CFR Working Paper NO. 26-05

Hard to Process: Atypical Firms and the Cross-Section of Expected Stock Returns

S. Weibels

centre for financial research
cologne

Hard to Process: Atypical Firms and the Cross-Section of Expected Stock Returns*

Sebastian Weibels†

Current version: January 2026

Link to Most Recent Version

Abstract

Theories of limited attention predict that investors rely on typical patterns to navigate high-dimensional firm characteristics, making atypical firms hard to process. To quantify this difficulty, we propose a data-driven measure of firm atypicality using an autoencoder (ATYP). The model learns typical patterns that describe most firms, and our measure aggregates the deviations those patterns cannot explain. Unlike proxies based on disclosure or organizational complexity, this approach captures the processing difficulty of the characteristics themselves. Empirically, we document that atypicality strongly predicts future returns. A decile portfolio that sells high-ATYP firms and buys low-ATYP firms earns 1.47% per month (equal-weighted) and 0.82% (value-weighted). The effect strengthens where investor attention is low and arbitrage is limited, suggesting mispricing as the explanation.

JEL classification: G10, G11, G12, G14, C45

Keywords: atypical firms, processing difficulty, return predictability, mispricing, machine learning

*I thank Tom Zimmermann, and Harm Schütt for their valuable feedback. I also thank seminar participants at the University of Tübingen, University of Cologne, and the 4th International Econometrics PhD Conference for helpful comments and suggestions.

†University of Cologne, Institute for Econometrics and Statistics and Centre for Financial Research, weibels@wiso.uni-koeln.de

1 Introduction

Information is fundamental to the functioning of financial markets and the allocation of capital. Standard asset pricing models typically assume that investors rapidly collect and interpret all available information, ensuring that prices fully reflect fundamental value. In practice, however, not all information is equally easy to process. Investors have bounded processing capacities and limited attention, preventing frictionless information processing. To cope, investors do not process every piece of information independently. Instead, they think in patterns and categories (e.g., Barberis and Shleifer, 2003; Peng and Xiong, 2006; Gabaix, 2014; Kacperczyk et al., 2016).

Difficulties arise when a firm presents information in unusual or idiosyncratic ways. We refer to such firms as *atypical firms*, meaning companies whose joint realization of observable information differs from that of most other firms. The key idea is that when a firm's information does not fit typical patterns, it imposes higher processing difficulty, making its firm-specific information more likely to be neglected or incorporated only gradually. A central question in asset pricing is whether such information-processing frictions have return predictive power.

Despite the centrality of information-processing frictions in theory, empirical work has lacked a general-purpose measure of how hard it is to process a firm's information at any given time. This gap is particularly striking given the prominence of information frictions in theoretical models of asset pricing. Models of bounded rationality, gradual information diffusion, and disagreement all rest on the premise that some information is more costly to process than others.[1]

This paper addresses this gap by proposing a new approach to capture information-processing frictions. Rather than treating processing difficulty as a property of disclosures or organizational scope, we define it as a statistical property of how well a firm fits within the joint distribution of characteristics, which represent the information investors use when forming expectations and making valuation decisions. We develop a data-driven measure, ATYP, that captures how atypical a firm appears relative to the typical patterns that describe most firms. An unsupervised machine learning model learns these patterns across a wide range of firm characteristics, and ATYP quantifies the residual component that the model cannot reconstruct. In other words, it captures the firm-specific component that typical patterns fail to explain. This approach conceptually

[1]Examples include Simon (1955); Miller (1977); Hong and Stein (1999); Peng and Xiong (2006); Van Nieuwerburgh and Veldkamp (2010); Gabaix (2014); Kacperczyk et al. (2016).

connects to the idea that investors rely on categorical and pattern-based reasoning when navigating high-dimensional information environments. Consequently, High-ATYP firms present unusual or idiosyncratic combinations of characteristics that are more difficult and therefore costly for investors to process.

Existing empirical proxies for information-processing and valuation difficulty capture only small slices of this broader concept.[2] Measures such as textual readability, segment counts, or accounting intricacy capture important aspects of information-processing frictions, but they are inherently narrow, infrequent, and tied to reporting choices. More fundamentally, they measure properties of disclosure or organizational scope rather than properties of the underlying economic reality that investors must interpret. ATYP differs in three important ways. First, whereas existing complexity measures focus on the reporting of information, ATYP focuses on the information itself. Second, it is fully data-driven and independent of managerial reporting choices. Third, it is available at high frequency, over a long sample, and for a broad set of firms, including international stocks. These features make ATYP a flexible tool for studying how processing frictions shape asset prices.

Empirically, we document a large and robust association between ATYP and future returns. Firms with higher ATYP, i.e., those for which information is harder to process, earn significantly lower subsequent returns. More specifically, a decile spread portfolio that sells high-ATYP firms and buys low-ATYP firms earns 1.47% per month when portfolios are equal-weighted and 0.82% when they are value-weighted. These results cannot be explained by standard risk factor models and are significantly stronger among firms with high limits to arbitrage and low investor attention, precisely where hard-to-process information is most likely to be neglected. Therefore, return predictability may be driven by mispricing rather than by compensation for risk.

By establishing ATYP as a predictive state variable, our paper makes four contributions. First, we develop and propose ATYP, a novel, general-purpose measure of processing frictions grounded in the statistical atypicality of a firm's characteristics. Second, we provide extensive validation showing that ATYP is not just a statistical curiosity but captures a meaningful, and previously unmeasured, dimension of information-processing frictions. Third, we document the pricing

[2] Important contributions include Li (2008); Loughran and McDonald (2014, 2024); Cohen and Lou (2012); García and Norli (2012); Barinov et al. (2024); Peterson (2012); Hoitash and Hoitash (2018).

power of ATYP for the cross-section of stocks, where we find that stocks with higher ATYP earn significantly lower future returns. Fourth, we show that these return patterns are concentrated among firms with higher limits to arbitrage and lower investor attention.

A key contribution of our study is validating that ATYP captures information-processing frictions, rather than repackaging existing measures. We show that ATYP is distinct from and only marginally negatively correlated with existing measures of firm complexity and information-processing frictions. Firms with many segments, long filings, or complex disclosures tend to be large, established companies with strong information environments. Despite their organizational complexity, such firms often exhibit low ATYP because they follow familiar characteristic patterns.

In contrast, ATYP is higher in firms with weak information environments. Consistent with this, we find that it is positively correlated with proxies of information uncertainty. However, ATYP remains conceptually and empirically distinct. Crucially, it has significant incremental explanatory power for outcomes linked to processing frictions. High-ATYP firms are associated with stronger analyst disagreement, higher return volatility, lower analyst accuracy, and stronger absolute earnings surprises.

Perhaps most compellingly, we document that high-ATYP firms experience slower price discovery following information events. Following earnings announcements, high-ATYP firms exhibit stronger and more persistent post-earnings-announcement drift. We find a similar pattern following industry information shocks according to Cohen and Lou (2012), where we compare how quickly atypical firms respond to news that has already been reflected in the prices of their simpler industry peers. High-ATYP firms adjust more slowly, consistent with investors requiring more time and effort to process information about atypical firms and incorporate it into valuations. This slow adjustment is difficult to reconcile with alternative explanations such as risk or data mining and instead points to information-processing frictions.

Having established that ATYP captures information-processing frictions, we turn to its implications for asset pricing. To the best of our knowledge, we are the first to demonstrate the direct cross-sectional pricing effects of processing difficulty induced by atypical information patterns. Previous studies by Cohen and Lou (2012) on organizational complexity and Zhang (2006) on information uncertainty have documented delayed reactions to information shocks, but they have not systematically explored the unconditional return predictability of information-processing frictions in the cross-section.

Empirically, sorting stocks into deciles by ATYP produces a striking monotonic pattern. Firms with high-ATYP earn significantly lower subsequent returns. The equal-weighted spread between the lowest and highest decile is -1.47% per month with a *t*-statistic of -4.54, while the value-weighted spread is -0.82% with a *t*-statistic of -2.50. Importantly, the return patterns remain highly robust after controlling for standard factors. Under the six factor model of Fama and French (2018), the equal-weighted spread continues to deliver significant alphas of -0.84% per month with a *t*-statistic of -4.52, while the value-weighted results weaken to -0.28%, consistent with the effect being concentrated in smaller firms, where investor attention and arbitrage possibilities are more limited.

Dependent double sorts and Fama-MacBeth regressions show that the ATYP premium remains large and significant after controlling for other commonly studied characteristics, including size, investment, profitability, momentum, reversal, illiquidity, analyst disagreement, and volatility. In fully saturated Fama-MacBeth specifications that include more than a dozen established predictors, ATYP continues to load significantly at -0.72 with a *t*-statistic of -3.21. The effect extends to different peer group definitions and holds within different clusters of characteristics, demonstrating that it is not confined to particular information domains. The pricing power extends to international markets, particularly in Europe where the equal-weighted spread is -1.21% per month, and applies robustly to alternative factor models including the q-factor model and behavioral factor frameworks. The results remain robust even when we change the measurement approach, use alternative data sources such as WRDS financial ratios, vary model specifications across 48 different hyperparameter combinations, or modify the sample construction. This pervasive robustness underscores that ATYP is not a repackaging of existing variables but an independent dimension of information frictions in asset pricing.

The negative pricing of ATYP raises a fundamental question: why do firms that are harder to process earn lower returns? The negative sign of the ATYP premium, combined with its concentration in settings where arbitrage and attention are limited, suggests a mispricing interpretation. Two well-established theoretical mechanisms explain why information-processing frictions might generate systematic overpricing followed by slow corrections.

High ATYP increases valuation uncertainty by making the mapping from fundamentals to value more ambiguous. Under such uncertainty, behavioral biases such as overconfidence become stronger (Daniel et al., 1998, 2001), which amplifies heterogeneous interpretations of signals and raises disagreement (Zhang, 2006). Combined with short-sale constraints, this disagreement shifts prices toward optimistic investors and creates overpricing among high-ATYP firms (Miller, 1977). At the same time, limited attention leads investors to focus on typical patterns rather than fully processing firm-specific information (Peng and Xiong, 2006). This selective processing causes signals about atypical firms to diffuse only gradually across investors, and in the spirit of Hong and Stein (1999), the resulting slow adjustment generates systematic underreaction and more persistent return predictability.

Our empirical evidence strongly supports this interpretation. We trace the economic channels of the ATYP premium to settings characterized by high limits to arbitrage and low investor attention. Conditioning on proxies for limits to arbitrage reveals that the ATYP premium is concentrated precisely where arbitrage is most difficult. For instance, among firms with high idiosyncratic volatility, the six-factor adjusted spread is -1.51% per month with a t-statistic of -–6.77, compared with an insignificant 0.13% among low-volatility firms. Parallel evidence arises when sorting firms into quintiles based on investor-attention proxies. The ATYP spread is large and highly significant in low-attention groups but shrinks drastically among firms with higher attention. For example, among small firms the six-factor adjusted spread is -1.54% per month (t = -6.90), while among large firms it is economically negligible at 0.20%.

The asymmetry of returns provides further evidence in favor of the mispricing interpretation. The return spread is driven almost entirely by the underperformance of high-ATYP firms, whereas low-ATYP firms earn only slightly above-average returns. High-ATYP firms deliver six-factor alphas of -0.76% per month in the equal-weighted specification, while low-ATYP firms deliver alphas of only 0.08%. Consistent with this view, double sorts with the mispricing factor of Stambaugh et al. (2015) show that the ATYP premium is concentrated among firms ex-ante classified as prone to mispricing. High-ATYP firms exhibit significantly higher mispricing scores on average, and the ATYP spread is most pronounced in the highest-mispricing quintile, with a six-factor alpha of -0.88% and a t-statistic of -3.98. Taken together, these results reinforce the conclusion that the ATYP premium reflects mispricing rather than compensation for risk.

Related literature

Our work relates to five strands of literature: (i) on the measurement of information-processing difficulty, where we provide an alternative to disclosure- or segment-based proxies, (ii) on the use of machine learning in finance, (iii) on the literature that catalogs cross-sectional predictors, (iv) on the literature examining the cross-sectional predictability of returns from information-processing frictions, and (v) on limited attention and behavioral biases in asset pricing.

First, our measure relates to other studies on information-processing frictions and firm complexity. The literature focuses on three big classes of proxies. Text-based approaches (e.g., Li, 2008; Loughran and McDonald, 2014, 2024) use readability or linguistic complexity of 10-K filings, finding that harder-to-process disclosures are associated with slower price adjustment, greater analyst forecast dispersion, and higher audit fees. Organizational measures focus on business segments or conglomerate structure (e.g., Cohen and Lou, 2012; Barinov et al., 2024), showing that multi-segment firms react more slowly to industry news and may be mispriced. Geographic dispersion has also been linked to delayed price reactions (e.g., García and Norli, 2012). Finally, accounting-based proxies emphasize the intricacy of financial reporting, such as the complexity of revenue-recognition rules or the breadth of accounting variables disclosed (e.g., Peterson, 2012; Hoitash and Hoitash, 2018). While these proxies capture important forms of disclosure or organizational scope, they do not necessarily reflect the complexity of a firm's underlying economic profile. A firm can have complex filings yet follow simple economic patterns. In contrast, ATYP is data-driven and captures information-processing difficulty in the underlying economic information that investors must interpret.

Second, recent work applies ML methods to capture high-dimensional and nonlinear relationships in returns. For instance, Gu et al. (2020) show that machine learning algorithms outperform linear models in out-of-sample return prediction, while Gu et al. (2021) develop a conditional asset pricing model based on autoencoder-derived latent factors. Closest to our approach, Bali et al. (2025) use an ensemble of random forest models to construct a measure of machine forecast disagreement (MFD), a proxy for investor disagreement. Our study differs in two key respects. First, we introduce ATYP, defined as the reconstruction error from an unsupervised machine learning model trained on firm characteristics, capturing the degree to which a firm deviates from the common, low-dimensional manifold of the market. Second, whereas MFD reflects the dispersion of beliefs after they have formed, ATYP is an ex-ante state variable describing the information environment before beliefs form. By increasing the cost of information-processing, ATYP interacts with investor attention and frictions, setting the stage for disagreement to persist. This upstream positioning in the information-processing channel, coupled with an unsupervised learning design, yields a persistent signal with pricing implications not subsumed by disagreement measures or standard predictors.

Third, a vast literature documents that firm characteristics predict the cross-section of expected returns. Early evidence has motivated large-scale compilations (e.g., Harvey et al., 2016; Hou et al., 2018; Chen and Zimmermann, 2022; Jensen et al., 2023) that catalog hundreds of predictors that investors actually use in creating coherent economic valuations. Our contribution is to move beyond marginal predictors by examining how unusual a firm's entire profile of characteristics appears, and to show that such atypicality is priced. This approach treats ATYP as a state variable that spans the joint distribution of characteristics, rather than as another stand-alone anomaly.

Fourth, to the best of our knowledge, only few studies have looked at the direct cross-sectional pricing implications of information-processing difficulty. Cohen and Lou (2012) look at the pricing effects of information shocks where they compare easy- and hard-to-process firms and find that the latter reacts more slowly to shocks. Yet, they do not explore the direct predictive power of their segment count measure in the cross-section of stocks. Barinov (2020) extend this analysis and finds that segment counts are weakly negatively priced in cross-section. Similarly, Zhang (2006) look at the cross-sectional predictions of several proxies of information uncertainty and also find a weak negative association between these proxies and future stock returns. However, they also focus on delayed price responses to news. ATYP however is a more direct proxy of information-processing frictions and a notably stronger predictor of stock returns. Also, we are the first to explore the pricing power across a variety of settings and specifications.

Fifth, our work builds on theoretical models of limited attention and bounded rationality (e.g., Simon, 1955; Peng and Xiong, 2006; Van Nieuwerburgh and Veldkamp, 2010; Gabaix, 2014; Kacperczyk et al., 2016), which emphasize that investors optimally allocate scarce cognitive resources across information signals. In these frameworks, costly or idiosyncratic information is more likely to be neglected or only partially processed. Our study also relates to behavioral theories showing that when the information environment is uncertain or difficult to interpret, investors overweight their own signals and become more overconfident, leading to greater belief dispersion and disagreement (e.g., Daniel et al., 1998, 2001; Zhang, 2006; Peng and Xiong, 2006). Empirically, limited attention slows the incorporation of firm-specific and related-firm information into prices (Hirshleifer and Teoh, 2003a; DellaVigna and Pollet, 2009; Cohen and Frazzini, 2008; Engelberg and Parsons, 2011), and information uncertainty amplifies behavioral biases in ways that generate disagreement and return predictability (Zhang, 2006). We contribute to this literature by operationalizing the idea that investors rely on typical patterns and neglect difficult-to-process firm-specific information. More specifically, we show that atypicality in the joint characteristic space systematically interacts with attention and arbitrage capacity.

The remainder of the paper proceeds as follows. Section 2 introduces the construction of ATYP. Section 3 describes the data and model specifications. Section 4 validates ATYP as a measure of information-processing difficulty. Section 5 presents the main empirical asset pricing results. Section 6 analyzes the economic channels through which ATYP affects prices. Section 7 runs a series of robustness checks. Section 8 concludes.

2 Constructing a Measure of Firm Atypicality

Our starting point is the idea that financial markets are populated by investors with limited attention and bounded information-processing capacity. In such an environment, firms differ not only in their characteristics, but also in how easy or hard it is to process them. A stock whose firm characteristics align with familiar information-processing patterns is easier for investors to interpret than one whose characteristics combine in an atypical or unfamiliar way.

For instance, a company may be long established and liquid, yet simultaneously display sharply negative profitability, weak sales and asset growth, high leverage, and persistently negative cash flows. It may invest heavily in research and development, but without corresponding improvements in earnings or operating performance. Such a mix of stable features and distressed fundamentals, combined with volatile performance indicators and repeated negative surprises, does not align with the typical patterns investors typically rely on. Profiles of this kind are inherently harder to process through standard heuristics, which illustrates the type of joint atypicality we aim to measure.

The challenge is to operationalize this idea in a way that remains faithful to its joint nature yet empirically tractable in large samples. To this end, we draw on machine learning methods for modeling high-dimensional data. We view the sample of firm characteristics observed over time as a cloud of points in a high-dimensional space, where most firms cluster around typical patterns in their characteristics. To capture these patterns, we employ an autoencoder, an unsupervised neural network that learns to compress firm characteristics into a compact latent representation and then reconstruct them. The model learns the typical patterns that characterize a typical firm, and ATYP reflects how well these patterns explain a given firm's profile. The greater the model's difficulty in explaining a firm's characteristics, the more atypical, and thus harder to process, the firm's profile is.

To estimate these latent patterns and assess how well they explain each firm's characteristics, we formalize the intuition using an autoencoder framework. Let i index firms and t index months in our sample. For each firm-month observation, let $\mathbf{X}_{i,t} \in \mathbb{R}^p$ denote the p-dimensional vector of standardized firm characteristics. An autoencoder consists of two parameterized functions: an encoder $f_\theta : \mathbb{R}^p \to \mathbb{R}^k$ with $k \ll p$, and a decoder $g_\phi : \mathbb{R}^k \to \mathbb{R}^p$. The encoder maps the high-dimensional characteristic vector into a k-dimensional latent representation,

$$\mathbf{z}_{i,t} = f_\theta(\mathbf{X}_{i,t}), \tag{1}$$

which summarizes the key patterns in the data. The decoder maps this latent vector z back into the original p-dimensional space to produce a reconstruction,

$$\hat{\mathbf{X}}_{i,t} = g_\phi(\mathbf{z}_{i,t}) = g_\phi(f_\theta(\mathbf{X}_{i,t})). \tag{2}$$

The parameters θ and ϕ are estimated jointly by solving

$$\min_{\theta,\phi} \sum_t \sum_i \|\mathbf{X}_{i,t} - g_\phi(f_\theta(\mathbf{X}_{i,t}))\|_2^2, \tag{3}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. This training criterion minimizes the sum of squared reconstruction errors across all observations. This forces the encoder to learn the systematic, low-dimensional patterns of the data and the decoder to accurately map these back into the original space.[3]

The reconstruction $\hat{\mathbf{X}}_{i,t}$ can be interpreted as the projection of the firm's characteristics onto the learned patterns of typical configurations. The reconstruction residual

$$\mathbf{R}_{i,t} = \mathbf{X}_{i,t} - \hat{\mathbf{X}}_{i,t} \tag{4}$$

is the residual part of the firm's characteristic vector that is not explained by these patterns. If the autoencoder captures the main nonlinear relationships among characteristics, $\mathbf{R}_{i,t}$ represents the firm-specific deviation from what is typical.

A key question is how to collapse this residual vector into a single scalar measure of atypicality. We define ATYP as the root mean squared reconstruction error (RMSE):

$$ATYP_{i,t} \;=\; \sqrt{\frac{1}{p}\sum_{j=1}^{p}\left(\mathbf{X}_{i,t}^{(j)} - \hat{\mathbf{X}}_{i,t}^{(j)}\right)^2}. \tag{5}$$

Intuitively, ATYP reflects the extent to which a firm's characteristics are not captured by the patterns learned by the autoencoder. A low score implies that the firm's characteristics conform to typical co-movements, whereas a high score signals an unusual or idiosyncratic combinations. Because the autoencoder is trained to reproduce typical patterns, it will assign large errors to combinations it has rarely seen.[4] Additionally, because we cross-sectionally standardize all input features to zero mean and unit variance, each component of the error vector $\mathbf{R}_{i,t}$ is expressed in units of one standard deviation.[5] Therefore, ATYP quantifies the average standardized deviation per feature between an observed firm and its reconstructed counterpart. This makes the measure directly interpretable in the same units as the input data and consistent with standard RMSE conventions in statistics.

Economically, ATYP captures the rarity of a firm's configuration of characteristics in a way that is inherently joint and context-dependent. Put differently, it measures the amount of non-redundant information an investor would need to process to fully understand the firm. This

[3]Figure IA.2.1 in the Internet Appendix provides a graphical illustration of an autoencoder model.
[4]We consider alternative measurements of ATYP in Section 7.8.
[5]See Section 3.1 for more information on feature preprocessing.

interpretation is directly linked to the bounded processing capacities. If processing costs increase with ATYP, then more attention is needed to fully incorporate firm-specific details into valuations. To further build intuition for the concept of processing difficulty, Figure IA.2.2 in the Internet Appendix visualizes the reconstruction accuracy of the autoencoder across all firms. The scatterplot contrasts realized characteristic values with their reconstructed counterparts. Firms that lie close to the diagonal represent typical firms whose characteristic patterns are well captured by the model. In contrast, atypical firms deviate more strongly from the diagonal, indicating that their joint configuration of characteristics cannot be easily reproduced by the typical patterns learned by the model.

Our motivation for employing a nonlinear framework rests on the idea that systematic combinations of firm characteristics are inherently nonlinear. The economic meaning of a given variable often depends on the configuration of others, in ways that cannot be captured by additive linear models. For example, a high book-to-market ratio signals financial distress when combined with low profitability and weak investment, but may reflect a traditional value firm when paired with strong profitability and conservative investment (Fama and French, 1992, 2015). Similarly, different industries may present different typical patterns. Ignoring such interactions and conditional effects means that firms with superficially similar attributes may be treated alike in linear models, even though they imply very different valuation profiles once context is taken into account.

To complement this, Stein (2009) documents that investors are becoming increasingly sophisticated. In our context, this means that they can understand nonlinear relationships between characteristics and factor them into their valuations. In fact, Li and Rossi (2021) show that the performance of mutual funds is nonlinearly related to the characteristics of the firms in which they invest. This demonstrates that sophisticated investors select stocks based on nonlinear combinations of characteristics.

So, while linear dimensionality-reduction methods such as PCA (e.g., Connor and Korajczyk, 1986; Kelly et al., 2019; Lettau and Pelger, 2020) provide a powerful way to summarize common variation across firms, they are inherently restricted to linear patterns. Autoencoders build on this idea. A linear autoencoder with mean squared error loss collapses to PCA, but by introducing nonlinear encoder and decoder functions, the autoencoder generalizes PCA by capturing nonlinear and heterogeneous interactions (Hinton and Salakhutdinov, 2006). This flexibility allows the model to better capture context-dependent patterns in firms.

Autoencoders are widely applied in anomaly detection, from identifying fraud to detecting irregular physiological signals in medicine. These successes highlight their suitability for identifying firms with unusual financial profiles that standard linear methods fail to flag. Recent work in finance shows that nonlinear ML methods capture economically meaningful interactions missed by linear approaches (Gu et al., 2020; Bryzgalova et al., 2023), and that replacing PCA with nonlinear architectures improves factor extraction (Gu et al., 2021). In our setting, the autoencoder provides a natural foundation for defining ATYP, measuring how far a firm's joint configuration of accounting and market variables lies from the manifold of typical firms.

Finally, our design is deliberately unsupervised. The construct we seek is information-processing frictions, not a new return forecaster. By separating the construction of ATYP from future returns, we ensure that any subsequent pricing power reflects the market's reaction to information-processing frictions, rather than mechanical overfitting to realized returns. Pricing tests then ask whether markets efficiently price these frictions, in the same spirit that Bali et al. (2025) first model beliefs and only subsequently study their pricing implications.

3 Data and Model

3.1 Data

We use the dataset from Jensen et al. (2023), a publicly available dataset of stock returns and characteristics. The underlying monthly return data are from the Center for Research in Security Prices (CRSP) and accounting information from Compustat. We restrict our sample on stocks trading at the NYSE, AMEX and NASDAQ. We also exclude financial and utility firms.[6] Finally, to reduce the effect of very small and illiquid stocks, we exclude low-priced stocks trading below $1 per share. Our sample covers the period from 1971 to 2023 of monthly data.

These characteristics are natural building blocks for our proposed measure because they summarize the economic fundamentals and market signals that investors are known to process when forming expectations (e.g., Harvey et al., 2016; Hou et al., 2018; Chen and Zimmermann,

[6] More specifically, we exclude financial firms (with one-digit SIC = 6) and utility firms (with two-digit SIC = 49).

2022; Jensen et al., 2023). Using these characteristics as inputs ensures that our measure is grounded in variables with demonstrated relevance for both investors and asset pricing models, rather than statistical artifacts.[7]

For firm characteristics, we start with the full set of 153 characteristics. Jensen et al. (2023) group these characteristics into 13 clusters. Our information set drops characteristics related to *Seasonality*, *Short-Term Reversal* and *Low Risk*.[8] We exclude short-horizon return-based risk measures (e.g., idiosyncratic volatility, lottery proxies, and short-term reversal) from the construction of the score to keep the measure focused on fundamentals and valuations. This choice keeps the economic interpretation of the score as information-processing difficulty of the firm's information set rather than ex-post return risk. Thus, in the baseline specification the input set comprises accounting and market-ratio variables spanning the clusters *Investment*, *Value*, *Profitability*, *Profit Growth*, *Accruals*, *Quality*, *Low Leverage*, *Momentum*, *Debt Issuance* and *Size*. The 117 characteristics make up our complete information set $\mathbf{X}$. Lastly, we require each observation to have at least 50% non-missing features.

We follow standard cross-sectional preprocessing of characteristics. First, we impute missing values with the median as suggested by Chen and McCoy (2024). We then winsorize all characteristics at the 1% level for both tails to mitigate the effect of outliers and data errors. Finally, we standardize all features to have mean zero and unit standard deviation.

3.2 Control Variables

To rigorously isolate the role of our main variable of interest in the cross-sectional return predictability framework, we include a comprehensive set of control variables that are widely recognized in the asset pricing literature for their ability to explain the cross-section of expected stock returns. These variables are constructed following established methodologies to ensure consistency with prior work and robustness of inference. The choice of variables follows partly Bali et al. (2025). All controls are taken from either Jensen et al. (2023) or, if not available there, from Chen and Zimmermann (2022).

The CAPM Beta (*BETA*) is calculated as the coefficient of a 60-month rolling window regression

[7]As a robustness check, we replicate our analysis using financial ratios from WRDS, which provide a more granular but less curated set of accounting-based predictors, in Section 7.5.

[8]For robustness we also replicate our results with the full information set in Section 7.5.

of monthly stock returns minus the risk-free rate on market return minus the risk-free rate as in Fama and MacBeth (1973). Firm size (*SIZE*) is defined as market capitalization at the end of month $t-1$, computed as the stock price times shares outstanding. The book-to-market ratio (*BM*) equals the book value of equity from the most recent fiscal year divided by market capitalization, both constructed following Fama and French (2008). To capture return dynamics, we include short-term reversal (*STR*) as the stock's return in month $t-1$ and momentum (*MOM*) measured as the cumulative return from months $t-12$ to $t-2$, consistent with Jegadeesh and Titman (1993). Profitability (*OP*) is operating profits over book equity, as in Fama and French (2015), and asset growth (*AG*) is the annual percentage change in total assets, following Cooper et al. (2008). Liquidity is proxied by turnover (*TURN*), measured as share volume over the past 126 trading days divided by shares outstanding as in Datar et al. (1998), and illiquidity (*ILLIQ*), the average ratio of absolute daily return to dollar volume over the same window, following Amihud (2002). Idiosyncratic volatility (*IVOL*) is the standard deviation of residuals from a daily market model regression over the past year, following Ali et al. (2003). Earnings surprises are captured via standardized unexpected earnings (*SUE*), computed as the earnings surprise scaled by the standard deviation of forecast errors, in line with Foster et al. (1984). To proxy for lottery-like payoffs, we include *MAX*, the average of the five highest daily returns in month $t-1$, as in Bali et al. (2011). Firm age (*AGE*) is the age of a firm in month, following Jiang et al. (2005). To control for disagreement, we use the forecast dispersion (*AFD*) measure of Diether et al. (2002) defined as the standard deviation of analysts' earnings estimates scaled by the mean earnings estimate. Lastly, we measure attention as the number of analysts following a firm (#*ANA*) as in Elgers et al. (2001).

For all variables we impute missing values with the cross-sectional median. Then, all variables are cross-sectionally winsorized at the 1st and 99th percentiles to mitigate the influence of outliers.

Table 1 shows the descriptive statistics for excess returns in $t+1$ and the specified control variables. Descriptives are calculated in each cross-section and displayed as the time-series average.

[TABLE 1 ABOUT HERE]

3.3 Autoencoder Model

Our goal is to measure how atypical a firm's configuration of characteristics is relative to the cross-section at a point in time. We operationalize atypicality with an autoencoder trained only on lagged characteristics and evaluated out of sample month by month. The model learns the low-dimensional patterns of typical co-movements, and the reconstruction residual reflects how well those patterns reproduce each firm's characteristics. We keep the design simple, regularized, and fully rolling to avoid look-ahead.

The autoencoder is trained on a rolling basis using the past 10 years of data. More specifically, at the end of month $t-1$ we construct ATYP for month t to $t+12$ using a model estimated only on the past 120 months $[t-120, \ldots, t-1]$. We re-estimate the model every 12 months so that the learned patterns can adapt over time. This is to ease the computational burden and is in line with other approaches in return prediction (e.g., Gu et al., 2020; Chen and McCoy, 2024). Therefore, we conduct out-of-sample tests for the period 1981 to 2023.

We implement a symmetric, fully connected autoencoder with ReLU activations in all hidden layers and a linear output layer. We begin with three hidden layers arranged in a pyramid structure, as described in Gu et al. (2020), such that the number of units is halved with each hidden layer. The bottleneck dimension in the baseline specification is set to $k = 16$, which corresponds to roughly 10–20% of the 117 input characteristics. Economically, larger k values absorb firm-specific noise and attenuate the cross-sectional variation in information-processing difficulty, while smaller k values blur typical structure and inflate noise. The selected k thus corresponds to the minimal latent dimensionality capturing the typical patterns of firm characteristics.[9]

This choice is further motivated by evidence that a relatively small number of latent factors is sufficient to organize the cross-section of stock returns. For instance, Barillas and Shanken (2018) show that most anomalies can be captured by a small set of benchmark factors, while Lettau and Pelger (2020) demonstrate that latent factor models extracted from a large panel of characteristics achieve high explanatory power with only a handful of factors. Our results are robust to alternative bottleneck dimensions in the range $\{12, 16, 20\}$. The network is trained to minimize mean squared reconstruction error using the Adam optimizer with a fixed learning

[9]We provide direct evidence for this choice by analyzing under- and overfitting in Section IA.1 in the Internet Appendix.

rate of 0.001. To guard against overfitting, we apply L_2 regularization and dropout before the first hidden layer. For our baseline specification, we select an L_2 penalty of 0.0001 and a dropout probability of 0.1.

Accounting data are noisy and asynchronously updated. To make the statistic robust, we use a denoising variant of the objective specified in Equation (3). More specifically, for each feature we add a small noise term, so that

$$\mathbf{X} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma), \tag{6}$$

which is known to help the model learn the true underlying patterns in the data and to filter idiosyncratic measurement noise. This choice echoes the concern in Bali et al. (2025) that prediction difficulty due to statistical noise can masquerade as information-processing difficulty, which they address by isolating estimation uncertainty. Our denoising step plays the analogous role on the characteristics side. For our baseline we start with $\mathcal{N}(0, 0.1)$.[10]

To ensure that ATYP captures genuine processing difficulty, we perform a series of validation exercises on the autoencoder fit detailed in Section IA.1 in the Internet Appendix. First, we analyze the trade-off between model flexibility and parsimony, confirming that our choice of bottleneck dimension balances the need to capture nonlinear patterns while avoiding the reproduction of noise. Second, we demonstrate that the network effectively filters out systematic economic patterns. While raw characteristics exhibit strong co-movements, the reconstruction residuals are virtually uncorrelated, indicating that ATYP isolates idiosyncratic firm-specific departures rather than common firm characteristics. Finally, we examine the economic drivers of the measure and find that conditional characteristics, such as *Investment* and *Profit Growth*, are the most difficult to reconstruct and most influential for the latent representation, whereas unconditional attributes like *Size* are easily captured. This confirms that ATYP reflects the difficulty of processing complex, context-dependent economic profiles rather than simply acting as a proxy for firm size or obscure reporting.

[10] The selection of suitable parameters may influence the results and therefore merits careful consideration. However, in Section 7.6, we perform a series of sensitivity analyses that suggest that our results are robust to a variety of hyperparameter selections.

3.4 Descriptive Statistics of ATYP

Figure 1 provides a first look at ATYP. The left panel plots its distribution across all firm-month observations. The measure is strictly nonnegative, highly right-skewed, and concentrated near zero. Most firms are well represented by the common patterns, while a nontrivial fraction exhibit unusually high ATYP values, reflecting atypical combinations of fundamentals. The dashed line marks the mean of 0.24, indicating that the bulk of firms cluster tightly around low values with a long right tail of hard-to-process firms.

To place ATYP in a broader context, the right panel plots the 12-month rolling mean of ATYP from 1981 to 2023. The measure varies meaningfully over time, rising during well-known periods of market stress such as the early 2000s tech bubble and the 2008 financial crisis, and declining in calmer periods. This pattern illustrates that ATYP is not a static characteristic but rather a relative measure of how atypical a firm's features are compared to the most other firms at a given time. The underlying patterns evolve in response to macroeconomic conditions, technological advancements, and regulatory shifts that shape the economic landscape. For instance, book leverage or inventory measures may be central in one decade while intangibles or R&D expenditures may dominate in another. As characteristic relevance shifts what was once typical becomes atypical and vice versa. Therefore, the time variation in ATYP partly reflects the natural evolution of the economy and the accounting landscape, reinforcing the notion that processing difficulty is an inherently dynamic construct.

[FIGURE 1 ABOUT HERE]

Next, we examine which firms exhibit high ATYP values. To this end, Table 2 reports cross-sectional correlations of ATYP with one-month-ahead returns and standard firm characteristics.[11] The first row shows that the average correlation between ATYP and future excess returns is negative. Additionally, ATYP is moderately positively correlated with idiosyncratic volatility (0.49), and maximum daily return (0.32), and negatively correlated with size (-0.14) as well as firm age (-0.25). Taken together, these patterns indicate that high-ATYP firms tend to be small, volatile, lottery-like. At the same time, correlations with book-to-market, 12–1 momentum, and short-term

[11] The correlations reported in our tables and figures are calculated using Pearson's correlation coefficient.

reversal are weak, suggesting that ATYP does not simply replicate standard anomaly variables.

These correlations are economically consistent with the interpretation of ATYP as a fundamental source of information uncertainty. Firms with atypical characteristics are inherently more difficult to model because they deviate from the standard patterns investors use to interpret economic signals. Because high-ATYP firms do not fit these patterns investors neglect the firm specific details. This neglect increases information uncertainty (Peng and Xiong, 2006). This model uncertainty naturally manifests as higher idiosyncratic volatility because market participants struggle to agree on a precise valuation for firms with ambiguous profiles. In other words, while idiosyncratic volatility captures the ex-post realization of this uncertainty, ATYP identifies its ex-ante structural origin.

Furthermore the results suggest that high ATYP firms operate in a weaker information environment. The negative association with analyst coverage suggests that information intermediaries allocate fewer resources to firms that are costlier to process and harder to explain to clients. Furthermore the link to younger and smaller firms aligns with the business lifecycle where early stage companies have not yet converged to the stable characteristic configurations of mature incumbents. Consequently, ATYP identifies the statistical root cause of the ambiguity found in high uncertainty environments.

These patterns align with prior evidence that investors face greater difficulty when processing these types of stocks (e.g., Hong et al., 2000; Baker and Wurgler, 2006; Bali et al., 2011). Importantly, the dispersion across percentiles shows that these correlations vary across the cross-section but remain economically modest, reinforcing that ATYP captures a distinct dimension of information-processing frictions rather than duplicating known anomalies.

[TABLE 2 ABOUT HERE]

To further characterize the types of firms driving the effect, we examine average firm characteristics across ATYP deciles. Again, high-ATYP firms are smaller, less profitable, more illiquid, and exhibit higher idiosyncratic volatility and lottery-like return profiles. For completeness, these univariate characteristic sorts are reported in Table IA.3.1 in the Internet Appendix.

4 ATYP as a Measure of Information-Processing Difficulty

In this section, we empirically validate ATYP as an ex-ante measure of information-processing difficulty. We begin by examining the relationship with some of the most notable existing proxies of processing difficulty. We further demonstrate that ATYP predicts future manifestations of information-processing frictions. Finally, we document that firms with higher ATYP exhibit slower price discovery.

4.1 Relation to Other Proxies of Information-Processing Difficulty

A central requirement for any proposed proxy of information-processing difficulty is that it captures a distinct underlying dimension rather than rebranding existing measures. To examine this, we systematically compare ATYP with four widely used proxies of processing difficulty or firm complexity: the textual complexity of 10-K filings, the number of business segments, the number of geographic segments, and the accounting reporting complexity (Cohen and Lou, 2012; García and Norli, 2012; Hoitash and Hoitash, 2018; Loughran and McDonald, 2024). Each of these measures reflects a different conceptual foundation. Textual complexity captures the readability and syntactic difficulty of disclosure, which proxies for the cognitive burden of parsing firm narratives. Segment counts capture organizational scope, which approximates the breadth of activities and the heterogeneity of cash-flow sources that investors must process. Accounting reporting complexity captures disclosure of more accounting items, which proxies for processing difficulty because it requires greater knowledge of authoritative accounting standards. Additionally, processing more items requires more resources. These proxies have proven empirically useful but remain confined to specific reporting dimensions and depend heavily on managerial disclosure choices. In contrast, ATYP is an outcome-based construct derived from firm characteristics rather than disclosures or organizational form. It measures the atypicality of a firm's information set and therefore embodies a statistical rather than representational notion of complexity.

To assess the empirical relation between these measures, we compute cross-sectional correlations between ATYP and each other proxy in every month and summarize the distribution of these correlations in Figure 2. The correlations are modest in magnitude and frequently negative. The mean correlation between ATYP and textual complexity is close to zero, while the correlations with business and geographic segment counts are slightly negative on average. These patterns show that information-processing difficulty, as captured by ATYP, is not merely a reformulation of disclosure or diversification breadth. Instead, it reflects a qualitatively distinct processing friction.

[FIGURE 2 ABOUT HERE]

The weak negative association is theoretically coherent. Large, diversified firms tend to exhibit high segment counts or lengthy disclosures, which inflate traditional measures of complexity. Yet their fundamentals typically follow well-understood patterns with stable characteristics that the model reconstructs easily, resulting in low ATYP scores. Hence, they have a strong information environment. Conversely, some firms that appear simple by conventional standards, because they have few segments or concise, readable filings, can exhibit unusual combinations of firm characteristics. The distinction highlights the difference between organizational complexity and information atypicality: the former concerns how many activities a firm undertakes, whereas the latter concerns how difficult its economic profile is to interpret given the prevailing structure of information in the market.

4.2 Incremental Explanatory Power

We next examine whether ATYP captures dimensions of information-processing difficulty that manifest in observable market outcomes following Loughran and McDonald (2014), Loughran and McDonald (2024) and Bali et al. (2025). Prior research establishes that processing difficulty is associated with greater disagreement among investors, lower forecast accuracy, higher sensitivity of prices to news, and larger deviations between realized and expected earnings. For example, Diether et al. (2002) and Johnson (2004) show that forecast dispersion rises with information uncertainty. Pástor and Pietro (2003) and Brogaard and Detzel (2015) link uncertainty and disagreement to elevated return volatility. DellaVigna and Pollet (2009) and Hirshleifer et al. (2009) document that processing difficulty and limited attention increase the likelihood of earnings surprises. Building on these insights, we investigate whether ATYP systematically predicts forecast dispersion, forecast accuracy, return volatility, and absolute earnings surprises. To test these associations, we regress each outcome in $t + 1$ on ATYP, controlling for the standard set of firm characteristics and including firm and month fixed effects. Standard errors are double-clustered by firm and month.

Table 3 reports the results. Consistent with the information-processing interpretation, ATYP is strongly and positively related to analyst forecast dispersion and subsequent return volatility. A one-unit increase in ATYP predicts a 0.161 rise in forecast dispersion and a 0.017 rise in return volatility in the following month. ATYP also loads positively and significantly on absolute earnings surprises, indicating that firms with high ATYP are more likely to deliver outcomes that deviate sharply from expectations. For forecast accuracy, we document a strong negative association, as analysts make more errors for high-ATYP firms. This strengthens the argument that ATYP is an ex-ante state variable that describes the information environment before beliefs are formed. By raising the cost of processing information, ATYP paves the way for persistent disagreement and mispricing. Note, that these effects are highly significant after controlling for known determinants, suggesting that ATYP provides incremental explanatory power beyond established predictors.

[TABLE 3 ABOUT HERE]

4.3 Post-Earnings-Announcement Drift

If ATYP captures an information-processing friction that impedes timely price adjustment, its pricing effects should be especially pronounced around major information events. The literature on post-earnings-announcement drift (PEAD) documents that stock prices underreact to earnings news, generating predictable returns in subsequent months (e.g., Bernard and Thomas, 1989; Engelberg et al., 2018). PEAD is commonly interpreted as evidence of limited attention and gradual information diffusion (Hirshleifer and Teoh, 2003b; Hirshleifer et al., 2009). If high-ATYP firms are more difficult to process, investors may require more time to incorporate earnings signals into prices, producing stronger and more persistent drift (Barinov et al., 2024).

We test this prediction using both bivariate portfolio sorts and Fama–MacBeth cross-sectional regressions. Table 4 reports dependent double sorts in which firms are first grouped into ATYP quintiles and then into earnings-surprise quintiles. We then compute the H–L earnings-surprise return spread over different horizons. Panel A shows that in month $t+1$, the spread is 0.27%

($t = 3.23$) among low-ATYP firms but rises to 0.89% ($t = 6.30$) among high-ATYP firms. Panels B and C extend the horizon to months $t + 2$ and $t + 3$, respectively.[12] In both cases, the spreads remain large and significant, with the strongest effects concentrated in the high-ATYP quintile. This persistence indicates that earnings signals are incorporated more gradually for atypical firms.

[TABLE 4 ABOUT HERE]

To corroborate, Table IA.3.2 in the Internet Appendix presents Fama–MacBeth regressions of post-announcement returns on earnings surprise, ATYP, and their interaction, controlling for standard firm characteristics. Across horizons $t + 1$ and $t + 2$, the interaction term is positive and significant. These regressions confirm the portfolio evidence in a continuous setting with controls.

4.4 Industry Information Diffusion

To examine whether the price discovery implications of ATYP extend beyond firm-specific earnings announcements, we also study lead-lag effects of industry information shocks following the approach of Hou (2007) and Cohen and Lou (2012). In their setting, a common industry shock affects two sets of firms: one that processes information straightforwardly, and another that requires more complex analysis to incorporate the same information into prices. In their study, they use the number of firm segments to characterize hard-to-process firms.

Analogously, we test whether ATYP influences firms' response to industry-wide signals using a two-step approach. First, we construct industry shocks as the average prior-month return of the simplest firms (lowest ATYP quintile) within each 2-digit SIC industry. We then examine how the most atypical firms (highest ATYP quintile) within the same industries respond to these shocks. If atypical firms process information more slowly, their prices should adjust more gradually to signals already reflected in their simpler peers. To test this, we sort high-ATYP firms into decile portfolios based on the industry shock and evaluate return patterns.

Table 5 shows that our findings mirror those of Cohen and Lou (2012). We find that high-ATYP firms exhibit a strong and predictable delayed response to information that has already been incorporated into the prices of their low-ATYP industry peers. In the equal-weighted specification

[12] The t-statistics reported in our tables are Newey and West (1987) adjusted with six lags to control for heteroskedasticity and autocorrelation.

in Panel A the portfolio of high-ATYP firms linked to the most negative industry shocks (Decile 1) generates a statistically significant six factor alpha of -1.54% ($t = -6.50$) while the spread between the positive and negative shock portfolios yields a monthly alpha of 1.62% ($t = 5.11$). The return pattern is monotonic confirming that information-processing difficulty impedes the timely incorporation of industry wide news. These results hold firmly in value-weighted portfolios in Panel B where the high minus low spread remains economically large at 1.10% per month ($t = 2.47$) after controlling for the six factor model. Collectively these estimates provide compelling evidence that ATYP captures a dimension of complexity that slows price discovery causing high-ATYP firms to lag behind the fundamental information revealed by their simpler peers.

[TABLE 5 ABOUT HERE]

5 Atypical Firms and the Cross-Section of Expected Returns

In this section, we conduct different tests to assess the predictive power of ATYP over future stock returns. First, we show results of the univariate portfolio sorts. Second, we run bivariate portfolio sorts to control for well-known stock characteristics. Third, we present firm-level Fama-MacBeth cross-sectional regression results. Fourth, we examine the persistence of the predictive power.

5.1 Univariate Portfolio Sorts

We begin our empirical analysis with univariate portfolio sorts. At the end of each month from 1981 to 2023, we sort all stocks into deciles according to their ATYP. Within each decile, we compute the one-month-ahead excess return, measured relative to the one-month Treasury bill rate, for both equal-weighted and value-weighted portfolios. To capture the return differential associated with information-processing difficulty, we form a long-short (H–L) portfolio that takes a long position in the highest-ATYP decile and a short position in the lowest-ATYP decile. This H–L spread provides a direct measure of the pricing impact of ATYP.

In addition to excess returns, we report alphas relative to standard factor models. Specifically, we estimate alphas under the capital asset pricing model (CAPM), the Fama and French (1993) three-factor model (FF3: market, size, and value), the Fama and French (2015) five-factor model

(FF5: market, size, value, profitability, and investment), and the Fama and French (2018) six-factor model (FF6: FF5 plus momentum). By controlling for these benchmarks, we assess whether the return patterns associated with ATYP are distinct from known risk factors.

Table 6 presents the results. A clear monotonic pattern emerges, with average returns declining steadily from the lowest- to the highest-ATYP decile. The equal-weighted H–L portfolio earns an average return of -1.47% per month ($t = -4.54$), while the value-weighted spread delivers -0.82% per month ($t = -2.50$). After controlling for risk factors, the spreads remain large and statistically significant. For example, under the FF6 model, the equal-weighted H–L alpha is –0.84% per month ($t = -4.52$). The value-weighted FF6 alpha is smaller in magnitude at –0.33% per month, with weaker significance ($t = -1.49$). This contrast indicates that the return predictability of ATYP is concentrated in smaller firms, where attention and arbitrage frictions are most binding and information-processing is less efficient.

Importantly, the negative association between ATYP and abnormal future returns is concentrated in the high-ATYP leg. The most atypical firms earn strongly negative alphas, while low-ATYP firms yield moderate positive alphas. This asymmetry suggests that investors systematically overvalue firms that are hard to process, leading to predictable reversals. The evidence is more consistent with mispricing than with compensation for risk, since high-ATYP firms appear overpriced relative to their fundamentals.

[TABLE 6 ABOUT HERE]

As an additional validation of the univariate portfolio sorts, Figure IA.2.3 in the Internet Appendix plots the cumulative performance of the low-minus-high (L–H) ATYP portfolios over the full sample period. Both equal- and value-weighted portfolios yield persistently positive returns, with the equal-weighted series showing particularly strong and steady outperformance of low-ATYP firms relative to high-ATYP firms. Importantly, this pattern is stable across decades and robust to different market environments, including recessions. This time-series evidence reinforces the cross-sectional sorts by showing that the premium associated with ATYP is a persistent and recurring feature of financial markets rather than a temporary anomaly.

5.2 Bivariate Portfolio Sorts

Next, we evaluate whether the predictive value of ATYP is subsumed by other well-established predictors of return. To do so, we perform dependent double sorts. At the start of each month, firms are first sorted into quintiles according to a given control characteristic, such as size, book-to-market ratio, 12–1 momentum, profitability, or volatility. Within each quintile, firms are further sorted into ATYP deciles. This procedure yields 50 portfolios (5 × 10), which are conditioned jointly on the control variable and ATYP. The H–L return spread is formed by combining the lowest-ATYP portfolios across quintiles into the low leg and the highest-ATYP portfolios into the high leg. For each control variable, we compute the H–L spread and its alpha under the FF6 model using equal- and value-weighted returns.

The findings shown in Table 7 are compelling. To keep the presentation concise, we focus on the H–L portfolio spreads and the Fama and French (2018) six-factor alphas. The evidence demonstrates that ATYP continues to generate substantial economic returns and maintains strong statistical significance even when accounting for various control factors. When portfolios are equally weighted, the six-factor alphas consistently fall between –0.57% and –0.89% monthly, accompanied by robust t-statistics ranging from –3.63 to –5.91. While value-weighted results show somewhat reduced magnitudes, they nonetheless remain statistically meaningful, with alphas spanning –0.13% to –0.74% per month and corresponding t-statistics between –0.89 and –5.08. The persistence of these effects indicates that the return premium associated with ATYP cannot be attributed to cross-sectional return predictors such as firm size, book-to-market ratios, momentum, or other established return predictors.

[TABLE 7 ABOUT HERE]

5.3 Fama–MacBeth Regressions

We complement the portfolio sorts with Fama and MacBeth (1973) cross-sectional regressions, which allow us to quantify the pricing implications of ATYP while simultaneously controlling for a broad set of firm characteristics. Specifically, at the end of each month we regress one-month-ahead excess stock returns on ATYP and standard predictors, and then average the coefficients over time.

Table 8 reports the results for five different specifications with increasing number of controls. Across all specifications, the loading on ATYP is negative, large in magnitude, and statistically significant. In the univariate specification, the ATYP coefficient is –2.33 ($t = -4.55$), implying that a one-unit increase in ATYP predicts a 2.33% lower excess return in the subsequent month. Importantly, the coefficient remains strongly negative even after sequentially including controls for size, value, momentum, profitability, investment, trading activity, illiquidity, idiosyncratic volatility, earnings surprises, firm age, disagreement, and attention. In the most saturated specification, which includes more than a dozen established predictors, ATYP continues to load significantly at –0.72 ($t = -3.21$). This demonstrates that the predictive power of ATYP is not subsumed by other well-known return determinants. Most importantly, the negative association is robust to characteristics related to illiquidity, idiosyncratic volatility, lottery-like return profiles and disagreement, suggesting that ATYP captures a distinct dimension of information-processing frictions.

[TABLE 8 ABOUT HERE]

The Fama–MacBeth results reinforce the portfolio evidence that stocks with atypical characteristic profiles systematically earn lower future returns, even after accounting for risk factors and anomaly controls. This robustness strengthens the interpretation of ATYP as an independent state variable linked to investor attention frictions rather than as a proxy for existing characteristics.

5.4 Return Persistence

We conclude our main results by examining the persistence of the ATYP premium. Since ATYP captures fundamental economic atypicality rather than transient noise, we expect both the signal and its associated return predictability to persist over longer horizons.

First, we verify the stability of the measure itself. Because ATYP is largely derived from fundamental accounting variables that evolve gradually, firms typically do not experience sharp month-to-month changes in their atypicality status. Figure IA.2.4 in the Internet Appendix confirms this stability. Transitions across ATYP deciles are infrequent, particularly at the extremes, where nearly half of the firms remain in the same decile after 12 months. This persistence suggests information-processing difficulty is a durable characteristic of a firm's information environment.

Next, we examine whether this persistent signal translates into long-lasting return predictability. Figure 3 plots the average high-minus-low (H–L) decile portfolio spread for horizons ranging from one to twelve months ahead. We report results for the full sample and distinguish between equal-weighted and value-weighted portfolios to assess the role of firm size, and for both raw returns and alphas.

[FIGURE 3 ABOUT HERE]

The return spread remains economically large and persistent across horizons. For equal-weighted portfolios, the H–L spread is consistently negative and lies close to -1.5% per month at short horizons, with only modest variation over time. Although there is some attenuation as the horizon increases, the spread remains around -1.25% even twelve months ahead. The corresponding alpha estimates closely track the raw returns, indicating that standard risk factor adjustments do not materially alter the magnitude or persistence of the effect. Value-weighted portfolios display a smaller but still robust pattern. The H–L spread starts at roughly -0.8% to -0.9% per month and declines to approximately -1.0% at intermediate horizons before partially rebounding toward the end of the year. As in the equal-weighted case, the alpha series closely mirrors the raw spread, suggesting that the negative return differential is not explained by conventional risk exposure. Overall, the figure indicates that the mispricing associated with atypical firms is slow to dissipate and is substantially stronger in equal-weighted portfolios, consistent with the interpretation that the effect is driven primarily by smaller firms and is amplified by limits to arbitrage and constrained investor attention.

6 Channels of Return Predictability

After documenting the negative relationship between ATYP and average stock returns, we examine the economic mechanisms that may explain this pattern. Based on the theoretical literature and the evidence presented thus far, we argue that information-processing difficulty may lead to mispricing. First, we examine whether ATYP covaries with established proxies of mispricing. Then, we explore potential channels. Specifically, we consider limits to arbitrage arising from short-sale constraints and limited investor attention.

6.1 Mispricing vs. Risk

The negative pricing of ATYP is best understood as a mispricing premium rather than compensation for risk. Two interacting mechanisms explain why information-processing difficulty leads to systematic overpricing and subsequent low returns.

First, high ATYP increases the ambiguity surrounding a firm's valuation, which raises information uncertainty. In environments with greater uncertainty, behavioral biases such as overconfidence become more pronounced, leading investors to overweight their own noisy and imperfect signals (Daniel et al., 1998, 2001; Zhang, 2006). This amplified confidence in private assessments generates stronger divergence of opinion across market participants. When such disagreement interacts with short-sale constraints, market prices increasingly reflect the valuations of optimistic investors rather than the full distribution of beliefs, resulting in systematic overpricing of firms with high ATYP (Miller, 1977). Consistent with this mechanism, Section 4.2 shows that ATYP is associated with higher forecast dispersion, forecast errors, and return volatility, all of which indicate greater disagreement and uncertainty in the processing of atypical firms.

Second, limited attention affects how quickly information about atypical firms is incorporated into prices. Investors with constrained processing capacity optimally allocate attention toward broad typical patterns and devote less effort to complex firm-specific details (Peng and Xiong, 2006). Because atypical firms deviate from these coarse patterns, their fundamentals receive less immediate scrutiny, and new information is only gradually absorbed across the investor population. In the spirit of Hong and Stein (1999), such slow diffusion produces systematic underreaction, where only a subset of sophisticated investors processes the news promptly and others update their valuations with delay. This mechanism explains the strong post-earnings-announcement drift and the sluggish response to industry shocks documented in Section 4.3 and Section 4.4. Moreover, limits to arbitrage, such as high idiosyncratic volatility and limited liquidity, prevent informed traders from correcting these pricing errors quickly (Pontiff, 2006), allowing mispricing in high-ATYP firms to persist for extended periods.

If the negative pricing of ATYP indeed reflects mispricing, it should covary with established mispricing measures. We therefore benchmark ATYP against the composite mispricing score of Stambaugh et al. (2015), where higher values indicate greater overpricing. Panel A of Table 9 shows that firms with high ATYP also have significantly higher mispricing scores, consistent with the idea that they are more prone to overpricing. Panel B reports dependent double sorts, where firms are first grouped by the mispricing score and then sorted into ATYP quintiles. The ATYP high-minus-low spread is most pronounced in the highest mispricing quintile, with a six-factor alpha of -0.88% ($t = -3.98$). These empirical results reinforce our interpretation that the ATYP premium reflects mispricing rather than compensation for risk and is concentrated in the most overpriced firms.

[TABLE 9 ABOUT HERE]

6.2 Limits to Arbitrage

When investors disagree, frictions that limit arbitrage can lead to overpricing. Since ATYP involves ex-ante information-processing frictions that increase investor disagreement, we should observe a similar effect with ATYP under high limits to arbitrage. The literature highlights several such frictions. First, high idiosyncratic volatility raises the residual risk of arbitrage positions, deterring arbitrageurs from betting against overpriced firms (Pontiff, 2006; Stambaugh et al., 2015). Second, illiquidity raises trading costs and price impact, further discouraging large corrective trades (Amihud, 2002; Pástor and Stambaugh, 2003). Third, young firms have shorter track records, higher uncertainty, and greater information asymmetry, making them especially difficult to value (Pástor and Pietro, 2003).

Table 10 explores these mechanisms by conditioning the ATYP premium on proxies for arbitrage frictions. The conditional sorts show that the negative ATYP premium is concentrated in the high-friction groups. For instance, in the high-IVOL group, the FF6-adjusted H–L return is -1.51% per month ($t = -6.77$), compared to a statistically insignificant 0.13% ($t = 1.90$) in the low-IVOL group. Similarly, the spread is -1.22% ($t = -6.18$) in illiquid stocks vs. 0.07% ($t = 0.73$)

in liquid stocks, and –1.18% ($t = -5.50$) in young firms vs. -0.06% ($t = -0.48$) in older firms. In all cases, the difference-in-spreads is large and highly significant.

[TABLE 10 ABOUT HERE]

These findings strongly support a mispricing interpretation. If ATYP represented a priced risk factor, the premium should persist regardless of arbitrage frictions, or potentially be stronger where arbitrage is easier. Instead, the effect is entirely absent in low-friction groups and survives standard risk factor controls only where limits to arbitrage are binding. We confirm this conclusion using Fama-MacBeth regressions in Table IA.3.3 in the Internet Appendix, where the interaction terms between ATYP and arbitrage proxies remain significant after controlling for firm characteristics. This evidence suggests that information-processing difficulty generates mispricing that persists precisely because arbitrageurs are unable or unwilling to correct it.

6.3 Limited Attention

Limited investor attention leads to a slower correction of mispricing, because constrained investors rely on simplified heuristics rather than full information processing. As Peng and Xiong (2006) demonstrate, investors cope with cognitive limits by allocating attention to broad typical patterns while neglecting firm-specific information. Because high-ATYP firms do not fit these standard categories, investors neglect their idiosyncratic details. In the spirit of Hong and Stein (1999), this neglect prevents signals from being immediately incorporated into prices. Instead, information diffuses only gradually across the investor population, leading to underreaction and persistent mispricing. However, this friction should vanish when attention is abundant. In high-attention environments, sophisticated investors rapidly process all available information regardless of its complexity and ensure that prices reflect fundamental value. To test this prediction, we condition the ATYP premium on three proxies for the information environment: firm size, analyst coverage, and short interest. Prior work shows that larger firms attract more informed attention Hong et al. (2000), low analyst coverage reflects weaker information efficiency Elgers et al. (2001), and high short interest signals a high share of sophisticated investors Boehmer and Wu (2012).

Table 11 reports the results of dependent double sorts. The findings show that the pricing effect of ATYP is sharply concentrated in low-attention environments, where investors are most likely to rely on coarse patterns. For example, within the low-size group, the six-factor adjusted high-minus-low return is -1.54% per month ($t = -6.90$), whereas in the high-size group the spread is economically and statistically negligible. We observe an identical pattern for analyst coverage, with a spread of -1.02% in low-coverage firms versus 0.27% in high-coverage firms, and for short interest with -1.39% versus -0.20%. In every specification, the difference in spreads between the low- and high-attention groups is large and statistically significant, ranging from 1.19% to 1.74% per month.

[TABLE 11 ABOUT HERE]

Table IA.3.4 in the Internet Appendix confirms these results using Fama-MacBeth cross-sectional regressions. We regress returns on ATYP, attention measures, and their interactions while controlling for standard characteristics. The interaction coefficients are consistently significant, indicating that the negative pricing of ATYP is conditional on the scarcity of investor attention.

These findings strongly support the attention-based mechanism. In low-attention segments of the market, pattern-based traders dominate. Because they process information through standard patterns, as modeled by Peng and Xiong (2006), they struggle to value atypical firms correctly. Consequently, information diffuses slowly, generating the systematic underreaction predicted by Hong and Stein (1999). In contrast, when attention is high, sophisticated investors allocate sufficient cognitive resources to decipher complex signals. In these environments, atypical information is incorporated efficiently and the return predictability disappears. Together with our evidence on limits to arbitrage, this confirms that information-processing difficulty generates overpricing precisely when investors lack the bandwidth to look beyond typical patterns.

7 Robustness Checks

Having established that ATYP is a valid measure of information-processing difficulty and that there is a negative cross-sectional relationship between ATYP and average stock returns that is driven by limited attention, limits to arbitrage and potentially mispricing, we next run a series of robustness checks to challenge the aforementioned findings.

7.1 Firm Characteristic Clusters

An important question is whether the ATYP premium generalizes across distinct domains of firm information or arises from a concentrated subset of characteristics. To address this, we re-calculate ATYP within clusters of related firm characteristics and examine whether the return predictability persists across these different information domains. The clusters are taken from Jensen et al. (2023). Within each cluster, we compute firm-specific reconstruction errors based solely on the variables in that group and form decile portfolios on the resulting ATYP measure.

Table IA.3.5 in the Internet Appendix reports the high-minus-low return spreads from these sorts, along with factor-adjusted alphas from the Fama and French (2018) six-factor model. The equal-weighted results show that the ATYP premium is present in nearly all clusters. For example, firms that are hard to process with respect to investment, profitability, and value characteristics exhibit economically large and statistically significant underperformance, with spreads ranging from -0.74% to -1.53% per month. The effect is somewhat weaker for the value-weighted portfolios, but still evident in investment and leverage-related clusters.

7.2 Within-Industry Construction

So far, our results have taken the typical patterns for the ATYP computation from the whole market. In this section, we examine whether the pricing of ATYP also emerges when the measure is estimated within industries. ATYP is re-estimated separately for each one digit SIC industry, and all preprocessing steps outlined in Section 3.1, including imputation, winsorization, and standardization, are performed within each industry in every month. The model is then trained separately for each industry, allowing it to learn the prevailing information structure within that group and to identify firms that are atypical relative to their immediate peers. The resulting industry-adjusted ATYP values are then combined across industries to form the complete sample.

Table IA.3.6 in the Internet Appendix presents the results. Panel A reports equal-weighted and Panel B value-weighted portfolio sorts for the full cross-section using the industry adjusted measure. The pattern of returns is clear and monotonic. Firms with higher ATYP continue to earn lower subsequent returns. The equal-weighted high minus low spread is -1.40 percent per month ($t = -4.73$) with an FF6 alpha of -0.93 percent ($t = -5.43$). Value-weighted portfolios show a smaller but consistent effect. The overall magnitude and statistical strength of the spreads remain comparable to the baseline results, showing that the pricing of information-processing difficulty persists even when ATYP is estimated entirely within industries.

Table IA.3.7 in the Internet Appendix further reports high minus low spreads for portfolio sorts conducted separately within each one digit SIC industry. The pricing effect of ATYP remains negative and significant across nearly all industries. Equal-weighted alphas range between -0.39 and -1.28 percent, with the strongest effects in *Public Administration*, *Services*, and *Trade*. Value-weighted results follow the same pattern. The persistence of the pattern across nearly all sectors demonstrates that information-processing difficulty is priced within the industrial structure of the economy.

7.3 Within-Size Group Construction

Next, we examine whether the pricing of ATYP varies across the size distribution. To ensure that the measure captures atypicality relative to firms with similar information environments, we re-estimate ATYP separately within each size group.[13] At the end of each month, firms are sorted into five groups based on their market equity: nano, micro, small, large, and mega. All stages of the ATYP construction, including preprocessing and the model training, are performed within each size group. This procedure removes any systematic relation between firm size and the ATYP measure. The boxplots in Figure IA.2.5 confirm that the distribution of ATYP is nearly identical across the five size groups, and that the unconditional correlation between size and ATYP, as documented in Section 3.4, effectively disappears under this design.

Table IA.3.8 presents the full sample portfolio sorts based on size-adjusted ATYP. Panel A reports equal-weighted portfolios and Panel B reports value-weighted portfolios. The return pattern across deciles is clear and monotonic. Firms in the lowest ATYP decile earn the highest subsequent returns, while firms in the highest ATYP decile earn substantially lower returns.

[13] The size groups are taken from Jensen et al. (2023). These groups are non-overlapping, and the breakpoints are based on the market equity of NYSE stocks at the end of each month. Specifically, mega caps are stocks with a market cap above the 80th percentile of NYSE stocks, large caps are all remaining stocks above the 50th percentile, small caps are all remaining stocks above the 20th percentile, micro caps are all remaining stocks above the 1st percentile, and nano caps are all remaining stocks.

The equal-weighted high minus low spread is -1.33 percent per month ($t = -4.81$), with a corresponding FF6 alpha of -0.80 percent ($t = -5.49$). As in the baseline results, value-weighted portfolios show a similar but somewhat attenuated pattern, consistent with the concentration of information-processing difficulty among smaller firms. Overall, the magnitude and statistical strength of the spreads remain comparable to the baseline results, showing that the pricing of processing difficulty persists even when ATYP is estimated entirely within size groups.

Despite the absence of a direct mechanical relation between ATYP and size, the return implications of the measure vary substantially across the size dimension. Table IA.3.9 reports the high minus low portfolio spreads computed separately within each size group. The pricing effect is concentrated almost entirely among the smaller firms. For nano and micro firms, the equal-weighted high minus low spreads are large and statistically significant. The spreads remain strongly negative after controlling for standard priced risk factors. For example, among nano firms, the FF6 alpha of the high minus low portfolio is approximately -1.59 percent per month with a t-statistic of -4.12. The pattern is nearly identical for micro firms. Small firms also exhibit a negative spread, though the magnitude is slightly smaller. In contrast, the pricing effect attenuates sharply among large and mega firms. For these firms, the high minus low spreads are close to zero, and the corresponding alphas are statistically indistinguishable from zero. Value-weighted results follow the same pattern and further underscore that the pricing of ATYP is concentrated among firms with lower size, consistent with interpretations related to investor attention and limits to arbitrage.

7.4 Alternative Factor Models

In this section, we examine whether the estimated ATYP premium survives when controlling for alternative factor models beyond the baseline Fama–French six-factor model. To do so, we re-estimate portfolio alphas using several widely employed frameworks that encompass both rational and behavioral sources of return variation. Specifically, we compare the equal- and value-weighted portfolio spreads under the six-factor model of Fama and French (2018), the four-factor q-model of Hou et al. (2014) that includes the market, size, investment, and profitability factors, the behavioral factor model (DHS) of Daniel et al. (2019) that employs market, post-earnings-announcement-drift, and financing factors, and the characteristic-efficient portfolio factor model (DMRS) of Daniel et al. (2020) which builds on the FF5 framework while neutralizing unpriced covariance structures. The corresponding results are reported in Table IA.3.10 in the Internet Appendix.

Panel A reports equal-weighted portfolio returns. When controlling for risk using the FF6 model, the alpha remains economically and statistically significant at -0.84 percent per month ($t = -4.52$). The magnitude of the alpha is only slightly reduced when applying the q-factor model (-0.73 percent, $t = -3.24$) or the behavioral DHS model (-0.67 percent, $t = -2.56$). The DMRS specification yields a comparable result (-0.93 percent, $t = -3.34$). The persistence of negative and statistically significant alphas across all model specifications indicates that the return differential associated with ATYP cannot be explained by standard investment-based or behavioral factor structures.

The somewhat smaller alpha under the DHS model is consistent with the interpretation that information-processing difficulty affects prices through behavioral channels related to limited attention and delayed information processing. Since the DHS framework explicitly includes a post–earnings-announcement-drift factor that captures investor underreaction to new information, it absorbs part of the same behavioral component that the information-processing difficulty measure identifies. Consequently, the weaker alpha under the DHS model reinforces the notion that the pricing effect of ATYP reflects potential mispricing arising from cognitive and attention-related frictions rather than compensation for systematic risk.

Panel B reports value-weighted results. The alphas across all factor models are smaller in magnitude and statistically weaker (-0.28, -0.23, -0.17, and -0.11 percent for the FF6, q, DHS, and DMRS models, respectively), which is expected given that high-ATYP firms are typically smaller and less liquid, causing value-weighted portfolios to dilute their influence. Nonetheless, the signs of all estimated alphas remain consistent with the equal-weighted results, and no specification fully eliminates the return differential between high- and low-ATYP firms.

7.5 Alternative Data Sources

To test the robustness of our findings to data construction choices, we replicate the main portfolio analyses using alternative sources of firm-level information. For each dataset, we apply the same data processing as in the baseline specification in Section 3.1: missing data are imputed with the cross-sectional median, variables are winsorized at the 1% level on both sides, standardized, and used as inputs in the autoencoder to construct ATYP as the firm-level reconstruction error. This approach ensures that any differences in results are due to the underlying data rather than methodological choices. The model is also specified as the baseline specification in Section 3.3.

As a first replication, we employ the WRDS Financial Ratio database, which provides a standardized set of 71 accounting and market-based variables. Table IA.3.11 in the Internet Appendix reports the results. The equal-weighted high-minus-low ATYP portfolio spread is large and statistically significant, with an average difference of –1.42% per month ($t = -5.68$) in excess returns and –0.95% ($t = -6.48$) after controlling for the Fama and French (2018) six-factor model. Value-weighted sorts yield a weaker but still meaningful spread of –0.50% ($t = -2.15$). The return pattern is robust across alternative factor models, including the CAPM, the Fama and French (1993) three-factor model, and the Fama and French (2015) five-factor model.

As a second replication, we expand the input space to the full set of 153 firm characteristics included in the dataset of Jensen et al. (2023) without excluding any clusters. Table IA.3.12 in the Internet Appendix presents the results. Equal-weighted sorts again show a pronounced ATYP effect, with the high-minus-low portfolio spread of –1.42% per month in excess returns ($t = -4.48$) and –0.83% ($t = -4.48$) after controlling for the Fama and French (2018) six-factor model. The spread remains highly significant across alternative factor models, including the three- and five-factor specifications. Value-weighted portfolios deliver somewhat weaker results, with a spread of –0.57% ($t = -1.82$) in excess returns and –0.94% ($t = -3.27$) relative to the CAPM. These findings show that the predictive power of ATYP is not an artifact of excluding seasonal or short-horizon return-based risk measures. Instead, ATYP continues to price the cross-section even in a maximally inclusive specification, reinforcing the robustness of the measure to dataset construction choices.

7.6 Different Model Specifications

The predictive power of ATYP may hinge on the choice of hyperparameters in the autoencoder or on the specific model architecture employed. To address this, we systematically vary the main tuning dimensions of the autoencoder, including the size of the bottleneck layer, the number of hidden layers, the number of units per layer, and the degree of regularization through L_2, dropout, and input noise added.

Table IA.3.13 in the Internet Appendix summarizes the results. Across all 48 hyperparameter choices, the equal-weighted high-minus-low ATYP portfolio spread remains large in magnitude and statistically significant, with t-statistics typically above four in both raw and factor-adjusted returns. The value-weighted spreads are smaller, but they are consistently negative and retain statistical significance for many specifications. This robustness underscores that the ATYP premium is not an artifact of a particular configuration of the network. Even when the dimensionality of the bottleneck or the strength of regularization is altered substantially, the return patterns associated with ATYP remain stable.

7.7 Autoencoder vs. PCA

In Section 2, we motivate the use of an autoencoder by arguing that systematic combinations of firm characteristics are inherently nonlinear. A potential concern, however, is whether the predictive power of ATYP truly stems from these nonlinear interactions or simply from linear deviations that could be captured by standard dimensionality reduction techniques like PCA. If ATYP is merely a proxy for linear atypicality, the additional complexity of the neural network would be unnecessary.

To isolate the contribution of nonlinear information processing, we construct a linear benchmark measure, ATYP_{PCA}, using a PCA model with the same bottleneck dimension ($k = 16$) as our baseline autoencoder. We then regress our original autoencoder-based measure, ATYP_{AE}, on ATYP_{PCA} for each firm-month and extract the orthogonal residual. This residual captures the component of information-processing difficulty that arises strictly from nonlinear deviations, i.e. information that the linear PCA model considers typical or cannot effectively compress.

Table IA.3.14 in the Internet Appendix reports the results of univariate portfolio sorts based on this orthogonalized measure. The equal-weighted H–L spread remains economically large and statistically significant at -0.53 percent per month ($t = -4.55$). Crucially, the spread retains a significant alpha of -0.42 percent ($t = -4.45$) under the Fama and French six-factor model, indicating that the nonlinear component of atypicality carries independent pricing information. While the value-weighted results are weaker, the persistence of a strong equal-weighted effect confirms that the autoencoder identifies a distinct set of hard-to-process firms. Those are likely cases with complex, interactive characteristic profiles that a linear model fails to capture. These findings validate our choice of a nonlinear architecture and suggest that information-processing difficulty is, at least in part, a function of complex, non-additive economic patterns.

7.8 Different Measurement Specifications

In constructing the ATYP measure as the root mean of squared reconstruction errors, it is possible that the measure places substantial weight on a small number of extreme observations within each cluster. To assess the influence of such observations, we develop alternative formulations of the ATYP measure that reduce sensitivity to outliers and evaluate the robustness of our findings across these specifications. We then replicate our univariate portfolio sorts using the alternative constructions in place of the mean-based ATYP definition in Equation (5).

First, we calculate *MedATYP* as

$$\text{MedATYP}_{i,t} \;=\; \text{median}_j \left[\left(\mathbf{X}_{i,t}^{(j)} - \hat{\mathbf{X}}_{i,t}^{(j)} \right)^2 \right]. \tag{7}$$

By construction, MedATYP is less sensitive to extreme deviations in individual characteristics and instead reflects the typical level of reconstruction error across dimensions. This construction ensures that the measure captures systematic mismatches between the observed and reconstructed firm profiles rather than being disproportionately affected by a single variable.

Second, we construct a rank-based measure, *RankATYP*, which normalizes reconstruction errors dimension by dimension. More specifically, for each month t and characteristic j, we assign firms to deciles based on the squared reconstruction error $\left(\mathbf{X}_{i,t}^{(j)} - \hat{\mathbf{X}}_{i,t}^{(j)} \right)^2$. We then compute each firm's RankATYP score as the average decile rank across all characteristics,

$$\text{RankATYP}_{i,t} \;=\; \frac{1}{p} \sum_{j=1}^{p} \text{RankDecile}_{i,t}^{(j)}. \tag{8}$$

This construction has two advantages: it eliminates the influence of scale differences across characteristics and ensures that no single feature with very large errors dominates the aggregate score. RankATYP thus provides a robust, distribution-free measure of how atypical a firm appears in the cross-section.

Tables IA.3.15 and IA.3.16 in the Internet Appendix show the results for MedATYP and RankATYP, respectively. Both specifications yield highly consistent results with our baseline measure. The decile spreads remain large, negative, and statistically significant, with equal-weighted excess returns of approximately –1.4% per month and significant FF6 alphas. The value-weighted spreads are somewhat weaker but follow the same monotonic pattern. These robustness exercises demonstrate that our findings are not artifacts of the specific aggregation method. Whether ATYP is computed as a mean, median, or average rank across reconstruction errors, the evidence consistently points to the same underlying mechanism.

7.9 Missing Data and Imputation

Section 3.1 explains how we handle missing data. In this section, we examine whether the ATYP measure reflects genuine information structure or mechanical effects arising from missing data and imputation procedures. Firms with less complete information could mechanically exhibit higher reconstruction errors because the autoencoder relies on imputed rather than observed characteristics. To address this concern, we implement two robustness checks designed to isolate any influence of missing or imputed characteristics on the estimated measure.

First, we recalculate the ATYP score using only the non-missing firm characteristics in each observation of $\mathbf{X}_{i,t}$. For every firm-month, we exclude imputed entries from the reconstruction and compute the reconstruction error in Equation (5) based on the number of available characteristics, denoted $p_{i,t}$. This procedure ensures that the measure reflects reconstruction difficulty for the actual information set reported by each firm rather than artifacts of imputation. The resulting portfolio sorts, reported in Table IA.3.17 in the Internet Appendix, show that the return pattern remains largely unchanged. The equal-weighted high-minus-low (H–L) spread is -1.54 percent per month in excess returns ($t = -4.79$) and remains economically and statistically significant at -0.90 percent ($t = -4.89$) under the FF6 model. The value-weighted results are smaller in magnitude (-0.77 percent, $t = -2.31$ in excess returns) but display the same monotonic decline across deciles.

These results demonstrate that the negative relation between ATYP and future returns is not driven by the imputation of missing data.

Second, we regress the original ATYP measure on the share of imputed characteristics for each firm-month and extract the orthogonal residual as a new measure that is purged of any linear association with data sparsity. This residual measure isolates the component of ATYP unrelated to missingness in the underlying features. Portfolio sorts based on this residual measure, presented in Table IA.3.18 in the Internet Appendix, yield qualitatively similar results. The equal-weighted H—L spread remains -1.10 percent per month ($t = -4.24$) and -0.70 percent ($t = -4.06$) under the FF6 specification, while the value-weighted spread is -0.40 percent ($t = -1.49$) in excess returns and directionally consistent with the baseline estimates. The persistence of economically meaningful and statistically significant return differentials after controlling for the share of imputed features confirms that the ATYP measure captures genuine variation in firms' information environments rather than artifacts of data completeness. Overall, both exercises show that the pricing effect of information-processing difficulty is robust to alternative treatments of missing data.

7.10 Impact of Extreme Observations

We argued in Section 2 that a firm may have entirely unremarkable values for each individual characteristic, yet the combination of those values may be so rare compared to other firms that it defies the patterns investors expect. A potential concern with that is that the ATYP measure may reflect the presence of extreme firm characteristics rather than genuine atypical combinations of otherwise regular features. Firms with highly unusual individual characteristics are often more difficult to process, and this difficulty may mechanically increase reconstruction errors even when the joint configuration of characteristics is not itself unusual. To examine this possibility, we test whether the pricing power of ATYP persists after removing the influence of extreme observations in the cross-section of characteristics.

For each month and characteristic, we assign firms to deciles based on their realized values. We then count the number of characteristics for each firm that fall into either the bottom or top decile of their respective cross-sectional distributions. This count represents the degree to which a firm displays extreme realizations of observable characteristics. We then follow the procedure in Section 7.9 and regress the original ATYP measure on this count and extract the residual component. This residual measure isolates the variation in ATYP that is orthogonal to extremeness in the underlying characteristics. By construction, it captures the atypicality that does not arise simply because a firm lies in extreme tails of individual characteristics, but rather reflects the distinctiveness of the joint configuration of these characteristics.

Table IA.3.19 in the Internet Appendix presents univariate portfolio sorts based on this orthogonalized ATYP measure. Panel A displays results for equal-weighted portfolios and Panel B for value-weighted portfolios. The return differentials decline in magnitude compared to the baseline results, which is expected because firms with many extreme characteristics are indeed harder to process and naturally fit into the ATYP measure. Nevertheless, the pattern of returns remains monotonic and economically meaningful. The equal-weighted high minus low spread is -0.81 percent per month in excess returns ($t = -2.53$) and remains significant after controlling for standard factor models. Under the Fama and French six factor specification, the spread is -0.47 percent ($t = -3.67$). The value-weighted results also show a negative and statistically meaningful return spread, though with smaller magnitudes consistent with the baseline findings.

Interestingly, after orthogonalizing ATYP with respect to the share of extreme characteristics, the statistical significance of the value-weighted spreads increases relative to the baseline. This pattern arises because the orthogonalization reduces the influence of very small firms, where both information-processing difficulty and extremeness are concentrated. As a result, variation in the residual ATYP measure becomes more pronounced among larger firms, increasing the ability of value-weighted portfolios to detect the pricing effect. The results indicate that firms with atypical combinations of otherwise moderate characteristics continue to exhibit strong predictive power in the cross-section, consistent with a broader information-processing frictions interpretation.

7.11 International Robustness

To examine whether the pricing of ATYP is a uniquely U.S. phenomenon or a broader feature of global markets, we extend the analysis to developed countries outside the U.S. This test provides an important check on the external validity of our findings, as market structure, disclosure regimes, and investor composition differ across regions. Global stock returns and firm characteristics are obtained from Jensen et al. (2023), and regional factor portfolios are sourced from the Kenneth R. French Data Library. We follow the Library's country classifications for both developed and European countries.[14] We begin our sample in 1990, since many firm characteristics are unavailable prior to that year for several developed markets. All data are processed following the baseline specification in Section 3.1, and we form portfolios and compute alphas exactly as in the U.S. baseline in Section 5.1.

Table IA.3.20 in the Internet Appendix reports results for all developed markets excluding the U.S. Equal-weighted sorts reveal a clear and statistically significant negative relation between ATYP and future returns. The high–low portfolio earns -0.65% per month in excess returns ($t = -1.96$) and remains significant after adjusting for standard factors, with CAPM and FF3 alphas of -0.99% ($t = -4.72$) and -0.88% ($t = -4.33$). Although magnitudes decline under FF5 and FF6, the spreads remain economically meaningful at -0.44% per month. Value-weighted portfolios exhibit the same pattern with smaller coefficients. When Japan is excluded, the ATYP premium strengthens, producing alphas comparable to those observed in the U.S. sample. This pattern aligns with prior evidence that several established anomalies, such as momentum (Asness, 2011) and profitability or investment effects (Fama and French, 2017), are notably weaker in Japan than in other developed markets.

Because Japan's distinct market structure can obscure cross-country patterns, we next examine Europe separately. Table IA.3.21 shows that ATYP remains strongly priced across European equities. The equal-weighted high–low spread averages about -1.21% per month in excess returns ($t = -4.40$) and retains statistical significance across all factor models, with FF6 alphas of roughly -0.55% ($t = -3.22$). Value-weighted portfolios yield similarly robust results. The European evidence thus confirms that ATYP commands a negative return premium across advanced markets, and that the weaker developed-market aggregate primarily reflects Japan's idiosyncratic dynamics rather than a failure of the ATYP measure. Overall, these findings underscore that the pricing of information-processing difficulty is not confined to U.S. data but represents a pervasive global pattern.

[14] See the French Data Library documentation for details on factor and portfolio construction.

8 Conclusion

This paper introduces ATYP, a new statistical measure of firm-level atypicality grounded in the joint structure of firm characteristics. ATYP is constructed as the root mean squared reconstruction error from an autoencoder trained to learn the typical patterns that describe most firms. By capturing the extent to which a firm deviates from these patterns, ATYP provides a continuous and data-driven proxy for information-processing frictions that is distinct from disclosure-based or organizational measures.

Our validation results show that ATYP reflects meaningful informational challenges. High-ATYP firms exhibit greater analyst disagreement, higher return volatility, larger absolute earnings surprises, and lower analyst accuracy—patterns consistent with greater uncertainty and more difficult interpretation. These firms also display slower price discovery, including stronger and more persistent post-earnings-announcement drift and delayed reactions to industry shocks, indicating that atypical information diffuses more slowly into prices.

We document a strong and robust cross-sectional relation between ATYP and future stock returns. A portfolio long in low-ATYP firms and short in high-ATYP firms earns 1.47% per month when equal-weighted and 0.82% when value-weighted. These return differences remain large after controlling for standard risk factors and established return predictors. In multivariate Fama–MacBeth regressions with an extensive set of controls, ATYP continues to load negatively and significantly, and the results hold across data sources, model specifications, and alternative constructions of the measure.

The evidence points to mispricing as the source of the ATYP premium. Atypical firms are harder to process, which increases disagreement and amplifies behavioral biases such as overconfidence. Combined with short-sale constraints, this disagreement pushes prices upward. Limited attention further slows the incorporation of firm-specific signals, while limits to arbitrage restrict the ability of sophisticated investors to correct the resulting mispricing. Consistent with this mechanism, the ATYP premium is concentrated in settings with high limits to arbitrage and low investor attention.

Taken together, these results indicate that ATYP is not merely a background feature of information environments but a systematic driver of prices. ATYP provides a unified way to capture information-processing difficulty in the cross-section, linking theoretical ideas about bounded information processing to empirical patterns in returns. For researchers, it offers a flexible tool for studying how the structure of firm information shapes investor behavior. For investors, it highlights a strong relation between atypicality and expected returns. For policymakers, it emphasizes that informational frictions have tangible consequences for capital allocation.

References

Ali, A., L.-S. Hwang, and M. A. Trombley (2003). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics 69*(2), 355–373.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Asness, C. (2011). Momentum in japan: The exception that proves the rule. *The Journal of Portfolio Management 37*(4), 67–75.

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance 61*(4), 1645–1680.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics 99*(2), 427–446.

Bali, T. G., B. T. Kelly, M. Mörke, and J. Rahman (2025). Machine forecast disagreement. *SSRN*.

Barberis, N. and A. Shleifer (2003). Style investing. *Journal of Financial Economics 68*(2), 161–199.

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *The Journal of Finance 73*(2), 715–754.

Barinov, A. (2020). Firm complexity and conglomerates expected returns. *SSRN*.

Barinov, A., S. Park, and Yıldızhan (2024). Firm complexity and post-earnings announcement drift. *Review of Accounting Studies 29*, 527–579.

Bernard, V. L. and J. K. Thomas (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research 27*, 1–36.

Boehmer, E. and J. J. Wu (2012, 09). Short selling and the price discovery process. *The Review of Financial Studies 26*(2), 287–322.

Brogaard, J. and A. Detzel (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science 61*(1), 3–18.

Bryzgalova, S., M. Pelger, and J. Zhu (2023). Forest through the trees: Building cross-sections of stock returns. Available at SSRN.

Chen, A. Y. and J. McCoy (2024). Missing values handling for machine learning portfolios. *Journal of Financial Economics 155*, 103815.

Chen, A. Y. and T. Zimmermann (2022). Open source cross-sectional asset pricing. *Critical Finance Review 27*(2), 207–264.

Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *Journal of Finance 63*(4), 1977–2011.

Cohen, L. and D. Lou (2012). Complicated firms. *Journal of Financial Economics 104*(2), 383–400.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*(3), 373–394.

Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance 63*(4), 1609–1651.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance 53*(6), 1839–1885.

Daniel, K., D. Hirshleifer, and L. Sun (2019, 06). Short- and long-horizon behavioral factors. *The Review of Financial Studies 33*(4), 1673–1736.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020, 04). The cross-section of risk and returns. *The Review of Financial Studies 33*(5), 1927–1979.

Daniel, K. D., D. Hirshleifer, and A. Subrahmanyam (2001). Overconfidence, arbitrage, and equilibrium asset pricing. *The Journal of Finance 56*(3), 921–965.

Datar, V. T., N. Y. Naik, and R. Radcliffe (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets 1*(2), 203–219.

DellaVigna, S. and J. M. Pollet (2009). Investor inattention and friday earnings announcements. *The Journal of Finance 64*(2), 709–749.

Diether, K. B., C. J. Malloy, and A. Scherbina (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance 57*(5), 2113–2141.

Elgers, P. T., M. H. Lo, and R. J. Pfeiffer (2001). Delayed security price adjustments to financial analysts' forecasts of annual earnings. *The Accounting Review 76*(4), 613–632.

Engelberg, J., R. D. McLean, and J. Pontiff (2018). Anomalies and news. *The Journal of Finance 73*(5), 1971–2001.

Engelberg, J. E. and C. A. Parsons (2011). The causal impact of media in financial markets. *The Journal of Finance 66*(1), 67–97.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*(2), 427–465.

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

Fama, E. F. and K. R. French (2008). Dissecting anomalies. *The Journal of Finance 63*(4), 1653–1678.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and K. R. French (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics 123*(3), 441–463.

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

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

Foster, G., C. Olsen, and T. Shevlin (1984). Earnings releases, anomalies, and the behavior of security returns. *The Accounting Review 59*(4), 574–603.

Gabaix, X. (2014, 09). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics 129*(4), 1661–1710.

García, D. and Norli (2012). Geographic dispersion and stock returns. *Journal of Financial Economics 106*(3), 547–565.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *Review of Financial Studies 33*(5), 2223–2273.

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*(1, Part B), 429–450.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Hinton, G. E. and R. R. Salakhutdinov (2006). Reducing the dimensionality of data with neural networks. *Science 313*(5786), 504–507.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance 64*(5), 2289–2325.

Hirshleifer, D. and S. H. Teoh (2003a). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics 36*(1-3), 337–386.

Hirshleifer, D. and S. H. Teoh (2003b). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics 36*(1), 337–386.

Hoitash, R. and U. Hoitash (2018). Measuring accounting reporting complexity with xbrl. *The Accounting Review 93*(1), pp. 259–287.

Hong, H., T. Lim, and J. C. Stein (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance 55*(1), 265–295.

Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance 54*(6), 2143–2184.

Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *The Review of Financial Studies 20*(4), 1113–1138.

Hou, K., C. Xue, and L. Zhang (2014, 09). Digesting anomalies: An investment approach. *The Review of Financial Studies 28*(3), 650–705.

Hou, K., C. Xue, and L. Zhang (2018, 12). Replicating anomalies. *The Review of Financial Studies 33*(5), 2019–2133.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance 48*(1), 65–91.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023). Is there a replication crisis in finance? *The Journal of Finance 78*(5), 2465–2518.

Jiang, G., C. M. C. Lee, and Y. Zhang (2005). Information uncertainty and expected returns. *Review of Accounting Studies 10*, pp. 185–221.

Johnson, T. C. (2004). Forecast dispersion and the cross section of expected returns. *The Journal of Finance 59*(5), 1957–1978.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A rational theory of mutual funds' attention allocation. *Econometrica 84*(2), 571–626.

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

Lettau, M. and M. Pelger (2020, 03). Factors that fit the time series and cross-section of stock returns. *The Review of Financial Studies 33*(5), 2274–2325.

Li, B. and A. G. Rossi (2021). Selecting mutual funds from the stocks they hold: A machine learning approach. *SSRN*.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics 45*(2), 221–247.

Loughran, T. and B. McDonald (2014). Measuring readability in financial disclosures. *The Journal of Finance 69*(4), 1643–1671.

Loughran, T. and B. McDonald (2024). Measuring firm complexity. *Journal of Financial and Quantitative Analysis 59*(6), 2487–2514.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance 32*(4), 1151–1168.

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

Peng, L. and W. Xiong (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics 80*(3), 563–602.

Peterson, K. (2012). Accounting complexity, misreporting, and the consequences of misreporting. *Review of Accounting Studies 17*, pp. 72–95.

Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics 42*(1), 35–52.

Pástor, and V. Pietro (2003). Stock valuation and learning about profitability. *The Journal of Finance 58*(5), 1749–1789.

Pástor, and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics 69*(1), 99–118.

Stambaugh, R. F., J. Yu, and Y. Yuan (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance 70*(5), 1903–1948.

Stein, J. C. (2009). Presidential address: Sophisticated investors and market efficiency. *The Journal of Finance 64*(4), 1517–1548.

Van Nieuwerburgh, S. and L. Veldkamp (2010, 04). Information acquisition and under-diversification. *The Review of Economic Studies 77*(2), 779–805.

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance 61*(1), 105–137.

Figures

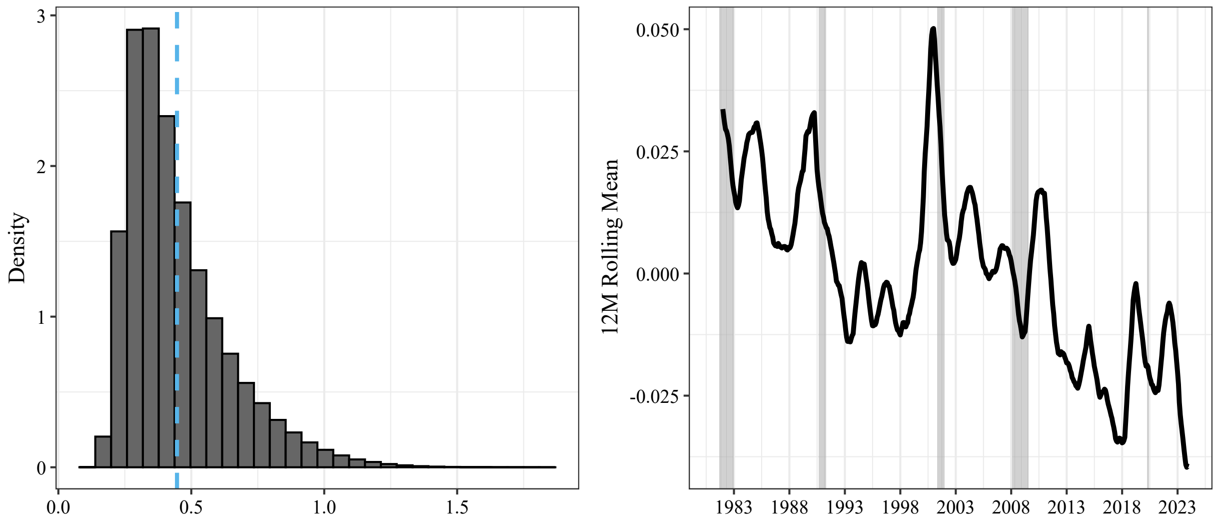

Figure 1: Distribution and Time-Series of ATYP
The left panel shows the cross-sectional distribution of ATYP across all firm-month observations. The y-axis shows density values. The dashed line indicates the mean. The right panel plots the 12-month rolling mean of ATYP, with shaded areas denoting NBER recessions. The sample is from 1981 to 2023.

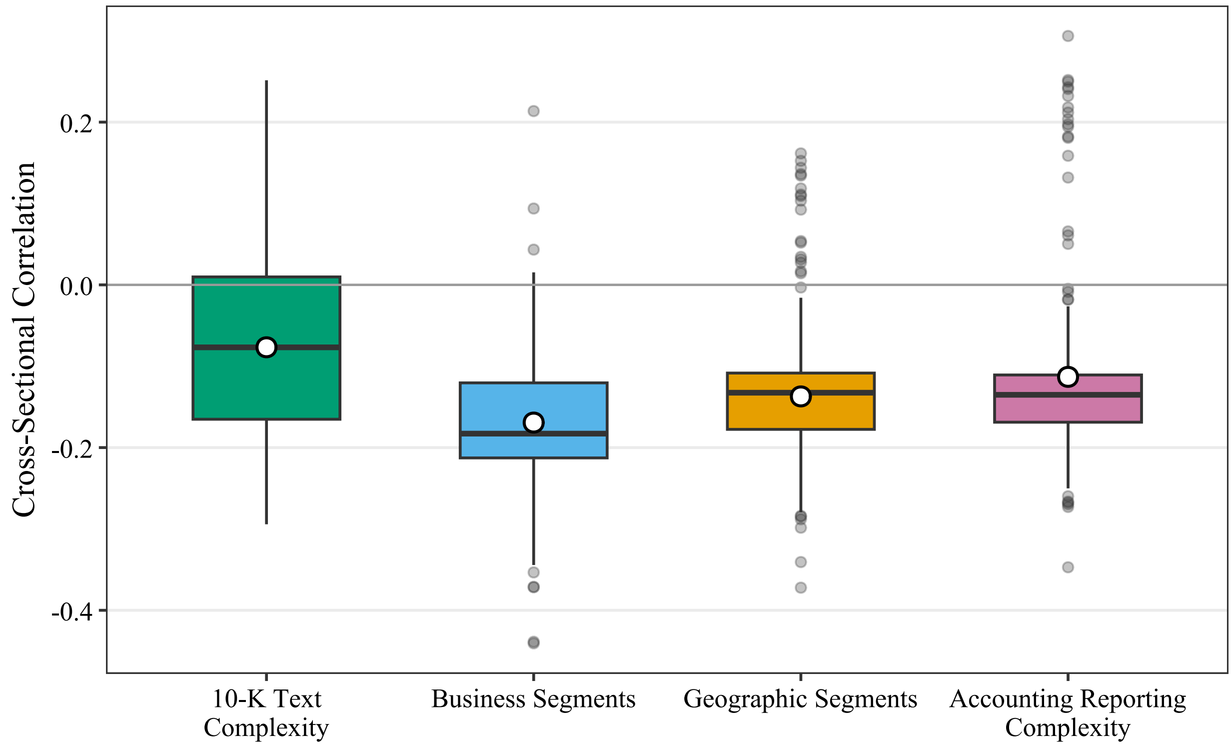

Figure 2: Correlation of ATYP with Established Proxies of Information-Processing Difficulty
The figure plots the distribution of cross-sectional correlations between ATYP and established proxies of information-processing difficulty. We consider four measures: 10-K text complexity, the number of business segments, the number of geographic segments, and accounting reporting complexity. For each month, we compute the Pearson correlation between ATYP and each proxy across firms. The boxplot summarizes the distribution of these correlations over time. The white circle marks the mean correlation for each proxy.

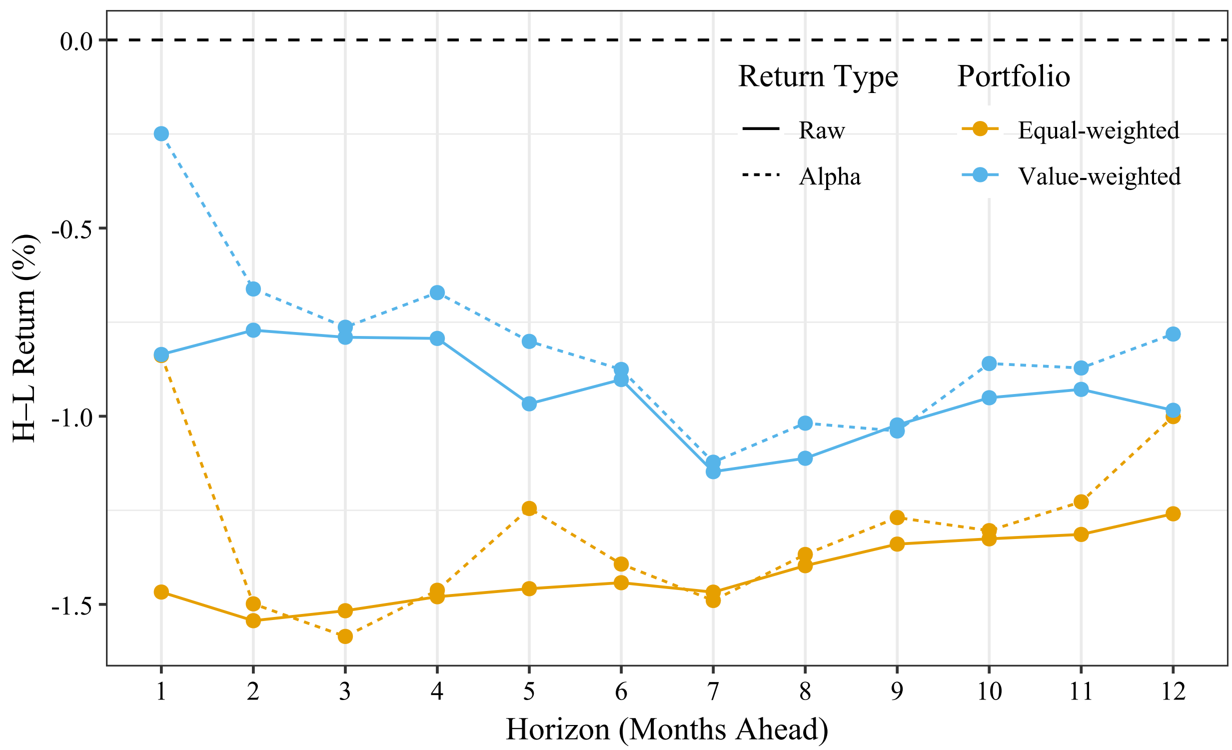

Figure 3: Long-Term Predictive Power
The figure plots average high-minus-low (H–L) returns of ATYP-sorted decile portfolios across different horizons up to 12 months ahead. At each month t, stocks are sorted into ATYP deciles, and the return spread between the highest and lowest deciles is computed for horizons $t+1$ through $t+12$. The figure reports results separately for equal-weighted and value-weighted portfolios, and for both raw returns and alphas. Solid lines represent raw spreads, and dotted lines represent alphas. The yellow line corresponds to equal-weighted portfolios and the blue line corresponds to value-weighted portfolios.

Tables

Table 1: Descriptive Statistics

	Mean	Sd	10th	Q1	Q2	Q3	90th
Ret_{t+1}	0.01	0.16	-0.15	-0.07	0.00	0.07	0.16
BETA	1.21	0.61	0.50	0.86	1.14	1.47	2.02
SIZE	2.69	7.94	0.03	0.09	0.35	1.44	5.58
BM	0.69	0.62	0.15	0.29	0.52	0.88	1.41
STR	0.01	0.14	-0.14	-0.07	0.00	0.08	0.17
MOM	0.14	0.54	-0.40	-0.18	0.05	0.33	0.75
OP	0.10	0.52	-0.27	0.04	0.18	0.29	0.44
AG	0.29	0.85	-0.15	-0.03	0.08	0.25	0.74
TURN	0.64	0.79	0.10	0.22	0.43	0.75	1.31
ILLIQ	2.94	9.39	0.00	0.04	0.18	1.14	6.40
IVOL	0.03	0.02	0.02	0.02	0.03	0.04	0.06
SUE	-0.11	1.61	-1.70	-0.57	-0.01	0.53	1.53
MAX	0.07	0.06	0.03	0.04	0.06	0.09	0.14
AGE	2.39	1.92	0.62	1.00	1.77	3.12	5.11
AFD	-0.13	0.29	-0.24	-0.08	-0.06	-0.04	-0.02
#ANA	5.79	6.46	0.23	1.40	3.58	7.58	14.94

Note: The table reports the descriptive statistics for the cross-sectional variables. Statistics are calculated in each cross-section and displayed as the time-series average. The variables are defined in Section 3.2. The mean, standard deviation (Sd), 10th percentile (10th), first up to third quartil (Q1-Q3), and the 90th percentile (90th) are shown. The sample is from 1981 to 2023.

Table 2: Cross-Sectional Correlations to ATYP

	Mean	Sd	10th	Q1	Q2	Q3	90th
Ret_{t+1}	-0.04	0.09	-0.16	-0.10	-0.04	0.03	0.08
BETA	0.13	0.08	0.04	0.07	0.13	0.17	0.23
SIZE	-0.14	0.03	-0.17	-0.16	-0.15	-0.13	-0.11
BM	0.00	0.09	-0.11	-0.08	-0.03	0.07	0.12
STR	-0.03	0.11	-0.17	-0.10	-0.03	0.03	0.09
MOM	-0.05	0.12	-0.21	-0.13	-0.05	0.02	0.11
OP	-0.37	0.04	-0.41	-0.40	-0.38	-0.35	-0.31
AG	0.28	0.10	0.16	0.24	0.31	0.35	0.38
TURN	0.14	0.12	0.00	0.05	0.12	0.21	0.31
ILLIQ	0.13	0.04	0.08	0.10	0.12	0.15	0.18
IVOL	0.49	0.09	0.37	0.44	0.50	0.56	0.60
SUE	-0.02	0.05	-0.08	-0.06	-0.02	0.01	0.04
MAX	0.32	0.06	0.24	0.28	0.31	0.36	0.39
AGE	-0.25	0.04	-0.30	-0.27	-0.24	-0.22	-0.20
AFD	0.07	0.03	0.03	0.05	0.06	0.09	0.12
#ANA	-0.23	0.06	-0.29	-0.28	-0.25	-0.19	-0.14

Note: The table reports summary statistics on the cross-sectional correlations of multiple variables with ATYP. Correlation is measured using Pearson's correlation coefficient. Statistics are calculated in the time-series of correlations. The variables are defined in Section 3.2. The mean, standard deviation (Sd), 10th percentile (10th), first up to third quartil (Q1-Q3), and the 90th percentile (90th) are shown. The sample is from 1981 to 2023.

Table 3: Explaining Forecast Dispersion, Forecast Accuracy, Return Volatility, and Abs. Earnings Surprise

	Dispersion	Accuracy	Volatility	Abs. SUE
Model:	(1)	(2)	(3)	(4)
Variables				
ATYP	0.087***	-1.49***	0.005***	1.01***
	(0.009)	(0.376)	(0.0002)	(0.058)
Controls	Yes	Yes	Yes	Yes
Fixed-effects				
Month FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Fit statistics				
Observations	1,139,347	1,282,279	1,747,739	1,599,628
R^2	0.24139	0.56263	0.55949	0.07316

Note: The table reports panel regressions of forecast dispersion, forecast accuracy, return volatility, and absolute standardized earnings surprise (SUE) in t+1 on ATYP. Specifications include firm and month fixed effects, with controls. The controls considered are defined in Section 3.2 excluding AFD and SUE. Standard errors are double-clustered by firm and month. *, **, and *** denote significance at the 10%, 5%, and 1% levels. The sample is from 1981 to 2023.

Table 4: ATYP and Post-Earnings-Announcement Drift

	Low	2	3	4	High	H-L	t-stat	FF6	t-stat
Panel A: Bivariate Sort on ATYP — $t+1$									
Low	0.77	0.92	0.90	1.03	1.04	0.27***	3.23	0.24***	2.99
2	0.71	0.88	0.77	1.11	1.14	0.42***	5.44	0.34***	4.33
3	0.50	0.85	0.55	0.96	1.07	0.57***	4.77	0.38***	4.05
4	0.19	0.57	0.16	0.86	0.88	0.69***	4.90	0.47***	3.89
High	-0.43	-0.18	-0.71	-0.16	0.46	0.89***	6.30	0.71***	5.06
H-L	-1.20	-1.10	-1.62	-1.19	-0.58	0.62***	4.36	0.47***	3.27
Panel B: Bivariate Sort on ATYP — $t+2$									
Low	0.88	0.95	0.97	1.01	1.06	0.18**	2.31	0.17**	1.99
2	0.71	0.89	0.76	0.97	1.10	0.39***	4.37	0.40***	3.87
3	0.60	0.81	0.58	1.04	1.05	0.45***	4.47	0.44***	3.90
4	0.26	0.61	0.17	0.72	0.92	0.66***	5.22	0.64***	5.03
High	-0.42	-0.06	-0.75	-0.16	0.36	0.79***	5.55	0.74***	4.80
H-L	-1.31	-1.02	-1.72	-1.17	-0.69	0.61***	4.27	0.57***	3.70
Panel C: Bivariate Sort on ATYP — $t+3$									
Low	0.88	1.00	0.93	1.01	1.03	0.15*	1.73	0.14	1.44
2	0.73	0.86	0.81	0.99	1.02	0.29***	3.64	0.30***	3.33
3	0.62	0.86	0.44	0.94	1.00	0.38***	3.49	0.35***	2.83
4	0.33	0.66	0.13	0.62	0.84	0.51***	4.20	0.36***	3.04
High	-0.36	-0.00	-0.83	-0.24	0.32	0.67***	4.56	0.59***	3.84
H-L	-1.24	-1.01	-1.76	-1.25	-0.71	0.53***	3.41	0.44***	2.68

Note: The table reports bivariate portfolio sorts examining the relation between ATYP and earnings surprises, linking post-earnings-announcement drift to information-processing difficulty. Panels A, B, and C show dependent double sorts, where firms are first partitioned into quintiles by ATYP and subsequently into quintiles by earnings surprises within each quintile. Panel A shows returns in month $t+1$, Panel B reports cumulative returns over months $t+1$ to $t+2$, and Panel C extends to months $t+1$ to $t+3$. Reported are high-minus-low return spreads (H–L), Newey and West (1987) adjusted t-statistics in parentheses, and Fama and French (2018) six-factor (FF6) adjusted alphas. *, **, and *** denote significance at the 10%, 5%, and 1% levels. The sample is from 1981 to 2023.

Table 5: Univariate Portfolio Sorts of High-ATYP Firms on Industry Information Shocks

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	-0.91**	(-2.07)	-1.84***	(-6.54)	-1.87***	(-7.35)	-1.75***	(-6.64)	-1.54***	(-6.50)
2	-0.62	(-1.32)	-1.57***	(-5.26)	-1.49***	(-6.06)	-1.14***	(-4.08)	-0.94***	(-3.58)
3	-0.30	(-0.71)	-1.29***	(-4.98)	-1.15***	(-5.40)	-0.84***	(-3.52)	-0.67***	(-3.17)
4	0.03	(0.06)	-0.95***	(-3.18)	-0.83***	(-3.37)	-0.38	(-1.57)	-0.23	(-0.93)
5	-0.07	(-0.15)	-1.01***	(-3.27)	-0.90***	(-3.99)	-0.64***	(-2.99)	-0.50**	(-2.42)
6	-0.18	(-0.41)	-1.11***	(-3.78)	-0.90***	(-4.04)	-0.50**	(-2.16)	-0.40	(-1.62)
7	-0.33	(-0.74)	-1.26***	(-4.41)	-1.14***	(-5.31)	-0.86***	(-4.12)	-0.69***	(-3.40)
8	0.17	(0.41)	-0.76***	(-2.71)	-0.63***	(-2.70)	-0.33	(-1.55)	-0.21	(-0.94)
9	0.54	(1.21)	-0.39	(-1.41)	-0.26	(-1.14)	0.01	(0.03)	0.12	(0.55)
10	0.62	(1.41)	-0.28	(-0.95)	-0.25	(-1.11)	-0.06	(-0.26)	0.08	(0.33)
H-L	1.53***	(5.09)	1.56***	(5.24)	1.62***	(5.19)	1.69***	(5.09)	1.62***	(5.11)
Panel B: Value-Weighted										
1	-0.08	(-0.20)	-1.05***	(-3.83)	-1.02***	(-3.54)	-0.96***	(-3.22)	-0.88***	(-3.13)
2	-0.00	(-0.01)	-1.03***	(-3.96)	-0.90***	(-3.50)	-0.61**	(-2.22)	-0.54**	(-2.03)
3	0.35	(0.83)	-0.71**	(-2.46)	-0.60**	(-2.04)	-0.29	(-0.95)	-0.13	(-0.45)
4	0.42	(1.07)	-0.51*	(-1.75)	-0.37	(-1.28)	0.03	(0.10)	-0.01	(-0.05)
5	0.28	(0.63)	-0.76**	(-2.35)	-0.54**	(-2.08)	-0.23	(-1.07)	-0.16	(-0.80)
6	0.40	(1.07)	-0.54**	(-2.06)	-0.35	(-1.53)	-0.16	(-0.66)	-0.17	(-0.70)
7	0.04	(0.09)	-0.93***	(-2.82)	-0.74**	(-2.51)	-0.42	(-1.59)	-0.33	(-1.23)
8	0.51	(1.38)	-0.45	(-1.61)	-0.37	(-1.34)	-0.19	(-0.72)	-0.07	(-0.25)
9	0.96**	(2.41)	0.05	(0.18)	0.27	(1.07)	0.62**	(2.36)	0.59**	(2.23)
10	0.87**	(2.34)	0.04	(0.16)	0.14	(0.52)	0.19	(0.69)	0.22	(0.74)
H-L	0.95**	(2.39)	1.09***	(2.68)	1.15***	(2.65)	1.16**	(2.50)	1.10**	(2.47)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts of high-ATYP firms based on industry information shocks following Cohen and Lou (2012). At the end of each month t, high-ATYP stocks are sorted into deciles according to returns of their corresponding low-ATYP industry peers in the previous month. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table 6: Univariate Portfolio Sorts on ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.94***	(4.25)	0.25*	(1.94)	0.18***	(2.64)	0.02	(0.38)	0.08	(1.52)
2	0.93***	(3.96)	0.21*	(1.67)	0.18***	(3.06)	0.07	(1.27)	0.12**	(2.29)
3	0.94***	(3.77)	0.19	(1.52)	0.18***	(3.10)	0.13**	(2.11)	0.19***	(3.44)
4	0.91***	(3.45)	0.14	(1.05)	0.15**	(2.44)	0.15**	(2.36)	0.22***	(3.74)
5	0.79***	(2.77)	-0.01	(-0.06)	0.02	(0.27)	0.06	(0.84)	0.15**	(2.02)
6	0.78***	(2.59)	-0.06	(-0.38)	-0.02	(-0.24)	0.09	(1.04)	0.20**	(2.40)
7	0.65*	(1.96)	-0.20	(-1.20)	-0.14	(-1.43)	-0.01	(-0.10)	0.10	(1.19)
8	0.41	(1.16)	-0.48**	(-2.51)	-0.40***	(-3.29)	-0.19*	(-1.73)	-0.06	(-0.54)
9	0.12	(0.31)	-0.82***	(-3.76)	-0.68***	(-4.91)	-0.33***	(-2.75)	-0.20	(-1.61)
10	-0.53	(-1.17)	-1.53***	(-5.47)	-1.35***	(-6.82)	-0.91***	(-5.21)	-0.76***	(-4.36)
H-L	-1.47***	(-4.54)	-1.78***	(-5.99)	-1.54***	(-6.50)	-0.93***	(-5.17)	-0.84***	(-4.52)
Panel B: Value-Weighted										
1	0.80***	(4.35)	0.15	(1.53)	0.10	(1.19)	-0.09	(-1.18)	-0.05	(-0.67)
2	0.66***	(3.65)	0.02	(0.23)	0.00	(0.03)	-0.15**	(-2.29)	-0.14**	(-2.09)
3	0.72***	(3.64)	0.06	(0.99)	0.07	(1.13)	-0.04	(-0.67)	-0.04	(-0.61)
4	0.77***	(3.77)	0.09	(1.20)	0.12*	(1.70)	0.06	(0.79)	0.06	(0.77)
5	0.72***	(3.16)	-0.00	(-0.02)	0.02	(0.28)	0.03	(0.42)	0.04	(0.54)
6	0.79***	(3.31)	0.02	(0.22)	0.08	(1.03)	0.15*	(1.81)	0.18**	(2.03)
7	0.85***	(3.21)	0.06	(0.57)	0.14	(1.56)	0.23***	(2.62)	0.26***	(2.78)
8	0.48*	(1.79)	-0.36***	(-3.04)	-0.25**	(-2.39)	-0.11	(-1.06)	-0.09	(-0.92)
9	0.38	(1.17)	-0.51***	(-2.73)	-0.37**	(-2.34)	-0.08	(-0.52)	-0.05	(-0.35)
10	-0.02	(-0.04)	-1.06***	(-4.21)	-0.82***	(-3.91)	-0.35**	(-2.00)	-0.33*	(-1.92)
H-L	-0.82**	(-2.50)	-1.21***	(-3.91)	-0.92***	(-3.65)	-0.26	(-1.38)	-0.28	(-1.49)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table 7: Bivariate Portfolio Sorts

	Equal-Weighted		Value-Weighted	
	H-L	FF6	H-L	FF6
BETA	-1.28***	-0.78***	-0.63***	-0.23
	(-4.68)	(-4.54)	(-2.67)	(-1.32)
SIZE	-1.31***	-0.79***	-1.28***	-0.76***
	(-4.71)	(-5.40)	(-4.53)	(-5.15)
BM	-1.34***	-0.72***	-0.85***	-0.28*
	(-4.49)	(-3.97)	(-2.87)	(-1.74)
STR	-1.42***	-0.89***	-0.62**	-0.16
	(-4.81)	(-4.96)	(-2.14)	(-0.89)
MOM	-1.21***	-0.74***	-0.89***	-0.47***
	(-4.40)	(-4.53)	(-3.28)	(-2.99)
OP	-1.05***	-0.77***	-0.57**	-0.22
	(-4.83)	(-4.84)	(-2.42)	(-1.32)
AG	-1.17***	-0.69***	-0.48**	-0.13
	(-4.71)	(-4.35)	(-2.32)	(-0.93)
TURN	-1.32***	-0.72***	-0.72***	-0.24
	(-4.65)	(-4.24)	(-2.93)	(-1.59)
ILLIQ	-1.33***	-0.81***	-1.18***	-0.74***
	(-4.60)	(-5.15)	(-4.38)	(-5.08)
IVOL	-0.96***	-0.67***	-0.70***	-0.39**
	(-6.30)	(-5.91)	(-3.09)	(-2.43)
SUE	-1.33***	-0.73***	-0.81***	-0.33*
	(-4.28)	(-3.90)	(-2.66)	(-1.80)
MAX	-1.00***	-0.57***	-0.80***	-0.38**
	(-4.26)	(-3.63)	(-3.21)	(-2.21)
AGE	-1.26***	-0.81***	-0.66***	-0.25
	(-4.71)	(-4.88)	(-2.66)	(-1.50)
AFD	-1.26***	-0.70***	-0.89***	-0.37**
	(-4.43)	(-4.39)	(-3.07)	(-2.11)
#ANA	-1.22***	-0.65***	-1.03***	-0.59***
	(-4.02)	(-3.98)	(-3.57)	(-3.22)

Note: The table reports average monthly return spreads (H–L) and Fama and French (2018) six-factor (FF6) alphas from bivariate portfolio sorts on ATYP and standard firm characteristics. In each month, stocks are independently sorted into quintiles by a given control variable and, within each quintile, further sorted into deciles by ATYP. Reported values are the average return differences between the highest and lowest ATYP deciles, averaged across the control quintiles. Panel columns display results for equal-weighted and value-weighted portfolios. H–L denotes the excess return spread, while FF6 reports the intercept from regressions of portfolio excess returns on the six-factor model. Newey and West (1987) t-statistics with six lags are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table 8: Fama-MacBeth Cross-Sectional Regressions

	Excess Return	Excess Return	Excess Return	Excess Return	Excess Return
(Intercept)	1.63***	0.87***	0.60***	1.04***	0.91***
	(7.57)	(4.08)	(2.85)	(4.58)	(3.93)
ATYP	-2.33***	-2.19***	-1.43***	-0.81***	-0.72***
	(-4.55)	(-5.02)	(-4.04)	(-3.51)	(-3.21)
BETA		0.10	0.09	0.22**	0.20**
		(0.94)	(0.88)	(2.29)	(2.18)
SIZE		0.00	-0.00	-0.00*	-0.00***
		(0.07)	(-0.42)	(-1.70)	(-3.35)
BM		0.63***	0.57***	0.41***	0.43***
		(5.54)	(5.25)	(4.23)	(4.62)
MOM		0.97***	0.91***	0.82***	0.83***
		(6.58)	(6.33)	(5.58)	(5.74)
AG			-0.35***	-0.33***	-0.34***
			(-6.09)	(-6.46)	(-6.67)
OP			0.53***	0.45***	0.44***
			(5.13)	(4.95)	(4.99)
STR				-2.09***	-2.13***
				(-5.73)	(-5.86)
TURN				-44.92***	-55.19***
				(-3.44)	(-4.10)
ILLIQ				0.13*	0.12*
				(1.96)	(1.95)
IVOL				-11.94***	-9.97**
				(-2.71)	(-2.38)
SUE				0.07***	0.07***
				(5.92)	(5.79)
MAX				-4.28***	-4.16***
				(-6.14)	(-6.16)
AGE					-0.00
					(-0.61)
AFD					-0.20**
					(-2.41)
#ANA					0.02***
					(3.88)

Note: The table reports the results of monthly Fama and MacBeth (1973) regressions of excess stock returns on ATYP and a broad set of firm characteristics. ATYP is included either individually or jointly with other predictors. The controls considered are defined in Section 3.2. Reported coefficients represent average slopes across months, with Newey and West (1987) adjusted t-statistics (six lags) shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table 9: Mispricing and ATYP

	Low	2	3	4	High	H-L	t-stat	FF6	tstat
Panel A: Average Mispricing									
	0.42	0.44	0.47	0.51	0.59	0.17***	47.54		
Panel B: Bivariate Sort on Mispricing									
Low	1.12	1.26	1.36	1.50	1.55	0.43***	2.64	0.56***	4.29
2	1.02	1.06	1.15	1.18	1.10	0.08	0.45	0.30**	2.39
3	0.82	0.92	0.97	0.88	0.67	-0.14	-0.65	0.23	1.50
4	0.55	0.52	0.45	0.35	0.18	-0.37	-1.42	0.01	0.08
High	-0.24	-0.48	-0.67	-0.87	-1.49	-1.25***	-4.74	-0.88***	-3.98
H-L	-1.36	-1.74	-2.03	-2.37	-3.04	-1.68***	-8.33	-1.44***	-7.60

Note: The table reports univariate and bivariate portfolio sorts analyzing the interaction between ATYP and the mispricing score of Stambaugh et al. (2015). Panel A displays average values of the mispricing score across ATYP quintiles. Panel B shows dependent double sorts in which firms are first grouped into quintiles based on the mispricing score and then sorted into quintiles by ATYP within each quintile. Reported are average excess returns, high-minus-low (H–L) spreads, associated t-statistics, and Fama and French (2018) six-factor (FF6) alphas. Newey and West (1987) adjusted t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table 10: Limits to Arbitrage and ATYP

	Low	2	3	4	High	H-L	t-stat	FF6	tstat
Panel A: Bivariate Sort on Idiosyncratic Volatility									
Low	0.91	0.90	0.89	0.87	0.88	-0.03	-0.41	0.13*	1.90
2	0.97	0.99	0.91	0.83	0.70	-0.27***	-2.76	-0.15	-1.60
3	0.98	0.87	0.75	0.48	-0.02	-1.00***	-5.70	-0.71***	-5.45
4	0.88	0.87	0.82	0.36	-0.06	-0.93***	-5.07	-0.63***	-4.88
High	0.72	0.50	0.08	-0.18	-1.06	-1.78***	-7.59	-1.51***	-6.77
H-L	-0.19	-0.40	-0.81	-1.05	-1.95	-1.76***	-7.56	-1.64***	-7.30
Panel B: Bivariate Sort on Illiquidity									
Low	0.83	0.83	0.78	0.71	0.45	-0.38*	-1.77	0.07	0.73
2	0.96	0.96	0.82	0.54	0.05	-0.91***	-3.56	-0.40***	-2.69
3	0.97	0.80	0.70	0.27	-0.28	-1.26***	-4.55	-0.66***	-3.63
4	1.05	0.96	0.65	0.33	-0.34	-1.39***	-4.66	-0.91***	-4.62
High	1.01	0.96	0.76	0.48	-0.41	-1.41***	-5.64	-1.22***	-6.18
H-L	0.17	0.13	-0.02	-0.22	-0.86	-1.03***	-4.81	-1.29***	-5.89
Panel C: Bivariate Sort on Firm Age									
Low	0.81	0.61	0.20	-0.18	-0.91	-1.73***	-6.20	-1.18***	-5.50
2	0.83	0.76	0.53	0.34	-0.36	-1.20***	-4.30	-0.67***	-3.50
3	1.05	1.02	0.94	0.72	0.17	-0.88***	-3.26	-0.41**	-2.49
4	0.97	1.03	0.91	0.84	0.30	-0.67***	-3.17	-0.31**	-2.22
High	0.91	0.88	0.88	0.91	0.68	-0.22	-1.26	-0.06	-0.48
H-L	0.09	0.27	0.68	1.10	1.60	1.51***	6.45	1.12***	5.23

Note: The table presents bivariate portfolio sorts examining the relation between ATYP and proxies for limits to arbitrage. Panels A, B and C show dependent double sorts in which firms are first partitioned into quintiles based on each limits to arbitrage proxy and then sorted into quintiles by ATYP within each quintile. Reported are average excess returns, high-minus-low (H–L) spreads, corresponding t-statistics, and Fama and French (2018) six-factor (FF6) alphas. Newey and West (1987) adjusted t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table 11: Attention Constraints and ATYP

	Low	2	3	4	High	H-L	t-stat	FF6	tstat
Panel A: Bivariate Sort on Size									
Low	1.13	0.87	0.71	0.28	-0.60	-1.73***	-6.43	-1.54***	-6.90
2	1.07	0.84	0.70	0.24	-0.39	-1.46***	-5.31	-0.94***	-5.00
3	1.00	0.94	0.72	0.38	-0.14	-1.13***	-4.10	-0.60***	-3.77
4	0.95	0.90	0.85	0.61	0.13	-0.82***	-3.41	-0.35***	-2.99
High	0.80	0.80	0.77	0.71	0.55	-0.26	-1.17	0.20*	1.96
H-L	-0.33	-0.07	0.06	0.42	1.15	1.48***	6.28	1.74***	7.90
Panel B: Bivariate Sort on #Analyst									
Low	0.92	1.04	0.69	0.28	-0.36	-1.29***	-4.62	-1.02***	-4.96
2	0.99	0.87	0.69	0.12	-0.67	-1.66***	-5.26	-1.20***	-5.81
3	1.00	0.88	0.84	0.53	-0.00	-1.00***	-3.75	-0.47***	-2.71
4	0.91	0.91	0.70	0.53	0.13	-0.78***	-2.98	-0.20	-1.25
High	0.86	0.82	0.88	0.72	0.56	-0.30	-1.23	0.27**	2.35
H-L	-0.06	-0.22	0.19	0.44	0.92	0.99***	4.25	1.29***	6.31
Panel C: Bivariate Sort on Short-interest									
Low	0.94	0.74	0.54	0.20	-0.63	-1.58***	-5.95	-1.39***	-6.41
2	0.95	1.03	0.97	0.47	0.01	-0.94***	-3.43	-0.50***	-2.81
3	1.07	0.98	0.93	0.59	0.12	-0.95***	-3.63	-0.50***	-3.23
4	0.88	0.77	0.55	0.38	-0.13	-1.01***	-3.66	-0.51***	-2.86
High	0.86	0.85	0.82	0.69	0.25	-0.60**	-2.46	-0.20	-1.47
H-L	-0.09	0.11	0.28	0.48	0.89	0.97***	4.09	1.19***	4.87

Note: The table presents bivariate portfolio sorts examining the relation between ATYP and proxies for investor attention constraints. Panels A, B and C show dependent double sorts in which firms are first partitioned into quintiles based on each attention proxy and then sorted into quintiles by ATYP within each quintile. Reported are average excess returns, high-minus-low (H–L) spreads, corresponding t-statistics, and Fama and French (2018) six-factor (FF6) alphas. Newey and West (1987) adjusted t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Internet Appendix for "Hard to Process: Atypical Firms and the Cross-Section of Expected Stock Returns"

Appendix IA.1 Autoencoder Fit

The bottleneck determines how much information the autoencoder compresses when capturing joint patterns in firm characteristics. A very small bottleneck underfits and produces large reconstruction errors, while a very large one overfits and begins to reproduce noise. The model must therefore balance flexibility and parsimony to avoid both under- and overfitting. To illustrate this tradeoff, we compare models with four, sixteen, and sixty four bottleneck dimensions. To make a more accurate comparison, we train the models without regularization and with only one hidden layer of 128 units. Figure IA.1.1 displays the distribution of ATYP across specifications. The dashed vertical lines mark the 10th and 90th percentile values of each distribution. The four dimensional model fails to capture the heterogeneity in the data and generates broad, right skewed errors, whereas the sixty four dimensional model nearly memorizes the data and yields errors close to zero. The intermediate model with sixteen dimensions achieves the most balanced performance. Its reconstruction errors fall in a moderate and symmetric range, indicating that it captures the essential nonlinear patterns without absorbing noise.[15] This specification provides a practical and economically meaningful basis for the ATYP measure.

For our measure to be valid, it is crucial to verify that the network learns the latent economic structure rather than simply memorizing the inputs. Figure IA.1.2 plots the pairwise correlations of the raw characteristics against the correlations of the corresponding residuals. In the raw data, characteristics exhibit strong systematic comovements that reflect standard economic forces. However, the figure shows that the residual correlations collapse toward zero and the fitted slope is nearly flat. This transformation demonstrates that the network effectively filters out the common variation present in the original data. The result confirms that the residual space contains almost no systematic structure. Consequently, ATYP isolates idiosyncratic firm specific departures from the typical patterns the network has learned. This distinction ensures that our measure captures unique processing difficulty rather than exposure to common firm characteristics.

We also examine which inputs are harder to reconstruct and which inputs most influence the bottleneck, i.e., the latent factors. At the feature level we rank characteristics by mean squared reconstruction error (MSE) and by importance in changing the latent features, defined as the mean

[15] Standardized magnitudes of about 0.2, 0.5, and 0.8 are commonly interpreted as small, moderate, and large differences in statistics (Cohen, 1988).

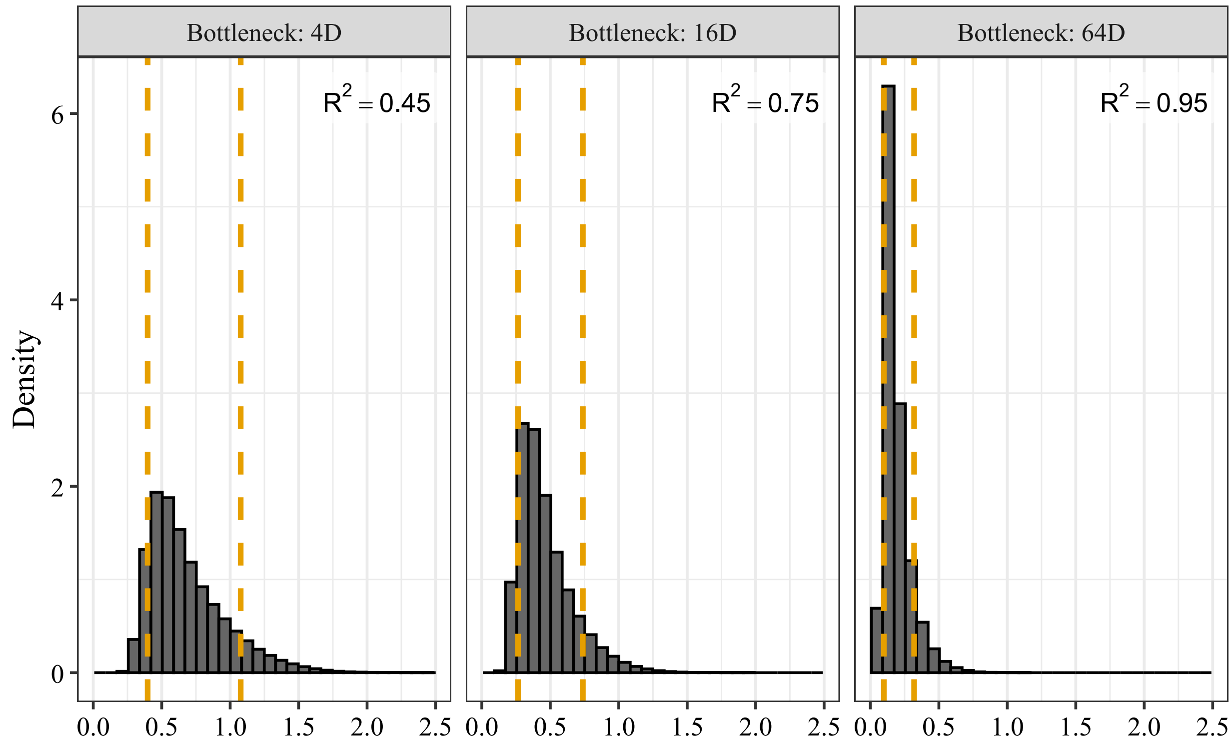

Figure IA.1.1: Distribution of ATYP across Bottleneck Dimensions
The figure plots the density of ATYP for autoencoders with four, 16, and 64 latent factors. Dashed vertical lines mark the 10th and 90th percentiles. The R^2 is the measure of overall reconstruction fit.

absolute gradient of the bottleneck with respect to each input. The most informative view is at the cluster level. Figure IA.1.3 summarizes mean squared error and importance by cluster. *Investment*, *Profit Growth*, *Value* and *Quality* are all relatively hard to reconstruct and highly influential for the latent representation, while *Size* and *Accruals* are easier to reconstruct and less influential. We observe that unconditional characteristics, most notably *Size*, exhibit the lowest reconstruction errors and bottleneck importance. This indicates that market capitalization acts as a primary, stable sorting category that follows linear, predictable patterns easily captured by the model. In sharp contrast, the model exhibits the greatest difficulty in reconstructing conditional characteristics, such as *Investment*, *Profit Growth* and *Value*. This distinction is economically fundamental, because the valuation implication of a variable like asset growth is inherently context-dependent, i.e. high growth may imply value creation for a highly profitable firm but financial distress or agency costs for a firm with poor cash flows. That the autoencoder assigns the highest importance and error to these features confirms that ATYP successfully isolates firms with complex, nonlinear economic

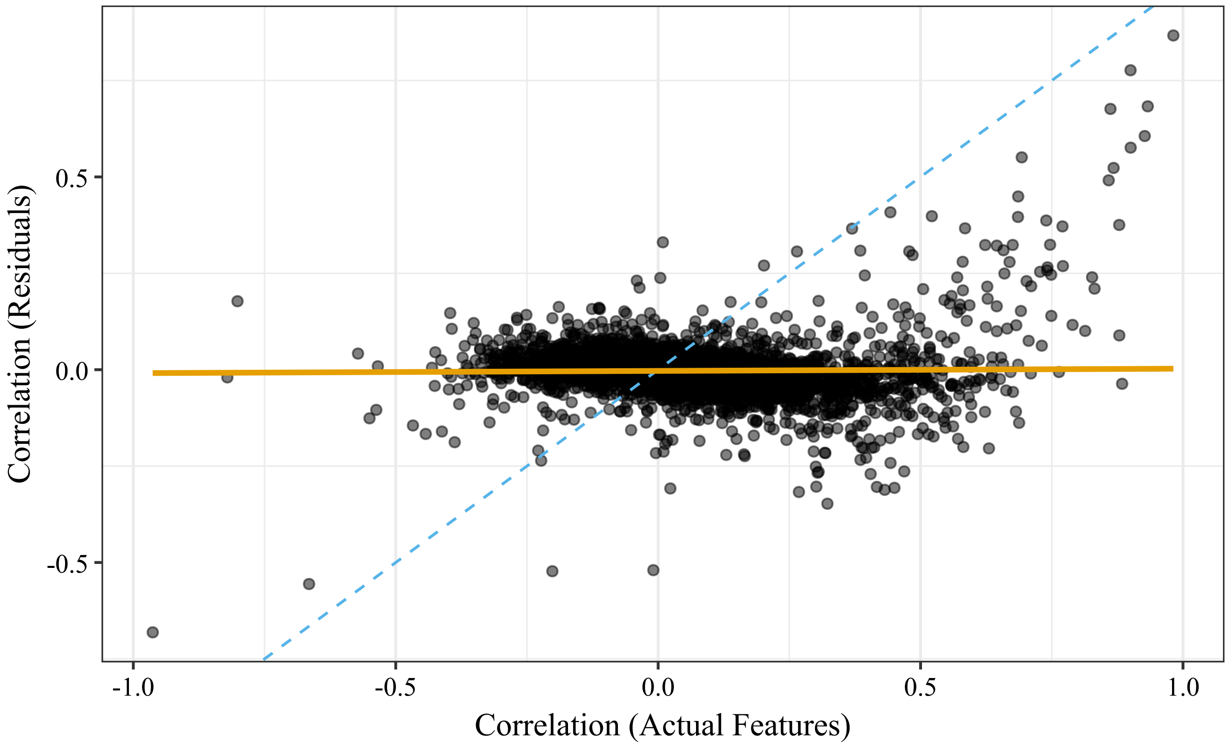

Figure IA.1.2: Pairwise correlations of raw characteristics vs. residuals
Pairwise correlations of raw characteristics versus correlations of residuals, i.e., the reconstructed characteristics. Correlation is measured using Pearson's correlation coefficient. The 45-degree dashed line indicates equality; the fitted line is the trend. The sample is from 1981 to 2023.

profiles, rather than simply identifying small or obscure firms.

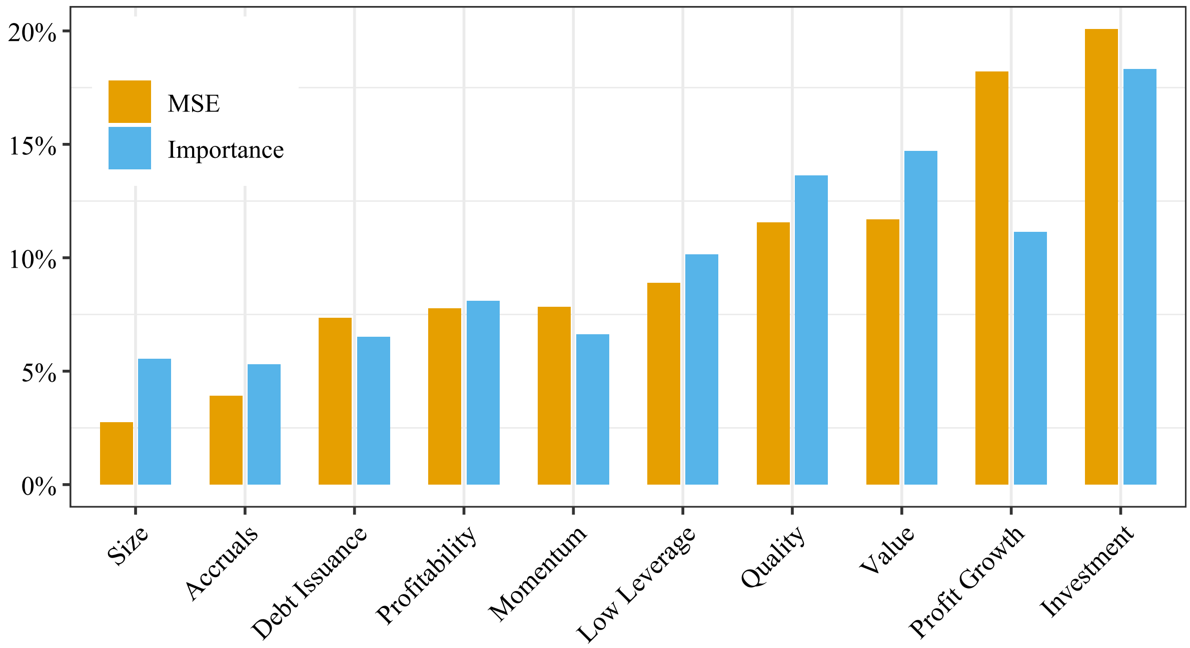

Figure IA.1.3: MSE and bottleneck importance by cluster
The figure summarizes mean squared error and importance by cluster. Mean squared error is calculated for the residuals, i.e., the reconstructed characteristics compared to the actual characteristics. The importance is measured as the mean absolute gradient of the characteristics w.r.t. the bottleneck layer, i.e., the output of the encoder in Equation (1). The clusters are taken from Jensen et al. (2023).

Appendix IA.2 Supplementary Figures

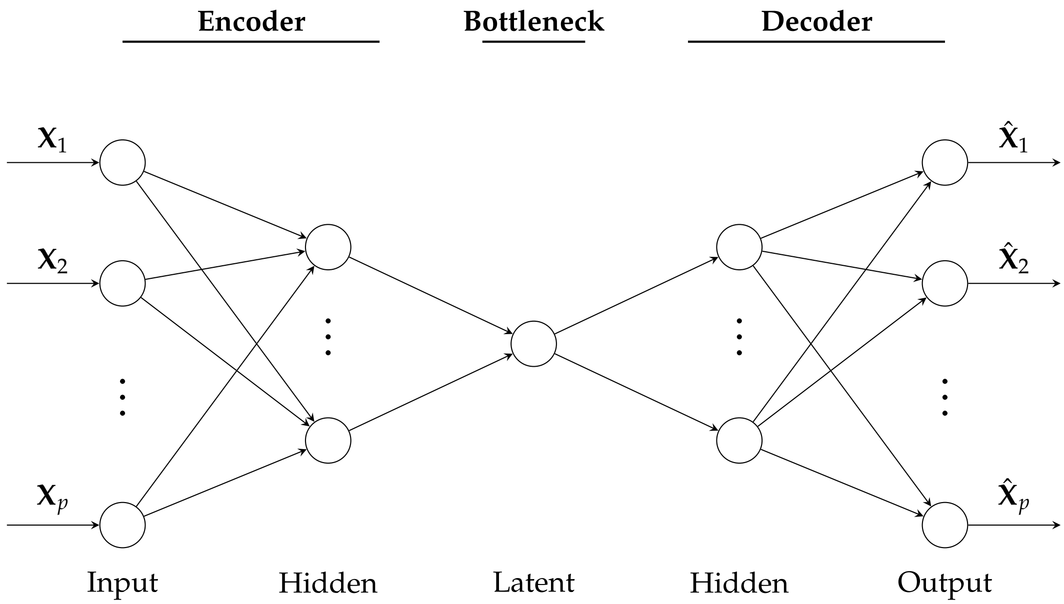

Figure IA.2.1: Graphical Illustration of an Autoencoder
This figure illustrates an autoencoder model graphically. The encoder compresses the inputs **X** into a small latent bottleneck space in multiple nonlinear hidden layers. The decoder then attempts to reconstruct the original inputs, **X**, from the latent space by predicting $\hat{\mathbf{X}}$.

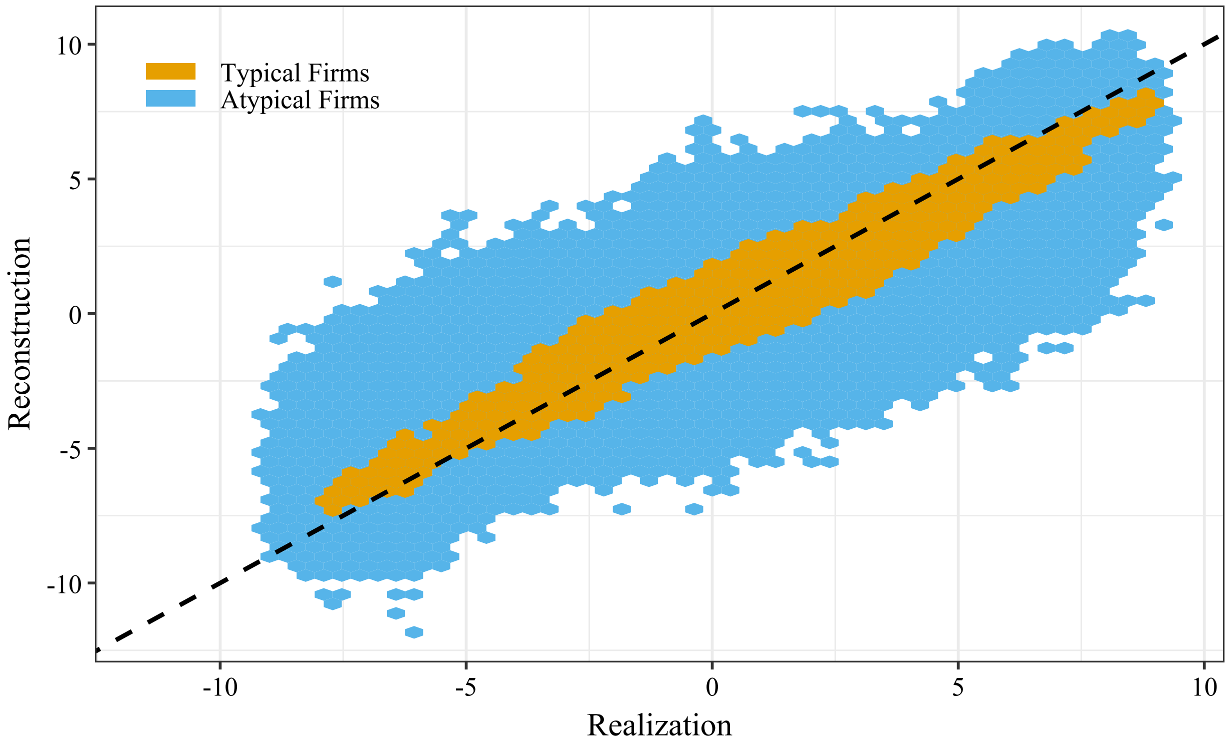

Figure IA.2.2: Reconstruction Accuracy and Atypical Firms
The figure plots realized firm characteristics on the horizontal axis and their reconstructed counterparts on the vertical axis. Each point represents a firm–characteristic observation. Firms that the autoencoder reconstructs accurately (lowest ATYP decile) are shown in orange and labeled as typical firms, while those with large reconstruction errors (highest ATYP decile) are shown in blue and labeled as atypical firms. The 45-degree dashed line indicates perfect reconstruction.

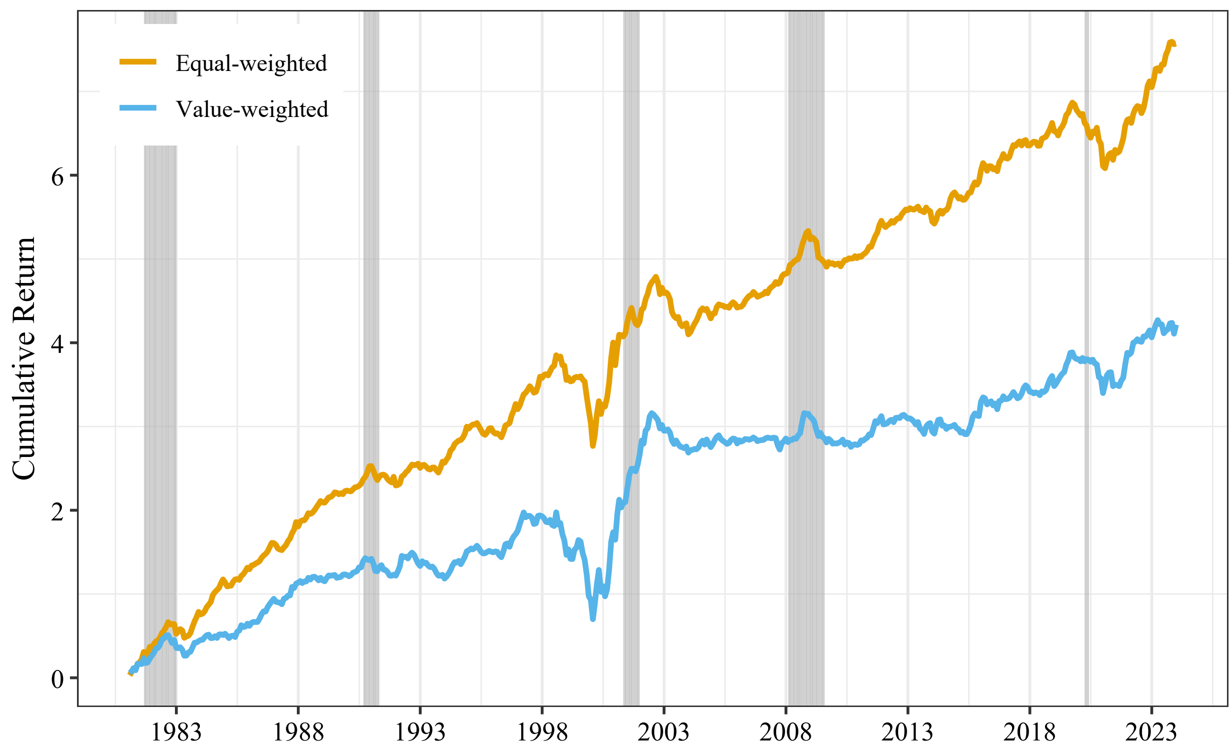

Figure IA.2.3: Cumulative performance over time for H–L ATYP portfolio sorts
The plot shows the cumulative sum of the equal- and value-weighted decile spread (H–L) portfolios' returns over time. Shaded areas represent NBER-dated recessions.

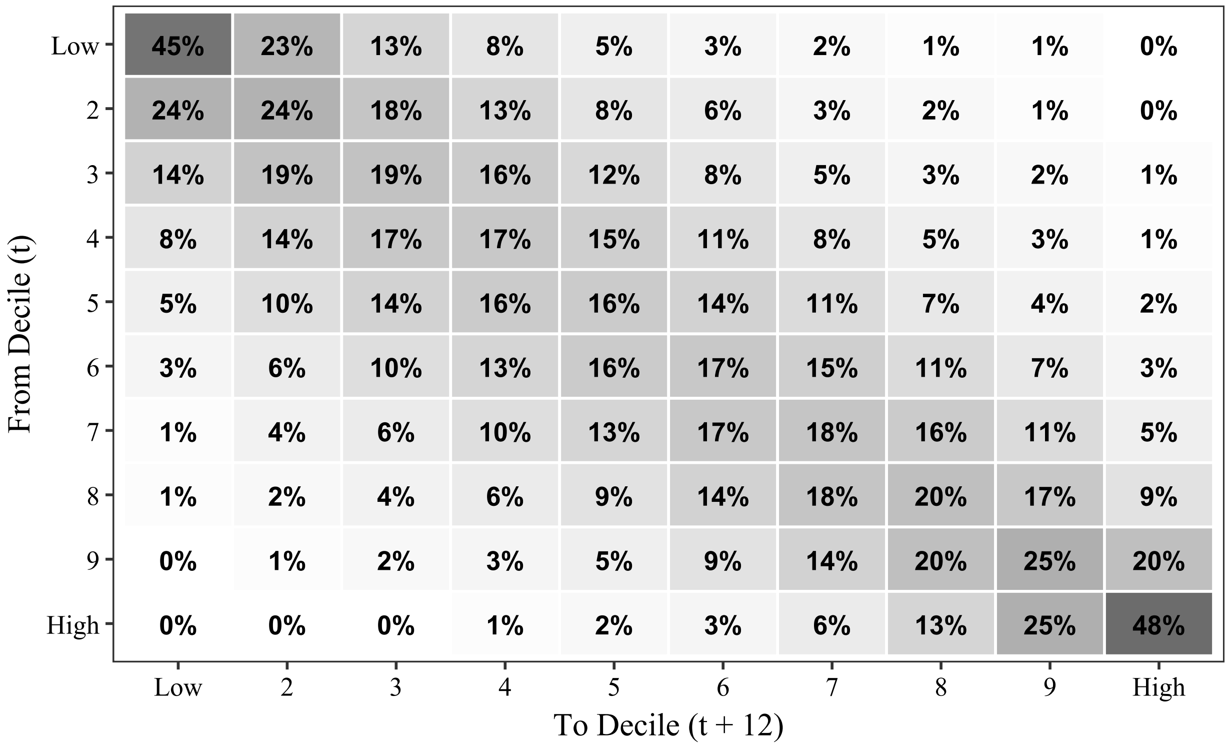

Figure IA.2.4: Transition Probabilities
The figure displays transition probabilities for ATYP deciles over a 12-month horizon. Each month, all firms are ranked by ATYP and assigned into decile portfolios from Low (lowest ATYP) to High (highest ATYP). The same sorting procedure is applied 12 months later. The entries show the average percentage of firms that move from a given ATYP decile at month t (rows) to each ATYP decile at month $t + 12$ (columns). Probabilities are time-series averages over the full sample period.

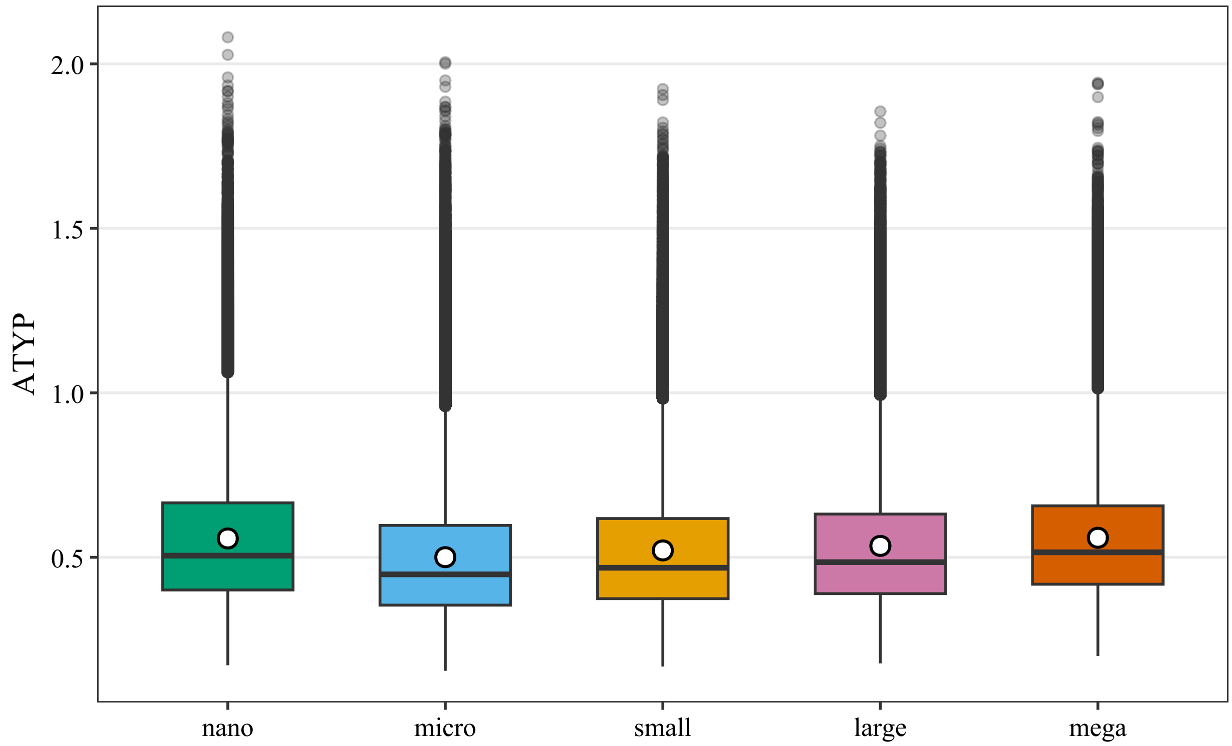

Figure IA.2.5: Distribution of ATYP Within Size Groups
The figure presents the distribution of ATYP scores across five size groups: nano, micro, small, large, and mega. ATYP is re-estimated within each size group following the full preprocessing and model training procedure. This construction removes any mechanical relation between firm size and ATYP. The boxplots show the distribution of ATYP across the size groups. The white circle denotes the mean ATYP value in each group.

Appendix IA.3 Supplementary Tables

Table IA.3.1: Average Stock Characteristics of ATYP-sorted Portfolios

	Low	2	3	4	5	6	7	8	9	High	H-L	t-stat
BETA	1.05	1.08	1.11	1.13	1.14	1.14	1.14	1.15	1.15	1.15	0.1**	(2.2)
SIZE	1.30	0.96	0.73	0.55	0.40	0.30	0.21	0.16	0.13	0.08	-1.22	(-0.77)
BM	0.53	0.55	0.55	0.56	0.56	0.55	0.53	0.52	0.47	0.40	-0.13***	(-3)
STR	0.01	0.01	0.01	0.01	0.00	0.00	0.00	-0.01	-0.01	-0.02	-0.03***	(-11.83)
MOM	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.02	0.00	-0.04	-0.15***	(-4.04)
OP	0.24	0.23	0.21	0.20	0.18	0.18	0.16	0.13	0.06	-0.05	-0.29***	(-12.37)
AG	0.06	0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.10	0.21	0.15*	(1.78)
TURN	0.39	0.42	0.43	0.43	0.43	0.43	0.43	0.45	0.48	0.54	0.16	(0.6)
ILLIQ	0.06	0.09	0.13	0.16	0.18	0.20	0.21	0.23	0.27	0.34	0.28*	(1.71)
IVOL	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.04	0.05	0.03***	(3.87)
SUE	-0.01	-0.01	-0.02	-0.02	-0.02	-0.02	-0.01	-0.02	-0.01	-0.02	0	(-0.19)
MAX	0.04	0.05	0.05	0.05	0.06	0.06	0.06	0.07	0.07	0.08	0.04***	(14.14)
AGE	2.74	2.38	2.20	2.05	1.87	1.73	1.58	1.42	1.25	1.10	-1.65***	(-18.71)
AFD	-0.04	-0.05	-0.05	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.02***	(-5.32)
#ANA	6.08	5.46	4.89	4.38	3.82	3.42	3.08	2.69	2.39	1.78	-4.3***	(-7.33)

Note: The table reports the time-series averages of the monthly cross-sectional median for stock characteristics across decile portfolios sorted on ATYP. Reported characteristics are defined in Section 3.2. The last two columns present high-minus-low differences and Newey and West (1987) adjusted t-statistics with six lags. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table IA.3.2: Fama-MacBeth Cross-Sectional Regressions with Post-Earnings-Announcement Drift

	Excess Return $t+1$	Excess Return $t+2$	Excess Return $t+3$
ATYP	-0.75***	-0.94***	-1.01***
	(-3.31)	(-4.51)	(-4.96)
SUE	-0.02	0.00	0.02
	(-0.62)	(0.11)	(0.92)
ATYP x SUE	0.18***	0.12**	0.04
	(3.44)	(2.46)	(0.84)
Controls	Yes	Yes	Yes

Note: The table reports results from Fama–MacBeth cross-sectional regressions linking post-earnings-announcement drift to information-processing difficulty. The dependent variable is the cumulative excess return over horizons $t+1$, $t+1$ to $t+2$, and $t+1$ to $t+3$. Independent variables include ATYP, standardized unexpected earnings (SUE), and their interaction. The controls considered are defined in Section 3.2. Standard errors are Newey and West (1987) adjusted with six lags. t-statistics are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table IA.3.3: Fama-MacBeth Cross-Sectional Regressions with Limits to Arbitrage Proxies

	Excess Return	Excess Return	Excess Return
ATYP	0.40	-0.72***	-1.35***
	(1.07)	(-3.09)	(-4.84)
ATYP x IVOL	-29.65***		
	(-4.22)		
ATYP x ILLIQ		-0.05	
		(-0.80)	
ATYP x AGE			0.29***
			(4.17)
Controls	Yes	Yes	Yes

Note: The table reports the results of monthly Fama and MacBeth (1973) regressions of excess stock returns on ATYP and a broad set of firm characteristics. ATYP is interacted with idiosyncratic volatility, illiquidity, and firm age, respectively. The controls considered are defined in Section 3.2. Reported coefficients represent average slopes across months, with Newey and West (1987) adjusted t-statistics (six lags) shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table IA.3.4: Fama-MacBeth Cross-Sectional Regressions with Attention Proxies

	Excess Return	Excess Return	Excess Return
ATYP	-0.85***	-1.27***	-0.96***
	(-3.64)	(-5.04)	(-4.25)
ATYP x SIZE	2.23***		
	(3.35)		
ATYP x #ANA		0.13***	
		(5.49)	
ATYP x SHORT			0.02***
			(4.16)
Controls	Yes	Yes	Yes

Note: The table reports the results of monthly Fama and MacBeth (1973) regressions of excess stock returns on ATYP and a broad set of firm characteristics. ATYP is interacted with firm size, analyst coverage, and short interest, respectively. The controls considered are defined in Section 3.2. Reported coefficients represent average slopes across months, with Newey and West (1987) adjusted t-statistics (six lags) shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table IA.3.5: Univariate Portfolio Sorts on ATYP by Firm Characteristic Clusters

	Equal-Weighted		Value-Weighted	
	H-L	FF6	H-L	FF6
Accruals	-0.74***	-0.48***	-0.07	0.19
	(-4.24)	(-3.67)	(-0.36)	(1.21)
Debt Issuance	-0.79***	-0.43***	-0.27	0.02
	(-5.07)	(-4.36)	(-1.62)	(0.20)
Investment	-1.53***	-0.96***	-0.80***	-0.34**
	(-5.41)	(-6.18)	(-2.93)	(-2.12)
Low Leverage	-0.84***	-0.42***	-0.43**	-0.23
	(-3.76)	(-2.84)	(-2.27)	(-1.44)
Momentum	-0.76***	-0.44***	-0.08	0.26
	(-3.37)	(-3.07)	(-0.35)	(1.31)
Profit Growth	-1.18***	-0.69***	-0.57**	-0.24
	(-4.77)	(-4.60)	(-2.34)	(-1.55)
Profitability	-1.28***	-0.78***	-0.28	0.07
	(-4.45)	(-4.60)	(-1.19)	(0.42)
Quality	-1.08***	-0.57***	-0.34	0.01
	(-4.22)	(-3.51)	(-1.34)	(0.04)
Size	0.22	0.45***	-0.03	0.16*
	(1.61)	(3.67)	(-0.21)	(1.96)
Value	-0.85***	-0.47***	-0.32	-0.06
	(-3.15)	(-2.69)	(-1.26)	(-0.41)

Note: The table reports average monthly excess returns and factor-adjusted alphas from univariate portfolio sorts on cluster-specific squared reconstruction errors used to construct ATYP. Each row corresponds to a firm characteristic cluster, with high-minus-low (H–L) decile return spreads reported for equal-weighted and value-weighted portfolios. FF6 denotes the intercept from a regression of portfolio excess returns on the Fama and French (2018) six-factor model. Newey and West (1987) adjusted t-statistics with six lags are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table IA.3.6: Univariate Portfolio Sorts on Industry-Adjusted ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.91***	(4.09)	0.21*	(1.72)	0.17***	(2.62)	0.06	(0.90)	0.11**	(2.07)
2	0.93***	(3.88)	0.18	(1.59)	0.18***	(3.06)	0.11*	(1.85)	0.18***	(3.43)
3	0.86***	(3.39)	0.10	(0.84)	0.11**	(2.03)	0.10*	(1.77)	0.15***	(2.93)
4	0.86***	(3.20)	0.06	(0.51)	0.09	(1.40)	0.11	(1.54)	0.18***	(2.72)
5	0.78***	(2.75)	-0.02	(-0.17)	0.00	(0.00)	0.06	(0.87)	0.16**	(2.32)
6	0.77**	(2.53)	-0.07	(-0.44)	-0.03	(-0.35)	0.07	(0.75)	0.17**	(2.05)
7	0.61*	(1.88)	-0.25	(-1.54)	-0.19**	(-2.18)	-0.06	(-0.68)	0.05	(0.53)
8	0.55	(1.57)	-0.34*	(-1.89)	-0.26**	(-2.52)	-0.07	(-0.67)	0.06	(0.61)
9	0.22	(0.57)	-0.70***	(-3.24)	-0.60***	(-4.26)	-0.31**	(-2.44)	-0.18	(-1.42)
10	-0.49	(-1.11)	-1.48***	(-5.49)	-1.33***	(-6.91)	-0.97***	(-5.87)	-0.81***	(-4.91)
H-L	-1.40***	(-4.73)	-1.69***	(-6.42)	-1.50***	(-6.89)	-1.02***	(-6.18)	-0.93***	(-5.43)
Panel B: Value-Weighted										
1	0.74***	(4.22)	0.12	(1.22)	0.08	(0.85)	-0.12	(-1.64)	-0.09	(-1.26)
2	0.74***	(4.11)	0.11	(1.44)	0.09	(1.14)	-0.09	(-1.26)	-0.09	(-1.33)
3	0.71***	(3.60)	0.04	(0.74)	0.06	(1.09)	-0.03	(-0.48)	-0.02	(-0.38)
4	0.70***	(3.13)	-0.02	(-0.32)	0.03	(0.43)	0.06	(0.68)	0.06	(0.70)
5	0.79***	(3.57)	0.05	(0.70)	0.08	(1.13)	0.04	(0.53)	0.10	(1.33)
6	0.73***	(3.16)	-0.03	(-0.37)	0.03	(0.41)	0.06	(0.80)	0.07	(0.91)
7	0.72***	(2.69)	-0.10	(-0.97)	-0.02	(-0.17)	0.11	(1.17)	0.10	(1.06)
8	0.71**	(2.54)	-0.12	(-0.93)	-0.04	(-0.33)	0.09	(0.76)	0.12	(1.11)
9	0.49	(1.50)	-0.39**	(-2.38)	-0.24*	(-1.77)	-0.02	(-0.19)	-0.01	(-0.09)
10	0.30	(0.74)	-0.74***	(-3.23)	-0.54***	(-2.66)	-0.12	(-0.66)	-0.12	(-0.65)
H-L	-0.45	(-1.40)	-0.86***	(-2.99)	-0.62**	(-2.46)	-0.00	(-0.00)	-0.03	(-0.14)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on the industry-adjusted measure of ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score estimated within one-digit SIC industries using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.7: Within-Industry Portfolio Sorts on Industry-Adjusted ATYP

	Equal-Weighted		Value-Weighted	
	H-L	FF6	H-L	FF6
Mining & Construction	-1.17***	-0.95***	-0.33	-0.19
	(-3.54)	(-2.91)	(-0.91)	(-0.57)
Manufacturing (Food, Textile, etc.)	-0.91**	-0.39	-0.92**	-0.53*
	(-2.38)	(-1.38)	(-2.39)	(-1.73)
Manufacturing (Machinery, Transport, etc.)	-1.60***	-1.03***	-0.87**	-0.21
	(-4.58)	(-4.37)	(-2.21)	(-0.72)
Transportation, Communications	-1.11***	-0.52*	-0.38	-0.09
	(-2.77)	(-1.68)	(-1.13)	(-0.25)
Wholesale & Retail Trade	-1.64***	-1.08***	-0.10	0.24
	(-4.72)	(-3.98)	(-0.27)	(0.74)
Services	-1.51***	-1.22***	-0.70*	-0.20
	(-4.29)	(-4.43)	(-1.66)	(-0.55)
Public Administration	-1.62***	-1.28***	-1.00**	-0.65*
	(-3.83)	(-3.69)	(-2.02)	(-1.65)

Note: The table reports average monthly excess returns and factor-adjusted alphas from univariate portfolio sorts conducted within each industry based on the industry-adjusted ATYP measure. Each row corresponds to a one-digit SIC industry, with high-minus-low (H–L) decile return spreads reported for equal-weighted and value-weighted portfolios. FF6 denotes the intercept from a regression of portfolio excess returns on the Fama and French (2018) six-factor model. Newey and West (1987) adjusted t-statistics with six lags are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table IA.3.8: Univariate Portfolio Sorts on Size-Adjusted ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	1.01***	(4.09)	0.33**	(2.23)	0.28***	(4.07)	0.17**	(2.46)	0.23***	(3.72)
2	0.92***	(3.58)	0.20	(1.41)	0.18***	(2.67)	0.10	(1.49)	0.18***	(2.79)
3	0.90***	(3.32)	0.14	(1.04)	0.13**	(2.06)	0.11	(1.56)	0.18***	(2.76)
4	0.87***	(3.26)	0.11	(0.78)	0.11*	(1.86)	0.11	(1.60)	0.20***	(3.20)
5	0.78***	(2.69)	-0.01	(-0.04)	0.01	(0.12)	0.05	(0.56)	0.14*	(1.70)
6	0.67**	(2.24)	-0.17	(-1.19)	-0.14*	(-1.77)	-0.07	(-0.91)	0.03	(0.39)
7	0.53*	(1.67)	-0.33**	(-2.02)	-0.28***	(-2.94)	-0.14	(-1.48)	-0.03	(-0.36)
8	0.46	(1.39)	-0.43**	(-2.53)	-0.34***	(-3.45)	-0.14	(-1.48)	-0.01	(-0.11)
9	0.11	(0.29)	-0.83***	(-4.31)	-0.70***	(-5.56)	-0.41***	(-3.77)	-0.29***	(-2.77)
10	-0.31	(-0.74)	-1.33***	(-5.37)	-1.13***	(-6.51)	-0.69***	(-4.91)	-0.57***	(-4.11)
H-L	-1.33***	(-4.81)	-1.66***	(-6.31)	-1.41***	(-6.98)	-0.86***	(-5.98)	-0.80***	(-5.49)
Panel B: Value-Weighted										
1	0.82***	(4.35)	0.20*	(1.94)	0.15*	(1.68)	-0.08	(-0.90)	-0.06	(-0.66)
2	0.78***	(4.15)	0.14	(1.49)	0.10	(1.25)	-0.10	(-1.54)	-0.09	(-1.25)
3	0.68***	(3.71)	0.06	(0.91)	0.03	(0.49)	-0.11*	(-1.74)	-0.13**	(-2.04)
4	0.68***	(3.47)	0.02	(0.30)	0.02	(0.30)	-0.07	(-1.06)	-0.07	(-1.05)
5	0.77***	(4.07)	0.12	(1.59)	0.13	(1.64)	0.01	(0.19)	-0.00	(-0.04)
6	0.74***	(3.54)	0.05	(0.58)	0.06	(0.80)	0.04	(0.46)	0.04	(0.48)
7	0.63***	(2.77)	-0.10	(-1.23)	-0.04	(-0.49)	-0.05	(-0.71)	-0.04	(-0.47)
8	0.90***	(3.93)	0.13*	(1.69)	0.19**	(2.28)	0.24***	(3.06)	0.28***	(3.44)
9	0.55*	(1.93)	-0.33***	(-3.03)	-0.21**	(-2.17)	-0.03	(-0.25)	0.02	(0.16)
10	0.53	(1.49)	-0.41*	(-1.88)	-0.18	(-1.10)	0.23*	(1.78)	0.21*	(1.64)
H-L	-0.30	(-1.03)	-0.61**	(-2.09)	-0.33	(-1.50)	0.31*	(1.80)	0.26	(1.59)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on the size-adjusted measure of ATYP. At the end of each month, firms are sorted into five groups based on their market equity: nano, micro, small, large, and mega. Then, at the end of each month t, all stocks are sorted into deciles according to their ATYP score estimated within size groups using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.9: Within-Size Group Portfolio Sorts on Size-Adjusted ATYP

	Equal-Weighted		Value-Weighted	
	H-L	FF6	H-L	FF6
nano	-1.84***	-1.59***	-1.80***	-1.42***
	(-4.85)	(-4.12)	(-4.12)	(-2.93)
micro	-1.85***	-1.32***	-1.63***	-0.99***
	(-5.94)	(-6.41)	(-4.97)	(-4.97)
small	-1.03***	-0.41***	-0.99***	-0.36**
	(-3.58)	(-2.70)	(-3.33)	(-2.34)
large	-0.49	0.06	-0.42	0.13
	(-1.54)	(0.34)	(-1.37)	(0.67)
mega	-0.33	0.31*	-0.01	0.58***
	(-1.11)	(1.93)	(-0.04)	(3.24)

Note: The table reports average monthly excess returns and factor-adjusted alphas from univariate portfolio sorts conducted within each size group based on the size-adjusted ATYP measure. At the end of each month, firms are sorted into five groups based on their market equity: nano, micro, small, large, and mega. Each row corresponds to a size group, with high-minus-low (H–L) decile return spreads reported for equal-weighted and value-weighted portfolios. FF6 denotes the intercept from a regression of portfolio excess returns on the Fama and French (2018) six-factor model. Newey and West (1987) adjusted t-statistics with six lags are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table IA.3.10: Univariate Portfolio Sorts on ATYP with Alternative Factor Models

	Excess Return	t-stat	FF6	t-stat	Q5	t-stat	DHS	t-stat	DMRS	t-stat
Panel A: Equal-Weighted										
1	0.94***	(4.25)	0.08	(1.52)	0.10	(1.07)	0.38***	(3.48)	0.42*	(1.72)
2	0.93***	(3.96)	0.12**	(2.29)	0.16**	(2.01)	0.44***	(3.72)	0.46*	(1.68)
3	0.94***	(3.77)	0.19***	(3.44)	0.23***	(2.94)	0.47***	(3.60)	0.51*	(1.82)
4	0.91***	(3.45)	0.22***	(3.74)	0.27***	(3.49)	0.54***	(3.77)	0.51*	(1.68)
5	0.79***	(2.77)	0.15**	(2.02)	0.22**	(2.51)	0.50***	(2.96)	0.44	(1.36)
6	0.78***	(2.59)	0.20**	(2.40)	0.28***	(3.03)	0.49***	(2.76)	0.47	(1.37)
7	0.65*	(1.96)	0.10	(1.19)	0.19**	(2.04)	0.46**	(2.35)	0.37	(1.02)
8	0.41	(1.16)	-0.06	(-0.54)	0.05	(0.43)	0.32	(1.45)	0.23	(0.60)
9	0.12	(0.31)	-0.20	(-1.61)	-0.09	(-0.63)	0.22	(0.89)	0.04	(0.11)
10	-0.53	(-1.17)	-0.76***	(-4.36)	-0.63***	(-3.26)	-0.29	(-0.95)	-0.51	(-1.11)
H-L	-1.47***	(-4.54)	-0.84***	(-4.52)	-0.73***	(-3.24)	-0.67**	(-2.56)	-0.93***	(-3.34)
Panel B: Value-Weighted										
1	0.80***	(4.35)	-0.05	(-0.67)	-0.04	(-0.45)	0.11	(1.35)	0.34	(1.53)
2	0.66***	(3.65)	-0.14**	(-2.09)	-0.12	(-1.62)	-0.02	(-0.22)	0.23	(1.05)
3	0.72***	(3.64)	-0.04	(-0.61)	-0.02	(-0.36)	0.03	(0.40)	0.32	(1.35)
4	0.77***	(3.77)	0.06	(0.77)	0.08	(1.08)	0.02	(0.20)	0.45*	(1.72)
5	0.72***	(3.16)	0.04	(0.54)	0.07	(0.93)	0.00	(0.01)	0.42*	(1.71)
6	0.79***	(3.31)	0.18**	(2.03)	0.22***	(2.65)	0.12	(1.37)	0.54*	(1.94)
7	0.85***	(3.21)	0.26***	(2.78)	0.36***	(3.77)	0.38***	(3.51)	0.64**	(2.05)
8	0.48*	(1.79)	-0.09	(-0.92)	-0.05	(-0.45)	-0.02	(-0.17)	0.40	(1.33)
9	0.38	(1.17)	-0.05	(-0.35)	-0.04	(-0.27)	0.17	(1.25)	0.46	(1.39)
10	-0.02	(-0.04)	-0.33*	(-1.92)	-0.27	(-1.25)	-0.06	(-0.32)	0.22	(0.47)
H-L	-0.82**	(-2.50)	-0.28	(-1.49)	-0.23	(-0.92)	-0.17	(-0.84)	-0.11	(-0.35)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the Fama and French (2018) six-factor model including momentum (FF6), the Hou et al. (2014) q-factor model (HXZ), the Daniel et al. (2019) behavioral factor model (DHS), and the Daniel et al. (2020) characteristic-efficient factor model (DMRS). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023 (December 2018 in case of DHS).

Table IA.3.11: Univariate Portfolio Sorts on ATYP for WRDS Financial Ratios

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	1.16***	(4.73)	0.42***	(2.97)	0.36***	(4.95)	0.25***	(3.26)	0.33***	(4.91)
2	1.07***	(4.19)	0.31**	(2.18)	0.28***	(3.92)	0.20**	(2.51)	0.29***	(4.39)
3	0.93***	(3.55)	0.15	(1.08)	0.14**	(2.20)	0.13*	(1.70)	0.23***	(3.55)
4	0.89***	(3.19)	0.10	(0.67)	0.11	(1.57)	0.13*	(1.66)	0.24***	(3.22)
5	0.84***	(2.90)	0.03	(0.18)	0.04	(0.59)	0.09	(1.06)	0.20***	(2.80)
6	0.68**	(2.27)	-0.15	(-0.99)	-0.10	(-1.21)	-0.03	(-0.36)	0.08	(1.02)
7	0.65**	(1.99)	-0.20	(-1.24)	-0.15*	(-1.72)	-0.01	(-0.06)	0.12	(1.28)
8	0.51	(1.54)	-0.35**	(-2.03)	-0.26***	(-2.69)	-0.11	(-1.10)	0.01	(0.14)
9	0.22	(0.59)	-0.68***	(-3.36)	-0.55***	(-4.39)	-0.29***	(-2.84)	-0.21*	(-1.93)
10	-0.25	(-0.64)	-1.19***	(-5.12)	-1.00***	(-6.38)	-0.66***	(-5.13)	-0.62***	(-4.58)
H-L	-1.42***	(-5.68)	-1.61***	(-6.68)	-1.37***	(-7.08)	-0.91***	(-6.06)	-0.95***	(-6.48)
Panel B: Value-Weighted										
1	0.75***	(3.62)	0.06	(0.74)	0.03	(0.41)	-0.10	(-1.35)	-0.05	(-0.73)
2	0.85***	(4.30)	0.19**	(2.30)	0.19**	(2.42)	0.05	(0.59)	0.07	(0.91)
3	0.76***	(4.15)	0.09	(1.34)	0.12*	(1.79)	0.03	(0.44)	0.07	(0.92)
4	0.76***	(4.05)	0.10	(1.39)	0.11	(1.61)	0.02	(0.28)	0.01	(0.19)
5	0.66***	(2.99)	-0.03	(-0.42)	0.02	(0.21)	-0.01	(-0.11)	-0.00	(-0.02)
6	0.67***	(2.79)	-0.07	(-0.58)	-0.02	(-0.20)	-0.01	(-0.14)	-0.01	(-0.05)
7	0.71***	(3.12)	-0.01	(-0.14)	0.03	(0.32)	0.02	(0.27)	0.05	(0.55)
8	0.72***	(2.85)	-0.06	(-0.63)	0.02	(0.23)	0.12	(1.26)	0.14	(1.47)
9	0.65**	(2.33)	-0.18	(-1.38)	-0.06	(-0.58)	0.04	(0.42)	0.01	(0.09)
10	0.25	(0.74)	-0.66***	(-3.57)	-0.47***	(-2.99)	-0.19	(-1.56)	-0.28**	(-2.38)
H-L	-0.50**	(-2.15)	-0.72***	(-3.03)	-0.51**	(-2.51)	-0.09	(-0.59)	-0.22	(-1.52)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP as a robustness check using the WRDS Financial Ratio dataset. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.12: Univariate Portfolio Sorts on ATYP for the Complete Jensen et al. (2023) Dataset

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.95***	(4.19)	0.25**	(1.97)	0.21***	(2.92)	0.08	(1.15)	0.13**	(2.24)
2	0.89***	(3.70)	0.16	(1.27)	0.14**	(2.10)	0.04	(0.62)	0.10*	(1.69)
3	0.88***	(3.52)	0.12	(1.00)	0.11*	(1.80)	0.06	(0.86)	0.13**	(2.21)
4	0.80***	(3.02)	0.01	(0.09)	0.02	(0.35)	0.01	(0.20)	0.09	(1.48)
5	0.84***	(2.93)	0.03	(0.21)	0.05	(0.65)	0.09	(1.16)	0.18**	(2.45)
6	0.77**	(2.47)	-0.06	(-0.42)	-0.03	(-0.41)	0.07	(0.76)	0.18**	(2.29)
7	0.67**	(2.10)	-0.19	(-1.18)	-0.12	(-1.39)	0.02	(0.18)	0.13	(1.53)
8	0.48	(1.36)	-0.41**	(-2.18)	-0.32***	(-2.73)	-0.11	(-1.02)	0.02	(0.19)
9	0.15	(0.39)	-0.77***	(-3.52)	-0.64***	(-4.50)	-0.31***	(-2.69)	-0.20*	(-1.68)
10	-0.47	(-1.04)	-1.47***	(-5.18)	-1.29***	(-6.41)	-0.86***	(-5.00)	-0.70***	(-4.05)
H-L	-1.42***	(-4.48)	-1.72***	(-5.94)	-1.50***	(-6.36)	-0.94***	(-5.27)	-0.83***	(-4.48)
Panel B: Value-Weighted										
1	0.80***	(4.29)	0.14	(1.52)	0.10	(1.16)	-0.07	(-0.89)	-0.03	(-0.41)
2	0.67***	(3.58)	0.03	(0.45)	-0.00	(-0.05)	-0.17***	(-2.67)	-0.14**	(-2.29)
3	0.76***	(4.20)	0.11*	(1.66)	0.11	(1.59)	-0.01	(-0.18)	-0.01	(-0.20)
4	0.70***	(3.42)	0.03	(0.49)	0.05	(0.87)	0.01	(0.19)	0.01	(0.16)
5	0.78***	(3.57)	0.08	(1.04)	0.12	(1.61)	0.12	(1.62)	0.13*	(1.65)
6	0.92***	(3.66)	0.12	(1.45)	0.18**	(2.18)	0.17**	(2.12)	0.22**	(2.56)
7	0.71***	(2.71)	-0.11	(-0.95)	-0.03	(-0.31)	0.08	(0.74)	0.11	(1.10)
8	0.72**	(2.53)	-0.12	(-0.87)	0.02	(0.19)	0.14	(1.13)	0.18	(1.45)
9	0.30	(0.97)	-0.58***	(-3.15)	-0.43***	(-2.87)	-0.17	(-1.24)	-0.15	(-1.14)
10	0.23	(0.58)	-0.79***	(-3.36)	-0.57***	(-2.71)	-0.17	(-0.88)	-0.17	(-0.90)
H-L	-0.57*	(-1.82)	-0.94***	(-3.27)	-0.67***	(-2.62)	-0.10	(-0.45)	-0.14	(-0.64)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP as a robustness check using the complete dataset of Jensen et al. (2023). At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.13: Univariate Portfolio Sorts on ATYP for Different Model Specifications

								Equal-Weighted				Value-Weighted			
Model	Bottleneck	#Layers	Units	L2	Dropout	Noise	R^2	H–L	t-stat	FF6	t-stat	H–L	t-stat	FF6	t-stat
Autoencoder	12	2	64	0	0	0	0.66	-1.48	-4.58	-0.86	-4.78	-0.59	-1.79	-0.13	-0.67
Autoencoder	12	3	128	0	0	0	0.69	-1.46	-4.42	-0.86	-4.73	-0.66	-2.03	-0.21	-1
Autoencoder	16	2	64	0	0	0	0.71	-1.42	-4.38	-0.78	-4.25	-0.68	-2.1	-0.15	-0.78
Autoencoder	16	3	128	0	0	0	0.74	-1.46	-4.4	-0.84	-4.45	-0.63	-1.9	-0.15	-0.71
Autoencoder	20	2	64	0	0	0	0.74	-1.48	-4.52	-0.83	-4.58	-0.75	-2.32	-0.22	-1.18
Autoencoder	20	3	128	0	0	0	0.79	-1.53	-4.63	-0.9	-4.93	-0.54	-1.6	-0.08	-0.37
Autoencoder	12	2	64	0.1	0	0	0.66	-1.43	-4.46	-0.8	-4.5	-0.57	-1.61	-0.1	-0.47
Autoencoder	12	3	128	0.1	0	0	0.69	-1.45	-4.39	-0.83	-4.52	-0.61	-1.88	-0.15	-0.75
Autoencoder	16	2	64	0.1	0	0	0.7	-1.46	-4.48	-0.83	-4.48	-0.71	-2.16	-0.22	-1.14
Autoencoder	16	3	128	0.1	0	0	0.74	-1.45	-4.4	-0.81	-4.35	-0.66	-2.01	-0.17	-0.9
Autoencoder	20	2	64	0.1	0	0	0.74	-1.5	-4.57	-0.85	-4.74	-0.87	-2.69	-0.4	-2.03
Autoencoder	20	3	128	0.1	0	0	0.78	-1.48	-4.46	-0.82	-4.28	-0.68	-1.91	-0.2	-0.9
Autoencoder	12	2	64	0	1e-04	0	0.66	-1.42	-4.38	-0.8	-4.37	-0.54	-1.58	-0.07	-0.34
Autoencoder	12	3	128	0	1e-04	0	0.7	-1.45	-4.4	-0.85	-4.6	-0.7	-2.13	-0.22	-1.12
Autoencoder	16	2	64	0	1e-04	0	0.71	-1.43	-4.37	-0.81	-4.3	-0.69	-2.18	-0.27	-1.37
Autoencoder	16	3	128	0	1e-04	0	0.75	-1.48	-4.52	-0.85	-4.59	-0.62	-1.82	-0.15	-0.73
Autoencoder	20	2	64	0	1e-04	0	0.74	-1.47	-4.46	-0.82	-4.54	-0.67	-1.97	-0.15	-0.79
Autoencoder	20	3	128	0	1e-04	0	0.79	-1.5	-4.54	-0.85	-4.71	-0.65	-1.92	-0.18	-0.84
Autoencoder	12	2	64	0.1	1e-04	0	0.66	-1.45	-4.55	-0.84	-4.73	-0.68	-2.12	-0.16	-0.86
Autoencoder	12	3	128	0.1	1e-04	0	0.69	-1.49	-4.49	-0.88	-4.74	-0.67	-2.06	-0.2	-1.01
Autoencoder	16	2	64	0.1	1e-04	0	0.7	-1.47	-4.5	-0.84	-4.62	-0.64	-1.96	-0.12	-0.63
Autoencoder	16	3	128	0.1	1e-04	0	0.74	-1.48	-4.55	-0.85	-4.57	-0.8	-2.42	-0.3	-1.55
Autoencoder	20	2	64	0.1	1e-04	0	0.74	-1.43	-4.33	-0.79	-4.24	-0.67	-1.98	-0.21	-1.02
Autoencoder	20	3	128	0.1	1e-04	0	0.78	-1.46	-4.42	-0.81	-4.4	-0.6	-1.8	-0.09	-0.48
Autoencoder	12	2	64	0	0	0.1	0.66	-1.43	-4.43	-0.81	-4.46	-0.59	-1.76	-0.13	-0.65
Autoencoder	12	3	128	0	0	0.1	0.69	-1.42	-4.31	-0.82	-4.54	-0.66	-2	-0.19	-0.91
Autoencoder	16	2	64	0	0	0.1	0.71	-1.44	-4.38	-0.8	-4.33	-0.6	-1.79	-0.13	-0.66
Autoencoder	16	3	128	0	0	0.1	0.75	-1.46	-4.44	-0.84	-4.53	-0.67	-1.96	-0.17	-0.81
Autoencoder	20	2	64	0	0	0.1	0.74	-1.48	-4.49	-0.83	-4.58	-0.72	-2.08	-0.18	-0.85
Autoencoder	20	3	128	0	0	0.1	0.79	-1.49	-4.56	-0.84	-4.53	-0.67	-2.08	-0.15	-0.77
Autoencoder	12	2	64	0.1	0	0.1	0.66	-1.44	-4.51	-0.81	-4.69	-0.62	-1.9	-0.13	-0.7

Continued on next page

Table IA.3.13: Univariate Portfolio Sorts on ATYP for Different Model Specifications *(continued)*

								Equal-Weighted				Value-Weighted			
Model	Bottleneck	#Layers	Units	L2	Dropout	Noise	R^2	H–L	t-stat	FF6	t-stat	H–L	t-stat	FF6	t-stat
Autoencoder	12	3	128	0.1	0	0.1	0.69	-1.45	-4.42	-0.84	-4.56	-0.58	-1.73	-0.06	-0.29
Autoencoder	16	2	64	0.1	0	0.1	0.7	-1.45	-4.43	-0.82	-4.46	-0.53	-1.56	-0.03	-0.17
Autoencoder	16	3	128	0.1	0	0.1	0.74	-1.48	-4.53	-0.84	-4.57	-0.76	-2.31	-0.31	-1.48
Autoencoder	20	2	64	0.1	0	0.1	0.74	-1.5	-4.54	-0.86	-4.68	-0.73	-2.16	-0.24	-1.2
Autoencoder	20	3	128	0.1	0	0.1	0.78	-1.47	-4.44	-0.84	-4.49	-0.74	-2.28	-0.26	-1.29
Autoencoder	12	2	64	0	1e-04	0.1	0.66	-1.43	-4.45	-0.81	-4.49	-0.54	-1.59	-0.01	-0.05
Autoencoder	12	3	128	0	1e-04	0.1	0.7	-1.46	-4.37	-0.85	-4.54	-0.59	-1.75	-0.1	-0.48
Autoencoder	16	2	64	0	1e-04	0.1	0.71	-1.47	-4.54	-0.85	-4.67	-0.56	-1.68	-0.1	-0.48
Autoencoder	16	3	128	0	1e-04	0.1	0.75	-1.49	-4.56	-0.85	-4.65	-0.63	-1.8	-0.15	-0.72
Autoencoder	20	2	64	0	1e-04	0.1	0.74	-1.45	-4.41	-0.81	-4.48	-0.71	-2.06	-0.21	-1.05
Autoencoder	20	3	128	0	1e-04	0.1	0.79	-1.49	-4.53	-0.84	-4.63	-0.64	-1.88	-0.17	-0.83
Autoencoder	12	2	64	0.1	1e-04	0.1	0.66	-1.45	-4.54	-0.83	-4.74	-0.56	-1.65	-0.06	-0.33
Autoencoder	12	3	128	0.1	1e-04	0.1	0.69	-1.45	-4.39	-0.85	-4.61	-0.64	-1.86	-0.14	-0.67
Autoencoder	16	2	64	0.1	1e-04	0.1	0.7	-1.44	-4.38	-0.81	-4.37	-0.53	-1.64	-0.08	-0.42
Autoencoder	16	3	128	0.1	1e-04	0.1	0.74	-1.52	-4.64	-0.89	-4.83	-0.82	-2.5	-0.29	-1.56
Autoencoder	20	2	64	0.1	1e-04	0.1	0.74	-1.47	-4.41	-0.83	-4.44	-0.82	-2.51	-0.34	-1.69
Autoencoder	20	3	128	0.1	1e-04	0.1	0.78	-1.52	-4.53	-0.85	-4.51	-0.7	-2.09	-0.18	-0.9

Note: The table reports average monthly excess returns and factor-adjusted alphas from univariate portfolio sorts on ATYP constructed using different model specifications. For autoencoders, we vary the bottleneck dimension, number of layers, number of hidden units, regularization (L2), dropout, and noise parameters. For PCA, we vary the number of components (Bottleneck). R^2 indicates the overall fit of each specification. Reported returns include equal-weighted and value-weighted high-minus-low (H–L) decile spreads, along with corresponding Fama and French (2018) six-factor (FF6) alphas and Newey and West (1987) adjusted t-statistics with six lags. The sample is from 1981 to 2023.

Table IA.3.14: Univariate Portfolio Sorts on $ATYP_{AE}$ Orthogonalized with Respect to $ATYP_{PCA}$

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.58*	(1.89)	-0.24*	(-1.65)	-0.18*	(-1.85)	-0.06	(-0.58)	0.07	(0.77)
2	0.65**	(2.29)	-0.14	(-1.04)	-0.12*	(-1.65)	-0.07	(-0.88)	0.03	(0.41)
3	0.77***	(2.82)	-0.01	(-0.07)	0.01	(0.08)	0.02	(0.30)	0.12*	(1.83)
4	0.78***	(2.90)	-0.00	(-0.03)	0.01	(0.21)	0.01	(0.09)	0.08	(1.37)
5	0.76***	(2.74)	-0.04	(-0.30)	-0.03	(-0.45)	-0.00	(-0.02)	0.08	(1.34)
6	0.74**	(2.58)	-0.07	(-0.48)	-0.05	(-0.70)	0.01	(0.11)	0.09	(1.32)
7	0.69**	(2.30)	-0.13	(-0.85)	-0.09	(-1.11)	0.01	(0.13)	0.10	(1.34)
8	0.54*	(1.71)	-0.32**	(-1.96)	-0.26***	(-3.12)	-0.13	(-1.43)	-0.03	(-0.35)
9	0.39	(1.16)	-0.49***	(-2.69)	-0.41***	(-4.12)	-0.24**	(-2.31)	-0.14	(-1.47)
10	0.04	(0.12)	-0.88***	(-4.05)	-0.76***	(-5.45)	-0.46***	(-3.75)	-0.35***	(-2.77)
H-L	-0.53***	(-4.55)	-0.63***	(-5.53)	-0.58***	(-5.89)	-0.41***	(-4.45)	-0.42***	(-4.45)
Panel B: Value-Weighted										
1	0.72***	(3.49)	0.04	(0.45)	0.11**	(2.13)	0.13**	(2.39)	0.14**	(2.53)
2	0.66***	(3.32)	0.00	(0.04)	0.01	(0.19)	-0.08	(-1.25)	-0.08	(-1.21)
3	0.71***	(3.36)	0.00	(0.08)	0.00	(0.04)	-0.12*	(-1.83)	-0.09	(-1.56)
4	0.72***	(3.48)	0.01	(0.11)	0.00	(0.02)	-0.09	(-1.59)	-0.09	(-1.52)
5	0.76***	(3.46)	0.01	(0.15)	0.02	(0.25)	-0.04	(-0.48)	-0.01	(-0.19)
6	0.68***	(3.01)	-0.10	(-1.23)	-0.09	(-1.25)	-0.10	(-1.48)	-0.07	(-1.18)
7	0.72***	(3.01)	-0.07	(-0.78)	-0.04	(-0.51)	-0.05	(-0.58)	-0.03	(-0.42)
8	0.61**	(2.47)	-0.19**	(-2.16)	-0.12*	(-1.79)	-0.05	(-0.75)	-0.02	(-0.35)
9	0.62**	(2.33)	-0.22**	(-2.14)	-0.14*	(-1.83)	-0.04	(-0.50)	-0.02	(-0.29)
10	0.49	(1.62)	-0.39***	(-2.61)	-0.28**	(-2.25)	-0.02	(-0.15)	-0.01	(-0.10)
H-L	-0.22	(-1.22)	-0.43***	(-2.63)	-0.38***	(-3.03)	-0.14	(-1.22)	-0.15	(-1.20)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on $ATYP_{AE}$ that has been orthogonalized with respect to $ATYP_{PCA}$. $ATYP_{PCA}$ is the ATYP measure derived by a linear PCA model with the same bottleneck size of 16. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.15: Univariate Portfolio Sorts on Median ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.97***	(4.42)	0.29**	(2.41)	0.23***	(3.77)	0.07	(1.26)	0.11**	(2.33)
2	0.89***	(3.92)	0.19	(1.49)	0.15***	(2.61)	0.04	(0.74)	0.09*	(1.71)
3	0.97***	(3.91)	0.23*	(1.79)	0.22***	(3.52)	0.15**	(2.43)	0.22***	(3.61)
4	0.87***	(3.28)	0.11	(0.79)	0.11*	(1.80)	0.13*	(1.89)	0.20***	(3.06)
5	0.83***	(2.89)	0.02	(0.17)	0.05	(0.65)	0.09	(1.24)	0.19***	(2.59)
6	0.75**	(2.51)	-0.07	(-0.48)	-0.03	(-0.42)	0.05	(0.55)	0.14*	(1.76)
7	0.62*	(1.90)	-0.26	(-1.58)	-0.19**	(-2.19)	-0.02	(-0.25)	0.10	(1.11)
8	0.43	(1.19)	-0.49***	(-2.69)	-0.39***	(-3.50)	-0.17	(-1.60)	-0.03	(-0.34)
9	0.05	(0.13)	-0.89***	(-3.95)	-0.76***	(-5.03)	-0.43***	(-3.64)	-0.29**	(-2.43)
10	-0.43	(-0.93)	-1.44***	(-4.94)	-1.26***	(-6.06)	-0.82***	(-4.41)	-0.67***	(-3.51)
H-L	-1.39***	(-4.27)	-1.72***	(-5.85)	-1.49***	(-6.26)	-0.89***	(-4.56)	-0.78***	(-3.86)
Panel B: Value-Weighted										
1	0.73***	(4.09)	0.10	(1.28)	0.04	(0.65)	-0.15**	(-2.53)	-0.13**	(-2.15)
2	0.73***	(4.08)	0.11	(1.51)	0.08	(1.17)	-0.08	(-1.08)	-0.05	(-0.77)
3	0.79***	(4.02)	0.12*	(1.88)	0.13**	(2.00)	0.08	(1.11)	0.05	(0.80)
4	0.79***	(4.03)	0.12*	(1.84)	0.14**	(2.21)	0.07	(1.03)	0.07	(0.99)
5	0.74***	(3.12)	-0.00	(-0.02)	0.06	(0.69)	0.06	(0.76)	0.07	(0.76)
6	0.67***	(2.63)	-0.13	(-1.55)	-0.07	(-0.80)	-0.05	(-0.47)	0.01	(0.09)
7	0.68**	(2.58)	-0.13	(-1.20)	-0.02	(-0.25)	0.11	(1.12)	0.12	(1.15)
8	0.69**	(2.29)	-0.23*	(-1.85)	-0.14	(-1.11)	-0.02	(-0.15)	0.01	(0.06)
9	0.40	(1.21)	-0.56***	(-3.24)	-0.37***	(-2.62)	-0.09	(-0.65)	-0.03	(-0.25)
10	0.14	(0.33)	-0.92***	(-3.54)	-0.66***	(-3.20)	-0.15	(-0.83)	-0.16	(-0.91)
H-L	-0.58*	(-1.77)	-1.02***	(-3.29)	-0.70***	(-3.00)	-0.00	(-0.01)	-0.03	(-0.16)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on the *median* ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.16: Univariate Portfolio Sorts on Rank ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.94***	(4.44)	0.27**	(2.30)	0.21***	(3.26)	0.04	(0.72)	0.09*	(1.71)
2	0.94***	(4.10)	0.24*	(1.86)	0.20***	(3.34)	0.07	(1.30)	0.12**	(2.29)
3	1.00***	(4.05)	0.26**	(2.01)	0.24***	(4.04)	0.18***	(2.86)	0.24***	(4.18)
4	0.84***	(3.16)	0.08	(0.58)	0.08	(1.36)	0.09	(1.42)	0.17***	(2.76)
5	0.83***	(2.95)	0.04	(0.27)	0.06	(0.83)	0.11	(1.46)	0.20***	(2.66)
6	0.74**	(2.42)	-0.09	(-0.57)	-0.04	(-0.55)	0.05	(0.60)	0.16*	(1.79)
7	0.62*	(1.91)	-0.25	(-1.52)	-0.19**	(-2.11)	-0.05	(-0.56)	0.07	(0.77)
8	0.38	(1.06)	-0.54***	(-2.87)	-0.43***	(-3.76)	-0.17*	(-1.67)	-0.04	(-0.41)
9	0.08	(0.19)	-0.88***	(-3.95)	-0.74***	(-4.97)	-0.40***	(-3.34)	-0.25**	(-2.14)
10	-0.44	(-0.95)	-1.45***	(-4.96)	-1.27***	(-6.10)	-0.84***	(-4.64)	-0.69***	(-3.75)
H-L	-1.38***	(-4.15)	-1.72***	(-5.72)	-1.48***	(-6.16)	-0.88***	(-4.64)	-0.78***	(-3.94)
Panel B: Value-Weighted										
1	0.70***	(3.97)	0.07	(0.90)	0.01	(0.17)	-0.17***	(-2.66)	-0.13**	(-2.09)
2	0.71***	(3.97)	0.11	(1.41)	0.08	(1.07)	-0.10	(-1.60)	-0.09	(-1.52)
3	0.79***	(3.91)	0.11*	(1.78)	0.12**	(2.01)	0.04	(0.54)	0.02	(0.38)
4	0.75***	(3.82)	0.06	(0.88)	0.09	(1.41)	0.03	(0.51)	0.03	(0.40)
5	0.73***	(3.08)	-0.01	(-0.18)	0.04	(0.52)	0.06	(0.67)	0.08	(0.86)
6	0.71***	(2.78)	-0.06	(-0.62)	0.02	(0.20)	0.08	(0.74)	0.09	(0.85)
7	0.80***	(2.87)	-0.07	(-0.59)	0.02	(0.25)	0.16	(1.64)	0.17*	(1.78)
8	0.64**	(2.10)	-0.26**	(-2.09)	-0.14	(-1.39)	0.00	(0.02)	0.06	(0.63)
9	0.35	(1.05)	-0.62***	(-3.56)	-0.43***	(-2.91)	-0.11	(-0.76)	-0.07	(-0.49)
10	0.15	(0.36)	-0.91***	(-3.61)	-0.66***	(-3.28)	-0.17	(-1.03)	-0.18	(-1.16)
H-L	-0.55*	(-1.75)	-0.98***	(-3.35)	-0.67***	(-2.96)	-0.00	(-0.03)	-0.05	(-0.31)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on the *rank* ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.17: Univariate Portfolio Sorts on ATYP using Non-Missing Characteristics

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.93***	(4.25)	0.24*	(1.93)	0.18***	(2.64)	0.02	(0.37)	0.08	(1.51)
2	0.96***	(4.11)	0.25*	(1.95)	0.21***	(3.61)	0.10*	(1.80)	0.15***	(2.95)
3	0.92***	(3.71)	0.17	(1.37)	0.16***	(2.71)	0.10*	(1.71)	0.16***	(2.91)
4	0.94***	(3.52)	0.17	(1.25)	0.17***	(2.77)	0.18***	(2.68)	0.25***	(3.94)
5	0.80***	(2.80)	0.00	(0.00)	0.02	(0.29)	0.06	(0.79)	0.16**	(2.09)
6	0.75**	(2.49)	-0.08	(-0.54)	-0.04	(-0.49)	0.06	(0.77)	0.18**	(2.17)
7	0.68**	(2.06)	-0.18	(-1.06)	-0.12	(-1.23)	0.01	(0.14)	0.12	(1.40)
8	0.42	(1.18)	-0.48**	(-2.50)	-0.38***	(-3.30)	-0.17	(-1.57)	-0.03	(-0.32)
9	0.13	(0.34)	-0.81***	(-3.66)	-0.68***	(-4.66)	-0.32***	(-2.65)	-0.19	(-1.59)
10	-0.60	(-1.35)	-1.60***	(-5.75)	-1.42***	(-7.15)	-0.96***	(-5.54)	-0.82***	(-4.69)
H-L	-1.54***	(-4.79)	-1.85***	(-6.21)	-1.60***	(-6.81)	-0.99***	(-5.56)	-0.90***	(-4.89)
Panel B: Value-Weighted										
1	0.79***	(4.38)	0.15	(1.57)	0.10	(1.23)	-0.08	(-1.06)	-0.04	(-0.52)
2	0.67***	(3.68)	0.03	(0.37)	0.01	(0.09)	-0.15**	(-2.30)	-0.14**	(-2.11)
3	0.73***	(3.74)	0.06	(1.03)	0.07	(1.20)	-0.03	(-0.58)	-0.03	(-0.47)
4	0.75***	(3.65)	0.07	(0.98)	0.11	(1.45)	0.06	(0.75)	0.06	(0.73)
5	0.75***	(3.19)	0.02	(0.28)	0.06	(0.79)	0.08	(1.04)	0.09	(1.25)
6	0.78***	(3.27)	0.00	(0.03)	0.08	(0.99)	0.14	(1.63)	0.17*	(1.91)
7	0.89***	(3.43)	0.09	(0.97)	0.17**	(1.96)	0.27***	(3.08)	0.29***	(3.17)
8	0.44	(1.61)	-0.40***	(-3.36)	-0.30***	(-2.79)	-0.17	(-1.58)	-0.14	(-1.39)
9	0.39	(1.18)	-0.53***	(-2.83)	-0.37**	(-2.39)	-0.05	(-0.37)	-0.03	(-0.20)
10	0.03	(0.06)	-1.03***	(-3.96)	-0.78***	(-3.64)	-0.30*	(-1.68)	-0.29	(-1.62)
H-L	-0.77**	(-2.31)	-1.18***	(-3.78)	-0.89***	(-3.49)	-0.23	(-1.18)	-0.25	(-1.32)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP recalculated using only non-missing firm characteristics in each observation of $\mathbf{X}i, t$. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.18: Univariate Portfolio Sorts on ATYP Orthogonalized with Respect to Imputation Intensity

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.72***	(2.83)	-0.03	(-0.19)	0.01	(0.08)	0.00	(0.04)	0.05	(0.63)
2	0.84***	(3.46)	0.11	(0.89)	0.09	(1.56)	0.02	(0.33)	0.08	(1.49)
3	0.85***	(3.43)	0.10	(0.85)	0.10*	(1.88)	0.05	(0.98)	0.11**	(2.37)
4	0.88***	(3.38)	0.11	(0.92)	0.11*	(1.89)	0.10	(1.51)	0.16***	(2.94)
5	0.88***	(3.14)	0.08	(0.55)	0.09	(1.25)	0.12	(1.45)	0.21***	(2.70)
6	0.69**	(2.35)	-0.12	(-0.83)	-0.09	(-1.20)	-0.03	(-0.33)	0.07	(0.88)
7	0.70**	(2.22)	-0.15	(-0.98)	-0.10	(-1.23)	0.02	(0.28)	0.14*	(1.79)
8	0.51	(1.45)	-0.38**	(-2.01)	-0.30**	(-2.56)	-0.11	(-1.03)	0.01	(0.09)
9	0.26	(0.67)	-0.67***	(-3.18)	-0.56***	(-4.14)	-0.27**	(-2.28)	-0.13	(-1.11)
10	-0.38	(-0.86)	-1.37***	(-5.12)	-1.22***	(-6.46)	-0.82***	(-4.85)	-0.66***	(-3.92)
H-L	-1.10***	(-4.24)	-1.35***	(-5.84)	-1.23***	(-5.95)	-0.83***	(-4.80)	-0.70***	(-4.06)
Panel B: Value-Weighted										
1	0.71***	(3.22)	-0.04	(-0.43)	-0.03	(-0.41)	-0.14*	(-1.81)	-0.12*	(-1.66)
2	0.78***	(4.19)	0.12	(1.40)	0.11	(1.29)	-0.04	(-0.57)	-0.02	(-0.32)
3	0.63***	(3.23)	-0.03	(-0.39)	-0.04	(-0.46)	-0.17**	(-2.34)	-0.15**	(-2.20)
4	0.73***	(3.84)	0.07	(1.34)	0.08	(1.50)	-0.02	(-0.35)	-0.02	(-0.33)
5	0.74***	(3.55)	0.05	(0.83)	0.08	(1.26)	0.01	(0.13)	0.02	(0.21)
6	0.76***	(3.24)	0.02	(0.27)	0.04	(0.55)	0.07	(0.87)	0.08	(1.01)
7	0.81***	(3.21)	0.03	(0.35)	0.12	(1.37)	0.21**	(2.41)	0.25***	(2.63)
8	0.76***	(2.88)	-0.04	(-0.43)	0.05	(0.58)	0.16*	(1.91)	0.19**	(2.26)
9	0.44	(1.51)	-0.43***	(-3.06)	-0.32***	(-2.60)	-0.12	(-1.08)	-0.10	(-0.97)
10	0.31	(0.81)	-0.69***	(-3.02)	-0.50**	(-2.54)	-0.08	(-0.44)	-0.05	(-0.28)
H-L	-0.40	(-1.49)	-0.65**	(-2.54)	-0.47**	(-2.06)	0.07	(0.36)	0.07	(0.42)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP that has been orthogonalized with respect to the share of imputed characteristics. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.19: Univariate Portfolio Sorts on ATYP Orthogonalized with Respect to Extreme Observations

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.75**	(2.44)	-0.09	(-0.54)	-0.03	(-0.29)	-0.01	(-0.15)	0.08	(0.81)
2	0.66**	(2.38)	-0.12	(-0.85)	-0.10	(-1.42)	-0.10	(-1.30)	-0.02	(-0.21)
3	0.77***	(2.84)	-0.01	(-0.09)	0.00	(0.02)	-0.01	(-0.10)	0.08	(1.17)
4	0.80***	(2.95)	0.01	(0.08)	0.02	(0.34)	0.03	(0.42)	0.12*	(1.95)
5	0.75***	(2.66)	-0.05	(-0.32)	-0.04	(-0.49)	-0.00	(-0.00)	0.09	(1.12)
6	0.76***	(2.69)	-0.03	(-0.22)	-0.01	(-0.18)	0.07	(0.91)	0.15**	(2.19)
7	0.70**	(2.35)	-0.13	(-0.88)	-0.09	(-1.32)	-0.01	(-0.13)	0.08	(1.19)
8	0.51	(1.62)	-0.33**	(-2.08)	-0.27***	(-3.15)	-0.12	(-1.21)	-0.01	(-0.10)
9	0.30	(0.87)	-0.57***	(-3.15)	-0.50***	(-4.58)	-0.26**	(-2.50)	-0.14	(-1.42)
10	-0.06	(-0.16)	-1.00***	(-4.38)	-0.86***	(-5.52)	-0.50***	(-3.55)	-0.39***	(-2.85)
H-L	-0.81***	(-5.03)	-0.91***	(-5.28)	-0.83***	(-5.83)	-0.49***	(-3.69)	-0.47***	(-3.67)
Panel B: Value-Weighted										
1	0.87***	(3.91)	0.17*	(1.81)	0.28***	(3.85)	0.21***	(2.60)	0.17**	(2.37)
2	0.70***	(3.75)	0.05	(0.62)	0.06	(0.80)	-0.03	(-0.35)	-0.02	(-0.25)
3	0.74***	(3.70)	0.05	(0.74)	0.04	(0.66)	-0.05	(-0.88)	-0.03	(-0.46)
4	0.76***	(3.64)	0.06	(0.78)	0.06	(0.73)	0.03	(0.32)	0.05	(0.66)
5	0.65***	(2.99)	-0.09	(-1.03)	-0.09	(-1.18)	-0.13	(-1.56)	-0.11	(-1.41)
6	0.61***	(2.69)	-0.13	(-1.62)	-0.11	(-1.45)	-0.08	(-1.14)	-0.05	(-0.64)
7	0.57**	(2.37)	-0.20**	(-2.43)	-0.19**	(-2.49)	-0.19**	(-2.57)	-0.18**	(-2.28)
8	0.58**	(2.14)	-0.22*	(-1.91)	-0.16	(-1.52)	-0.05	(-0.51)	-0.01	(-0.10)
9	0.41	(1.49)	-0.41***	(-3.20)	-0.32***	(-2.89)	-0.10	(-1.02)	-0.05	(-0.54)
10	0.27	(0.82)	-0.64***	(-3.40)	-0.49***	(-3.08)	-0.13	(-1.02)	-0.10	(-0.79)
H-L	-0.60***	(-2.97)	-0.81***	(-4.14)	-0.77***	(-4.18)	-0.33**	(-2.25)	-0.27*	(-1.87)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP that has been orthogonalized with respect to the share of extreme observations. Extreme observations are those that fall within the bottom or top decile of each characteristic in the cross-section. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table IA.3.20: Univariate Portfolio Sorts on ATYP in Developed Countries (excl. USA)

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.62**	(2.34)	0.34**	(2.11)	0.27**	(1.99)	0.06	(0.45)	0.07	(0.49)
2	0.72**	(2.50)	0.38**	(2.51)	0.30***	(2.60)	0.09	(0.82)	0.10	(0.88)
3	0.69**	(2.26)	0.31**	(2.20)	0.23**	(2.28)	0.02	(0.17)	0.02	(0.25)
4	0.70**	(2.16)	0.30**	(2.15)	0.21**	(2.24)	0.03	(0.32)	0.03	(0.38)
5	0.75**	(2.20)	0.32**	(2.29)	0.23**	(2.48)	0.06	(0.62)	0.06	(0.68)
6	0.69*	(1.93)	0.24*	(1.69)	0.15*	(1.66)	-0.01	(-0.08)	0.02	(0.18)
7	0.69*	(1.76)	0.20	(1.32)	0.12	(1.24)	0.00	(0.05)	0.02	(0.27)
8	0.56	(1.36)	0.04	(0.25)	-0.03	(-0.28)	-0.10	(-1.17)	-0.08	(-0.97)
9	0.46	(1.03)	-0.10	(-0.60)	-0.15	(-1.34)	-0.13	(-1.20)	-0.11	(-1.01)
10	-0.03	(-0.05)	-0.64***	(-3.21)	-0.61***	(-4.27)	-0.37***	(-2.83)	-0.37***	(-2.77)
H-L	-0.65*	(-1.96)	-0.99***	(-4.72)	-0.88***	(-4.33)	-0.43**	(-2.12)	-0.44**	(-2.11)
Panel B: Value-Weighted										
1	0.43	(1.55)	0.08	(0.63)	0.00	(0.01)	-0.22**	(-2.07)	-0.21**	(-1.98)
2	0.51*	(1.68)	0.11	(1.01)	0.04	(0.38)	-0.18*	(-1.87)	-0.19**	(-1.98)
3	0.51*	(1.69)	0.07	(0.70)	0.02	(0.16)	-0.22**	(-2.21)	-0.25**	(-2.50)
4	0.47	(1.49)	0.02	(0.16)	-0.03	(-0.31)	-0.21**	(-2.30)	-0.23***	(-2.60)
5	0.39	(1.23)	-0.08	(-0.79)	-0.11	(-1.23)	-0.23**	(-2.50)	-0.25***	(-2.70)
6	0.41	(1.29)	-0.05	(-0.52)	-0.07	(-0.71)	-0.22**	(-2.19)	-0.23**	(-2.17)
7	0.47	(1.36)	-0.02	(-0.20)	-0.03	(-0.32)	-0.10	(-0.86)	-0.12	(-1.01)
8	0.24	(0.64)	-0.26**	(-2.25)	-0.24**	(-2.19)	-0.24**	(-2.01)	-0.27**	(-2.33)
9	0.44	(1.12)	-0.12	(-0.91)	-0.06	(-0.54)	-0.02	(-0.20)	-0.03	(-0.24)
10	-0.11	(-0.23)	-0.69***	(-3.09)	-0.60***	(-3.09)	-0.44**	(-2.30)	-0.44**	(-2.36)
H-L	-0.53*	(-1.95)	-0.77***	(-3.34)	-0.61***	(-2.80)	-0.22	(-0.99)	-0.24	(-1.06)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP using stocks from developed countries (excl. USA). At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the local one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, the Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). The regional factor portfolios and country classifications for developed markets are taken from the Kenneth R. French Data Library. Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample period is from 2000 to 2023.

Table IA.3.21: Univariate Portfolio Sorts on ATYP in Europe

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.88**	(2.26)	0.47***	(3.83)	0.35***	(5.23)	0.21***	(2.97)	0.22***	(2.88)
2	0.77**	(1.98)	0.36***	(3.29)	0.23***	(4.49)	0.15***	(2.79)	0.16***	(2.67)
3	0.80**	(2.01)	0.39***	(3.44)	0.26***	(5.11)	0.21***	(4.26)	0.24***	(4.57)
4	0.76*	(1.88)	0.34***	(3.04)	0.21***	(3.75)	0.21***	(3.46)	0.24***	(3.96)
5	0.63	(1.51)	0.20*	(1.72)	0.07	(1.41)	0.15**	(2.54)	0.20***	(3.61)
6	0.58	(1.34)	0.15	(1.18)	0.02	(0.28)	0.11	(1.52)	0.17**	(2.38)
7	0.45	(1.00)	0.01	(0.04)	-0.11	(-1.54)	0.06	(0.85)	0.12*	(1.80)
8	0.33	(0.71)	-0.12	(-0.76)	-0.23**	(-2.50)	-0.01	(-0.09)	0.06	(0.68)
9	0.11	(0.24)	-0.35**	(-1.98)	-0.43***	(-3.84)	-0.08	(-0.69)	-0.02	(-0.17)
10	-0.33	(-0.61)	-0.81***	(-3.31)	-0.86***	(-5.44)	-0.36**	(-2.53)	-0.34**	(-2.28)
H-L	-1.21***	(-4.40)	-1.28***	(-4.94)	-1.21***	(-6.38)	-0.57***	(-3.49)	-0.55***	(-3.22)
Panel B: Value-Weighted										
1	0.75**	(2.17)	0.36***	(3.32)	0.34***	(3.58)	0.12	(1.07)	0.10	(0.92)
2	0.59*	(1.77)	0.20**	(2.54)	0.19***	(2.69)	-0.01	(-0.13)	-0.04	(-0.47)
3	0.56	(1.64)	0.15*	(1.73)	0.15*	(1.77)	-0.06	(-0.67)	-0.08	(-0.96)
4	0.56	(1.62)	0.14	(1.43)	0.19*	(1.82)	0.02	(0.19)	0.01	(0.08)
5	0.59*	(1.66)	0.17*	(1.85)	0.16*	(1.74)	0.07	(0.78)	0.08	(0.86)
6	0.60	(1.64)	0.20**	(2.09)	0.20**	(1.98)	0.14	(1.20)	0.14	(1.14)
7	0.48	(1.40)	0.06	(0.64)	0.08	(0.82)	0.12	(1.02)	0.09	(0.72)
8	0.56	(1.49)	0.13	(1.15)	0.18*	(1.85)	0.19	(1.54)	0.21	(1.50)
9	0.19	(0.44)	-0.29**	(-2.05)	-0.24*	(-1.70)	-0.15	(-0.94)	-0.18	(-1.07)
10	0.04	(0.10)	-0.43**	(-2.49)	-0.40**	(-2.54)	-0.29*	(-1.85)	-0.29*	(-1.82)
H-L	-0.71***	(-3.72)	-0.78***	(-4.23)	-0.74***	(-4.31)	-0.42**	(-2.27)	-0.40**	(-2.11)

Note: The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP using stocks from European countries. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the local one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, the Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). The regional factor portfolios and country classifications for European markets are taken from the Kenneth R. French Data Library. Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample period is from 2000 to 2023.

References

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge.

Daniel, K., D. Hirshleifer, and L. Sun (2019, 06). Short- and long-horizon behavioral factors. *The Review of Financial Studies 33*(4), 1673–1736.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020, 04). The cross-section of risk and returns. *The Review of Financial Studies 33*(5), 1927–1979.

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

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

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

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

Hou, K., C. Xue, and L. Zhang (2014, 09). Digesting anomalies: An investment approach. *The Review of Financial Studies 28*(3), 650–705.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023). Is there a replication crisis in finance? *The Journal of Finance 78*(5), 2465–2518.

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

centre for financial research cologne

cfr working papers are available for download from www.cfr-cologne.de.

2026

No.	Author(s)	Title
26-05	S. Weibels	Hard to Process: Atypical Firms and the Cross-Section of Expected Stock Returns
26-04	B. Ballensiefen, F. Somogyi, H. Winterberg	Demand for Dollars: Evidence from Survey Expectations
26-03	J. Fausch, M. Frigg, S. Ruenzi, F. Weigert	Machine Learning Mutual Fund Flows
26-02	T. G. Bali, A. Goyal, M. Mörke, F. Weigert	In Search of Seasonality in Intraday and Overnight Option Returns
26-01	S. Müller, N. Pugachyov, F. Weigert	Forecasting Mutual Fund Performance – Combining Return-Based with Portfolio Holdings-Based Predictors

2025

No.	Author(s)	Title
25-11	D. Aobdia, G. Köchling, P. Limbach, A. Yoon	Emissions Restatements after the SEC's Request for Public Input on Climate-Related Disclosures: Evidence from Carbon Disclosure Project Filings
25-10	V. Agarwal, J.-P. Gomez, K. Hosseini, M. Jha	ESG Metrics in Executive Compensation: A Multitasking Approach
25-09	V. Agarwal, P. Ghosh, N. Prabhala, H. Zhao	Animal Spirits on Steroids: Evidence from Retail Options Trading in India
25-08	V. Agarwal, S. Cao, S. Huang, M. Kim	Incentive Realignment: Mutual Funds' Influence on Executive Compensation Contracts
25-07	V. Agarwal, Y. E. Arisoy, T. Trinh	Eponymous Hedge Funds
25-06	D. Hess, F. Simon, S. Weibels	Interpretable Machine Learning for Earnings Forecasts: Leveraging High-Dimensional Financial Statement Data
25-05	B. F. Ballensiefen	Collateral Choice
25-04	C. Andres, F. Brochet, P. Limbach, N. Schumacher	Sell-Side Analysts with Accounting Experience
25-03	W. Bazley, G. Cici, J. Liao	Conflicts of Interest among Affiliated Financial Advisors in 401(k) Plans: Implications for Plan Participants
25-02	A. T. Maître, N. Pugachyov, F. Weigert	Twitter-Based Attention and the Cross-Section of Cryptocurrency Returns

25-01	N. Käfer, M. Mörke, F. Weigert, T. Wiest	A Bayesian Stochastic Discount Factor for the Cross-Section of Individual Equity Options

2024

No.	Author(s)	Title
24-06	V. Beyer, T. Bauckloh	Non-Standard Errors in Carbon Premia
24-05	C. Achilles, P. Limbach, M. Wolff, A. Yoon	Inside the Blackbox of Firm Environmental Efforts: Evidence from Emissions Reduction Initiatives
24-04	Ivan T. Ivanov, T. Zimmermann	The “Privatization” of Municipal Debt
24-03	T. Dyer, G. Köchling, P. Limbach	Traditional Investment Research and Social Networks: Evidence from Facebook Connections
24-02	A. Y. Chen, A. Lopez-Lira, T. Zimmermann	Does Peer-Reviewed Research Help Predict Stock Returns?
24-01	G. Cici, P. Schuster, F. Weishaupt	Once a Trader, Always a Trader: The Role of Traders in Fund Management

2023

No.	Author(s)	Title
23-08	A. Braun, J. Braun, F. Weigert	Extreme Weather Risk and the Cost of Equity
23-07	A. G. Huang, R. Wermers, J. Xue	“Buy the Rumor, Sell the News”: Liquidity Provision by Bond Funds Following Corporate News Events
23-06	J. Dörries, O. Korn, G. J. Power	How Should the Long-term Investor Harvest Variance Risk Premiums?
23-05	V. Agarwal, W. Jiang, Y. Luo, H. Zou	The Real Effect of Sociopolitical Racial Animus: Mutual Fund Manager Performance During the AAPI Hate
23-04	V. Agarwal, B. Barber, S. Cheng, A. Hameed, H. Shanker, A. Yasuda	Do Investors Overvalue Startups? Evidence from the Junior Stakes of Mutual Funds
23-03	A. Höck, T. Bauckloh, M. Dumrose, C. Klein	ESG Criteria and the Credit Risk of Corporate Bond Portfolios
23-02	T. Bauckloh, J. Dobrick, A. Höck, S. Utz, M. Wagner	In partnership for the goals? The level of agreement between SDG ratings
23-01	F. Simon, S. Weibels, T. Zimmermann	Deep Parametric Portfolio Policies

This document only covers the most recent CFR working papers. A full list can be found at www.cfr-cologne.de.

centre for financial research
cfr/university of cologne
Albertus-Magnus-Platz
D-50923 cologne
fon +49(0)221-470-6995
fax +49(0)221-470-3992
kempf@cfr-cologne.de
www.cfr-cologne.de