of Finance*, *68*(3), 1267–1300.

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, *78*(1), 35–71.

Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *Review of Financial Studies*, *30*(12), 4389–4436.

Hansen, A. L., & Kazinnik, S. (2023). Can ChatGPT decipher Fedspeak? *Working Paper.*

Hillert, A., & Ungeheuer, M. (2019). The value of visibility. *Working Paper.*

Holmström, J. (2022). From AI to digital transformation: The AI readiness framework. *Business Horizons*, *65*(3), 329–339.

Hou, K., Mo, H., Xue, C., & Zhang, L. (2021). An augmented q-factor model with expected growth. *Review of Finance*, *25*(1), 1–41.

Johnk, J., Weibert, M., & Wyrtki, K. (2021). Ready or not, AI comes–An interview study of organizational AI readiness factors. *Business Information Systems Engineering*, *63*, 1–16.

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*(3), 501–524.

Korinek, A. (2023). *Language models and cognitive automation for economic research* (Working Paper No. 30957). National Bureau of Economic Research.

Kozak, S., Nagel, S., & Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, *135*(2), 271–292.

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *Review of Financial Studies*, *34*(7), 3265–3315.

Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT forecast stock price movements? Return predictability and large language models. *Working Paper*.

Lui, A. K. H., Lee, M. C. M., & Ngai, E. W. T. (2022). Impact of artificial intelligence investment on firm value. *Annals of Operations Research*, *308*(1), 373–388.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mullainathan, S., & Shleifer, A. (2005). The market for news. *American Economic Review*, *95*(4), 1031–1053.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Working Paper*.

Peters, R. H., & Taylor, L. A. (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics*, *123*(2), 251–272.

Rajan, R. G., & Zingales, L. (1995). What do we know about capital structure? some evidence from international data. *Journal of Finance*, *50*(5), 1421–1460.

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004.

Shiller, R. J. (2020). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.

Stambaugh, R. F., & Yuan, Y. (2017). Mispricing factors. *Review of Financial Studies*, *30*(4), 1270–1315.

van Binsbergen, J. H., Bryzgalova, S., Mukhopadhyay, M., & Sharma, V. (2022). (Almost) 200 years of news-based economic sentiment. *Working Paper*.

Yang, K.-C., & Menczer, F. (2023). Large language models can rate news outlet credibility.

Zhang, M. B. (2019). Labor-technology substitution: Implications for asset pricing. *Journal of Finance*, *74*(4), 1793–1839.

Zolas, N., Kroff, Z., Brynjolfsson, E., McElheran, K., Beede, D. N., Buffington, C., Goldschlag, N., Foster, L., & Dinlersoz, E. (2020). *Advanced technologies adoption and use by U.S. firms: Evidence from the annual business survey* (Working Paper No. 28290). National Bureau of Economic Research.

Figure 1. Estimation Scheme

This figure plots the rolling estimation scheme to construct the AI score. Every month t, news articles in the previous 120 months (including month t) are used to compute word vectors via *word2vec*. These word vectors are used to construct a dynamic 100-word AI dictionary. Articles in month $t+1$ are then used to construct the real-time AI score.

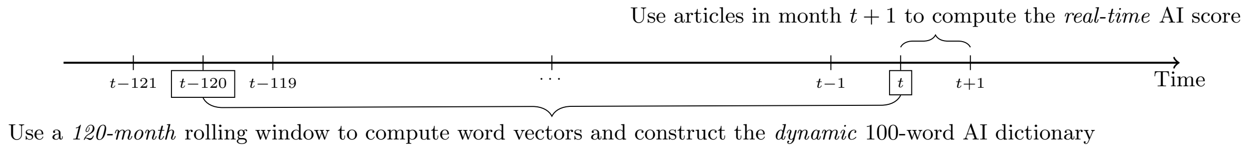

Figure 2. AI Word Cloud

This figure plots the word cloud for terms related to *artificial intelligence (AI)* in the New York Times articles from 1985 to 2020. The bigger the size, the more frequently the term appears in news articles. See Section 2 for details on extracting these terms.

Figure 3. Time Series AI Index

This figure plots the time series of the daily artificial intelligence (AI) index constructed from the New York Times articles. The sample is from 1985 to 2020. See Section 2 for details on constructing the AI index.

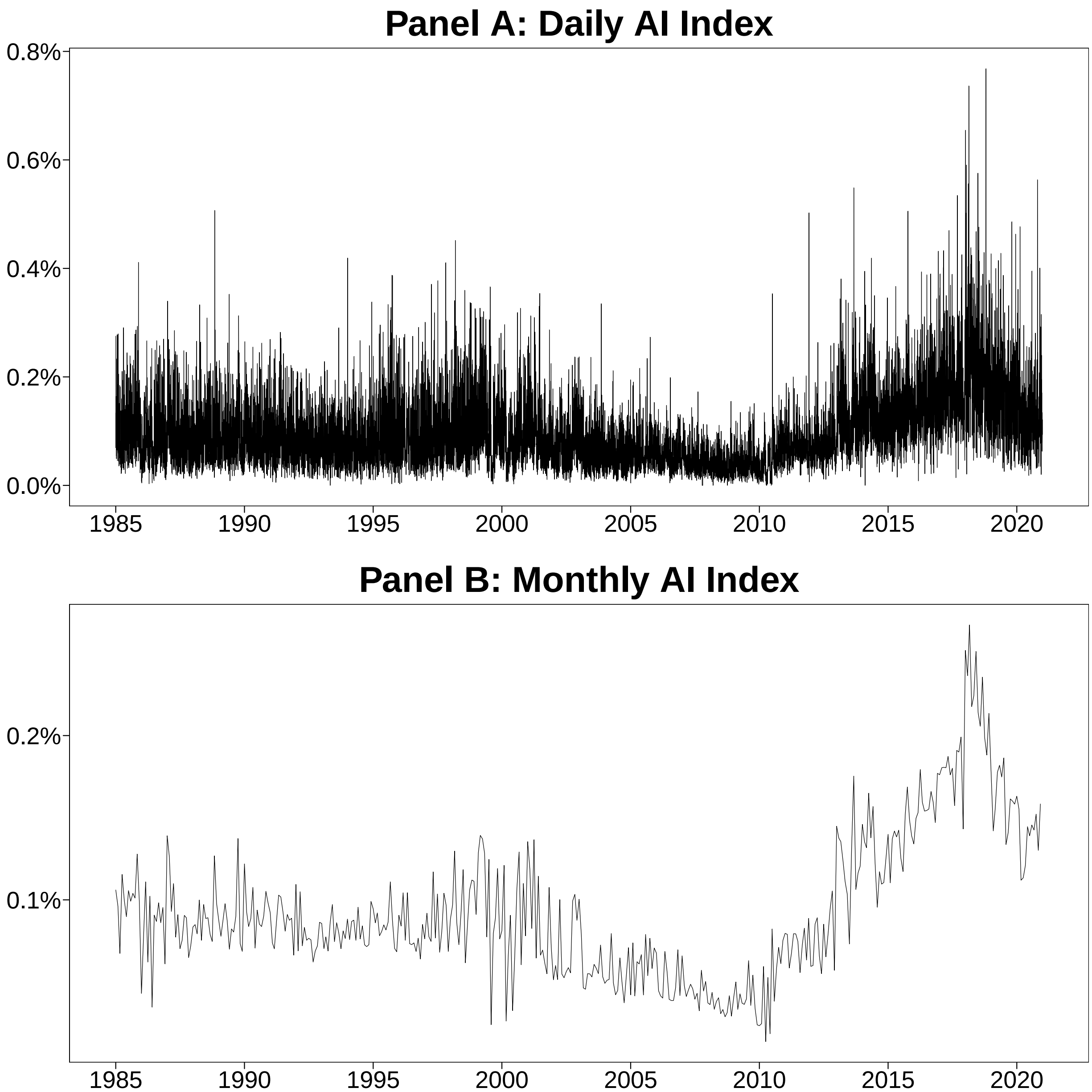

Figure 4. Time Series of AI Betas at Industry Level

This figure plots the time series of artificial intelligence (AI) betas for 11 Fama-French industries. We compute monthly AI betas for every stock by regressing daily stock returns onto the daily AI index and six Fama-French factors using a rolling six-month window. Every month, we cross-sectionally scale AI betas from -0.5 to 0.5. Scaled stock-level AI betas are then averaged into industry-level betas weighted by stock market capitalization. The sample is from 1985 to 2020. See Section 2 for details on constructing the AI index.

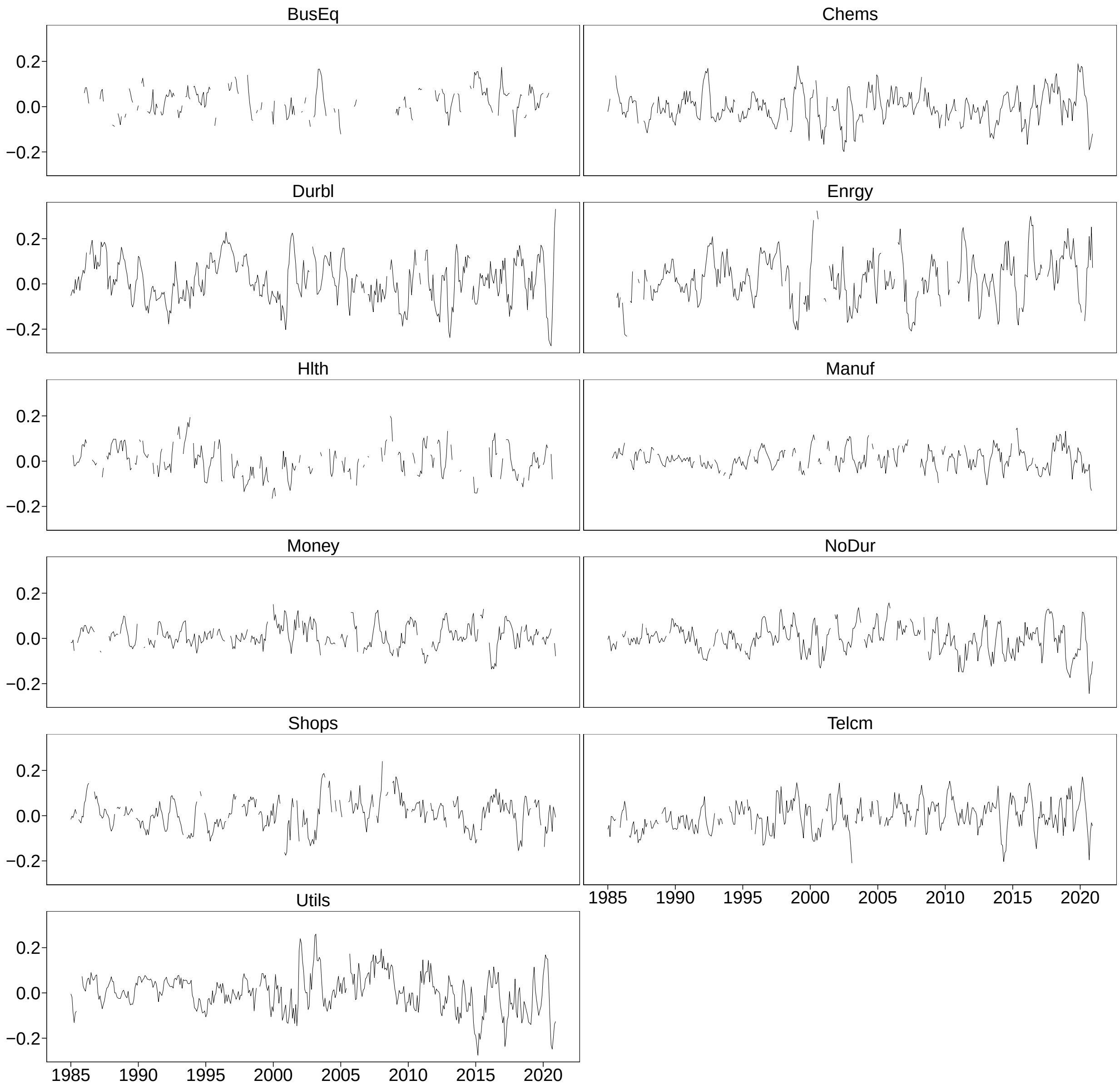

Figure 5. Predicting Next 12-Month Returns

This figure plots the factor risk premium for AI betas estimated via Fama-Macbeth regressions. Every month, we run the following cross-sectional regression:

$$r_{it+h} = a + \lambda_t \times AI\_beta_{it} + \gamma_t^{'} X_{it} + \epsilon_{it+h},$$

where r_{it+h} is excess return of stock i in month $t+h$ in annualized percentage point (x axis), AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index and Fama-French six factors over a rolling six-month window, and X_{it} is a set of control variables of stock i in month t: market equity (mve), book-market (bm), market beta (beta), momentum (mom12m), annual change in investment (invest), return on equity (roeq), and standard deviation of daily returns from previous month (retvol). Every month, all independent variables are ranked and rescaled to fall within the range from -0.5 to 0.5. Solid lines are time-series averages of λ_t and shades indicate the 90^{th} confidence interval using Newey-West standard errors with six lags. Panel A reports results for all stocks; Panel B (C) reports results for stock below (above) the 50^{th} percentile NYSE market value break point. The sample is from 1985 to 2020.

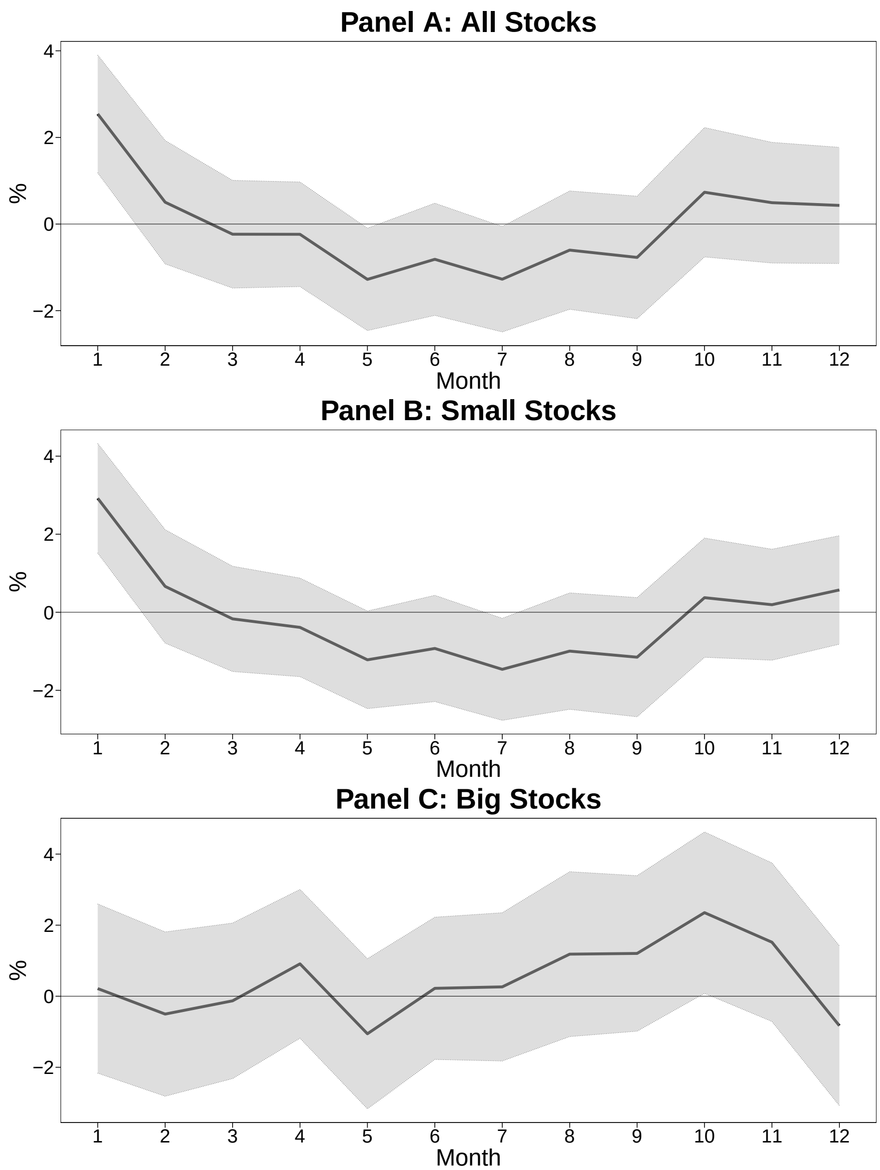

Table 1
Characteristics of AI Betas

This table reports results from the following panel regression

$$AI\_beta_{it} = \alpha + \beta X_{it} + e_{it},$$

AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index and Fama-French six factors over a rolling six-month window and X_{it} is a vector of industry dummies in Panel A and firm characteristics in Panel B. The baseline industry in Panel A is Business Equipment; the vector of characteristics in Panel B includes age, market equity (mve), book-market (bm), market beta (beta), momentum (mom12m), standard deviation of daily returns from previous month (retvol), return on equity (roeq), gross profit divided total asset (gma), leverage (lev), two-year percent change in capital expenditures (grcapx), R&D divided by sales (rd_sale), annual change in investment (invest), industry sales concentration (herf), and sales growth (sgr). Every month, both AI betas and firm characteristics are ranked and rescaled to fall within the range from -0.5 to 0.5. In Panel A, the regression includes year-month fixed effects; in Panel B, the regression includes both firm and year-month fixed effects. Standard errors in both panels are clustered by firms and year-months. The sample is from 1985 to 2020.

Panel A: Industry Dummies

	β	t
Chems	0.002	(0.44)
Durbl	0.003	(0.88)
Enrgy	-0.012	(-2.63)
Hlth	-0.016	(-5.76)
Manuf	0.001	(0.58)
Money	-0.001	(-0.60)
NoDur	-0.006	(-1.96)
Other	-0.006	(-2.25)
Shops	-0.005	(-1.81)
Telcm	-0.004	(-1.04)
Utils	0.003	(0.82)

Panel B: Firm Characteristics

	β	t
age	-0.005	(-0.48)
mve	0.025	(2.25)
bm	0.010	(1.79)
beta	0.011	(2.30)
mom12m	-0.001	(-0.29)
invest	0.001	(0.27)
roeq	0.006	(1.93)
retvol	-0.016	(-4.63)
gma	0.005	(0.70)
lev	0.013	(1.76)
grcapx	-0.004	(-1.01)
rd_sale	0.042	(3.45)
herf	0.005	(0.62)
sgr	0.007	(1.86)

Table 2
Predicting Next One-Month Returns

This table report factor risk premia estimated via Fama-Macbeth regressions. Every month, we run the following cross-sectional regression:

$$r_{it+1} = a_t + \lambda_t \times AI\_beta_{it} + \gamma_t^{'} X_{it} + \epsilon_{it+1},$$

where r_{it+1} is the excess return of stock i in month $t+1$ in annualized percentage point, AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index and Fama-French six factors over a rolling six-month window, and X_{it} is a set of control variables of stock i in month t: market equity (mve), book-market (bm), market beta (beta), momentum (mom12m), standard deviation of daily returns from previous month (retvol), return on equity (roeq), gross profit divided total asset (gma), leverage (lev), two-year percent change in capital expenditures (grcapx), R&D divided by sales (rd_sale), annual change in investment (invest), industry sales concentration (herf), and sales growth (sgr). Every month, all independent variables are ranked and rescaled to fall within the range from -0.5 to 0.5. Reported are time-series averages of λ_t and γ_t with their t-statistics corrected for Newey-West standard errors with six lags in parentheses. Panel A reports results for all stocks; Panel B (C) reports results for stock below (above) the 50$^{\text{th}}$ percentile NYSE market value break point. The sample is from 1985 to 2020.

	Panel A: All Stocks			Panel B: Small Stocks			Panel C: Big Stocks		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
AI_beta	2.77	2.54	1.90	2.93	2.92	2.25	1.83	0.22	0.11
	(2.92)	(3.09)	(2.03)	(3.04)	(3.44)	(2.22)	(1.04)	(0.15)	(0.07)
mve		-16.20	-15.24		-21.32	-19.07		-7.91	-10.73
		(-4.90)	(-3.93)		(-4.71)	(-3.82)		(-1.00)	(-1.25)
bm		1.47	5.82		2.04	6.72		-1.08	4.01
		(0.50)	(2.32)		(0.68)	(2.55)		(-0.35)	(1.35)
beta		5.53	2.00		7.52	3.40		1.65	-0.97
		(1.50)	(0.61)		(1.92)	(0.99)		(0.40)	(-0.25)
mom12m		7.31	2.76		7.84	2.41		7.38	6.03
		(2.02)	(0.81)		(2.22)	(0.69)		(1.79)	(1.68)
invest		-8.32	-6.30		-9.15	-7.74		-3.02	0.31
		(-5.54)	(-4.70)		(-5.59)	(-5.06)		(-2.36)	(0.17)
roeq		6.85	7.68		7.29	8.36		3.60	4.20
		(3.10)	(4.67)		(2.99)	(4.61)		(1.90)	(2.17)
retvol		-5.29	-5.53		-5.06	-5.61		-6.41	-4.77
		(-1.93)	(-2.04)		(-1.80)	(-2.03)		(-2.08)	(-1.70)
gma			7.38			7.37			5.97
			(3.16)			(2.91)			(1.90)
lev			8.08			8.45			3.48
			(2.68)			(2.57)			(1.17)
grcapx			-2.96			-2.49			-4.28
			(-2.50)			(-1.95)			(-2.47)
rd_sale			15.86			16.85			8.68
			(4.19)			(4.26)			(2.37)
herf			-0.60			-0.83			-0.42
			(-0.36)			(-0.43)			(-0.26)
sgr			-1.77			-1.93			1.72
			(-1.42)			(-1.39)			(0.80)
Intercept	11.97	12.33	15.38	12.56	11.39	14.68	9.97	10.42	13.99
	(3.42)	(3.56)	(3.88)	(3.34)	(3.36)	(3.72)	(3.65)	(2.33)	(2.66)
R^2(%)	0.05	4.08	4.65	0.05	3.67	4.05	0.29	10.09	12.23

Table 3
Predicting Next One-Month Returns:
Decile-Transformed AI Beta

This table report factor risk premia estimated via Fama-Macbeth regressions. Every month, we run the following cross-sectional regression:

$$r_{it+1} = a_t + \lambda_t \times AI\_beta_{it} + \gamma_t^{'} X_{it} + \epsilon_{it+1},$$

where r_{it+1} is the excess return of stock i in month $t+1$ in annualized percentage point, AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index and Fama-French six factors over a rolling six-month window, and X_{it} is a set of control variables of stock i in month t: market equity (mve), book-market (bm), market beta (beta), momentum (mom12m), standard deviation of daily returns from previous month (retvol), return on equity (roeq), gross profit divided total asset (gma), leverage (lev), two-year percent change in capital expenditures (grcapx), R&D divided by sales (rd_sale), annual change in investment (invest), industry sales concentration (herf), and sales growth (sgr). Every month, AI betas are ranked into 10 deciles and all other independent variables are normally ranked and rescaled to fall within the range from -0.5 to 0.5. Reported are time-series averages of λ_t and γ_t with their t-statistics corrected for Newey-West standard errors with six lags in parentheses. Panel A reports results for all stocks; Panel B (C) reports results for stock below (above) the 50$^{\text{th}}$ percentile NYSE market value break point. The sample is from 1985 to 2020.

	Panel A: All Stocks			Panel B: Small Stocks			Panel C: Big Stocks		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
AI_beta	2.89	2.66	1.92	3.07	3.08	2.31	1.82	0.07	-0.12
	(2.83)	(2.98)	(1.89)	(2.96)	(3.35)	(2.09)	(0.96)	(0.04)	(-0.07)
mve		-16.20	-15.21		-21.32	-19.01		-7.88	-10.45
		(-4.90)	(-3.93)		(-4.71)	(-3.82)		(-1.00)	(-1.21)
bm		1.47	5.83		2.04	6.74		-1.05	4.06
		(0.50)	(2.32)		(0.68)	(2.56)		(-0.34)	(1.37)
beta		5.54	2.00		7.52	3.37		1.67	-0.91
		(1.50)	(0.62)		(1.92)	(0.99)		(0.40)	(-0.23)
mom12m		7.32	2.78		7.85	2.43		7.37	6.06
		(2.02)	(0.81)		(2.22)	(0.70)		(1.78)	(1.69)
invest		-8.32	-6.36		-9.16	-7.81		-3.02	0.34
		(-5.54)	(-4.73)		(-5.60)	(-5.07)		(-2.35)	(0.18)
roeq		6.85	7.63		7.29	8.33		3.60	4.17
		(3.09)	(4.63)		(2.99)	(4.59)		(1.90)	(2.16)
retvol		-5.30	-5.50		-5.08	-5.59		-6.41	-4.77
		(-1.93)	(-2.02)		(-1.80)	(-2.02)		(-2.08)	(-1.69)
gma			7.38			7.40			5.98
			(3.15)			(2.92)			(1.91)
lev			7.96			8.35			3.45
			(2.63)			(2.53)			(1.15)
grcapx			-2.98			-2.51			-4.23
			(-2.52)			(-1.96)			(-2.46)
rd_sale			15.65			16.69			8.52
			(4.21)			(4.27)			(2.40)
herf			-0.54			-0.81			-0.38
			(-0.32)			(-0.42)			(-0.24)
sgr			-1.78			-1.96			1.72
			(-1.43)			(-1.42)			(0.80)
Intercept	11.97	12.33	15.38	12.56	11.39	14.69	9.98	10.42	13.87
	(3.43)	(3.56)	(3.88)	(3.34)	(3.36)	(3.72)	(3.65)	(2.32)	(2.64)
R^2(%)	0.05	4.08	4.65	0.04	3.66	4.04	0.29	10.08	12.22

Table 4
Predicting Next One-Month Returns:
Industry-Adjusted Returns

This table report factor risk premium estimated via Fama-Macbeth regressions. Every month, we run the following cross-sectional regression:

$$r_{it+1} = a_t + \lambda_t \times AI\_beta_{it} + \gamma_t^{'} X_{it} + \epsilon_{it+1},$$

where r_{it+1} is the return of stock i over industry value-weighted average return in month $t+1$ in annualized percentage point, AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index and Fama-French six factors over a rolling six-month window, and X_{it} is a set of control variables of stock i in month t: market equity (mve), book-market (bm), market beta (beta), momentum (mom12m), standard deviation of daily returns from previous month (retvol), return on equity (roeq), gross profit divided total asset (gma), leverage (lev), two-year percent change in capital expenditures (grcapx), R&D divided by sales (rd_sale), annual change in investment (invest), industry sales concentration (herf), and sales growth (sgr). Every month, all independent variables are ranked and rescaled to fall within the range from -0.5 to 0.5. Reported are time-series averages of λ_t and γ_t with their t-statistics corrected for Newey-West standard errors with six lags in parentheses. Panel A reports results for all stocks; Panel B (C) reports results for stocks below (above) the 50$^{\text{th}}$ percentile NYSE market value break point. The sample is from 1985 to 2020.

	Panel A: All Stocks			Panel B: Small Stocks			Panel C: Big Stocks		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
AI_beta	2.52	2.39	1.99	2.65	2.65	2.11	1.55	1.09	1.60
	(2.89)	(3.10)	(2.13)	(2.88)	(3.17)	(2.07)	(1.25)	(0.94)	(1.06)
mve		-15.06	-15.24		-20.22	-19.36		-5.81	-8.96
		(-4.74)	(-3.97)		(-4.56)	(-3.86)		(-0.91)	(-1.21)
bm		2.47	5.51		2.85	6.50		0.18	2.28
		(1.02)	(2.03)		(1.10)	(2.29)		(0.10)	(0.86)
beta		5.50	3.00		7.53	4.31		1.29	0.79
		(2.04)	(1.18)		(2.39)	(1.46)		(0.67)	(0.36)
mom12m		5.97	1.49		6.83	1.52		4.24	3.00
		(1.88)	(0.48)		(2.17)	(0.47)		(1.27)	(0.98)
invest		-7.15	-5.90		-8.19	-7.33		-1.07	0.94
		(-5.86)	(-4.26)		(-6.06)	(-4.63)		(-1.10)	(0.52)
roeq		6.81	8.06		7.56	8.86		2.32	4.04
		(3.32)	(4.79)		(3.38)	(4.78)		(1.51)	(2.19)
retvol		-5.30	-5.52		-4.91	-5.38		-6.94	-5.33
		(-2.30)	(-2.22)		(-2.02)	(-2.03)		(-2.81)	(-2.08)
gma			5.26			5.61			2.54
			(2.05)			(2.03)			(0.83)
lev			7.65			7.79			4.36
			(2.69)			(2.45)			(1.63)
grcapx			-2.54			-2.05			-3.79
			(-2.20)			(-1.58)			(-2.56)
rd_sale			12.07			13.51			3.65
			(4.45)			(4.51)			(1.54)
herf			-3.38			-3.92			-2.16
			(-1.93)			(-1.93)			(-1.57)
sgr			-0.49			-0.69			3.25
			(-0.41)			(-0.52)			(1.65)
Intercept	2.39	2.70	5.00	2.94	1.77	4.21	0.56	0.65	3.20
	(1.46)	(1.67)	(2.36)	(1.48)	(1.11)	(2.02)	(1.16)	(0.21)	(0.88)
R^2(%)	0.04	2.96	3.63	0.04	2.90	3.40	0.13	4.44	6.24

Table 5
Predicting Next One-Month Returns in Panel Regressions

This table report factor risk premia estimated via the panel regressions:

$$r_{it+1} = a_t + \lambda \times AI\_beta_{it} + \gamma^{'} X_{it} + \epsilon_{it+1},$$

where r_{it+1} is the excess return of stock i in month $t+1$ in annualized percentage point, AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index and Fama-French six factors over a rolling six-month window, and X_{it} is a set of 101 control variables of stock i in month t from Green, Hand, and Zhang (2017). Every month, all independent variables are ranked and rescaled to fall from -0.5 to 0.5. We include firm and month fixed effects and cluster standard errors by firm and month. Column (1) reports results for all stocks; column (2)/(3) reports results for stocks below/above the 50$^{\text{th}}$ percentile NYSE market value break point. Panel A reports results for excess returns while Panel B reports results for industry-adjusted returns. The sample is from 1985 to 2020.

	All Stocks (1)	Small Stocks (2)	Big Stocks (3)
Panel A: Excess Returns			
AI_beta	5.59	7.39	2.67
	(2.15)	(2.19)	(0.79)
$R^2(\%)$	20.19	20.40	24.75
Controls	101 Firm Characteristics		
Fixed Effects	Firm and Year-Month		
Cluster S.E.	Firm and Year-Month		
Panel B: Industry-Adjusted Returns			
AI_beta	5.57	5.75	5.84
	(2.29)	(1.76)	(1.92)
$R^2(\%)$	4.79	6.91	3.10
Controls	101 Firm Characteristics		
Fixed Effects	Firm and Year-Month		
Cluster S.E.	Firm and Year-Month		

Table 6
Predicting Next 12-Month Returns

This table report factor risk premium estimated via Fama-Macbeth regressions. Every month, we run the following cross-sectional regression:

$$r_{it+h} = a + \lambda_t \times AI\_beta_{it} + \gamma_t^{'} X_{it} + \epsilon_{it+h},$$

where r_{it+h} is the excess return of stock i in month $t+h$ in annualized percentage point (each h in each column header), AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index and Fama-French six factors over a rolling six-month window, and X_{it} is a set of control variables of stock i in month t: market equity (mve), book-market (bm), market beta (beta), momentum (mom12m), annual change in investment (invest), return on equity (roeq), and standard deviation of daily returns from previous month (retvol). Every month, all independent variables are ranked and rescaled to fall within the range from -0.5 to 0.5. Reported are time-series averages of λ_t and γ_t with their t-statistics corrected for Newey-West standard errors with six lags in parentheses. Panel A reports results for all stocks; Panel B (C) reports results for stocks below (above) the 50^{th} percentile NYSE market value break point. The sample is from 1985 to 2020.

Return	(t+1)	(t+2)	(t+3)	(t+4)	(t+5)	(t+6)	(t+7)	(t+8)	(t+9)	(t+10)	(t+11)	(t+12)
						Panel A: All Stocks						
AI_beta	2.54	0.50	-0.23	-0.24	-1.28	-0.81	-1.27	-0.60	-0.77	0.73	0.49	0.43
	(3.09)	(0.58)	(-0.31)	(-0.32)	(-1.78)	(-1.04)	(-1.73)	(-0.73)	(-0.90)	(0.81)	(0.58)	(0.53)
Intercept	12.33	12.21	12.30	12.50	12.55	12.54	12.62	12.79	12.99	13.09	12.98	12.92
	(3.56)	(3.51)	(3.53)	(3.59)	(3.59)	(3.59)	(3.61)	(3.65)	(3.70)	(3.72)	(3.69)	(3.67)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2(%)	4.08	3.96	3.84	3.75	3.66	3.58	3.50	3.42	3.34	3.30	3.26	3.25
						Panel B: Small Stocks						
AI_beta	2.92	0.66	-0.17	-0.39	-1.22	-0.93	-1.46	-1.00	-1.15	0.37	0.19	0.57
	(3.44)	(0.75)	(-0.21)	(-0.51)	(-1.61)	(-1.12)	(-1.85)	(-1.10)	(-1.25)	(0.40)	(0.22)	(0.68)
Intercept	11.39	11.47	11.76	11.84	11.90	11.89	11.97	12.04	12.38	12.47	12.38	12.37
	(3.36)	(3.36)	(3.44)	(3.46)	(3.48)	(3.47)	(3.49)	(3.52)	(3.60)	(3.60)	(3.58)	(3.56)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2(%)	3.67	3.54	3.42	3.33	3.25	3.17	3.09	3.01	2.92	2.89	2.85	2.83
						Panel C: Big Stocks						
AI_beta	0.22	-0.50	-0.13	0.91	-1.06	0.22	0.26	1.18	1.20	2.35	1.52	-0.83
	(0.15)	(-0.36)	(-0.10)	(0.72)	(-0.83)	(0.18)	(0.21)	(0.84)	(0.91)	(1.71)	(1.12)	(-0.61)
Intercept	10.42	12.04	12.46	12.30	10.96	12.92	13.99	13.42	13.14	12.89	11.42	12.33
	(2.33)	(2.67)	(2.73)	(2.75)	(2.38)	(2.82)	(3.00)	(2.88)	(2.79)	(2.88)	(2.42)	(2.58)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2(%)	10.09	9.84	9.65	9.36	9.05	8.82	8.72	8.49	8.34	8.16	8.12	7.97

Table 7
Predicting Next 12-Month Returns: Industry-Adjusted Returns

This table report factor risk premium estimated via Fama-Macbeth regressions. Every month, we run the following cross-sectional regression:

$$r_{it+h} = a_t + \lambda_t \times AI\_beta_{it} + \gamma_t^{'} X_{it} + \epsilon_{it+h},$$

where r_{it+h} is the return of stock i over industry value-weighted average return in month $t+h$ in annualized percentage point (each h in each column header), AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index and Fama-French six factors over a rolling six-month window, and X_{it} is a set of control variables of stock i in month t: market equity (mve), book-market (bm), market beta (beta), momentum (mom12m), annual change in investment (invest), return on equity (roeq), and standard deviation of daily returns from previous month (retvol). Every month, all independent variables are ranked and rescaled to fall from -0.5 to 0.5. Reported are time-series averages of λ_t and γ_t with their t-statistics corrected for Newey-West standard errors with six lags in parentheses. Panel A reports results for all stocks; Panel B (C) reports results for stocks below (above) the 50$^{\text{th}}$ percentile NYSE market value break point. The sample is from 1985 to 2020.

Return	(t+1)	(t+2)	(t+3)	(t+4)	(t+5)	(t+6)	(t+7)	(t+8)	(t+9)	(t+10)	(t+11)	(t+12)
Panel A: All Stocks												
AI_beta	2.39	0.50	-0.04	-0.03	-0.76	-0.46	-0.84	-0.54	-1.00	0.36	0.14	0.14
	(3.10)	(0.62)	(-0.06)	(-0.04)	(-1.07)	(-0.65)	(-1.17)	(-0.71)	(-1.22)	(0.42)	(0.18)	(0.17)
Intercept	2.70	2.60	2.62	2.75	2.92	2.95	3.01	3.11	3.15	3.30	3.36	3.38
	(1.67)	(1.60)	(1.61)	(1.68)	(1.79)	(1.80)	(1.83)	(1.89)	(1.92)	(2.00)	(2.03)	(2.04)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2(%)	2.96	2.87	2.78	2.72	2.66	2.59	2.53	2.47	2.41	2.36	2.33	2.32
Panel B: Small Stocks												
AI_beta	2.65	0.52	-0.12	-0.27	-0.88	-0.64	-1.12	-0.88	-1.31	0.01	-0.16	0.27
	(3.17)	(0.60)	(-0.14)	(-0.34)	(-1.14)	(-0.83)	(-1.46)	(-1.05)	(-1.48)	(0.01)	(-0.20)	(0.32)
Intercept	1.77	1.86	2.08	2.08	2.25	2.26	2.36	2.37	2.55	2.66	2.73	2.83
	(1.11)	(1.16)	(1.28)	(1.28)	(1.39)	(1.40)	(1.45)	(1.46)	(1.56)	(1.61)	(1.66)	(1.71)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2(%)	2.90	2.79	2.70	2.64	2.57	2.51	2.45	2.38	2.31	2.27	2.23	2.21
Panel C: Big Stocks												
AI_beta	1.09	0.64	0.91	1.75	0.55	1.04	1.28	0.97	0.63	1.73	0.80	-1.33
	(0.94)	(0.57)	(0.88)	(1.64)	(0.55)	(1.07)	(1.24)	(0.86)	(0.63)	(1.60)	(0.70)	(-1.15)
Intercept	0.65	2.42	2.70	2.60	1.39	3.44	3.91	3.54	3.19	3.69	2.66	2.82
	(0.21)	(0.78)	(0.85)	(0.83)	(0.43)	(1.06)	(1.23)	(1.12)	(0.99)	(1.23)	(0.79)	(0.83)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2(%)	4.44	4.33	4.20	4.02	3.84	3.69	3.67	3.59	3.51	3.35	3.36	3.30

Table 8
Predicting Financial Performances

This table reports the results of the following panel regression

$$y_{i,t\to t+12} = \alpha + \beta AI\_beta_{it} + \epsilon_{i,t\to t+12},$$

where $y_{i,t\to t+12}$ is the next 12 month value of stock i's characteristics: sales growth (sgr), capital expenditure growth (grcapx), return on asset (roaq), and return on equity (reoq) and AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index and Fama-French six factors over a rolling six-month window. All regressions include firm and year-month fixed effects and t-statistics are computed with standard errors double-clustered by both firms and year-months. Panel A reports results for all stocks; Panel B (C) reports results for stocks below (above) the 50$^{\text{th}}$ percentile NYSE market value break point. The sample is from 1985 to 2020. The sample is from 1985 to 2020.

	Panel A: All Stocks			
	sgr	grcapx	roaq	roeq
AI_beta	0.33	0.17	0.28	0.37
	(1.63)	(0.87)	(1.86)	(2.39)
R^2(%)	21.23	12.64	42.79	34.00
	Panel B: Small Stocks			
	sgr	grcapx	roaq	roeq
AI_beta	0.42	0.24	0.25	0.32
	(1.98)	(1.14)	(1.55)	(1.96)
R^2(%)	21.78	13.42	39.82	31.00
	Panel C: Big Stocks			
	sgr	grcapx	roaq	roeq
AI_beta	-0.45	-0.55	0.46	0.65
	(-0.97)	(-1.26)	(1.33)	(1.70)
R^2(%)	28.72	18.17	48.48	31.89

Table 9
Portfolios Sorted on AI Beta

This table reports the results of portfolios sorted based on AI betas. Every month, we sort stocks into three equal buckets based on their AI betas in the previous month. AI beta is constructed from the rolling regression of daily stock returns onto the daily AI index and Fama-French six factors over a rolling six-month window. The results for each portfolio are reported in columns 1 to 3; the last column reports results for the high-minus-low AI beta portfolio. The first two rows report time-series averages of AI betas (scaled to range -0.5 to 0.5) and market value (in millions) for each portfolio. Panel A (B) reports equal-weighted (value-weighted) portfolio returns results. The first row of each panel reports time-series average portfolio returns over risk-free rate in annualized percentage points. The remaining rows report alphas against benchmark factor models, including six-factor model (FF6) of Fama and French (2018), q-factor model (Q5) of Hou et al. (2021), mispricing model of Stambaugh and Yuan (2017), and behavioral model of Daniel, Hirshleifer, and Sun (2020). t-statistics corrected for Newey-West standard errors with six lags are reported in parentheses. The sample is from 1985 to 2020.

	1	2	3	3-1
AI Beta	-0.33	0.00	0.33	
Size	1982	5073	2295	
Panel A: Equal Weighted				
Mean Return	11.42	11.15	13.33	1.91
	(2.92)	(3.80)	(3.56)	(3.07)
FF6	3.92	2.26	6.06	2.14
	(2.70)	(2.87)	(4.34)	(3.15)
Q5	5.07	2.96	7.46	2.39
	(2.78)	(2.64)	(4.26)	(3.57)
Mispricing	5.55	3.01	8.05	2.50
	(3.07)	(2.88)	(4.38)	(3.52)
Behavioral	7.42	4.74	9.55	2.14
	(2.66)	(2.89)	(3.44)	(2.72)
Panel B: Value Weighted				
Mean Return	8.53	9.09	10.19	1.66
	(2.85)	(3.86)	(3.82)	(1.39)
FF6	0.02	-0.34	1.39	1.37
	(0.03)	(-0.87)	(1.82)	(0.98)
Q5	-0.36	-0.85	1.17	1.53
	(-0.44)	(-1.64)	(1.64)	(1.18)
Mispricing	0.33	-0.62	2.17	1.83
	(0.37)	(-1.13)	(2.61)	(1.25)
Behavioral	0.52	-0.75	2.25	1.73
	(0.57)	(-1.56)	(2.64)	(1.15)

Table 10
Portfolios Double-Sorted on Size and AI Beta

This table reports the results of portfolios double-sorted on size and AI betas. Every month, stocks are sorted into two buckets using the NYSE 50^{th} size break point in the previous month (Size 1 and Size 2); then, within each size bucket, stocks are sorted into three equal buckets based on their AI betas in the previous month. AI beta is constructed from rolling regression of daily stock returns onto the daily AI index and Fama-French six factors over a rolling six-month window. Time-series average excess returns in annualized percentage points for each portfolio are reported in columns 1 to 3; columns 3-1 report results for the high-minus-low AI beta portfolio. The remaining columns report alphas against benchmark factor models, including six-factor model (FF6) of Fama and French (2018), q-factor model (Q5) of Hou et al. (2021), mispricing model of Stambaugh and Yuan (2017), and behavioral model of Daniel, Hirshleifer, and Sun (2020). t-statistics corrected for Newey-West standard errors with six lags are reported in parentheses. Panel A (B) reports equal-weighted (value-weighted) portfolio returns results. The sample is from 1985 to 2020.

AI beta	1	2	3	3-1	FF6	Q5	Mispricing	Behavioral
Panel A: Equal Weighted								
Size 1	11.91	11.86	13.92	2.01	2.42	2.62	2.82	2.43
	(2.86)	(3.68)	(3.44)	(3.00)	(3.33)	(3.65)	(3.72)	(2.84)
Size 2	9.60	9.67	10.60	1.00	0.92	0.92	1.11	0.51
	(3.18)	(3.81)	(3.83)	(1.22)	(0.93)	(0.95)	(0.98)	(0.48)
Panel B: Value Weighted								
Size 1	8.87	11.09	11.41	2.55	2.84	2.55	3.22	3.07
	(2.37)	(3.69)	(3.29)	(2.86)	(2.97)	(2.85)	(3.06)	(2.78)
Size 2	8.49	8.77	10.55	2.06	1.20	1.07	1.38	2.06
	(3.02)	(3.81)	(4.10)	(1.94)	(0.94)	(0.88)	(0.97)	(1.47)

Appendix
AI Narrative and Stock Mispricing

Appendix A Additional Tables

Table A.1
Top Ten Companies Having Highest AI Exposures from 2011 to 2020

This table lists the top 10 companies having the highest AI betas for each industry from 2011 to 2020. Companies are selected based on the time-series average of monthly AI betas; companies must have at least 60 non-missing AI beta observations from 2011 to 2020 to be included.

Industry	PERMNO	Name	Industry	PERMNO	Name
BusEq	15119	WORKIVA INC	Money	15064	FIRST FOUNDATION INC
BusEq	80185	PLANTRONICS INC	Money	43617	AMERICAN INDEPENDENCE CORP
BusEq	90914	WEBMD HEALTH CORP	Money	13400	WAGEWORKS INC
BusEq	87160	DATALINK CORP	Money	86793	YADKIN FINANCIAL CORP
BusEq	81165	MIND TECHNOLOGY INC	Money	80569	PULASKI FINANCIAL CORP
Chems	13556	TARONIS TECHNOLOGIES	NoDur	68718	TOFUTTI BRANDS INC
Chems	89597	FLEXIBLE SOLUTIONS INTL INC	NoDur	85951	OMEGA PROTEIN CORP
Chems	82699	ELIZABETH ARDEN INC	NoDur	14781	TRIBUNE PUBLISHING CO
Chems	12579	GEVO INC	NoDur	89471	LEAPFROG ENTERPRISES INC
Chems	13610	OLIN CORP	NoDur	92449	REEDS INC
Durbl	10667	FUEL SYSTEMS SOLUTIONS INC	Other	83875	INDUSTRIAL SERVICES AMER INC
Durbl	92589	KANDI TECHNOLOGIES GROUP	Other	50606	TIDEWATER INC
Durbl	54704	MODINE MANUFACTURING CO	Other	13267	CAESARS ENTERTAINMENT CORP
Durbl	60506	PACCAR INC	Other	24441	COEUR MINING INC
Durbl	12503	NAVISTAR INTERNATIONAL CORP	Other	83734	WILHELMINA INTERNATIONAL INC
Enrgy	12389	TRIANGLE PETROLEUM CORP	Shops	14983	WAYFAIR INC
Enrgy	63335	PIONEER ENERGY SERVICES CORP	Shops	75489	STAPLES INC
Enrgy	92619	GRAN TIERRA ENERGY INC	Shops	92471	TITAN MACHINERY INC
Enrgy	14634	PARSLEY ENERGY INC	Shops	89447	CALAVO GROWERS INC
Enrgy	13124	VITAL ENGY INC	Shops	14812	EL POLLO LOCO HOLDINGS INC
Hlth	92805	TIANYIN PHARMACEUTICAL CO	Telcm	15226	ANTERIX INC
Hlth	14468	PHIO PHARMACEUTICALS CORP	Telcm	92454	INTELIQUENT INC
Hlth	91348	ALPHATEC HOLDINGS INC	Telcm	89916	NEXSTAR MEDIA GROUP
Hlth	86444	INOVIO PHARMACEUTICALS INC	Telcm	14060	VISLINK TECHNOLOGIES INC
Hlth	14033	ORGANOVO HOLDINGS INC	Telcm	90763	NEUSTAR INC
Manuf	93149	ZAGG INC	Utils	13019	GENIE ENERGY LTD
Manuf	14982	VIVINT SOLAR INC	Utils	12476	TARGA RESOURCES CORP
Manuf	75272	SEVCON INC	Utils	29285	DELTA NATURAL GAS CO INC
Manuf	41320	SL INDUSTRIES INC	Utils	26470	ENERGEN CORP
Manuf	81857	CAMERON INTERNATIONAL CORP	Utils	25590	NATIONAL FUEL GAS CO

Table A.2
Predicting Next One-Month Returns: Three-Month Construction Window

This table report factor risk premia estimated via Fama-Macbeth regressions. Every month, we run the following cross-sectional regression:

$$r_{it+1} = a_t + \lambda_t \times AI\_beta_{it} + \gamma_t^{'} X_{it} + \epsilon_{it+1},$$

where r_{it+1} is excess return of stock i in month $t+1$ in annualized percentage point, AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index over a rolling three-month window, and X_{it} is a set of control variables of stock i in month t: market equity (mve), book-market (bm), market beta (beta), momentum (mom12m), standard deviation of daily returns from previous month (retvol), return on equity (roeq), gross profit divided total asset (gma), leverage (lev), two-year percent change in capital expenditures (grcapx), R&D divided by sales (rd_sale), annual change in investment (invest), industry sales concentration (herf), and sales growth (sgr). Every month, all independent variables are ranked and rescaled to fall from -0.5 to 0.5. Reported are time-series averages of λ_t and γ_t with their t-statistics corrected for Newey-West standard errors with six lags in parentheses. Panel A reports results for all stocks; Panel B (C) reports results for stocks below (above) the 50$^{\text{th}}$ percentile NYSE market value break point. The sample is from 1985 to 2020.

	Panel A: All Stocks			Panel B: Small Stocks			Panel C: Big Stocks		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
AI_beta	2.17	2.31	2.92	2.57	2.91	3.79	-0.09	-1.60	-2.14
	(2.19)	(2.54)	(2.50)	(2.45)	(3.01)	(2.93)	(-0.05)	(-1.16)	(-1.31)
mve		-16.09	-15.16		-21.13	-19.05		-8.24	-10.71
		(-4.88)	(-3.92)		(-4.70)	(-3.84)		(-1.04)	(-1.24)
bm		1.51	5.88		2.11	6.81		-1.13	4.09
		(0.51)	(2.34)		(0.70)	(2.57)		(-0.37)	(1.38)
beta		5.49	1.92		7.45	3.32		1.32	-1.14
		(1.49)	(0.59)		(1.90)	(0.97)		(0.32)	(-0.30)
mom12m		7.21	2.68		7.72	2.34		7.27	5.92
		(1.99)	(0.78)		(2.17)	(0.67)		(1.76)	(1.66)
invest		-8.36	-6.22		-9.17	-7.61		-3.12	0.04
		(-5.57)	(-4.67)		(-5.63)	(-5.00)		(-2.43)	(0.02)
roeq		6.87	7.75		7.30	8.40		3.61	4.26
		(3.11)	(4.73)		(3.00)	(4.65)		(1.91)	(2.20)
retvol		-5.24	-5.46		-5.03	-5.57		-6.18	-4.67
		(-1.91)	(-2.01)		(-1.78)	(-2.02)		(-2.04)	(-1.67)
gma			7.36			7.36			6.01
			(3.15)			(2.91)			(1.93)
lev			8.01			8.41			3.48
			(2.66)			(2.56)			(1.16)
grcapx			-3.01			-2.53			-4.22
			(-2.54)			(-1.98)			(-2.42)
rd_sale			15.95			16.96			8.80
			(4.21)			(4.28)			(2.42)
herf			-0.59			-0.86			-0.33
			(-0.35)			(-0.44)			(-0.20)
sgr			-1.76			-1.94			1.94
			(-1.42)			(-1.41)			(0.90)
Intercept	11.98	12.34	15.39	12.58	11.41	14.68	9.99	10.60	14.02
	(3.43)	(3.56)	(3.88)	(3.34)	(3.36)	(3.72)	(3.66)	(2.36)	(2.66)
R^2(%)	0.05	4.08	4.67	0.05	3.67	4.07	0.29	10.06	12.19

Table A.3
Predicting Next One-Month Returns: One-Month Construction Window

This table report factor risk premia estimated via Fama-Macbeth regressions. Every month, we run the following cross-sectional regression:

$$r_{it+1} = a_t + \lambda_t \times AI\_beta_{it} + \gamma_t^{'} X_{it} + \epsilon_{it+1},$$

where r_{it+1} is excess return of stock i in month $t+1$ in annualized percentage point, AI_beta is AI beta of stock i in month t constructed from rolling regression of daily stock returns onto daily AI index over a rolling one-month window, and X_{it} is a set of control variables of stock i in month t: market equity (mve), book-market (bm), market beta (beta), momentum (mom12m), standard deviation of daily returns from previous month (retvol), return on equity (roeq), gross profit divided total asset (gma), leverage (lev), two-year percent change in capital expenditures (grcapx), R&D divided by sales (rd_sale), annual change in investment (invest), industry sales concentration (herf), and sales growth (sgr). Every month, all independent variables are ranked and rescaled to fall from -0.5 to 0.5. Reported are time-series averages of λ_t and γ_t with their t-statistics corrected for Newey-West standard errors with six lags in parentheses. Panel A reports results for all stocks; Panel B (C) reports results for stocks below (above) the 50$^{\text{th}}$ percentile NYSE market value break point. The sample is from 1985 to 2020.

	Panel A: All Stocks			Panel B: Small Stocks			Panel C: Big Stocks		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
AI_beta	3.95	2.85	3.60	4.22	2.99	4.18	3.13	1.07	-1.08
	(2.38)	(2.29)	(2.75)	(2.55)	(2.39)	(3.13)	(1.11)	(0.65)	(-0.56)
mve		-16.19	-15.20		-21.27	-19.09		-7.58	-10.08
		(-4.91)	(-3.91)		(-4.75)	(-3.85)		(-0.96)	(-1.19)
bm		1.46	5.86		2.05	6.79		-0.87	4.03
		(0.49)	(2.33)		(0.68)	(2.56)		(-0.28)	(1.36)
beta		5.51	2.06		7.44	3.45		1.65	-0.71
		(1.49)	(0.63)		(1.89)	(1.00)		(0.40)	(-0.19)
mom12m		7.52	2.97		8.00	2.62		7.72	6.48
		(2.09)	(0.87)		(2.27)	(0.76)		(1.88)	(1.81)
invest		-8.33	-6.37		-9.19	-7.85		-3.04	0.26
		(-5.53)	(-4.78)		(-5.60)	(-5.16)		(-2.40)	(0.14)
roeq		6.81	7.62		7.25	8.28		3.64	4.42
		(3.09)	(4.65)		(2.98)	(4.57)		(1.93)	(2.24)
retvol		-5.13	-5.36		-5.01	-5.57		-5.92	-4.36
		(-1.88)	(-2.00)		(-1.78)	(-2.03)		(-1.99)	(-1.59)
gma			7.42			7.42			5.90
			(3.17)			(2.94)			(1.90)
lev			7.95			8.37			3.91
			(2.63)			(2.54)			(1.31)
grcapx			-3.05			-2.59			-4.44
			(-2.56)			(-2.02)			(-2.55)
rd_sale			15.93			16.90			8.86
			(4.21)			(4.28)			(2.44)
herf			-0.57			-0.84			-0.62
			(-0.34)			(-0.44)			(-0.39)
sgr			-1.67			-1.84			1.94
			(-1.34)			(-1.34)			(0.90)
Intercept	11.98	12.34	15.36	12.57	11.41	14.65	10.03	10.29	13.79
	(3.43)	(3.56)	(3.87)	(3.33)	(3.36)	(3.71)	(3.73)	(2.27)	(2.61)
R^2(%)	0.15	4.10	4.67	0.13	3.68	4.07	0.87	10.10	12.18