# Contextualizing Profitability

Alex G. Kim[*] Valeri V. Nikolaev[†]

June 2, 2024

**Abstract**

We study the role of context in asset pricing by focusing on the narratives surrounding profitability. Using a large language model, we incorporate narrative context into the measurement of profitability. Contextualized profitability outperforms conventional profitability measures both in statistical and economic terms. Its predictive power stems from the model's ability to learn transitory vs. persistent variation in profits. Furthermore, the factor based on contextualized profitability is superior in pricing portfolios of assets and eliminates alpha in small extreme growth portfolios – the biggest challenge facing the five-factor model (Fama and French, 2015). Our results imply that incorporating narrative context not only improves investment strategies but also enhances the asset pricing tests.

**Keywords:** Contextual information, asset pricing, machine learning, neural networks, large language models, BERT, Transformer, operating profitability, factor models

**JEL Codes:** C13, C45, C55, C58, G11, G12, M41

[*]The University of Chicago, Booth School of Business, alex.kim@chicagobooth.edu

[†]The University of Chicago, Booth School of Business, valeri.nikolaev@chicagobooth.edu

This paper was previously circulated under the title "Profitability Context and the Cross-Section of Stock Returns." We appreciate helpful comments from Eugene Fama, Kevin Kim (discussant), Ralph Koijen, Jongsub Lee, Jonathan Lewellen, Sanjog Misra, Stefan Nagel, and seminar participants at the University of Calgary, and the University of Chicago, and 2023 Korean Accounting Association International Summer Conference. Both authors gratefully acknowledge financial support from the University of Chicago Research Support Center.

# 1 Introduction

Characteristics in empirical asset pricing models are generally taken out of context. For instance, when two companies report identical earnings, the interpretation of these figures can significantly diverge depending on a variety of qualitative factors not captured by numeric information alone. Recognizing this, the Security and Exchange Commission (SEC) mandates that companies provide narrative disclosures alongside numerical reports to provide context to the reported numbers. These disclosures encompass discussions of significant economic events, including changes in strategic goals, product demand fluctuations, competitive shifts, production delays, supply chain disruptions, future plans, and many others. In this study, we examine the role of the narrative context for asset pricing. Focusing on operating profitability, we show that incorporating context substantially enhances the value of numeric characteristics and risk factors. Notably, contextualizing profitability goes a long way in addressing the five-factor model's biggest challenge – the pricing of small stocks with extreme growth (Fama and French, 2015).

The primary goal of the SEC's reporting requirement provided in the MD&A section of the annual and quarterly reports is to assist investors in forming more accurate expectations of future earnings and cash flows. These expectations are commonly represented in asset pricing models by firm characteristics. More specifically, in Fama and French (2015) five-factor model, profitability and investment characteristics aim to capture expected future profitability and investment. As Fama and French (2015) note, measuring these expectations poses an important challenge (see also Fama and French, 2006). To address it, recent studies have made adjustments to reported net income, aiming to eliminate transitory components. Notably, Novy-Marx (2013) proposes the use of gross profit as a proxy for future profitability, while Ball et al. (2015), among others, advocate for the use of operating profitability. More recently, Rouen et al. (2021) leverage proprietary data to construct a measure of "core earnings."

While refining profitability based on disclosed quantitative data is a logical step, the adjustments generally overlook the rich narrative context vital for interpreting the reported figures. For example, Li et al. (2013) show that companies' return on assets is more likely to mean revert when 10-Ks feature more extensive discussions of the competition. More recently, Kim and Nikolaev (2023) show that contextual information helps financial statements' users form expectations about a firm's future prospects. At the same time, asset pricing methodologies generally disregard this type of qualitative insight when measuring characteristics.

Consider the case of GameStop, an electronics and video game retailer, as an illustrative example. The company, known for selling game consoles and being a part of the S&P 400 index, enjoyed robust financial performance during the late 2000s and early 2010s, driven by the surging popularity of console games. In 2010 and 2011, GameStop was in the fourth quintile of operating profitability within its industry. However, the operating profitability declined sharply to −6% of total assets in 2012, placing the company in the first quintile. Nevertheless, the company's profits rebounded in the subsequent year. In such a situation, empirical asset pricing models would likely predict a low (or even negative) expected return following the release of the 2012 earnings. The company defied such expectations, however, delivering strong stock market performance in subsequent years.[1]

GameStop's numeric disclosures shed little light on its dip in profitability in 2012. Yet, a close examination of the narrative discussion in the firm's 2012 MD&A provides considerable insight. The company disclosed the conclusion of the present-generation console cycle and revealed the anticipated release of next-generation consoles (e.g., Xbox) in the near future. In fact, some console hardware had been on the market for over six years, leading customers to postpone purchases in anticipation of new consoles. This context explains a sharp but transitory decline in the 2012 product demand, which rebounded in 2013. The company also forecasted a new revenue stream anticipated from high-margin trade-ins of older consoles triggered by the release of new hardware. See Appendix A for the snippets excerpted from the MD&A.[2]

This example illustrates the value of context when analyzing reported performance (see Figure 2 for a summary). Note that these narrative disclosures are of limited value on their own; their primary value is in helping one to interpret and interpolate changes in reported profits. Importantly, the numeric information in GameStop's income statement does not provide insight into the decline in 2012 profitability. In this case, there is also no accounting adjustment or refinement that would allow one to eliminate the transitory shock from earnings.

In this paper, we aim to improve our understanding of the importance of context related to firm characteristics in the cross-section of stock returns by addressing the following questions. Does contextual information present opportunities to improve the asset pricing models? How important are the interactions between numeric characteristics and their context for asset pricing? Third, is it feasible for researchers to use language models and machine learning to construct meaningful contextualized characteristics and risk factors?

To address these questions, we focus on the narrative discussions of operating performance in the MD&A section of annual reports. We focus on operating profitability (rather than net income) because it already largely excludes the transitory (quantitative) components and can be viewed as an effective proxy for core earnings. Incorporating context into a measure of operating profitability (*OP*) poses a challenge, which we address in two steps. First, we quantify the inherently unstructured and highly multifaceted nature of narrative context related to operating performance

[1] Taking a long position in GameStop's stock on February 2nd, 2013, shortly after the release of its 2012 10-K report, and holding it until February 3rd, 2014, when the 2013 10-K report came out, yields a 96% buy-and-hold return.

[2] We also present other snippets from MD&As in Appendix B.

using Bidirectional Encoder Representations from Transformers (BERT). As a Large Language Model (LLM) that relies on Transformer architecture (akin to GPT), BERT encodes textual data into vectors (embeddings) and is effective in modeling contextual information (Devlin et al., 2018).[3]

In the second step, we develop a context-based proxy for future operating profitability (contextualized profitability), which incorporates the narrative content from the MD&A discussion pertaining to reported profits. Our approach utilizes the methodology developed in Farrell et al. (2021a), which employs artificial neural networks to model rich heterogeneity in the persistence of profits. Each year, we run a linear cross-sectional regression using the current operating profit to predict its own value one year ahead. Distinctively, the regression's parameters, including the intercept and slope, are not static but modeled as flexible functions of time-dependent textual input. In particular, these parameters are represented by deep neural networks that process textual vectors embedding the narrative context related to profitability.

The advantage of this approach over conventional densely connected neural networks is that it forces the model to focus on learning the *interactions* between the narrative information and numeric input known to be critical in forming expectations about the future (Kim and Nikolaev, 2023). This approach effectively teaches the model to discern and adjust for the less relevant (e.g., transitory) aspects of operating profits by leveraging the narrative information.

In addition to the contextualized profitability proxy, which capitalizes on the interactions between numeric and textual information, we also construct a 'text-only proxy' for future profitability. It follows the same regression-based approach but exclusively models the intercept as a function of text, excluding the operating profit and its interactions with contextual information. The intercept captures any direct information revealed by the narrative alone and can be viewed as an indicator of profitability sentiment. This is akin to forming expectations based on GameStop's 2012 discussions about product demand without incorporating numeric data and its context-dependent interpretations.

We train the models annually using a five-year moving window. We then combine the parameters from each training step with the available textual and numeric information up to time $t$ to generate out-of-sample forecasts of $t + 1$ operating profits. These forecasts rely on current profits as the primary input but, importantly, are refined to adjust for them for the related narrative context.

As a starting point for our empirical investigation, we verify the efficacy of contextual information by horseracing the three proxies for future profitability (scaled by total assets): (1) conventional operating profitability ($OP$), which uses the current value to proxy for future values; (2) a 'text-only' proxy ($OP^C$), which solely relies on contextual information; and (3) the contextualized profitability proxy ($OP^{CN}$), which incorporated operating profitability and its contextual narrative in a way that

[3]BERT is pre-trained on a large corpus of Wikipedia texts. Importantly, corporate communications or business press is not a part of BERT's training set, which ensures that the model did not see our textual data during its training.

allows for highly non-linear interactions between them. We find that $OP^{CN}$ distinctly excels at this task. Not only does it significantly outperform both measures, but it also subsumes them when the three variables are used to predict a year ahead profits out-of-sample. This result underscores that contextual information is likely valuable from an asset pricing perspective, not merely as a sentiment indicator but because it interacts with and thus alters the interpretation of reported profitability.

Equipped with our contextualized measure of profitability, we perform several asset pricing tests. We start with Fama and MacBeth (1973) regressions that evaluate the profitability proxies in explaining the cross-section of stock returns. When evaluated individually, contextualized profitability outperforms both the conventional $OP$ and the text-only proxy in predicting stock returns. Contextualized profitability also subsumes the predictive power of the other two proxies when the three variables are used simultaneously. This result is particularly noteworthy because we originally designed and trained the model to predict profitability, not returns, and yet it captures largely all return-relevant information contained in the other proxies. These results hold both in the entire sample and after excluding Microcaps. We provide further out-of-sample evidence on the robustness of these results in Section 9.1.

In addition to the standard definition of operating profitability (Fama and French, 2015), we explore eight other proxies for expected profitability, ranging from net income to gross profit and also allowing for a non-linear relation between current and expected future profits, multiple lags to filter out transitory components, and an alternative (even more flexible) way of incorporating context. In all cases, we find that our proposed contextualized profitability proxy dominates these refinements.

Besides statistical power, it is important to judge the economic significance of these results. To do so, we follow the approach used in Fama and French (2006). We use the predicted values from the Fama-MacBeth regressions to allocate stocks into above-median (high) and below-median (low) portfolios and then compute the spread in average returns between them. We observe an economically significant increase in the spread in both predicted and actual average returns between high-minus-low portfolios when we use contextualized profitability instead of unadjusted $OP$. In fact, the increment in the spread attributable to incorporating profitability context exceeds the increment from introducing unadjusted $OP$ into the model in the first place (conditional on other characteristics).

We hypothesize that the source of $OP^{CN}$'s superior performance lies in the model's ability to leverage the narrative information to distinguish between transitory vs. permanent fluctuations in profitability. When the narrative context predicts that changes in profitability are transitory (permanent), we show that the conventional $OP$ loses (retains) its ability to predict stock returns. In contrast, contextualized profitability remains statistically significant irrespective of the $OP$'s transitory nature, and in line with our hypothesis.

So far we have relied on Fama and MacBeth (1973) regressions, which assume linearity and equally weight observations. We relax these constraints by analyzing value-weighted portfolios. We sort the stocks annually, at the end of June, into ten portfolios based on (1) contextualized profitability and (2) unadjusted $OP$. Using CAPM, three-, four-, and five-factor models, we observe that alphas for $OP^{CN}$-based portfolios consistently dominate the corresponding alphas for $OP$-based portfolios. We also perform a double sort into quintiles based on contextualized profitability and, independently, on operating profitability, creating 5×5 portfolios. Implementing the three-factor model for each portfolio reveals that, within every $OP$ quintile, high-minus-low alpha for strategies based on contextualized profitability is significantly positive. Conversely, within each $OP^{CN}$ quintile, the $OP$-based high-minus-low strategy generates insignificant alphas. Collectively, the results corroborate the regression-based evidence of the dominance of contextualized profitability.

Next, we consider whether incorporating context when measuring characteristics can improve the asset pricing models. In particular, Fama and French (2015) state: "The five-factor model's main problem is its failure to capture the low average returns on small stocks whose returns behave like those of firms that invest a lot despite low profitability." As we discuss further, it is plausible that taking the context associated with low profitability when constructing factor portfolios can help explain these anomalous findings. To test this, we construct a factor $RMW^{CN}$ based on contextualized profitability and contrast its performance with the conventional profitability-based factor ($RMW$).[4]

Our spanning regressions show that $RMW$ is fully subsumed by the context-adjusted factor but not the other way around. This confirms that $RMW^{CN}$ contains at least as much information as $RMW$, in line with our prior findings. We also show that the five-factor model that relies on the contextualized profitability factor is more effective at pricing profitability-based portfolios.

More importantly, we then revisit the most problematic portfolio identified in Fama and French (2015) (see pp. 12-13). Specifically, we construct $5 \times 5$ portfolios based on *Size* and $B/M$ characteristics. Closely in line with their findings, we observe that the three-factor model yields the highest alpha of -0.55 for small, high-growth stocks and that the inclusion of the standard profitability and growth factors reduces this alpha to -0.31, which remains notably anomalous. Importantly, with the incorporation of the contextualized profitability factor, the alpha further decreases to -0.13 and its statistical significance dissipates, thereby significantly mitigating the anomaly.

This paper contributes to the literature in three ways. First, we establish an interplay between

[4]Adopting Fama and French (2015) methodology, we sort stocks into 2×3 portfolios. We sort stocks into two groups based on size and independently sort them into three groups on contextualized profitability. The factor returns are the difference in returns between the two high portfolios and two low $OP^{CN}$-based portfolios.

numeric information and its corresponding narrative context. Accounting for narrative context not only substantially enhances the predictive value of profitability relative to an array of proxies. Second, we show that the contextualized profitability factor significantly enhances the ability of the five-factor model to price securities. Importantly, it is successful in addressing the key challenge facing the five-factor model Fama and French (2015). Thrid, we introduce and empirically validate a novel approach that researchers can use to measure contextualized characteristics. Our context-based approach to measuring operating profitability consistently outperforms the conventional proxies and also yields a superior profitability factor. We also find that the suggested approach is more effective compared to an unrestricted feed forward ANN. Finally, our study extends the rapidly growing asset pricing literature that leverages machine learning (e.g., Ke et al., 2019; Gu et al., 2020, 2021). For instance, Gu et al. (2020) show the superiority of machine learning models compared to traditional methodologies due to non-linear interactions among the variables. While a number of prior studies used textual information to measure sentiment and predict stock returns (see Gentzkow et al., 2019, for review), our study pushes the boundary further by demonstrating the unique value of interacting textual and numeric data through large language models.

## 2 Profitability Context and Expected Stock Returns

Fama and French (2006) propose a unifying framework that helps to interpret the link between expected returns and firm characteristics. It relies on the following valuation equation:

$$M_t = \sum_{\tau=1}^{\infty} \frac{\mathbb{E}_t(Y_{t+\tau} - \Delta B_{t+\tau})}{(1+r)^{\tau}} \tag{1}$$

where $M_t$ is the market value of equity at the end of year $t$, $Y_{t+\tau}$ is the year $t+\tau$ earnings, $\Delta B_{t+\tau}$ is the change in book value of equity over the year $t+\tau$, which measures investment.

The equation predicts that expected stock returns $r$ will vary predictably with the $B_t/M_t$ ratio, future expected profits, and future expected investments. In particular, controlling for $B_t/M_t$ and expected investment, higher expected profitability implies higher expected returns. This framework also implies that variables that are helpful in predicting expected profitability or expected investment should also be useful in predicting stock returns.

While Fama and French (2006) find support for these predictions in cross-sectional regressions, the economic magnitudes of the combined effect of profitability and investment is relatively small economically. Furthermore, they find that *simple proxies*, such as the current level of profitability, dominate more complex proxies, namely, those based on the fitted values from a linear regression of future profitability on a broad set of firm characteristics.[5]

[5]Explanations for this finding boil down to considerable noise in the fitted values relative to true expectations.

Novy-Marx (2013) made an important step forward by noting that the above expression is stated in terms of firms' "true profitability," whereas the accounting earnings (net income) reported at the bottom of the income statement is often a misleading proxy for true earnings. Among other things, accounting earnings contain transitory components uninformative about future profits; earnings treat R&D as an expense whereas economically, it is an investment that increases future profits. To address this issue, he proposes to focus on the top profitability number in the income statement and shows that Gross Profit ($GP$) is a superior predictor of returns than net income. A number of studies have shown that profitability proxy can be further improved by making more refined accounting adjustments (e.g., Ball et al., 2015, 2016; Rouen et al., 2021).

In contrast to these studies, the focus on adjustments for 'softer' (non-quantitative) factors that nevertheless alter the interpretation of quantitative characteristics, e.g., profitability, has been missing from the literature. However, a narrative discussion that allows investors to evaluate, e.g., how rapidly the profits are likely to mean-revert, should be helpful in assessing the implications of current profits for future profits. In particular, a drop in current profits due to a temporary demand shock (like in the GameStop example) should be interpreted differently from a similar drop accompanied by significant changes in the competitive landscape. Such information is inherently hard to communicate quantitatively due to its non-systematic and unstructured nature. At the same time, it provides the context that affects the interpretation of numeric profitability values and what they imply for future stock returns.

## 2.1 Constructing a Context-Based Measure of Profitability

The challenge we are facing is to quantify and compress highly multidimensional narrative information in a way that allows us to refine the interpretation of numeric data. We turn to machine learning and the recent advances in language modeling to tackle this challenge.

We break the problem into two steps. Step one employs a pre-trained BERT to encode the portion of management's discussion related to operating performance (located in the MD&A part of 10-Ks) in the form of 768-dimensional vectors, also called as (contextualized) embeddings. Unlike uni-dimensional proxies such as tone or sentiment, text vectors encode the multifaceted meaning of the textual information.

One way to proceed from here could be to feed the profitability figures and textual embeddings as inputs into a standard densely connected feed-forward neural network, $f(c_{it}, OP_{it})$, and to train it to predict future profitability while allowing for rich interactions between textual and numeric data. In the absence of unlimited data, however, this approach is likely to face noise-related issues akin to those associated with more complex proxies in Fama and French (2006).[6] Intuitively, the dense

[6]Recall that they find that expected profitability proxies based on multiple predictors are dominated by the levels of (past) profitability presumably due to considerable estimation errors.

network would have to be very deep and would require vast amounts of data to learn the interactions between textual and numeric inputs. Thus, we follow an approach that imposes additional structure on the model, as we discuss next.

Specifically, in step two, we combine the simplicity of a univariate regression model with the power of a neural network by following the approach in Farrell et al. (2021a,b). Their methodology treats parameters in a regression model as deep neural nets, allowing them to be non-parametric functions of multidimensional input. Accordingly, we take a simple predictive regression that uses $OP$ to predict itself and hence generates a measure of expected future profits:

$$OP_{it+1} = \theta_0 + \theta_1 OP_{it} + \epsilon_{it}. \tag{2}$$

We use operating profitability because it already incorporates key quantitative adjustments to net income that aim to improve its accuracy as a proxy for future profits.

However, we further allow the model to "interpret" the reported $OP_t$ values differently, depending on their narrative context $c_{it}$ (text vector):

$$OP_{it+1} = \theta_0(c_{it}) + \theta_1(c_{it}) OP_{it} + \epsilon_{it} \tag{3}$$

where $\boldsymbol{\theta(c_{it})}$ is a flexible nonparametric function that represents a deep ANN.

The primary advantage of this structure, compared to a general functional form $f(c_{it}, OP_{it})$, is that, during the training, the model is effectively "forced" to focus on learning the interactions between text and numbers, i.e., whether and to what extent the level of $OP_{it}$ needs to be "qualified" (relative to its average value) when it comes to predicting $OP_{it+1}$. For example, when the context suggests the transitory nature of reported numbers, the model will learn to place less weight on the current level of OP and vice versa. Additionally, the model also adjusts the intercept, i.e., the level we expect unconditionally.[7] (In Section 9.2, we explore an alternative approach involving Ridge Regressions to implement this model.)

To estimate $\boldsymbol{\theta} = \{\theta_0, \theta_1\}$, we minimize the the solve following and obtain $\widehat{\boldsymbol{\theta}}$:

$$\widehat{\boldsymbol{\theta}} = \underset{\tilde{\theta} \in \mathcal{F}_{DNN}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left(OP_{it+1} - \theta_0(c_{it}) - \theta_1(c_{it}) OP_{it}\right)^2 \tag{4}$$

where $\mathcal{F}_{DNN}$ is the parameter space spanned by a neural network. The solution to this problem enables us to estimate $\mathbb{E}_t[OP_{it+1}|\Omega_t]$, where $\Omega_t$ is the information set that includes numeric profitability values and their context available to the public at time $t$.[8]

[7]Alternatively, we can interpret intercepts as adjusting the level of OP and slopes as adjusting the sensitivity of OP to the deviations from the intercept.

[8]It is important to note that the fact that equation 3 includes a lot more parameters than equation 2 does not imply

Based on these steps, we generate *contextualized operating profitability*, $OP^{CN}$, where the superscript "CN" highlights the interplay between context and numbers:

$$OP_{it}^{CN} \equiv \mathbb{E}(OP_{it+1}|OP_{it}, c_{it}; \widehat{\theta}(\cdot)) = \widehat{\theta}_0(c_{it}) + \widehat{\theta}_1(c_{it})OP_{it} \tag{5}$$

Returning to our example, the model indeed places less weight on GameStop's negative profitability based on the disclosed context and elevates the company from the 1st to 4th quintile of operating profitability, on par with its historical trends.

Note that $OP_{it}^{CN}$ can be decomposed into two parts, both of which depend on $c_{it}$: one that varies with $OP_{it}$ and one that does not. The invariant part can be interpreted as the sentiment that is extracted by the model from the $c_{it}$ about future profits, whereas the variable part captures the $interactions$ between the $OP$ characteristic and its context (i.e., concerns the interpretation of numeric information). Because the latter is of ultimate interest for this study, we also control for sentiment in some specifications. To do so, we directly construct a sentiment proxy by dropping the $OP_t$ term and re-training the model with the intercept only, such that:

$$OP_{it}^{C} \equiv \mathbb{E}(OP_{it+1}|c_{it}; \widehat{\theta}'(\cdot)) = \widehat{\theta}_0'(c_{it}) \tag{6}$$

where the superscript "C" highlights that the expectation is formed based on contextual information only.

# 3 Implementation and Training

## 3.1 Extracting Context Information with BERT

We extract narrative information about profitability from MD&A sections of all available annual reports. The objective of MD&A is to provide the relevant context that helps investors interpret operating performance. Using a keyword list from Kothari et al. (2009), we identify sentences that explicitly discuss profitability and retrieve these sentences. We also retrieve the sentence that precedes and the one that follows each sentence that explicitly mentions profitability to expand the context.

For each firm-year, we pool and process the identified sentences with BERT – a large language model developed by Google (Devlin et al., 2018). The model is particularly well-suited for our task as, in addition to encoding the meaning of words themselves, it is also designed to capture a deeper context within which words are used.

We use the base uncased model of BERT, which is pre-trained on a corpus of Wikipedia text.[9]

its will have a better predictive power out-of-sample, which is in line with the findings in Fama and French (2006).

[9]While there are domain-specific BERT models such as FinBERT (Huang et al., 2022), we refrain from their use

Although BERT's pre-trained parameters can be fine-tuned for a variety of applications, doing so is unnecessary in our case (and is effectively achieved) due to several layers of trained parameters added on top of the final BERT layer. The last-stage-hidden-state processed by BERT contains a 768-dimensional numeric vector that encodes contextual information. This vector is associated with a special classification (CLS) token, which is always placed at the beginning of the input text; this vector effectively summarizes the contextualized information in a given input text.

We pre-process the text by dropping special characters, tabular information, and graphics. Whenever a sentence contains numbers, we replace them with placeholder tokens to enable the model to learn about the presence of numeric information and its position. BERT can process 512 tokens (”words”) simultaneously. The profitability-related sentences, combined, are often longer than 512 tokens. In such cases, we divide the text into multiple 512-token passages (chunks) and run them through BERT independently. Following conventional practice, we then take the average of the embeddings associated with the CLS tokens to aggregate information. We refer to these embedding vectors as $c_{it}$.

## 3.2 Model Design

Textual vectors $\boldsymbol{c_{it}}$ serve as input into a deep neural net that we train to determine the adjustments for profitability context. The model's design is illustrated in Figure 1. Following the input layer with 768 neurons, the model includes three hidden layers with 256, 64, and 16 neurons, respectively, and one output layer with two neurons.[10] The output layer produces the two structural parameters in the predictive regression (4). The model is trained using root mean squared error (RMSE), which is given by the difference between $(\widehat{\theta}_{0it}(c_{it}) + \widehat{\theta}_{1it}(c_{it})OP_{it})$ and $OP_{it+1}$. All neurons except those in the output layer are activated with a ReLU activation function.

We train the model to predict one year ahead profitability based on a rolling window spanning four years ($t-4$, $t-3$, $t-2$, and $t-1$). We next use time $t$ data to construct an out-of-sample forecast of $OP_{i,t+1}$. This ensures that our model does not see any test data during training. Specifically, as our sample period is 1995-2020, we first train the model based on 1995-1998 data; subsequently, we use the profitability and its context information from 1999 to predict $OP$ in 2000.[11] We randomly assign 30% of each training sample as a validation set. To reduce overfitting, we use a 20% dropout rate. We stop the training when the loss function tested on the validation set does not improve by at

because their training set may overlap with corporate filings.

[10]The number of hidden layers and the number of neurons in each layer can be changed. Our results are qualitatively the same with several other specifications.

[11]Similarly, our last training session is based on the 2015-2018 period. We then take the data from 2019 to predict the profitability in 2020.

least 0.001 for ten consecutive epochs.[12] We use a learning rate of 0.001 and the Adam optimizer.[13]

This procedure generates the estimated functions $\theta_{0,it}(\cdot)$, $\theta_{1,it}(\cdot)$, and $\theta'_{0,it}(\cdot)$ used to construct time $t$ forecasts, $OP_{it}^{CN}$ and $OP_{it}^{C}$, of $\mathbf{E}_t(OP_{it+1})$ as discussed previously. See Figure 3 for the visual illustration of the model.

# 4 Data

Our sample consists of US-listed securities for the period 1995-2020; the SEC mandated the electronic submission of corporate filings in 1994. We include all firms with a valid MD&A extracted from the 10-K filings. Firm characteristics are taken from Compustat, and stock returns are from CRSP. We require that the stocks are traded in NYSE, Amex, and Nasdaq, and only include ordinary common shares. We exclude financial sector firms (firms with one-digit Standard Industry Classification code six). When delisting returns are missing on CRSP, we set them to -30% for NYSE and Amex stocks and -55% for Nasdaq stocks (Shumway, 1997; Shumway and Warther, 1999). The sample consists of firms with non-missing book value of equity, market value of equity, previous one-month return, previous one-year buy-and-hold return excluding the most recent month, total assets growth, and operating profit. Lastly, we require that each MD&A contains at least one sentence that discusses the results of operations, i.e., how profitable the company is.[14] We use firm characteristics from Compustat with a six-month lag relative to stock returns to ensure that accounting information is public and is incorporated into stock prices. To reduce the influence of outliers, we drop observations when the independent variables are below 1% or 99% in the distribution.

We use the conventional definition of operating profitability ($OP$): revenues less cost of goods sold, selling, general, and administrative expenses, and, following Fama and French (2015), we also subtract interest expense. We also use several alternative profitability proxies including gross profit (Novy-Marx, 2013), operating profitability that excludes R&D (Ball et al., 2015), and net income. The book value of equity is the total shareholders' equity, plus deferred taxes, investment tax credits, post-retirement benefit liabilities, and less preferred stock. When shareholders' equity is missing, we impute common share value or total assets minus total liabilities. When other balance sheet items (deferred taxes and investment credits) are missing on Compustat, we set their values to zero. Preferred stock is measured, depending on data availability, based on its redemption value, liquidation value, or carrying value, using this order.

[12]We set the maximum training to 250 epochs. However, our model does not reach the maximum epoch allowed for any of the training samples.

[13]These parameters can also be changed at the discretion of researchers. Slightly changing these parameters yields similar results.

[14]This restriction aims to ensure the validity of the extracted information. It is highly unlikely to be binding for valid observations since the objective of MD&A is to discuss the results of operations

For Fama and MacBeth (1973) regressions, we update momentum ($r_{0,1}$ and $r_{2,12}$) and market capitalization ($\log(ME)$) each month, and book value of equity ($\log(BE/ME)$) and investment ($Investment$) each quarter. Profitability measures are updated annually because the corresponding MD&A texts are published annually. We perform the regression analysis for the samples of all and all-but-microcaps. Microcaps refer to stocks that are below the 20th percentile of the NYSE market value, obtained from Ken French's data repository. To mitigate the influence of outliers, we truncate one percent of independent variables in our analyses (Ball et al., 2015, 2016). To facilitate comparisons, we employ a constant sample throughout the paper. In portfolio analysis, we rebalance portfolios on June 30th of each year.

Table 1 presents descriptive statistics for the variables used in the analyses. Operating profitability ($OP$) has a mean of 0.19 and a standard deviation of 0.29. On the other hand, $OP^{CN}$, which is our contextualized profitability measure, has a mean of 0.16 and a considerably lower standard deviation of 0.29. This fact is consistent with context-based adjustments removing transitory variation from the reported profits. The three profitability measures are positively correlated with each other (untabulated). $OP$ and $OP^{CN}$ exhibit a Pearson correlation of 0.67 (statistically significant under 1% significance level), whereas $OP$ and $OP^{C}$ have a much lower correlation of 0.06 (still statistically significant under 1% significance level).

## 5 Validating Contextualized Profitability

In this section, we validate the predictive ability of the three profitability measures. To do so, we run simple ordinary least squares regressions of future operating profitability on our proxies and a constant term. Note that the analysis is performed out-of-sample as the model trained without seeing the values of the target variable. We cluster standard errors two ways at the firm and year levels and report $t$-statistics.

Table 2 presents the results. The first column shows that the standard (unadjusted) $OP$ exhibits a coefficient of 0.514. Unsurprisingly, it is highly significant but is relatively far from unity, which is in line with the presence of significant transitory components in the unadjusted measure. The adjusted $R^2$ for this model is 25.49%. The second column features the text-only proxy of profitability $OP^{C}$, which captures the overall sentiment in contextual information. The slope coefficient in this column is even lower, 0.384, and it has an order of magnitude lower t-value.[15] Accordingly, the adjusted $R^2$ in the second column drops to 3.88% only. The third column evaluates our contextualized profitability measure, which delivers the strongest predictive power. Despite the forecasts being out-of-sample, the coefficient on $OP^{CN}$ is close to unity and the adjusted $R^2$ is

[15]Note that $OP^{C}$ does not consider any numeric information in its construction and that $OP^{C}$ and $OP$ are not mechanically correlated. Even so, we observe a significant positive association between current $OP^{C}$ and future $OP$, which implies that text information on its own has (some) predictive power with respect to future profitability as well.

30.03%.

Furthermore, when we include all three profitability proxies in one regression, the contextualized proxy, $OP^{CN}$, subsumes the other two measures and remains highly significant. $OP$ and $OP^{C}$ exhibit a large drop in their slope coefficients, which become statistically insignificant. The adjusted $R^2$ improves only slightly compared to column (3) (from 30.03% to 31.48%) in line with $OP^{CN}$ capturing largely all relevant information about future profits.

In sum, the contextualized profitability, $OP^{CN}$, is a more powerful predictor of future profitability, subsuming the unadjusted profitability measure in the prediction task. This result, in itself, is not surprising as the model was trained to excel in this task. Still, it showcases and validates the value of contextual information when interacting with the numeric data and also sets the stage for subsequent asset pricing tests.

# 6 The Cross-Section of Stock Returns

## 6.1 Fama-MacBeth Regressions

Our main analysis starts by examining the three measures of profitability using Fama and MacBeth (1973) regressions. Following prior studies (e.g., Novy-Marx, 2013; Ball et al., 2015), we regress monthly stock returns on our proxies for profitability and the following control variables: stock return for the prior month ($r_{0,1}$), stock return for the prior year excluding the most recent month($r_{2,12}$), log of the market value of equity ($\log(ME)$), log of book-to-market ratio ($\log(BE/ME)$), and investment ($Investment$). We multiply the regression coefficients by 100 to ease interpretation.

Table 3, Panel A, presents the results for the entire sample. In line with prior literature, Column 1 shows that operating profitability ($OP$) is a statistically significant predictor of stock returns with a slope of 0.91 ($t$=3.01). We note that the t-values throughout the table are somewhat lower compared to prior studies because our sample starts in 2000 (electronic filings are available since 1995, and we require five years of data to pre-train the first model), whereas many studies use data starting in 1963.

Column 2 replaces $OP$ with the text-only proxy, $OP^{C}$. We find a statistically insignificant coefficient on $OP^{C}$ and a noticeable drop in the $R^2$ from 3.94 % to 3.28%. This implies that the context alone does not have predictive power with respect to stock returns. In contrast, column 3 is based on contextualized profitability, $OP^{CN}$, which has a coefficient of 1.95 with $t$-value=4.23. Consistent with the notion that $OP^{CN}$ is a less noisy measure of expected profitability, its coefficient is considerably higher than that in column 1 (0.91 vs. 1.95), and so is the corresponding t-value (3.01 vs. 4.23). R-squared also exhibits a noticeable increase (from 3.94% to 4.45%). The signs and significance of the coefficients on other characteristics remain quantitatively similar and in line with prior studies.

Column 4 performs a horserace between $OP$ and $OP^{CN}$ by including them simultaneously. The coefficient on $OP$ drops sharply in absolute value and loses its statistical significance ($t$=-0.67), whereas the coefficient on $OP^{CN}$ increases in value and preserves statistical significance ($t$=2.29). Furthermore, when we include all three proxies of profitability in the fifth column, only the coefficient on $OP^{CN}$ remains statistically significant ($t$=2.34). In other words, we find that contextualized profitability subsumes information in the other two measures.

Panel B repeats the analysis excluding microcaps. We find a set of very similar results. As in Panel A, $OP^{CN}$ dominates the other profitability measures based on statistical significance and in terms of explanatory power. It also subsumes information in the other two measures.

Overall, our findings imply that taking the profitability context into account yields a superior proxy for expected profitability. It is noteworthy that the contextualized proxy subsumes information in the other proxies because the model was not trained to predict returns. To the extent that each predictor is subject to its own (independent) measurement error, one would expect to see some incremental value added by each of the profitability proxies. In contrast, we observe that $OP^{CN}$ captures largely all relevant information.

**Sub-periods Analysis** Figure 4 plots $t$-values associated with operating profitability and contextualized profitability based on Fama and MacBeth (1973) regressions estimated over a 10-year rolling window. The horizontal axis shows the end of the rolling window. Since our contextualized profitability is available starting from 2000, our first estimation period ends in January 2010. As of that time, both operating profitability and contextualized profitability show relatively high levels of statistical significance with $t$-values around 4. However, as time goes by, the $t$-values associated with operating profitability decline in magnitude, reaching 1.30 in 2020. At the same time, the $t$-values associated with contextualized profitability stay at approximately the same level throughout the sample period.

Overall, our evidence suggests that the results documented in Table 3 hold for different subperiods in our sample. Furthermore, the evidence suggests that while operating profitability is losing its predictive power over time, this is not the case for contextualized profitability.

**Alternative Profitability Measures** In the analysis above *OP* follows the definition in Fama and French (2015). It is interesting to explore whether contextualized profitability dominates a broader set of profitability measures and refinements. To this end, we explore eight additional proxies, ranging between net income and gross profit and also allowing for a more general functional form. In the order of sophistication, proxies *OP1*, *OP2*, and *OP3* are net income, gross profit (Novy-Marx, 2013), and operating profit adjusted for R&D (Ball et al., 2015). *OP4* is a 2-digits SIC industry adjusted (demeaned) *OP* (e.g., Soliman, 2004).

The remaining proxies rely on statistical procedures to estimate expected profitability. *OP5* is an expected profitability proxy that allows for differential persistence of positive vs. negative profitability based on the following regression: $OP_{t+1} = \beta_0 + \beta_1 OP_t + \beta_2 OP_t \times Neg_t + \beta_3 Neg_t + \varepsilon$, where $Neg$ is an indicator for negative profitability.[16] *OP6* is the predicted $OP_{it+1}$ value from a linear regression using current *OP* and its five lagged values as predictors. *OP7* uses the same five lags of *OP* but in a deep feed-forward neural net trained to predict $OP_{it+1}$ – a specification that allows for arbitrary functional form of the expectation function conditional on current and lagged profits.[17] Finally, the last proxy, *OP8*, allows for an arbitrary functional form in the relation between expected profitability, context vector ($c$), and current *OP*: $f(c, OP)$. This is achieved using a deep ANN, similar to $OP7$. Finally, we also explore the performance of Cash-based *OP* (Ball et al., 2016) vs. Cash-based $OP^{CN}$ in Appendix C.

In Table 4, we compare the performance of $OP^{CN}$ with the eight proxies. Our benchmark is the Fama-MacBeth regressions in Table 3. We find that all eight proxies exhibit positive and statistically significant associations with future returns (columns 1 to 8). However, once these proxies are included along with $OP^{CN}$ (columns 9 to 16), they generally lose statistical significance, whereas $OP^{CN}$ preserves its predictive power. Overall, our evidence suggests that the superiority of contextualized profitability generalizes to a number of proxies.

## 6.2 Economic Magnitudes

To place our results in perspective, we evaluate the economic significance of the superior information content of $OP^{CN}$ by following the approach in Fama and French (2006). We first partition our samples based on the fitted values from the Fama-MacBeth regressions. Specifically, we use the slopes reported in columns 1-3 of Table 3, Panel A, to obtain predicted monthly returns over the following year for each of the three models (each featuring a different profitability proxy). The predicted monthly returns are then used to assign stocks into high (above-median) vs. low (below-median) expected return portfolios. We then calculate the difference (spread) between the average *predicted* returns on high vs. low portfolios. We also calculate the same high-low return spread for the average actual returns.

Table 5 presents the high minus low spreads on equally-weighted (EW) and value-weighted

[16] We use rolling regressions of five years to estimate regressions each year (Fama and French, 2006).

[17] We use ANN model that takes five lags of profitability measures as its input. Similar to our baseline model, the ANN has two hidden layers with 64 and 16 neurons each and an output layer with a single neuron. The model predicts a one-period-ahead profitability value by allowing full non-linear interactions among lagged values of profitability. ReLU activation function is used for the first two layers and the sigmoid function is used in the last layer. The model minimizes the RMSE values. We use a rolling five-year training window and test it on the subsequent period. The resulting profitability is a fitted value considering all potential non-linear interactions of past profitability information and is expected to serve as a better proxy of future profitability.

(VW) portfolio returns. The first two columns are based on predicted returns, followed by two columns based on the actual returns. In the last two columns, we report $t$-values associated with the spreads based on the actual returns. Our discussion here will focus on value-weighted portfolios, although the results for equally-weighted portfolios are analogous.

Raw "0" in the table reports a benchmark Fama-MacBeth regression *without* profitability but with all other controls. For this model, the value-weighted predicted spread between high and low portfolios is 0.40, whereas the actual value-weighted spread is 0.35. Row 1 adds operating profitability ($OP$) into the model and re-computes the value-weighted spreads. The spread on high minus low predicted returns increases from 0.40 to 0.45, whereas the spread in actual returns increases from 0.35 to 0.42. The increase in spreads is economically significant and is in line with $OP$ being an important determinant of expected returns. Raw 2 replaces $OP$ with the text-only profitability proxy, $OP^C$. Not surprisingly, we observe a decline in spread across both predicted and actual portfolios. This happens because the text-only measure is a weak proxy for expected profitability.

The results differ starkly in row 3, where we employ $OP^{CN}$ as a proxy. We observe that the value-weighted spread in high minus low portfolio returns predicted by the cross-sectional model increases from 0.45 (the case of $OP$) to 0.54. In the case of actual returns, the spread in the high minus low portfolio increases from 0.42 (the case of $OP$) to 0.51. In both cases, the spread increment gained from contextualization (i.e., moving from $OP$ to $OP^{CN}$) exceeds the corresponding increment (from 0.40 to 0.45) from adding $OP$ into the model. Overall, the evidence indicates that adjusting operating profitability to incorporate its narrative context yields not only a statistically significant but also economically considerable gain in a large and well-diversified portfolio.

### 6.3 Sources of Predictive Ability of $OP^{CN}$

It is interesting to understand the factors responsible for the superior performance of contextualized profitability. We do not find that unidimensional textual attributes, such as tone, sentiment, or readability, can explain our results (untabulated). Our primary hypothesis is that textual information surrounding the disclosure of earnings is helpful for understanding differential persistence of *OP*, in line with our GameStop example. In particular, when the context suggests that *OP* will persist into the future, *OP* will perform relatively better in predicting future stock returns compared to when *OP* contains transitory shocks. In contrast, $OP^{CN}$, which already incorporates context, is expected to perform consistently regardless of contextual information about persistence.

To test this explanation, as a first step, we design an ANN model that uses the narrative context to predict whether the current *OP* will persist or reverse. Specifically, the input layer has 768 neurons (corresponding to context vectors) and the two hidden layers have 256 and 16 neurons each. The target variable is an indicator that equals one when the current $OP$ change has the same sign as $OP$ change over the subsequent year, and zero otherwise. This variable aims to capture whether the change in earnings is permanent or transitory.[18] When the estimated value is close to one, contextual information predicts that earnings will be persistent. When the estimated value is close to zero, contextual information predicts that $OP$ will be transitory and hence $OP$ should be dominated by $OP^{CN}$.

The results of this analysis are reported in Table 6. High persistence denotes the top quartile of predicted persistence values and low persistence denotes the bottom quartile of predicted persistence values. When predicted persistence is high, both $OP$ and $OP^{CN}$ can predict future stock returns. When they are simultaneously included in the same regression, only $OP^{CN}$ retains its statistical significance. In contrast, when persistence is expected to be low, $OP$ loses its statistical significance, while $OP^{CN}$ retains its significance. These findings suggest that the informativeness of the narrative context about the transitory vs. permanent variation in profitability is responsible for $OP^{CN}$'s superior performance.

### 6.4 Portfolio Sorts

Portfolio tests provide an alternative, potentially more robust, way of evaluating the predictive ability of profitability proxies. Specifically, portfolio sorts do not impose linearity nor make parametric assumptions as compared to Fama-MacBeth regressions. Additionally, portfolio sorts do not over-emphasize economically small stocks like linear regression does (due to equal weighting).

**Single Sorts** Table 7 reports the analysis based on single sorts and compares contextualized profitability to the standard operating profitability measure. We tabulate excess returns over the risk-free rate, CAPM alphas, three-factor model alphas (FF3), four-factor model alphas (FF4), and five-factor alphas (FF5) all of which are based on decile portfolios sorted on operating profitability. Portfolios are value-weighted to mitigate the undue influence of small stocks.[19] We also present long-short portfolio returns and their corresponding $t$-values.

Panel A shows portfolios based on contextualized profitability. The monthly excess return on the high-minus-low portfolio is positive and statistically insignificant (61 basis points with a $t$-value of 2.51); however, the average excess returns exhibit non-monotonic pattern.[20] In contrast, CAPM alphas, three-factor alphas, four-factor alphas, and five-factor alphas all display monotonically

[18]The model design is analogous to that used earlier. In particular, we use the ReLU activation function for the input layer and the first hidden layer and the sigmoid activation function for the second hidden layer. We use binary cross-entropy loss to train the model. The model also uses rolling training windows of four years. We use a learning rate of $1e$-5 and a batch size of 64.

[19]We find similar results using equal-weighted portfolios (untabulated).

[20]Note that Equation (1) does not imply that profitability necessarily predicts returns without controlling for other factors, such as book-to-market. Therefore, a sort based on only contextualized profitability may not produce a spread.

increasing patterns. High-minus-low portfolio monthly alphas range from 43 basis points (five-factor model) to 63 basis points (three-factor model), and they are all statistically significant ($t$-values ranging from 2.48 to 3.00). It is noteworthy that we observe a positive and economically significant alpha even after controlling for the profitability factor, which is constructed based on operating profitability (Fama-French five factors model). This result indicates that contextualized profitability contains useful information about expected returns not reflected by the unadjusted operating profitability.

Panel B tabulates the same analysis except that portfolios are formed on the conventional $OP$ proxy. In this panel, the high-minus-low portfolio excess return increases in magnitude (0.64) and reaches statistical significance (there is a large wedge between the 10th and 9th portfolios, however). We also observe statistically and economically significant alphas in the case of CAPM, three-factor, and four-factor models. However, in line with our prior findings, the alphas across these models are considerably lower as compared to those based on $OP^{CN}$ (Panel A). In particular, the high-minus-low portfolio CAPM alpha is 52 basis points (versus 45 basis points in Panel A), the three-factor alpha is 44 basis points (versus 63 basis points in Panel A), and the four-factor alpha is 42 basis points (versus 57 basis points in Panel A). Finally, the five-factor model's alpha is statistically insignificant, which is expected since the profitability factor is included in the model.

**Double Sorts** To shed further light on contextualized profitability, we also provide independent two-way sorts on $OP^{CN}$ and $OP$, tabulated in Table 8. In this analysis, we focus on the three-factor model alphas for 5×5 portfolios. For each $OP$-based quintile, we observe that the trading strategy based on contextualized profitability generates positive high-minus-low returns. These range from 58 basis points ($t$=2.68) for the highest operating profitability quintile to 73 basis points ($t$=3.57) for the lowest operating profitability quintile. Interestingly, we see a negative trend in the high-minus-low returns as we move from the lowest to the highest operating profitability quintile.[21] This result suggests that context information plays a more important role when profitability is low.

In contrast, for every quintile based on contextualized profitability, an $OP$-based trading strategy does not generate significant high-minus-low alpha. Specifically, high-minus-low alphas range from 5 to 33 basis points and all of them are statistically insignificant ($t$-values ranging from 0.03 to 1.47).

Overall, the portfolio analysis corroborates our Fama and MacBeth (1973) regression results. contextualized profitability carries useful information about the cross-section of stock returns beyond the conventional profitability proxies but not the other way around.

[21] In fact, the difference of the high-minus-low returns between the highest quintile and the lowest quintile is 15 basis points ($t$=2.03).

# 7 Context-Based Profitability Factor

As a next step, we construct a profitability factor that relies on contextualized profitability and explore its ability to price assets. We follow the methodology in Fama and French (2015) and form $2\times3$ portfolios as of the end of June each year. Specifically, we sort stocks into two groups (big and small) based on the median NYSE market capitalization breakpoints provided in Ken French's data library.[22] We independently sort stocks into three groups based on contextualized profitability. The high (low) group consists of stocks with contextualized profitability above (below) the 70th (30th) percentile breakpoint. The factor return is the difference in the value-weighted average return for two high portfolios and that for two low portfolios.

Panel A of Table 9 presents the annualized returns, annualized standard deviations, and the corresponding $t$-values for the factors used in our analysis. We include the five factors suggested by Fama and French (2015): return less risk-free rate ($MKT$), size ($SMB$), value ($HML$), investment ($CMA$), and profitability ($RMW$). We also include the contextualized profitability factor, which is denoted as $RMW^{CN}$. Compared to $RMW$, $RMW^{CN}$ has a higher annualized return (6.62% versus 4.80%) and a higher standard deviation (9.23% versus 10.18%). All factor returns are statistically significant at conventional levels. Interestingly, $RMW^{CN}$ has the highest t-statistics across the six factors ($t$=3.32), which is equivalent to having the highest Sharpe ratio.

Panel B of Table 9 presents the correlations among the factor returns. As one would expect, $RMW$ and $RMW^{CN}$ are highly correlated with a correlation coefficient of 0.71. However, it is also clear that contextual information makes the correlation between $RMW^{CN}$ and $RMW$ far from one. $RMW^{CN}$ and $RMW$ share similar relations with other factor returns. Both are positively correlated with $CMA$ and $HML$ and negatively correlated with $MKT$ and $SMB$.

## 7.1 Spanning Tests

To examine whether $RMW^{CN}$ and $RMW$ carry sufficiently different sets of information, we perform spanning tests reported in Panel C of Table 9. Models in columns 1 and 2 use monthly $RMW$ as the dependent variable and regress it on the other factor returns (with and without the context-based factor). When $RMW$ is regressed on factor returns excluding the $RMW^{CN}$ factor, we observe statistically significant factor loadings on $MKT$, $SMB$, and $HML$ factors. We also observe a statistically significant alpha of 0.56 (with *t-value*=4.54). This result is intuitive and implies that $RMW$ cannot be priced by the other four factors. This changes, however, in column 2, where we include the context-based factor $RMW^{CN}$ in addition to the other four factors. The alpha in this regression loses statistical and economic significance (-0.04 with $t$=-0.36), suggesting that the model now fully prices the risks associated with $RMW$. Additionally, while the coefficient

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

on $RMW^{CN}$ is highly significant, the loadings on other factors drop both in value and statistical significance.

In contrast, columns 3 and 4 use $RMW^{CN}$ as the dependent variable. When we regress $RMW^{CN}$ on the four factors, excluding the profitability factor, we also observe a significant alpha (0.61 with *t*=5.51). In fact, both economic and statistical significance is substantially higher compared to the case of $RMW$. Furthermore, when we include $RMW$ as an additional independent variable in the fourth column, we now still observe both statistically and economically a significant alpha (0.60 with *t*=5.35). This contrasts sharply with column 2 and indicates that the five conventional factors do not price the contextualized profitability factor.

We conclude that $RMW^{CN}$, in combination with other factors, span the $RMW$ factor but not the other way around. That is, $RMW^{CN}$ contains useful information about asset prices not reflected by the standard five factors, including the conventional profitability factor $RMW$.

## 7.2 Pricing of Profitability Portfolios

We next examine the ability of the constructed factor to price portfolios at a more granular level. We start with $5 \times 5$ portfolios based on operating profitability and contextualized profitability. Portfolio sorts are performed independently, in the same manner as in Table 8. Table 10 reports the results. For each portfolio, Panel A presents alphas based on (1) the five-factor model (upper part) and (2) based on the conventional five-factor model that replaces $RMW$ with the contextualized profitability factor $RMW^{CN}$.

When we consider the five-factor model, high-minus-low portfolios based on $OP$ generally do not generate returns that are significantly different from zero. One exception is the first quintile of contextualized profitability, where we observe a high-minus-low return of 37 basis points (*t*=2.35). In contrast, high-minus-low portfolio returns based on contextualized profitability are significantly positive across all profitability quintiles. As in Table 8, we see a declining pattern in contextualized profitability high-minus-low portfolio returns as we move across OP quintiles (56 basis points for the lowest profitability quintile versus 33 basis points for the highest profitability quintile), which suggests the importance of context for companies with lower reported profits.

The lower part of Panel A presents alphas from the *modified* five-factor model that replaces $RMW$ with $RMW^{CN}$. Now, in contrast to the above, none of the five high-minus-low portfolio returns based on contextualized profitability sorts are statistically significant. At the same time, operating profitability-based strategies also do not generate positive alphas conditional on either quintile of contextualized profitability. Overall, this suggests that, in line with the spanning tests, the context-based profitability factor prices portfolios more accurately.

To test this conclusion more formally, Panel B of Table 10 presents Gibbons et al. (1989) statistics and the averaged absolute values of alphas corresponding to each set of the 25 portfolios in Panel

A. Gibbons et al. (1989) (hereafter, GRS) statistics tests the probability that alphas generated from the factor models are jointly different from zero. For the five-factor model, the average absolute alpha ($A(|\alpha|)$) across the 25 portfolios is 17 basis points, and the GRS statistic is 2.86, which is significantly different from zero. In contrast, for the modified model that relies on contextualized $RMW^{CN}$ and the other four factors, the average absolute alpha decreases by about 60% to 0.08. The corresponding GRS statistic also decreases to 1.25, which is statistically indistinguishable from zero. Overall, the evidence implies that the five-factor model based on context-adjusted $OP^{CN}$ factor is more effective at pricing the profitability portfolios.

### 7.3 Pricing of the Problematic Portfolios

We next put our profitability proxy to the test by evaluating its performance in pricing the most challenging-to-price portfolios. Fama and French (2015) show that the smallest stocks with the highest growth present a major challenge for both three- and five-factor models. Specifically, they find for stocks in the smallest size and lowest *B/M* quintiles, three- and five-factor model alphas are -0.55 and -0.31, respectively. Based on their findings, it is plausible that measuring profitability factors with error is responsible for the failure of the five-factor model to price this portfolio. Indeed, the extreme portfolio has negative exposure to the profitability factor and to the extent this exposure is underestimated (in absolute terms), a significant negative alpha is expected. Furthermore, the accounting profitability for small firms with extreme growth is likely to be context-dependent due to the accounting challenges these firms are likely to face. For example, firms in the problematic portfolio are likely to invest more extensively in R&D projects, whereas accounting rules assume that such investments have no future value and consequently are treated as reductions in profit.[23]

To probe our model, we re-construct $5 \times 5$ *size* and *B/M* portfolios based on these characteristics. Specifically, as of June 30 each year, we sort stocks into five groups using NYSE market capitalization breakpoints. We independently sort stocks into five groups based on NYSE book-to-market breakpoints. For each of the 25 resulting portfolios, we estimate alphas using the three-factor model, five-factor model, and the modified five-factor model that relies on $RMW^{CN}$ factor.

For each of the three models, Table 11, Panel A, presents the estimated alphas and their corresponding t-statistics. Despite our sample period being considerably shorter, we very closely replicate the problematic portfolio results in Fama and French (2015). The three-factor model shows a large negative alpha of -0.55 in the small and low book-to-market portfolio (*t*-value -4.50). This alpha decreases by about 40% to -0.26 (*t*-value -3.26) when we add investment and profitability factors into the model. However, as we replace the $RMW$ factor with our $RMW^{CN}$ factor, the alpha in the problematic portfolio further declines to -0.08 and loses statistical significance (*t*-value

[23]Further, R&D projects are heterogeneous in terms of realization of future cash flows and their effect on long-term profitability is context-specific. Standard profitability measures cannot capture this context.

-1.00). This implies that incorporating the profitability context largely eliminates the well-known miss-specification of the five-factor models.

Panel B evaluates the model's performance across a broader set of portfolios by computing GRS statistic (Gibbons et al., 1989) and the average absolute value of alphas across each set of 25 portfolios. For the three-factor alphas, the average absolute value of alphas across portfolios, $A(|\alpha|)$, is 0.12, and the corresponding GRS test statistic is 3.80, which rejects the null that alphas are jointly zero. However, the average absolute alphas decrease to 0.09 when we use the five-factor model, and GRS statistic also declines to 2.73. Finally, the model with a contextualized profitability factor again delivers a considerable improvement, with $A(|\alpha|)$ of 0.06 and GRS statistic of 1.69, which is at the borderline of statistical significance. In fact, both the relative and absolute improvement in these performance metrics from incorporating profitability context is at least as large as the improvement associated with going from three to five factors in the model.

### 7.4 Ex Post Maximum Sharpe Ratio

We also evaluate how much an investor could gain in mean-variance efficiency from incorporating the context into profitability measurement. We examine ex post mean-variance efficient portfolios by comparing Sharpe ratios associated with different models. Sharpe ratios characterize a feasible set of investment opportunities that investors are facing. We construct ex post portfolios that maximize the mean-to-variance ratio using the results from Table 9. We begin with $MKT$ and gradually expose our portfolio to the five factors and $RMW^{CN}$. Specifically, we include $RMW^{CN}$ on top of the five factors to show an increase in investment opportunities by our new factor.

Table 12 presents optimal factor loadings along with the ex-post Sharpe ratios. The annualized Sharpe ratio for the market portfolio during our sample period is 0.46. When we add $SMB$, $HML$, and $CMA$ into the model, the maximum Sharpe ratio increases to 0.77. Now, when we include $RMW$ into the model, the Sharpe ratio improves further to 1.04. That is, an investor who trades the five factors can achieve an annualized mean-to-variance efficiency ratio of 1.04. When we replace $RMW$ with $RMW^{CN}$, the annualized Sharpe ratio improves to 1.27, which constitutes a 60% improvement relative to the four-factors model and a 19% improvement on top of the conventional five-factor model. Finally, jointly incorporating $RMW$ and $RMW^{CN}$ into the investment strategy increments the Sharpe by another 0.21 points to 1.48, or a 42% increase relative to the traditional five-factor model (from 1.04 to 1.48).[24]

Overall, this analysis suggests that investors benefit from incorporating profitability context into their investment decisions.

[24] The difference of the bias-adjusted squared Sharpe ratios between the traditional five-factor model (model 5) and the model with contextualized profitability (model 6) 0.23 with a p-value of less than 0.01 (Barillas et al., 2020).

# 8 Extending the Prediction Horizon

Our final analysis extends the prediction horizon beyond one year. Specifically, we rerun the Fama-MacBeth regressions presented earlier in Table 3, columns 1 and 3, while replacing the corresponding profitability proxies with their lags. Lags use six-month increments and go back up to five years.[25] The sample starts in January 2006 to ensure that all observations have at least five years of lagged data. We present the results graphically in Figure 5. Panels A and B plot regression coefficients and their 95% confidence intervals. Panel C plots the corresponding $t$-values.

Panel A shows the plot based on the contextualized profitability, which corresponds to the model specification in column 3 of Table 3. The zero-lag coefficient on $OP^{CN}$ is 1.95 with a $t$-value of 4.23. As profitability lag increases, i.e., information becomes stale over time, the coefficients decay in magnitude. Nevertheless, even after five years, we still observe a positive coefficient on $OP^{CN}$, implying that the profitability context (from five years ago) still appears to carry useful information not captured by the *current* characteristics. Panel C shows $t$-values, which also decay over time, remain statistically significant even at the four-and-a-half-year mark.

Panel B plots the coefficients for operating profitability, corresponding to the specification in column 1 of Table 3. Similar to Panel A, the coefficients decay over time as the information becomes less current. However, the coefficients are lower in magnitude and reach the borderline of statistical significance after about two years. After four years, they are virtually equal to zero. Panel C also shows that the corresponding $t$-values decay faster than those of contextualized profitability. [26]

Finally, we observe that the $t$-values corresponding to $OP^{CN}$ are generally higher than those corresponding to $OP$. Further, at lags zero and one, the $t$-values for the two measures are reasonably close. However, as the lag increases, the gap widens, suggesting that the value of information in unadjusted profitability decays more quickly. This further highlights the value added by incorporating the narrative context when measuring profitability or numeric characteristics more generally.

[25] Note that we keep all other regressors updated as in Table 3 while lagging operating profitability or contextualized profitability only. Doing so corresponds to a scenario where investors do not know the current value of profitability while having access to all other information. In conjunction with other equity characteristics, we examine whether stale profitability information can still predict future stock returns.

[26] In Ball et al. (2015), the coefficients on operating profitability remain statistically significant for about five years. The difference with the current result stems from the fact that the predictive power of operating profitability became lower in the 2010s. This can also be confirmed by the results in Figure 4. Our sample is also shorter, which reduces the power of the tests.

# 9 Additional Analyses

## 9.1 Post-BERT Out-of-Sample Analysis

In our main analysis, we follow several other studies and do not retrain BERT by year (e.g., see Jiang et al., 2022) because doing so is computationally very expensive. Instead, our approach is to add an ANN model on top of BERT and train it by year to learn the regression parameter functions $\theta_{0,t}(\cdot)$ and $\theta_{1,t}(\cdot)$. BERT's output thus serves as an input to our ANN.

Our tests are out-of-sample and should be immune to "in-sample" bias for the following reasons. We run BERT on corporate disclosures, which are markedly different from BERT's training data. BERT is trained on English Wikipedia articles and English textbooks with the objective of predicting masked words in sentences. By doing so, the model learns the general structure of language. This means that our model is tested on data that was *never* seen during the training. BERT is not trained on tasks such as sentiment analysis, which could have an ex-post correlation with the stock market. Instead, it is trained to uncover masked words in a sentence in a non-financial domain.

Nevertheless, one practical concern is that an investor, back in, e.g., 2010, cannot use BERT like we do because the language model is trained on data extending up to 2018. To address this, we conduct a post-BERT-training out-of-sample test implementable by an investor who has access to a pre-trained BERT model. While the exact content of BERT's training corpus remains undisclosed, we can deduce that it does not encompass text data post-October 2018.[27] Consequently, we confine the training period of ANN we use on top of BERT to 2019-2021, and we use 2022 as a test period. Such analysis represents a genuine out-of-sample test implementable by an investor.

We compute $OP^{CN}$ for operating profitability values published in 2022 (representing the fiscal year 2021's values). All other parameters remain unchanged. We replicate the single-sort portfolio analysis (portfolios created in June 2022 and retained until June 2023) and the Fama-MacBeth monthly regressions (from June 2022 to June 2023). This ensures all BERT outputs employed in our model are purely out-of-sample.

Our findings are presented in Table 13. Panel A displays monthly Fama and MacBeth (1973) regressions. In line with Table 3, we observe significant positive coefficients for both $OP$ (column (1)) and $OP^{CN}$ (column (3)). As previously, the coefficient on $OP^{CN}$ is substantially higher (0.46 vs. 2.68) and exhibits a higher $t$-value. The adjusted R-squared increases from 4.02% to 4.26% when switching from $OP$ to $OP^{CN}$. Additionally, when both $OP$ (column (1)) and $OP^{CN}$ are simultaneously included in the model (column (4)), $OP$ loses statistical significance while $OP^{CN}$ retains its. As in Table 3, $OP^{C}$ lacks significant predictive power, and when all three variables are used in one regression (column (5)), only $OP^{CN}$ retains statistical significance.

We also complement regression analysis by performing portfolio sorts based on $OP^{CN}$ and $OP$

[27]BERT became publicly available in October 2018, and its training took four days on multiple TPUs.

in Panels B and C, respectively. Panel B reveals that portfolios (established on June 30th) based on $OP^{CN}$ consistently yield positive long-short returns. In Panel C, portfolios formed based on conventional profitability proxy generally yield lower returns. Notably, the long-short alpha from the Fama and French (2015)'s five-factor model reported in the last column is only significant in the case of contextualized profitability (0.77 in Panel B vs. 0.44 in Panel C).

In summary, even with a limited dataset that strictly excludes BERT's training period, our results imply the superiority of the contextualized proxy for profitability and indicate the high relevance of contextual information for asset pricing.

### 9.2 Ridge Regression Benchmark

We also explore an alternative strategy to measuring $OP^{CN}$ that relies on ridge regressions instead of ANN. Our goal is to implement equation (3) in the paper using the Ridge regression methodology. Accordingly, we estimate $\theta_0(c_{it})$ and $\theta_1(c_{it})$ as linear functions that take BERT embeddings $c_{it}$ as inputs and regularize the estimated parameters using L2 penalty.[28] Training and testing datasets, along with all other training parameters, are identical to those of the original model. This setup makes it easy to compare the Ridge estimates with those based on the Farrell et al. (2021b) approach.

We employ the same dataset as in our primary analyses, replicating Table 3 (Fama-MacBeth regressions) and Table 7 (single-sort portfolios). Appendix D, Panel A shows regression results using the new $OP^{CN}$, derived from the Ridge model. Columns (1)-(3) repeat Table 3's findings for reference. Column (4) exhibits a positive and statistically significant coefficient on $OP^{CN}$. However, it has a lower t-value compared to the unadjusted $OP$. Further, when both $OP$ and $OP^{CN}$ are included in column (5), the $OP^{CN}$ loses its statistical significance.

Panel B reports portfolio tests using the Ridge-based $OP^{CN}$. For readers' convenience, we repeat baseline outcomes (employing traditional operating profitability) and ANN-based outcomes. While the Ridge-based measure yields positive long-short returns, the alphas are comparable to the baseline and are smaller than the ANN-based alphas. Furthermore, after adjusting for the five risk factors, the Ridge-based long-short return becomes statistically insignificant.

In sum, we find that Ridge-based estimation is less effective at incorporating profitability context, which is not unexpected given that it does not incorporate deep non-linear interactions between textual (and numeric) characteristics.

[28] We set the regularization parameter to 0.005, which we believe to be a reasonable value; however, we have not performed a formal analysis to ensure its optimality. Note, however, that choosing optimal hyperparameters of ANN architecture is outside of the scope of this paper either.

## 10 Conclusion

Empirical asset pricing models generally overlook the narrative context of the firm (stock) characteristics they rely on. At the same time, companies produce vast amounts of qualitative information that shapes the interpretation of quantitative characteristics, such as discussions on competitive pressures or fluctuations in demand. We study whether integrating the contextual narratives surrounding reported figures enhances our ability to measure firm characteristics and, as a result, helps to explain the cross-section of stock returns more effectively. Leveraging recent breakthroughs in language modeling and deep learning, we develop a contextualized measure of operating profitability. This approach moves beyond the 'one-size-fits-all' method of evaluating firm characteristics, capturing the rich contextual diversity found in corporate communications.

The construction of our measure involves two components. First, we use a large language model (BERT) to systematically encode multidimensional contextual information from the Management Discussion and Analysis (MD&A) sections of annual reports. Subsequently, we use the approach in Farrell et al. (2021a), which relies on a deep neural network to discern how narrative context affects the interpretation of numeric data. The model enables us to incorporate the rich heterogeneity in the context surrounding the reporting of operating profits across firms and over time.

We perform Fama-MacBeth regression analysis and carry out portfolio sorts to demonstrate that contextualized profitability dominates and subsumes the traditional measures of operating profitability in explaining the cross-section of future stock returns. Our measure also carries more information and hence provides more insights compared to the conventional measure of operating profitability when it comes to predicting longer-term returns. Beyond statistical significance, we also demonstrate the economic relevance of incorporating context into profitability measurement. In particular, the impact of incorporating contextual information is at least as important as the initial inclusion of the profitability characteristic itself.

We also show that the risk factor $RMW^{CN}$ constructed based on contextualized profitability delivers superior performance as compared to the traditional $RMW$ factor. Most notably, the contextualized factor goes a long way in resolving the biggest challenge facing the five-factor model identified in Fama and French (2015).

Overall, our study is the first to demonstrate the value and significance of incorporating context when measuring asset pricing characteristics and constructing factors. The methodology we developed for contextualized profitability is not only robust but readily adaptable, offering a valuable tool for asset pricing researchers. Furthermore, it provides a framework that can be extended to explore the importance of the context of other corporate characteristics and can be used outside the asset pricing field.

## References

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. Nikolaev (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics 121*(1), 28–45.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. V. Nikolaev (2015). Deflating profitability. *Journal of Financial Economics 117*(2), 225–248.

Barillas, F., R. Kan, C. Robotti, and J. Shanken (2020). Model comparison with Sharpe ratios. *Journal of Financial and Quantitative Analysis 55*(6), 1840–1874.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805*.

Fama, E. F. and K. R. French (2006). Profitability, investment and average returns. *Journal of Financial Economics 82*(3), 491–518.

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

Farrell, M. H., T. Liang, and S. Misra (2021a). Deep learning for individual heterogeneity: An automatic inference framework.

Farrell, M. H., T. Liang, and S. Misra (2021b). Deep neural networks for estimation and inference. *Econometrica 89*(1), 181–213.

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–74.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, 1121–1152.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The 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), 429–450.

Huang, A. H., H. Wang, and Y. Yang (2022). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*.

Jiang, J., B. T. Kelly, and D. Xiu (2022). Expected returns and large language models. *Available at SSRN*.

Ke, Z. T., B. T. Kelly, and D. Xiu (2019). Predicting returns with text data. Technical report, National Bureau of Economic Research.

Kim, A. G. and V. V. Nikolaev (2023). Context-based interpretation of financial information. *The University of Chicago Working Paper*.

Kothari, S. P., X. Li, and J. E. Short (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review 84*(5), 1639–1670.

Li, F., R. Lundholm, and M. Minnis (2013). A measure of competition based on 10-K filings. *Journal of Accounting Research 51*(2), 399–436.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108*(1), 1–28.

Rouen, E., E. C. So, and C. C. Wang (2021). Core earnings: New data and evidence. *Journal of Financial Economics 142*(3), 1068–1091.

Shumway, T. (1997). The delisting bias in CRSP data. *The Journal of Finance 52*(1), 327–340.

Shumway, T. and V. A. Warther (1999). The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *The Journal of Finance 54*(6), 2361–2379.

Soliman, M. T. (2004). Using industry-adjusted DuPont analysis to predict future profitability. *Available at SSRN 456700*.

## Appendix A. Snippets from 2012 GameStop 10-K Report

This table displays snippets from GameStop's 10-K for the fiscal year 2012. GameStop identified the causes of its abrupt decline in profitability and explained why it was expecting a quick recovery based on mainly three topics: the aging of gaming consoles, investment in e-commerce, and its unique trade-in policy. In this table, we report selected texts excerpted from GameStop's 2012 10-K pertinent to each topic.

| Topic | Text |
|---|---|
| Console aging | Historically, new hardware consoles are typically introduced every four to five years. However, the current generation of hardware consoles is now over six years old and consumer demand is declining. We have seen declines in new hardware and software sales in fiscal 2012 due to the age of the current console cycle. The introduction of new consoles, like the Wii U, or further price cuts on the current generation of consoles could partially offset these declines. |
| E-Commerce bisiness | We expect that future growth in the electronic game industry will also be driven by the sale of video games delivered in digital form and the expansion of other forms of gaming. We currently sell various types of products that relate to the digital category, including digitally downloadable content, Xbox LIVE, PlayStation and Nintendo network points cards, as well as prepaid digital and online timecards. We expect our sales of digital products to increase in fiscal 2013. We have made significant investments in e-commerce, digital kiosks and in-store and Web site functionality to enable our customers to access digital content easily and facilitate the digital sales and delivery process. |
| Trade-in products | Our buy-sell-trade program offers consumers the opportunity to trade-in pre-owned video game products in exchange for store credits applicable to future purchases, which, in turn, drives more sales. [. . . ] Our trade-in program also allows us to be one of the only suppliers of previous generation platforms and related video games. [. . . ] Pre-owned video game products generate significantly higher gross margins than new video game products [. . . ] We expect that trade-ins will be negatively impacted in the time period between recent and expected announcements of next-generation consoles and the launch of these consoles. We expect the launch of next-generation consoles and software to drive trade-ins of older video game products, thereby expanding our supply of pre-owned video game products. |

# Appendix B. Other Examples of Context

This table displays a number of snippets from corporate 10-K disclosures where the conventional and contextualized measures of profitability disagree. Current $OP$ denotes the quintile rank of current operating profitability, Current $OP^{\mathrm{CN}}$ is the quintile rank of current contextualized profitability, Future $OP$ is the quintile rank of one-year-ahead operating profitability, and Text is the actual snippet from MD&A sessions. Texts that are deemed relevant are boldfaced and underlined.

| Current OP | Current $\mathrm{OP}^{\mathrm{CN}}$ | Future OP | Text |
|---|---|---|---|
| 2 | 5 | 5 | [. . . ] As outlined in Notes 2 and 14 of Notes to Consolidated Financial Statements, included in Item 8. "Financial Statements and Supplementary Data," pursuant to our Second Amended Plan of Reorganization, or our Plan, **we emerged from chapter 11 bankruptcy on July 6, 2006** with all of our fabricated products facilities and operations and a 49% interest in Anglesey, which owns a smelter in the United Kingdom. [. . . ] |
| 2 | 5 | 5 | [. . . ] Beginning in the second quarter of 2005, the Anglesey-related operating results were adversely affected by an approximate 20% increase in contractual alumina costs. However, **contractual pricing for alumina is expected to improve approximately 20%** (versus 2006) **beginning in the second quarter of 2007**. [. . . ] |
| 2 | 5 | 5 | [. . . ] **Margins for our rental business have recently averaged 59% to 62%, while margins for the compressor sales business have recently averaged approximately 25% to 29%**. Our strategy for growth is focused on our compressor rental business, and **we intend to use the additional fabrication capacity available from our acquisition of SCS to expand our rental fleet** while continuing the core custom fabrication business that SCS established. As our rental business grows and contributes a larger percentage of our total revenues, **we expect our overall margins to improve from those experienced in 2007**. [. . . ] |
| 2 | 5 | 4 | [. . . ] In general, we expect our overall business activity and revenues to **track the level of activity in the natural gas industry**. [. . . ]We believe demand will remain strong throughout 2008 due **to high oil and natural gas prices and increased demand for natural gas**. [. . . ] we believe the long-term trend of activity in our markets is favorable. |
| 2 | 5 | 5 | [. . . ] For 2006, we reported a 5% increase in annual production volumes over the 2005 period. We also replaced 97% of our total 2006 production [. . . ]On January 31, 2007, we completed the purchase. of certain oil and natural gas properties located in 13 counties in south and southeast Texas and other assets from Smith Production Inc. [. . . ]We believe this area to be **a continued target of exploitable potential** for us in future years. |
| 5 | 2 | 2 | [. . . ] Due to the uncertain nature of litigation in general, we are unable to estimate a range of possible loss related to lawsuits filed against us, but based on our historical experience and consultation with counsel, we typically reserve an amount we believe will be sufficient to cover any damages assessed against us. We have accrued $483 and $1,958 for potential arbitration and lawsuit losses as of December 31, 2006 and 2005, respectively. However, we have in the past been assessed damages that exceeded our reserves. If we misjudged the amount of damages that may be assessed against us from pending or threatened claims, **or if we are unable to adequately estimate the amount of damages that will be assessed against us** from claims that arise in the future and reserve accordingly, our operating income would be reduced. **Such costs may have a material adverse effect** on our future financial position**, results of operations or liquidity**. [. . . ] |

| | | | |
|---|---|---|---|
| 5 | 2 | 2 | [. . . ] We do not currently have any commercial products for sale. To date, **our revenues have been derived principally from our collaboration** and license agreements and Small Business Innovative Research grants. In the future, we believe our revenues will consist of milestone payments, technology licensing fees and sponsored research fees under existing and future collaborative arrangements, royalties from collaborations with current and future strategic partners and commercial product sales. **Because a substantial portion of our revenues for the foreseeable future will depend on achieving development and clinical milestones**, our results of operations may vary substantially from year-to-year and quarter-to-quarter. We believe that **period-to-period comparisons of our operating results are not meaningful** and you should not rely on them as indicative of our future performance. [. . . ] |
| 5 | 2 | 2 | [. . . ]We have depended upon equity and debt financings and to a lesser degree license fee, **research and development and milestone payments** from our collaborative partners and licensees to fund our research and product development programs and **expect to do so for the foreseeable future**.[. . . ] Our operating expenses consist primarily of royalty expense, costs associated with research and development and general and administrative costs associated with our operations. [. . . ] **We expect research and development expense to increase substantially in the foreseeable future**. We expect that a large percentage of this will be incurred in support of our clinical trial programs and toxicology studies for bicifadine, ocinaplon, DOV diltiazem, DOV 216,303 and DOV 21,947, as well as our product candidates in our preclinical program if they progress into clinical trials. [. . . ] |
| 5 | 2 | 2 | [. . . ] On January 16, 2009, we entered into an amendment to the securities class action settlement agreement. Under the terms of the settlement agreement in the securities class action, as amended, which is subject to notice to the shareholder claimants and court approval, our insurance carrier will make a cash payment, we will make cash payments of $466,667 in January 2009 and $233,333 in May 2009, and we will contribute one million shares of our common stock. The shares being contributed to the settlement will be distributed to the settlement claimants if and when the court grants final approval of the settlement and the settlement becomes effective. [. . . ] **There is no assurance that the settlements described above will be achieved**, and if not achieved, there can be no assurance that our insurance will be adequate to cover our costs relating to the litigation. **Any expenses incurred in connection with the litigation not covered by available insurance** or any adverse resolution of such litigation could have a **material adverse effect on our financial condition** and future viability. [. . . ] |
| 5 | 2 | 2 | [. . . ]We recognize revenue in accordance with the Financial Accounting Standards Board (FASB) Subtopic ASC 605-25, ”Revenue Recognition—Multiple-Element Arrangements.” As of January 1, 2011, we adopted on a prospective basis the accounting updates to guidance ASC 605 ”Revenue Recognition”, subtopic ASC 605-25 ”Revenue with Multiple Element Arrangements” and subtopic ASC 605-28 ”Revenue Recognition-Milestone Method” [. . . ] However, **these updates will result in different accounting treatment for future new collaboration arrangements** and substantive milestones earned after the dates of adoption. [. . . ] For multiple-element arrangements entered into prior to January 1, 2011, **we determined that the deliverables under our collaboration agreements** with GSK and Astellas **did not meet the criteria required for separate accounting units for the purposes of revenue recognition**.[. . . ] |

# Appendix C. Cash-Based Operating Profitability

This table reports the predictive ability of cash-based operating profitability and context-adjusted cash-based operating profitability. We obtain cash-based operating profitability as $OP – \Delta RECT – \Delta INVT – \Delta XPP + \Delta(DRC + DRLT) + \Delta(AP) + \Delta(XACC)$ where $\Delta$ denotes the changes in each variable, *RECT* is accounts receivable, *INVT* is inventory, *XPP* is prepaid expenses, *DRC+DRLT* is deferred revenue, *AP* is trad accounts payable, and *XACC* is accrued expenses. We include lagged one-month returns, lagged one-year returns skipping prior month return, firm size, book-to-market ratio, investment as additional stock characteristics. We use a consistent sample as in Table 3.

| | (1) | (2) | (3) |
|---|---|---|---|
| Cash-Based *OP* | 2.36 | | 1.80 |
| | [6.52] | | [2.15] |
| Cash-Based $OP^{CN}$ | | 3.05 | 2.53 |
| | | [7.36] | [6.17] |
| Controls | Yes | Yes | Yes |
| Average Adjusted $R^2$ | 4.05% | 4.85% | 5.13% |

# Appendix D. Ridge Benchmark

This table reports contextualized profitability obtained from Ridge-style estimates of equation 3. We parametrize $\theta_0(c_{it}))$ and $\theta_1(c_{it}))$ as linear functions and add L2 penalty to regularize their parameters. The regularization parameter is set to 0.005. Training and testing splits and all other training parameters are identical to the original model. In Panel A, we repeat Fama and MacBeth (1973) regressions with the new $OP^{CN}$. In Panel B, we form portfolios by sorting on Ridge-based $OP^{CN}$ and present long-short returns. We also present long-short portfolio returns of the portfolios sorted on baseline operating profitability (columns denoted as Base) and our original contextualized profitability (columns denoted as ANN).

**Panel A**. Fama-MacBeth Regressions

| | Baseline | ANN | | Ridge Benchmark | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $r_{0,1}$ | -2.26 | -2.43 | -2.46 | -2.35 | -2.32 |
| | [-3.09] | [-3.42] | [-3.49] | [-3.25] | [-3.16] |
| $r_{2,12}$ | -0.10 | -0.13 | -0.15 | -0.11 | -0.12 |
| | [-0.30] | [-0.38] | [-0.44] | [-0.35] | [-0.37] |
| log(*ME*) | -0.16 | -0.18 | -0.19 | -0.16 | -0.17 |
| | [-2.98] | [-3.57] | [-3.74] | [-2.99] | [-3.35] |
| log(*BE/ME*) | 0.02 | -0.05 | -0.04 | -0.03 | -0.02 |
| | [0.17] | [-0.39] | [-0.34] | [-0.21] | [-0.22] |
| *Investment* | -1.70 | -1.68 | -1.70 | -1.81 | -1.85 |
| | [-4.52] | [-4.45] | [-4.47] | [-4.62] | [-4.70] |
| *OP* | 0.91 | | -0.32 | | 0.85 |
| | [3.01] | | [-0.67] | | [2.96] |
| $OP^{CN}$ | | 1.95 | 3.26 | 1.13 | 1.32 |
| | | [4.23] | [2.29] | [2.05] | [1.46] |
| Average Adj $R^2$ | 3.94% | 4.45% | 4.68% | 4.02% | 4.40% |

**Panel B**. Portfolio Sorts

| | | | | Alphas | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Excess Return | | | CAPM | | | FF3 | | | FF4 | | | FF5 | | |
| | Base | ANN | Ridge | Base | ANN | Ridge | Base | ANN | Ridge | Base | ANN | Ridge | Base | ANN | Ridge |
| Low | 0.34 | 0.35 | 0.46 | -0.38 | -0.22 | -0.25 | -0.30 | -0.35 | -0.30 | -0.26 | -0.31 | -0.31 | -0.04 | -0.21 | -0.13 |
| High | 0.98 | 0.96 | 1.05 | 0.14 | 0.23 | 0.11 | 0.14 | 0.28 | 0.21 | 0.16 | 0.26 | 0.16 | 0.26 | 0.22 | 0.25 |
| H-L | 0.64 | 0.61 | 0.59 | 0.52 | 0.45 | 0.36 | 0.44 | 0.63 | 0.51 | 0.42 | 0.57 | 0.47 | 0.30 | 0.43 | 0.38 |
| *t-value* | [1.80] | [2.51] | [1.65] | [2.63] | [2.61] | [1.69] | [2.48] | [3.00] | [2.72] | [2.38] | [2.90] | [2.60] | [1.24] | [2.48] | [1.56] |

## Figure 1. Operating Profitability of GameStop

This figure shows the trend in GameStop's operating profitability from 2010 to 2016. Bars represent the level of operating profitability, and the solid line represents the quintile rank of operating profitability. Quintile rank is obtained based on quarter-industry two-digit SIC groups.

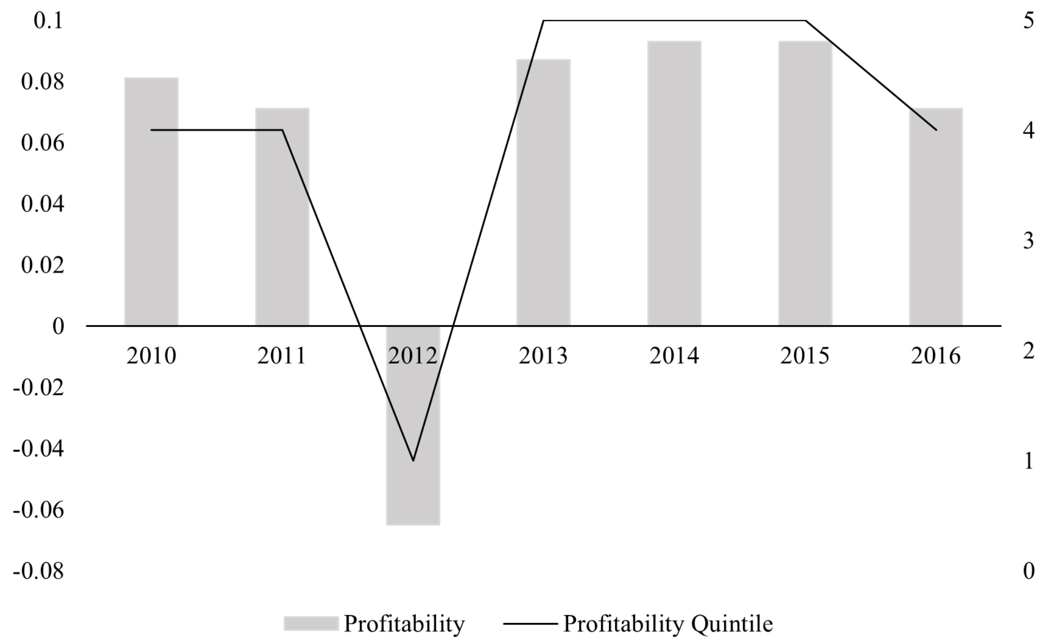



Figure 1. Operating Profitability of GameStop from 2010 to 2016

## Figure 2. Return Trends and Event Timeline of GameStop in 2013

This figure illustrates the stock price movement of GameStop in 2013, along with some notable events. The solid line represents the movement in stock prices. The figure spans from February 2013 (shortly after the release of the 10-K) to December 2013. Pertinent corporate events are marked with dashed lines and further elaborated in text boxes.

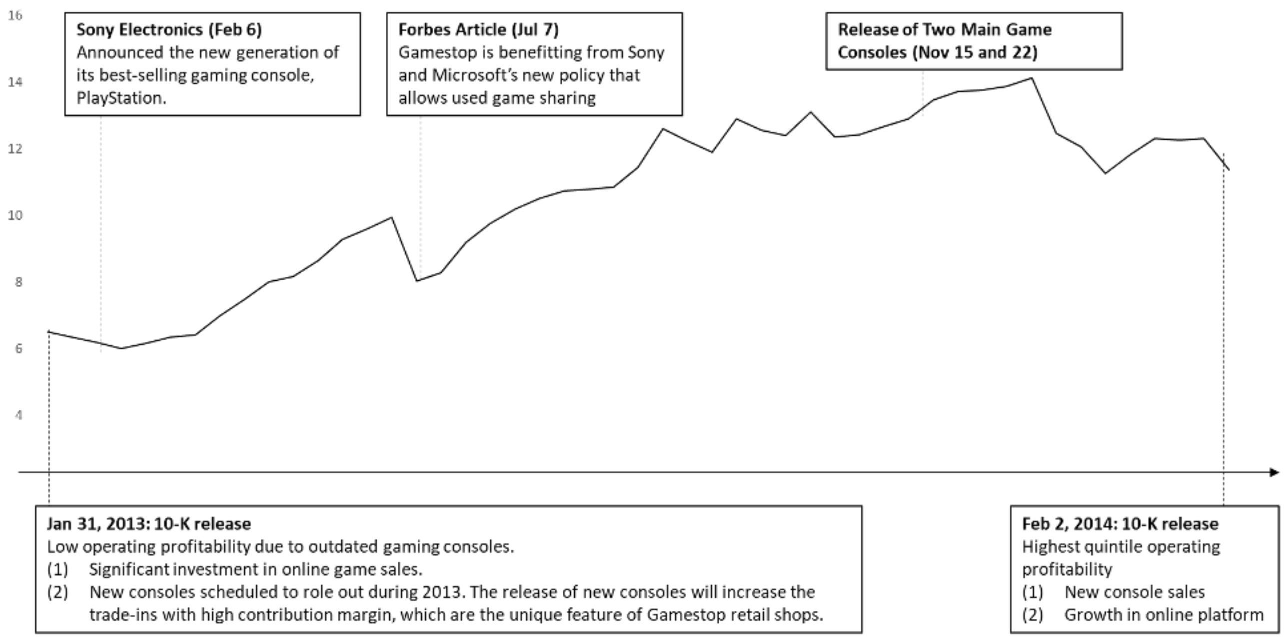



Figure 2. Stock Returns and Notable Events of GameStop in 2013

## Figure 3. Feedforward Deep Neural Network

This figure illustrates the feedforward deep neural network model that we use to obtain contextualized profitability. The input layer comprises context neurons extracted from the last hidden stage of the BERT base-uncased model. There are three hidden layers with 256, 64, and 16 neurons each. The input layer and the three hidden layers are fully connected with each other. The parameter layer comprises two parameters from the ordinary least squares regression $OP_{it+1} = \theta_0 + \theta_1 OP_{it} + \varepsilon_i t$. We get contextualized parameters $\widehat{\theta}_{0it}$ and $\widehat{\theta}_{1it}$ and combine them with the real data. We use RMSE as a loss function. The model iterates training following the rules that we introduce in Section 2.

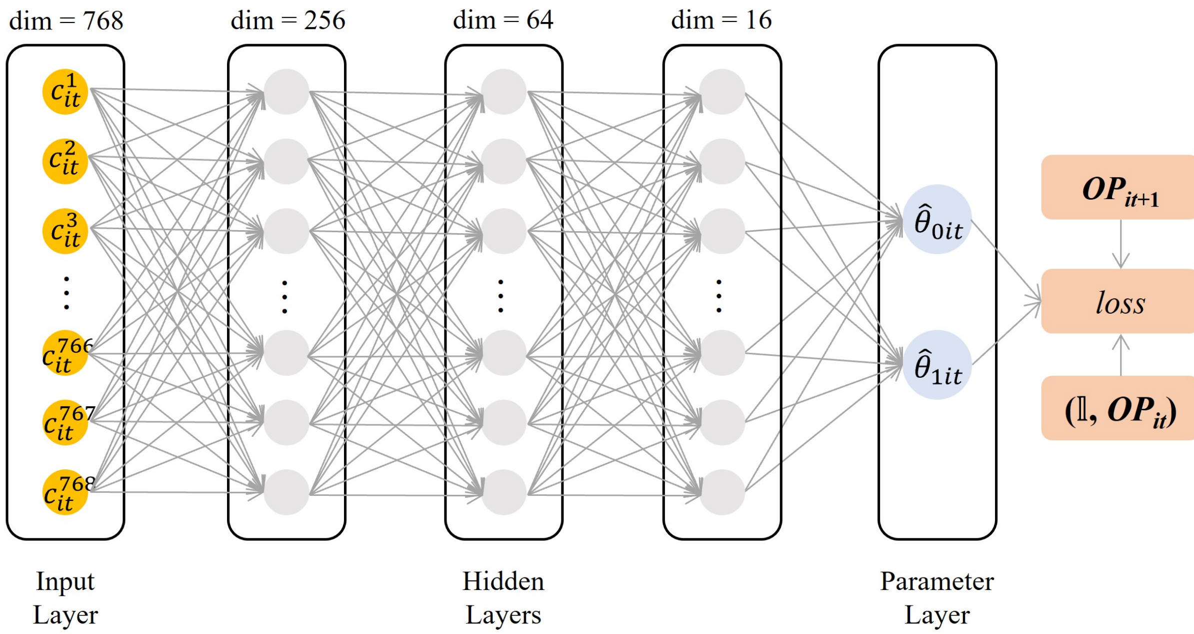



Figure 3. Feedforward Deep Neural Network

## Figure 4. Subperiod Analysis of Fama and MacBeth (1973) Regressions

This figure plots $t$-values associated with operating profitability and contextualized profitability based on Fama and MacBeth (1973) regressions estimated over a 10-year rolling window. We regress future stock returns on operating profitability or contextualized profitability while controlling for other equity characteristics used in the main analysis (columns (1) and (2) of Table 3). The horizontal axis shows the end of the window. Since our contextualized profitability is available starting from 2000, our first estimation period ends in January 2010.

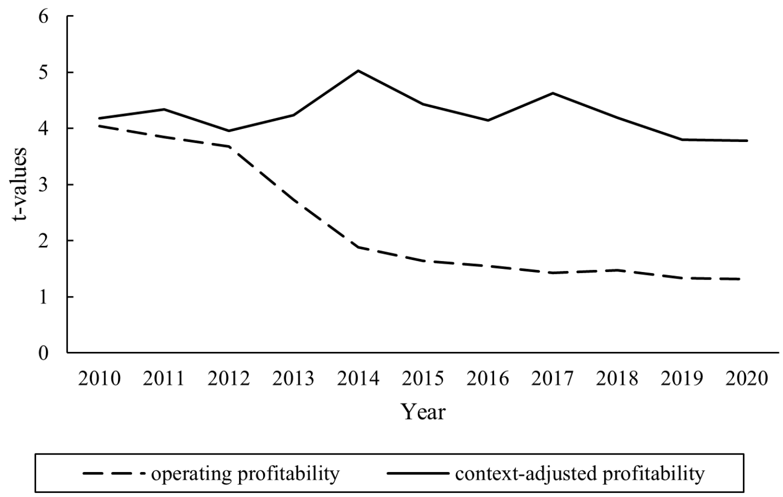
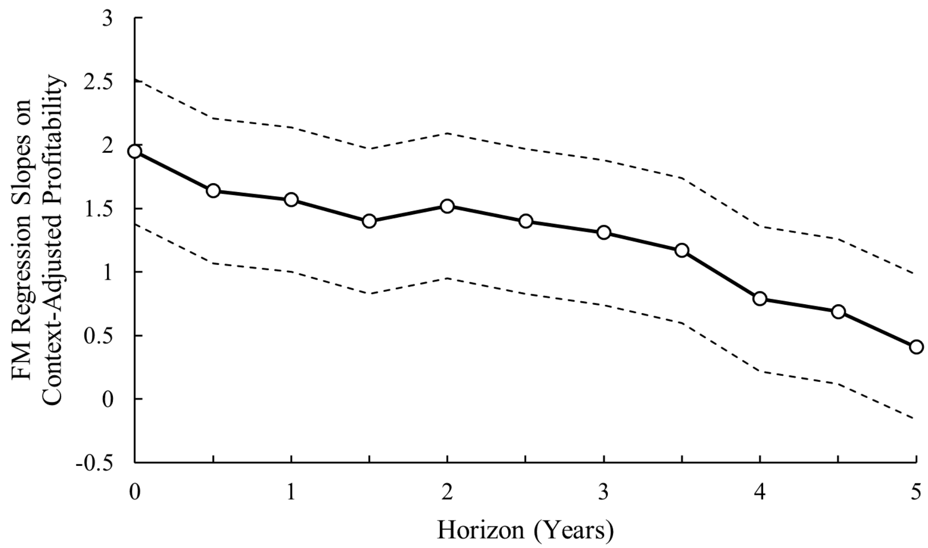


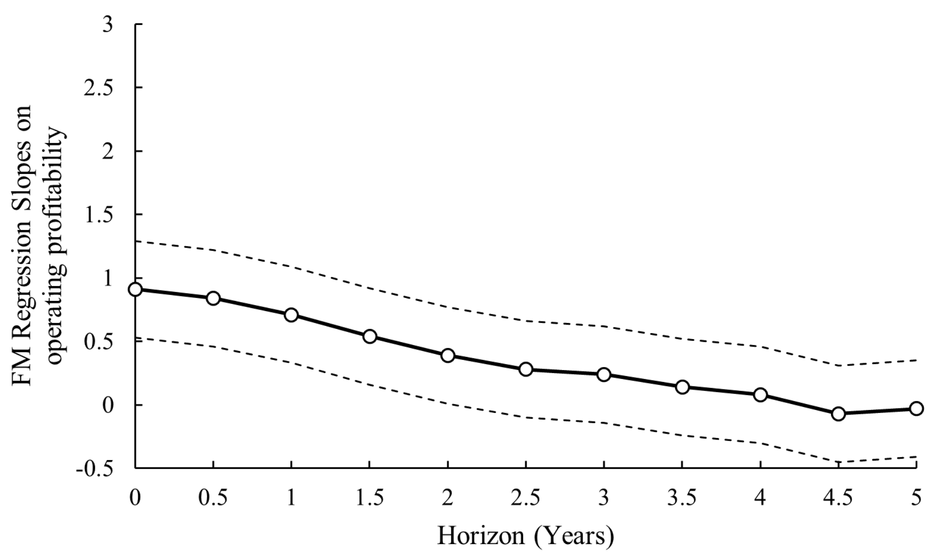

Figure 4. Subsample Analysis of Fama and MacBeth (1973) Regressions

# Figure 5. Extending Predictive Horizons

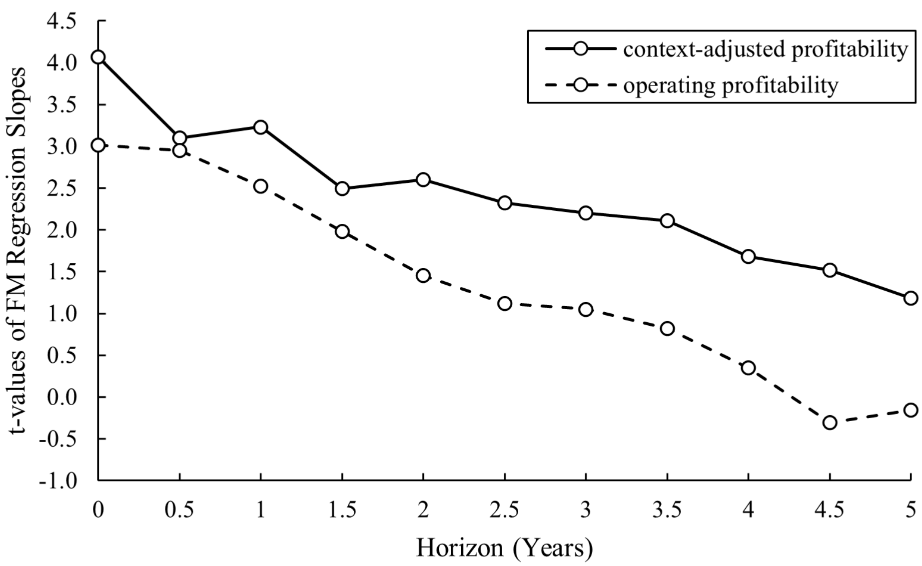
In this figure, we repeat the Fama and MacBeth (1973) regression analysis presented in Table 3, columns 1 and 3, while replacing the corresponding profitability proxies with their lags using six-month increments up to five years into the past. Regressions are estimated using observations from January 2006 to ensure that all observations have at least five years of lagged data. Panels A and B plot regression coefficients and their 95% confidence intervals. Panel A shows the plot based on contextualized profitability, which corresponds to the specification in column 3 of Table 3. Panel B plots the coefficients for operating profitability, corresponding to the specification in column 1 of Table 3. Panel C plots the corresponding $t$-values of the regressions.




Panel A. Fama and MacBeth (1973) Coefficients on Contextualized Profitability




Panel B. Fama and MacBeth (1973) Coefficients on Operating Profitability




Panel C. *t*-values of Fama and MacBeth (1973) Regression Coefficients

# Table 1. Descriptive Statistics

This table reports the descriptive statistics of variables used in further analyses. Our sample period starts in 1995 and ends in 2020. $r_{0,1}$ is a one-month lagged return; $r_{2,12}$ is a one-year lagged return calculated after skipping one month; log(*ME*) is the natural logarithm of the monthly market capitalization; log(*BE/ME*) is the natural logarithm of the book-to-market ratio; *Investment* is asset growth ratio; *OP* is calculated by subtracting cost of goods sold and sales expense less R&D expense from revenue, deflated by total assets; $OP^{CN}$ is the expectation of profitability in t+1 formed at time t using the approach of Farrell et al. (2021a). We require our observations to be traded on NYSE, Amex, and Nasdaq, and be a common share. We exclude financial firms and observations without valid financial information.

| | | | | Percentile | | |
|---|---|---|---|---|---|---|
| Variables | N | Mean | SD | $25^{th}$ | $50^{th}$ | $75^{th}$ |
| $r_{0,1}$ | 62,153 | 0.01 | 0.17 | -0.07 | 0.00 | 0.07 |
| $r_{2,12}$ | 62,153 | 0.10 | 0.58 | -0.21 | 0.04 | 0.30 |
| log(*ME*) | 62,153 | 5.82 | 1.94 | 4.39 | 5.79 | 7.21 |
| log(*BE/ME*) | 62,153 | -0.68 | 0.76 | -1.20 | -0.64 | -0.15 |
| *Investment* | 62,153 | 0.05 | 0.38 | 0.00 | 0.02 | 0.07 |
| *OP* | 62,153 | 0.19 | 0.29 | 0.02 | 0.20 | 0.34 |
| $OP^{C}$ | 62,153 | 0.10 | 0.02 | 0.08 | 0.10 | 0.12 |
| $OP^{CN}$ | 62,153 | 0.16 | 0.28 | 0.06 | 0.17 | 0.25 |

## Table 2. Predictive Ability of Expectations about Profitability

This table validates the predictive ability of the three profitability measures: traditional operating profitability (*OP*), text-only profitability ($OP^C$), and contextualized profitability ($OP^{CN}$). We run simple ordinary least squares regressions of future operating profitability on our proxies and a constant term. The analyses are out-of-sample as the model is trained without seeing the predicted values. The dependent variable is the realized value of operating profitability of year *t*+1. All independent variables are for year *t*. *t*-values of are reported within parentheses, and standard errors are clustered two ways at firm and year levels. Continuous variables are winsorized at 1% level.

| | Dependent Variable = $OP_{it+1}$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $OP_{it}$ | 0.5136 | | | 0.1010 |
| | [72.18] | | | [1.01] |
| ${OP^C}_{it}$ | | 0.3838 | | 0.2013 |
| | | [3.06] | | [0.52] |
| ${OP^{CN}}_{it}$ | | | 0.9453 | 0.9056 |
| | | | [65.13] | [40.35] |
| Adjusted $R^2$ | 25.49% | 3.88% | 30.03% | 31.48% |

# Table 3. Fama-MacBeth Regressions

This table reports monthly Fama and MacBeth (1973) regressions and Newey-West $t$-values (with lag=3). We regress monthly stock returns on our proxies for profitability and the following control variables: stock return for the prior month, stock return for the prior year (skipping the most recent month), log of the market value of equity, log of book-to-market ratio, and investment. We use returns from January 2001 to October 2021. $r_{0,1}$ is the lagged monthly return, $r_{2,12}$ is the lagged yearly return after skipping a month. *Investment* is the assets growth ratio. *OP* is (total revenue – cost of goods sold – sales, administrative expense – interest expense) scaled by total assets. log(*ME*) is the natural logarithm of the market value, log(*BE/ME*) is the natural logarithm of the book-to-market ratio, $OP^{CN}$ is contextualized profitability. Panel A presents the results for the entire sample. Panel B repeats the same analysis after we exclude microcaps from our sample. A stock is considered a microcap if its market value is below the 20th percentile of the NYSE market value distribution. All independent variables are trimmed at 1% level.

| **Panel A**. All Stocks | | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $r_{0,1}$ | -2.26 | -2.02 | -2.43 | -2.46 | -2.49 |
| | [-3.09] | [-2.73] | [-3.42] | [-3.49] | [-3.53] |
| $r_{2,12}$ | -0.10 | -0.05 | -0.13 | -0.15 | -0.14 |
| | [-0.30] | [-0.14] | [-0.38] | [-0.44] | [-0.42] |
| log(*ME*) | -0.16 | -0.12 | -0.18 | -0.19 | -0.18 |
| | [-2.98] | [-1.88] | [-3.57] | [-3.74] | [-3.75] |
| log(*BE/ME*) | 0.02 | -0.05 | -0.05 | -0.04 | -0.03 |
| | [0.17] | [-0.39] | [-0.39] | [-0.34] | [-0.22] |
| *Investment* | -1.70 | -1.35 | -1.68 | -1.70 | -1.73 |
| | [-4.52] | [-3.55] | [-4.45] | [-4.47] | [-4.60] |
| *OP* | 0.91 | | | -0.32 | -0.41 |
| | [3.01] | | | [-0.67] | [-0.83] |
| $OP^{C}$ | | 4.17 | | | 5.94 |
| | | [0.34] | | | [0.48] |
| $OP^{CN}$ | | | 1.95 | 3.26 | 3.47 |
| | | | [4.23] | [2.29] | [2.34] |
| Average Adj $R^2$ | 3.94% | 3.28% | 4.45% | 4.68% | 4.73% |

| **Panel B**. All but Microcap Stocks | | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $r_{0,1}$ | -1.26 | -1.02 | -1.48 | -1.65 | -1.51 |
| | [-1.96] | [-1.55] | [-2.35] | [-2.64] | [-2.40] |
| $r_{2,12}$ | 0.01 | 0.04 | -0.01 | -0.02 | -0.01 |
| | [0.03] | [0.13] | [-0.04] | [-0.05] | [-0.04] |
| log(*ME*) | -0.16 | -0.10 | -0.17 | -0.17 | -0.17 |
| | [-2.85] | [-1.55] | [-3.51] | [-3.51] | [-3.68] |
| log(*BE/ME*) | -0.02 | -0.01 | -0.10 | -0.04 | -0.06 |
| | [-0.14] | [-0.04] | [-0.85] | [-0.36] | [-0.52] |
| *Investment* | -2.02 | -1.59 | -2.01 | -2.02 | -2.01 |
| | [-5.03] | [-3.98] | [-5.00] | [-5.05] | [-5.02] |
| *OP* | 0.99 | | | -0.17 | -0.36 |
| | [3.11] | | | [-0.36] | [-0.72] |
| $OP^{C}$ | | -1.16 | | | -1.18 |
| | | [-0.10] | | | [-0.09] |
| $OP^{CN}$ | | | 2.13 | 3.14 | 3.52 |
| | | | [3.27] | [2.31] | [2.47] |
| Average Adj $R^2$ | 4.64% | 3.79% | 5.18% | 5.39% | 5.66% |

# Table 4. Alternative Measures of Profitability

This table reports monthly Fama and MacBeth (1973) regressions and Newey-West *t*-values (with lag=3). We regress monthly stock returns on our proxies for profitability and the following control variables: stock return for the prior month, stock return for the prior year (skipping the most recent month), log of the market value of equity, log of book-to-market ratio, and investment. We use returns from January 2001 to October 2021. $r_{0,1}$ is the lagged monthly return, $r_{2,12}$ is the lagged yearly return after skipping a month. *Investment* is the assets growth ratio. *OP* is (total revenue – cost of goods sold – sales, administrative expense – interest expense) scaled by total assets. log(*ME*) is the natural logarithm of the market value, log(*BE/ME*) is the natural logarithm of the book-to-market ratio, $OP^{CN}$ is contextualized profitability. *OP1* is net income, *OP2* is gross profit, *OP3* is R&D adjusted profitability, *OP4* is industry-adjusted profitability, *OP5* is operating profitability that adjusts for differential persistence of positive and negative earnings, *OP6* is expected profitability based on a linear model with five lags of *OP*, *OP7* is the predicted value of *OP* based on an ANN model that takes five lags of *OP* values as its input and allows for an arbitrary functional form, and *OP8* is the fitted value of *OP* using an ANN model that takes embedding vectors and current *OP* as its input and allows for unrestricted interactions between the two inputs. We include lagged one-month returns, lagged one-year returns skipping prior month return, firm size, book-to-market ratio, investment as additional stock characteristics. All continuous independent variables are truncated at 1% and 99%.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *OP1* | 1.98 | | | | | | | | -0.25 | | | | | | | |
| | [3.16] | | | | | | | | [-0.69] | | | | | | | |
| *OP2* | | 1.56 | | | | | | | | -0.11 | | | | | | |
| | | [3.35] | | | | | | | | [-0.19] | | | | | | |
| *OP3* | | | 1.01 | | | | | | | | -0.23 | | | | | |
| | | | [3.15] | | | | | | | | [-0.51] | | | | | |
| *OP4* | | | | 2.66 | | | | | | | | -0.56 | | | | |
| | | | | [3.86] | | | | | | | | [-1.80] | | | | |
| *OP5* | | | | | 1.19 | | | | | | | | 0.67 | | | |
| | | | | | [2.88] | | | | | | | | [1.65] | | | |
| *OP6* | | | | | | 0.65 | | | | | | | | 0.15 | | |
| | | | | | | [1.98] | | | | | | | | [0.23] | | |
| *OP7* | | | | | | | 1.88 | | | | | | | | -0.23 | |
| | | | | | | | [3.69] | | | | | | | | [-0.51] | |
| *OP8* | | | | | | | | 1.78 | | | | | | | | -1.03 |
| | | | | | | | | [2.65] | | | | | | | | [-1.95] |
| $OP^{CN}$ | | | | | | | | | 3.01 | 3.22 | 3.06 | 3.66 | 2.20 | 2.15 | 3.06 | 2.80 |
| | | | | | | | | | [2.56] | [2.60] | [2.28] | [2.80] | [2.92] | [2.93] | [2.28] | [3.11] |
| Ctrls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adj R2 | 4.01% | 4.23% | 4.13% | 4.33% | 4.18% | 3.46% | 4.20% | 4.39% | 4.36% | 4.49% | 4.93% | 5.02% | 4.64% | 4.21% | 4.93% | 4.90% |

# Table 5. Economic Magnitudes

This table presents the high minus low spreads on equally-weighted (EW) and value-weighted (VW) portfolio returns. Using the averaged monthly Fama and MacBeth (1973) regression slopes from Table 3, we calculate monthly predicted returns for each stock. We then partition the sample into two groups, high and low, based on monthly predicted returns. The first two columns are based on the high minus low spreads in predicted returns, followed by two columns of high-low spreads in the actual returns. The last two columns report *t*-values associated with the actual spreads. Raw "0" in the table reports a benchmark model that is based on Fama-MacBeth regression without profitability but with all other controls (one-month momentum, one-year momentum, size, book-to-market, and investment). Then for Raw 1, 2, and 3, we use regression specifications from Panel A, Table 3, columns (1), (2), and (3), respectively. We also report t-statistics of actual return spreads based on their time-series standard errors.

| | Predicted Spread | | Actual Spread | | t(Actual Spread) | |
|---|---|---|---|---|---|---|
| Regression | EW | VW | EW | VW | EW | VW |
| 0 Benchmark (no $OP$) | 0.38 | 0.40 | 0.44 | 0.35 | 2.78 | 2.13 |
| 1 ($OP$) | 0.43 | 0.45 | 0.50 | 0.42 | 3.52 | 3.25 |
| 2 ($OP^C$) | 0.39 | 0.42 | 0.44 | 0.36 | 3.30 | 2.16 |
| 3 ($OP^{CN}$) | 0.51 | 0.54 | 0.59 | 0.51 | 4.48 | 3.59 |

# Table 6. Testing the Sources of Predictability

This table reports the predictive ability of conventional vs. context-adjusted profitability proxies depending on predicted persistence of *OP*. We predict *OP*'s persistence based on contextual information only. To do so, we train an ANN model that uses narrative context embeddings as input. The predicted variable is an indicator that equals one when earnings change in the subsequent year has the same sign as the earnings change over the current year (i.e., persists vs. reverses), and zero otherwise. The input layer has 768 neurons and the two hidden layers have 256 and 16 neurons each. The output layer has one neuron. We use ReLU activation function for the input layer and the first hidden layer, and sigmoid activation function for the second hidden layer. We use binary cross-entropy loss to train the model. The model also uses rolling training windows of four years. We use a learning rate of 1e-5 and a batch size of 64. When the estimated value is close to one, contextual information predicts that earnings will be persistent. High persistence denotes the top quartile of predicted persistence values and low persistence denotes the bottom quartile of predicted persistence values. All independent variables are trimmed at 1% level.

| | High Persistence | | | Low Persistence | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *OP* | 1.55 | | 0.12 | -0.36 | | -1.63 |
| | [4.18] | | [0.54] | [-0.99] | | [-1.58] |
| $OP^{CN}$ | | 1.99 | 1.67 | | 1.76 | 2.37 |
| | | [3.26] | [3.83] | | [2.55] | [3.52] |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Average Adj $R^2$ | 4.45% | 4.60% | 4.98% | 3.85% | 4.50% | 4.77% |

# Table 7. Single Sort Portfolios

This table reports portfolio analysis based on single sorts and compares contextualized profitability to operating profitability. We tabulate excess returns over the risk-free rate, CAPM alphas, three-factor model alphas (FF3), four-factor model alphas (FF4), and five-factor alphas (FF5), based on decile portfolios sorted on contextualized profitability (Panel A) and conventional operating profitability (Panel B). We construct value-weighted portfolios on June 30th of each year based on $OP^{CN}$ (Panel A) and $OP$ (Panel B) and hold the portfolio for one year. The sample starts in January 2001 and ends in October 2021. We also present long-short portfolio returns and their corresponding $t$-values in parentheses.

**Panel A**. Sorted on contextualized Profitability

| | | Alphas | | | |
|---|---|---|---|---|---|
| | Excess Return | CAPM | FF3 | FF4 | FF5 |
| Low | 0.35 | -0.22 | -0.35 | -0.31 | -0.21 |
| 2 | 0.43 | -0.16 | -0.33 | -0.24 | -0.16 |
| 3 | 0.48 | -0.14 | -0.21 | -0.18 | -0.17 |
| 4 | 0.53 | -0.18 | -0.20 | -0.22 | -0.18 |
| 5 | 0.62 | -0.14 | -0.15 | -0.13 | -0.16 |
| 6 | 0.68 | -0.03 | -0.14 | -0.02 | 0.02 |
| 7 | 0.65 | 0.00 | -0.06 | -0.00 | 0.10 |
| 8 | 0.84 | 0.14 | 0.18 | 0.16 | 0.18 |
| 9 | 0.83 | 0.21 | 0.29 | 0.24 | 0.16 |
| High | 0.96 | 0.23 | 0.28 | 0.26 | 0.22 |
| High - Low | 0.61 | 0.45 | 0.63 | 0.57 | 0.43 |
| $t$-value | [2.51] | [2.61] | [3.00] | [2.90] | [2.48] |

**Panel B**. Sorted on Profitability

| | | Alphas | | | |
|---|---|---|---|---|---|
| | Excess Return | CAPM | FF3 | FF4 | FF5 |
| Low | 0.34 | -0.38 | -0.30 | -0.26 | -0.04 |
| 2 | 0.35 | -0.40 | -0.36 | -0.38 | -0.05 |
| 3 | 0.38 | -0.27 | -0.28 | -0.30 | -0.15 |
| 4 | 0.51 | -0.02 | -0.15 | -0.13 | -0.15 |
| 5 | 0.44 | -0.08 | -0.08 | -0.07 | 0.06 |
| 6 | 0.61 | 0.13 | 0.03 | 0.04 | 0.05 |
| 7 | 0.58 | 0.16 | 0.08 | 0.05 | 0.18 |
| 8 | 0.70 | 0.15 | 0.12 | 0.10 | 0.21 |
| 9 | 0.64 | 0.15 | 0.11 | 0.11 | 0.04 |
| High | 0.98 | 0.14 | 0.14 | 0.16 | 0.26 |
| High - Low | 0.64 | 0.52 | 0.44 | 0.42 | 0.30 |
| $t$-value | [1.80] | [2.63] | [2.48] | [2.38] | [1.24] |

## Table 8. Double-Sort Portfolios and Three-Factor Alphas

This table reports independent two-way sorts on contextualized profitability and operating profitability. We focus on three-factor model alphas for two-way sorted 5×5 portfolios. We construct value-weighted portfolios on June 30th of each year and hold the portfolio for one year. The sample starts in January 2001 and ends in October 2021. We also present long-short portfolio returns and their corresponding $t$-values in parentheses.

| Contextualized | Profitability | | | | | | |
|---|---|---|---|---|---|---|---|
| Profitability | Low | 2 | 3 | 4 | High | H - L | $t$-value |
| Low | -0.50 | -0.42 | -0.40 | -0.35 | -0.40 | 0.10 | [0.51] |
| 2 | -0.23 | -0.10 | -0.05 | -0.25 | 0.10 | 0.33 | [1.45] |
| 3 | -0.10 | -0.03 | -0.10 | 0.12 | 0.22 | 0.32 | [1.47] |
| 4 | -0.16 | 0.12 | 0.20 | 0.23 | 0.11 | 0.27 | [1.26] |
| High | 0.23 | 0.20 | 0.22 | 0.16 | 0.18 | 0.05 | [0.03] |
| High - Low | 0.73 | 0.62 | 0.62 | 0.51 | 0.58 | | |
| $t$-value | [3.57] | [2.87] | [2.73] | [2.46] | [2.68] | | |

# Table 9. Spanning Regressions

This table reports the descriptive statistics, correlation matrix, and spanning regression results for the context-adjust profitability factor. We follow the steps in Fama and French (2015) and perform 2 × 3 independent sorts each year on June 30th. We sort stocks into two groups based on the median NYSE market capitalization breakpoints from Ken French's data library. We independently sort stocks into three groups based on contextualized profitability. The high group corresponds to stocks with contextualized profitability higher than 70th percentile breakpoints, and the low group corresponds to those with contextualized profitability lower than 30th percentile breakpoint. The factor return is the difference in the value-weighted average return for two high portfolios and that for two low portfolios. In Panel A, we present the annualized returns, annualized standard deviations, and the corresponding t-values for the factors used in the analysis: market return minus the risk-free rate (*MKT*), size (*SMB*), value (*HML*), investment (*CMA*), profitability (*RMW*), and contextualized profitability ($RMW^{CN}$). Panel B reports Pearson correlations among the factors. Finally, in Panel C, we tabulate spanning regression results. Models in columns 1 and 2 use monthly *RMW* as the dependent variable and regress it on the other factor returns (with and without context-based factor). In contrast, columns 3 and 4 use $RMW^{CN}$ as the dependent variable. $t$-values are reported in parentheses.

**Panel A**. Descriptive Statistics

| | *MKT* | *SMB* | *HML* | *CMA* | *RMW* | $RMW^{CN}$ |
|---|---|---|---|---|---|---|
| Annualized Returns (%) | 7.20 | 1.68 | 3.36 | 3.00 | 4.80 | 6.62 |
| Annual Standard deviation (%) | 15.69 | 11.60 | 10.88 | 6.96 | 10.18 | 9.23 |
| $t$-value | 1.94 | 2.20 | 1.70 | 1.99 | 2.00 | 3.32 |

**Panel B**. Correlation Matrix

| | *MKT* | *SMB* | *HML* | *CMA* | *RMW* | $RMW^{CN}$ |
|---|---|---|---|---|---|---|
| *MKT* | 1 | | | | | |
| *SMB* | 0.27 | 1 | | | | |
| *HML* | -0.00 | 0.05 | 1 | | | |
| *CMA* | -0.24 | 0.07 | 0.57 | 1 | | |
| *RMW* | -0.37 | -0.49 | 0.39 | 0.24 | 1 | |
| $RMW^{CN}$ | -0.47 | -0.46 | 0.36 | 0.39 | 0.71 | 1 |

**Panel C**. Spanning Regressions

| Dep Var= | *RMW* | | $RMW^{CN}$ | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Alpha | 0.56 | -0.04 | 0.61 | 0.60 |
| | [4.54] | [-0.36] | [5.51] | [5.35] |
| *MKT* | -0.21 | -0.05 | -0.35 | -0.32 |
| | [-6.55] | [-1.76] | [-6.56] | [-6.32] |
| *SMB* | -0.23 | -0.05 | -0.31 | -0.28 |
| | [-4.76] | [-1.41] | [-4.28] | [-3.84] |
| *HML* | 0.19 | -0.13 | 0.43 | 0.40 |
| | [3.36] | [-2.71] | [5.26] | [4.83] |
| *CMA* | -0.01 | 0.11 | -0.16 | -0.21 |
| | [-0.05] | [1.51] | [-1.70] | [-2.01] |
| *RMW* | | | | 1.23 |
| | | | | [14.57] |
| $RMW^{CN}$ | | 0.32 | | |
| | | [13.68] | | |
| *N* | 249 | 249 | 249 | 249 |
| Adjusted $R^2$ | 29.39% | 62.85% | 36.52% | 60.33% |

# Table 10. Double Sorts and Five-Factor Alphas

This table reports performance for 5×5 portfolios based on operating profitability and contextualized profitability. Portfolio sorts are performed independently in the same manner as in Table 6. We construct value-weighted portfolios on June 30th of each year and hold the portfolio for one year. The sample starts in January 2001 and ends in October 2021. Panel A presents alphas based on (1) the five-factor model (upper part) and (2) the five-factor model that relies on the contextualized profitability factor $RMW^{CN}$ instead of $RMW$. Panel B reports test statistics. A($|\alpha|$) is the averaged regression intercepts; GRS is the Gibbons et al. (1989) test statistic which measures the probability that all alphas are jointly zero for a given model.

**Panel A**. Alphas

| | Monthly Alphas | | | | | | *t*-values | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $OP$ | | | | | | OP | | | | | |
| $OP^{CN}$ | Low | 2 | 3 | 4 | High | H-L | Low | 2 | 3 | 4 | High | H-L |
| | Five-Factor Model | | | | | | | | | | | |
| Low | -0.25 | -0.18 | -0.21 | -0.17 | -0.12 | 0.37 | -1.88 | -1.33 | -0.85 | -0.78 | -1.63 | 2.35 |
| 2 | -0.22 | 0.12 | 0.23 | 0.15 | 0.03 | 0.25 | -1.76 | 1.18 | 1.72 | 1.56 | 1.58 | 1.80 |
| 3 | 0.12 | 0.11 | 0.12 | 0.10 | 0.12 | 0.00 | 1.21 | 1.10 | 1.39 | 1.08 | 1.09 | 0.00 |
| 4 | 0.26 | 0.15 | 0.11 | 0.23 | 0.19 | -0.07 | 2.15 | 1.65 | 1.18 | 1.58 | 1.40 | -0.06 |
| High | 0.31 | 0.22 | 0.23 | 0.21 | 0.21 | -0.10 | 2.52 | 1.88 | 1.90 | 1.56 | 1.51 | -0.15 |
| H-L | 0.56 | 0.40 | 0.44 | 0.38 | 0.33 | | 3.35 | 2.28 | 2.57 | 2.30 | 2.21 | |
| | Four-Factor Model + Contextualized Profitability Factor | | | | | | | | | | | |
| Low | -0.20 | -0.10 | -0.03 | -0.07 | -0.04 | 0.16 | -1.65 | -1.23 | -0.42 | -0.86 | -0.35 | 1.30 |
| 2 | -0.04 | 0.12 | 0.01 | 0.13 | 0.05 | 0.09 | -0.15 | 1.35 | 0.03 | 1.36 | 0.62 | 0.98 |
| 3 | 0.01 | 0.02 | 0.11 | 0.12 | 0.11 | 0.10 | 0.01 | 0.06 | 1.29 | 1.30 | 1.00 | 1.21 |
| 4 | -0.10 | 0.10 | 0.03 | 0.06 | 0.07 | 0.17 | -1.30 | 1.29 | 0.36 | 0.60 | 0.88 | 1.33 |
| High | -0.12 | -0.11 | 0.05 | 0.03 | 0.06 | 0.18 | -1.36 | -1.33 | 0.45 | 0.36 | 0.72 | 1.51 |
| H-L | 0.08 | -0.01 | 0.08 | 0.10 | 0.10 | | 0.18 | -0.08 | 0.95 | 1.02 | 0.83 | |

**Panel B**. Test Statistics

| Model | GRS | A($|\alpha|$) |
|---|---|---|
| Four Factors + $RMW$ | 2.86 | 0.17 |
| Four Factors + $RMW^{CN}$ | 1.25 | 0.08 |

# Table 11. Size-B/M Portfolios, $OP$, and $OP^{CN}$

This table reports the alphas of the portfolios sorted on size and book-to-market. We construct 5 × 5 size and B/M portfolios based on these characteristics using the NYSE breakpoints. Specifically, as of June 30 each year, we sort stocks into five groups using NYSE market capitalization breakpoints. We independently sort stocks into five groups based on NYSE book-to-market breakpoints. For each of the 25 resulting portfolios, we estimate alphas using Fama and French's three-factor model, five-factor model, and the five-factor model that relies on $RMW^{CN}$ factor. In Panel A, we regress the monthly returns for each portfolio on monthly factors to estimate alphas. We report alphas for the three-factor model, the five-factor model, and the four-factor model with contextualized profitability factor. We also report t-statistics for each alpha estimate. In Panel B, we calculate the GRS statistics and the averaged absolute alphas for the three models.

**Panel A**. Portfolio Sorts

| | Monthly Alphas | | | | | t-values | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Book-to-Market | Low | 2 | 3 | 4 | High | Low | 2 | 3 | 4 | High |
| | Three Factor Model | | | | | | | | | |
| Small | -0.55 | 0.03 | 0.04 | 0.17 | 0.13 | -4.50 | 0.30 | 0.45 | 2.30 | 1.94 |
| 2 | -0.26 | -0.06 | -0.12 | 0.11 | -0.11 | -3.26 | -0.60 | -1.71 | 1.62 | -1.60 |
| 3 | -0.10 | 0.04 | 0.02 | 0.04 | 0.11 | -1.10 | 0.99 | 0.03 | 0.35 | 1.62 |
| 4 | 0.16 | -0.12 | 0.07 | 0.15 | -0.05 | 2.11 | -1.54 | 0.85 | 2.31 | -0.48 |
| Big | 0.22 | 0.06 | -0.10 | -0.12 | -0.12 | 3.00 | 0.55 | -1.59 | -2.10 | -2.13 |
| | Five Factor Model | | | | | | | | | |
| Small | -0.31 | 0.00 | 0.04 | 0.12 | 0.16 | -3.13 | 0.00 | 0.42 | 1.30 | 2.00 |
| 2 | -0.15 | -0.02 | -0.13 | 0.03 | -0.04 | -1.80 | -0.23 | -1.46 | 0.42 | -0.41 |
| 3 | -0.13 | -0.15 | 0.06 | 0.06 | 0.03 | -1.61 | -1.68 | 0.15 | 0.89 | 0.36 |
| 4 | 0.10 | 0.13 | 0.03 | 0.11 | -0.05 | 1.45 | 1.63 | 0.06 | 1.22 | -0.55 |
| Big | 0.12 | 0.17 | -0.08 | 0.04 | -0.11 | 1.59 | 1.96 | -0.19 | 0.08 | -1.31 |
| | Four-Factor Model + Contextualized Profitability Factor | | | | | | | | | |
| Small | -0.08 | 0.08 | 0.05 | 0.07 | 0.09 | -1.00 | 0.92 | 0.62 | 0.80 | 1.02 |
| 2 | -0.03 | -0.02 | -0.12 | 0.03 | -0.03 | -0.31 | -0.29 | -1.36 | 0.33 | -0.29 |
| 3 | -0.11 | -0.05 | 0.03 | 0.04 | 0.05 | -1.29 | -0.66 | 0.32 | 0.45 | 0.50 |
| 4 | 0.05 | 0.06 | 0.02 | 0.06 | -0.04 | 0.62 | 0.83 | 0.26 | 0.89 | -0.46 |
| Big | 0.11 | 0.12 | -0.04 | 0.02 | -0.06 | 1.31 | 1.46 | -0.40 | 0.21 | -0.93 |

**Panel B**. GRS Statistics

| | Three-Factor | Five-Factor | Four-Factor + $OP^{CN}$ |
|---|---|---|---|
| GRS | 3.80 | 2.73 | 1.69 |
| $A(\lvert\alpha\rvert)$ | 0.12 | 0.09 | 0.06 |

# Table 12. Ex-Post Sharpe Ratios

This table presents optimal factor loadings along with the ex-post Sharpe ratios. We start from the market return less risk-free returns (*MKT*) and include the factors sequentially. In the sixth iteration, we also add $RMW^{CN}$ factor. The sample starts in January 2001 and ends in October 2021.

| | Optimal Loadings | | | | | | Sharpe |
|---|---|---|---|---|---|---|---|
| | *MKT* | *SMB* | *HML* | *CMA* | *RMW* | $RMW^{CN}$ | Ratio |
| 1 | 100% | | | | | | 0.46 |
| 2 | 93% | 7% | | | | | 0.47 |
| 3 | 48% | 3% | 49% | | | | 0.55 |
| 4 | 26% | 3% | 17% | 54% | | | 0.77 |
| 5 | 24% | 2% | 8% | 36% | 30% | | 1.04 |
| 6 | 29% | 1% | 9% | 38% | | 23% | 1.27 |
| 7 | 27% | 1% | 5% | 29% | 20% | 18% | 1.48 |

# Table 13. Out-of-Sample Tests

This table reports out-of-sample test results. We use financial reports released from 2019 to 2021 to train our model and then use data released in 2022 to estimate contextualized profitability. For Fama and MacBeth (1973) regressions, we use returns from June 2022 to June 2023. Other specifications are identical to those in Table 3. We report regression results with a full sample in Panel A. For single sorts, we form portfolios on June 30, 2022 and hold it for one year until June 30, 2023. All other specifications are identical to those in Table 5. For simplicity, we report returns of only top and bottom deciles and their long-short returns. Panel B reports portfolios sorted on contextualized profitability and Panel C reports portfolios sorted on operating profitability.

**Panel A**. Fama-MacBeth Regressions

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $r_{0,1}$ | -3.91 | -3.90 | -3.99 | -2.49 | -2.50 |
| | [-1.67] | [-1.68] | [-1.71] | [-1.84] | [-1.85] |
| $r_{2,12}$ | 0.26 | 0.25 | 0.19 | -0.37 | -0.36 |
| | [0.20] | [0.18] | [0.14] | [-1.54] | [-1.60] |
| log(*ME*) | 0.30 | 0.29 | 0.27 | 0.24 | 0.25 |
| | [1.75] | [1.72] | [1.63] | [0.31] | [0.36] |
| log(*BE/ME*) | -0.03 | -0.03 | -0.04 | -0.07 | -0.03 |
| | [-0.07] | [-0.08] | [-0.11] | [-0.31] | [-0.22] |
| *Investment* | -0.36 | -0.37 | -0.42 | -0.87 | -0.73 |
| | [-0.25] | [-0.24] | [-0.30] | [-1.83] | [-1.88] |
| *OP* | 0.46 | | | -0.58 | -0.60 |
| | [2.97] | | | [-1.08] | [-1.06] |
| $OP^C$ | | 5.06 | | | 4.88 |
| | | [0.22] | | | [0.25] |
| $OP^{CN}$ | | | 2.68 | 3.75 | 3.96 |
| | | | [3.82] | [3.57] | [3.55] |
| Average Adj $R^2$ | 4.02% | 3.65% | 4.26% | 5.08% | 5.16% |

**Panel B**. Single Sorts on Contextualized Profitability

| | | Alphas | | | |
|---|---|---|---|---|---|
| | Excess Return | CAPM | FF3 | FF4 | FF5 |
| Low | 0.41 | -0.44 | -0.28 | -0.45 | 0.05 |
| High | 1.75 | 0.50 | 0.35 | 0.40 | 0.82 |
| High - Low | 1.34 | 0.94 | 0.63 | 0.85 | 0.77 |
| *t*-value | [1.68] | [1.95] | [2.40] | [3.21] | [2.54] |

**Panel C**. Single Sorts on Operating Profitability

| | | Alphas | | | |
|---|---|---|---|---|---|
| | Excess Return | CAPM | FF3 | FF4 | FF5 |
| Low | 0.85 | -0.51 | -0.05 | -0.23 | 0.15 |
| High | 2.04 | 0.48 | 0.54 | 0.58 | 0.59 |
| High - Low | 1.19 | 0.99 | 0.59 | 0.81 | 0.44 |
| *t*-value | [1.36] | [1.97] | [2.02] | [2.61] | [1.69] |