Information Uncertainty and Deep Learning*

Suk-Joon Byun †　　Sangheum Cho‡

Preliminary Draft

Abstract

This paper finds that the nonlinearity of deep learning captures behavioral mispricing in firms with high information uncertainty. Using a measure of capturing the nonlinearity, the return predictability of the variable becomes stronger in firms with higher information uncertainty. Moreover, as behavioral biases like investor inattention can induce common mispricing, the long-short strategy by the variable also has significant exposures to behavioral factors representing the common mispricing. Combining the cross-section and time-series results, the economic mechanism of deep learning models is to capture the commonality of behavioral mispricing and covariances with the commonality.

JEL classification: To be included
Keywords: Deep Learning, Conditional Autoencoder, Machine Learning, Information Uncertainty, Mispricing, Behavioral Factors, Empirical Asset Pricing

*This draft: May 6, 2021
†Korea Advanced Institute of Science and Technology (e-mail: sjbyun99@kaist.ac.kr)
‡Korea Advanced Institute of Science and Technology (e-mail: sh.cho@kaist.ac.kr)

1. Introduction

Deep learning is the heart of artificial intelligence. Models with deep learning methods show good performance in many areas, from image recognition to natural language translation. The origin of performance is its nonlinearity. The nonlinear structure of deep learning models enables them to capture useful latent features in complex and uncertain types of information (LeCun et al. (2015)). In empirical asset pricing, the nonlinearity also makes them have better performance in predicting returns compared to other traditional econometric techniques in predicting stock returns (Gu et al. (2020)). However, at the same time, the nonlinearity makes it hard to interpret the economic meanings from them. To use deep learning models in empirical asset pricing with credible out-of-sample performance, understanding their economic mechanism would be very important (Karolyi and Van Nieuwerburgh (2020)).

This paper studies the economic mechanism of deep learning models. We find that the nonlinear structure of deep learning models captures behavioral mispricing by investor inattention, particularly concentrated on firms with high information uncertainty. Using a variable representing the distinctive feature of the nonlinear structure, the variable predicts returns and shows higher return predictability in firms with higher information uncertainty. Moreover, from the empirical literature, information uncertainty amplifies mispricing by behavioral biases like investor inattention (Hirshleifer (2001), Hirshleifer et al. (2018)), and the effect can induce common mispricing (Daniel et al. (2020)). Therefore, if the variable of the distinctive role of nonlinear structure actually represents mispricing by investor inattention, we might expect it also has an association with investor inattention. Consistent with the argument, the long-short portfolio constructed by the variable has strong exposures to behavioral factors proxying common mispricing by investor inattention constructed by Daniel et al. (2020). Therefore, we suggest the economic mechanism of deep learning models is to capture the commonality of behavioral mispricing and covariances to factors proxying the commonality.

Why the nonlinear structure in deep learning models capture such behavioral mispricing in firms with higher information uncertainty? In general, nonlinear structure in deep learning models is good at capturing latent features in complex and uncertain types of information (LeCun et al. (2015)). Similarly, in empirical asset pricing, we might also expect deep learning models to work better in complex and uncertain information by discovering the latent features. In other words, in cross-section, deep learning models would work better in predicting returns for firms with more complex and uncertain information by exploiting hidden but valuable things there. What would be such latent features in predicting returns? Investigating it would answer the economic mechanism of deep learning models.

The empirical literature about the effects of complex and uncertain information on asset prices is well-established. With the psychological evidence that individuals tend to be less attentive to complex and uncertain information (Tversky and Kahneman (1974)), the behavioral finance literature shows that investors are also less attentive to such complex and difficult-to-process information (e.g., Hirshleifer and Teoh (2003)). As a result, the behavioral bias of investor inattention induces mispricing, particularly stronger in firms with more complex and uncertain information (e.g., Hirshleifer et al. (2018)).

Based on the literature, the latent features in complex and uncertain information might be the signals related to mispricing incurred by investor inattention. In other words, the hidden but valuable things behind the hard-to-process information are the mispricing signals related to behavioral biases. Therefore, the nonlinear structure of deep learning models would discover the mispricing signal, particularly concentrated in firms with complex and uncertain information.

To test the hypothesis empirically, we primarily need a variable representing the distinctive feature of the nonlinear structure in deep learning models. We use the Conditional Autoencoder (CA) model of Gu et al. (2021) as our deep learning model. The CA model produces predicted returns for each firm-month observations as functions of firm characteristics. We believe that using the CA model is appropriate for studying the economic mechanism of deep learning models. Because the model has an economic structure as the latent factor model, outputs from the model can be interpreted as the covariance between individual stock returns and latent factors. Moreover, the covariance terms, the outputs from the CA model, are functions of firm characteristics. Therefore, comparing the difference between covariance terms across firms is suitable for the cross-sectional study of deep learning models.

We construct nonlinear and linear signals using the CA model. The nonlinear CA model with three layers of nonlinear activation function produces a $CA3$ signal. The $CA3$ is our benchmark of deep learning signal. The linear CA model with no layer of nonlinear activation function produces a $CA0$ signal. It is our benchmark of the linear machine learning signal. We take the difference between $CA3$ and $CA0$ signal and denote is as NML, *nonlinear-minus-linear*. Since each the $CA3$ and the $CA0$ signal is an aggregation of multiple anomalies, the NML signal is also a composite signal of anomalies. Notably, the NML signal is mainly attributable to the nonlinear structure of the deep learning model. Therefore, it is our main variable of interest. We investigate the return predictability of NML and test whether it has stronger predictability in firms with higher information uncertainty. We also examine the time-series properties of the NML-sorted portfolios.

Besides the NML signal, we also need proxies of information uncertainty to study the effect of the NML signal in the cross-section of information uncertainty. Zhang (2006) defines information uncertainty as to the ambiguity of new information on the firm value or poor information quality. In other words, information uncertainty means how intricate given information about a firm is for investors to extract valuable signals. This study argues that the nonlinear structure of deep learning models is good at capturing some profitable signals even when the given information about a firm's valuation is seemingly hard to process. Therefore, the definition of information uncertainty fits our purpose. He suggests proxies of information uncertainty as stock return volatility, cash flow volatility, or inverse of firm age. Similar measures are also used in Jiang et al. (2005) and Lam and Wei (2011).

With the NML signal and proxies of information uncertainty, we test our hypothesis. Empirical results support our hypothesis. At first, the NML signal predicts future returns. We run the Fama-Macbeth regression of future returns on the NML signal. The NML signal shows statistically significant return predictability. A 1 standard deviation increase of the NML signal predicts 1.20% higher returns. The significance remains after controlling $CA0$, the linear machine learning signal, and several determinants of returns (e.g., size, book-to-market, and momentum). It implies that the NML signal has separate return predictability than the linear signal or common firm characteristics.

More importantly, the return predictability of the NML signal becomes stronger in firms with higher information uncertainty. We run the Fama-Macbeth regression of returns on the interaction of the NML signal and the dummy of high information uncertainty. When the information uncertainty is proxied by stock return volatility, in firms with higher volatility, the return predictability of the NML signal is 0.80% higher than lower volatility firms. Using the proxy of information uncertainty as inverse of firm age, cash flow volatility, and analyst dispersion provide similar results. Based on Daniel et al. (2020), the significantly higher return predictability of NML on firms with higher information uncertainty is the supporting evidence of NML to represent behavioral mispricing argument rather than conventional risk-based argument.

However, there is an important thing to address about the NML signal. In the estimated CA model, the NML signal is the "systematic" part of the model's latent factors. However, our main results show that the NML signal is actually behavioral mispricing. Then, how can we reconcile these seemingly contradictory results? We test two explanations. First, our main results might be driven by the misspecified idiosyncratic mispricing part because we do not include the idiosyncratic mispricing part in our baseline specification. To test the argument, we include the pure idiosyncratic mispricing part and construct two NML signals: NML^{α} that is the difference between the alpha parts, and NML^{β} that is the difference between beta parts. With those NML signals, we conduct the same analysis. We find that NML^{α} itself has no return predictability, and the interaction of NML^{α} with information uncertainty is insignificant as well. However, NML^{β} still shows significant return predictability, and the interaction with information uncertainty remains significant. This result is consistent with the results of Kelly et al. (2019) and Gu et al. (2021) that including the pure idiosyncratic part in the linear or nonlinear machine learning models does not lead to improvement in performance. These results imply that the NML signal, the systematic part of the latent factor models, captures the "systematic" mispricing that is called the commonality of behavioral mispricing in the literature (Baker and Wurgler (2006), Kozak et al. (2018), Daniel et al. (2020)). Related to the common mispricing literature, we focus on the behavioral factor model of Daniel et al. (2020) for two reasons. First, their factors are traded and theoretically motivated. Second, their factors are traded factors. They show that the behavioral factor model prices various anomalies well. The factor model has three factors: market, $PEAD$ (investor inattention), and FIN (overconfidence).

To test whether the NML signal captures the commonality of mispricing, we form deciles of value-weighted NML-sorted portfolios using NYSE breakpoints. The long-short portfolio would exploit behavioral mispricing. The time-series regression of the returns of the long-short portfolio on the behavioral factors supports our argument. At first, the risk-adjusted returns of the long-short portfolio produce significant returns that are 0.68% per month ($t = 3.92$).[1] More importantly, it has statistically significant exposure to the behavioral factors. By 1% change of $PEAD$ (FIN) factor, the long-short portfolio by NML covaries with 0.45% (0.23%). The higher exposure of the decile long-short portfolio by NML on $PEAD$ than FIN further supports our argument because the $PEAD$ factor is more related to information uncertainty through the investor inattention channel. Moreover, the common mispricing literature suggests that the short leg should be more associated with the behavioral mispricing due to short-sale constraints (Stambaugh et al. (2012), Daniel et al. (2020)). Our results are also consistent with the literature. The short-leg of the NML-sorted portfolio shows significant exposures to the behavioral factors. However, the long-leg does not have exposure to the factors.

As NML-sorted portfolios exploit behavioral mispricing, the magnitude should be greater

[1] Risk-adusted returns are calculated using five-factor model by Fama and French (2015) augmented by momentum factor of Carhart (1997).

in firms with higher information uncertainty because behavioral mispricing can be more severe in those firms. The results from double-sorted portfolios by information uncertainty-NML also support our argument. Using stock return volatility as a proxy of information uncertainty, in the highest quintile of the volatility, the long-short portfolio by NML earns 2.54% per month ($t = 7.36$). However, in the lowest quintile, it shows only 0.08% per month ($t = 0.44$). Then, the returns on the group of the highest information uncertainty should also have higher exposure on the behavioral factors, while should not for the group of the lowest information uncertainty. The time series of regressing the returns from a long-short portfolio of NML on behavioral factors supports the arguments. In the highest quintile of information uncertainty proxied by stock return volatility, the decile long-short portfolio by NML covaries significantly covaries with $PEAD$. With 1% change of $PEAD$, the NML long-short portfolio shows 0.67% of covariation. However, in the lowest quintile, it shows insignificant -0.08% of covariation. Using other measures of information uncertainty also shows consistent results.

In summary, this paper finds that the nonlinear structure of deep learning models is good at exploiting behavioral mispricing by investor inattention, particularly concentrated in firms with high information uncertainty. We show the result using the NML signal, representing the distinctive feature of deep learning models. Importantly, as the NML signal is the "systematic" part of the latent factor model by the deep learning method, the NML signal would represent the covariances with factors related to the commonality of behavioral mispricing. Therefore, the economic mechanism of deep learning models might be to capture covariance between individual stock returns with the economy-wide behavioral mispricing by instrumenting the covariance using firm characteristics proxying information uncertainty.

This paper contributes to the emerging literature on understanding the economic mechanism of deep learning models. We show that deep learning models capture the commonality of behavioral mispricing and covariance with the commonality. Recently, there have been many studies on machine learning methods (e.g., Gu et al. (2020), Gu et al. (2021), Chen et al. (2019)). Although those models show good performance in predicting asset returns, it still needs to improve their economic interpretability for their more robust and credible usage (Karolyi and Van Nieuwerburgh (2020)). Therefore, Avramov et al. (2021) suggest deep learning models would perform well by identifying difficult-to-arbitrage stocks and market states.

As Avramov et al. (2021) provide a comprehensive study on the economic interpretability of deep learning models, comparing how our results are different from their study is important. Our research differs from theirs in the motivation and the methodology. First, we focus more on the behavioral mechanism of incurring mispricing, but Avramov et al. (2021) focus on the general cross-sectional properties related to mispricing, which is the firm size. They find that, in the cross-section, after excluding difficult-to-arbitrage stocks (e.g., microcaps or no credit rating), deep learning models' performance exacerbates a lot. Although size is important, it proxies too many things (e.g., information uncertainty (Zhang (2006)), short sale constraints Israel and Moskowitz (2013), trading frictions (Gu et al. (2020))). Therefore, more detailed cross-sectional characteristics are needed to improve our understanding of the economic meaning of deep learning models. As an answer, we propose that information uncertainty be related to deep learning models' nature. Furthermore, our motivation for time-series analysis is also different from Avramov et al. (2021). Concerning the time-series implication of behavioral finance, we focus on the commonality of behavioral mispricing literature (Daniel et al. (2020)). However, they focus on market frictions such as volatile market periods to relate deep learning models to difficult-to-arbitrage states.

Furthermore, our research differs from Avramov et al. (2021) in methodology. We focus on the distinctive feature of the nonlinear structure of deep learning models using our NML signal because the nonlinear structure is key to deep learning models. Notably, the NML signal is constructed by taking the difference between the nonlinear and linear CA models that only differ in their nonlinearity. This approach rules out any alternative source of explanation, such as differences of asset pricing assumption, training sample, or estimation techniques. However, Avramov et al. (2021) compare various nonlinear deep learning models (neural network model of Gu et al. (2020), SDF of Chen et al. (2019), conditional autoencoder (CA) of Gu et al. (2021)) with a linear machine learning model (instrumented principal component analysis (IPCA) of Kelly et al. (2019)). Their differences not only come from nonlinearity but also come from asset pricing assumptions, input data, and training methods. For example, nonlinear neural network models are pure prediction-based models using macro variables, but the linear IPCA model uses latent factor structure. Therefore, besides nonlinearity, the outputs from those machine learning models have different economic meanings. Furthermore, although the nonlinear CA model and IPCA model share their beta-pricing structure, they are different in terms of their latent factors, treating of input data, and training procedure Gu et al. (2021).

Besides the literature on the economic interpretability of deep learning models, this paper also contributes to the literature on estimating latent factor models. The literature discusses whether the pure idiosyncratic mispricing part is important in estimating latent factor models. Kelly et al. (2019) show that when using enough number of latent factors (say 5 or 6), there is no additional explanatory power of the pure idiosyncratic mispricing part. Gu et al. (2021) provide similar evidence with their deep learning models. Our results are also consistent with these empirical findings. The distinctive feature attributable to the nonlinear idiosyncratic mispricing part of the deep learning model plays no important role. On the other hand, the distinctive nonlinear systematic part is important. However, our results differ from the literature in terms of interpretation. The literature interprets their results as supportive evidence of the conditional beta model. In contrast, we argue that our results are supportive evidence of the commonality of behavioral mispricings, such as Daniel et al. (2020).

Therefore, our finding also contributes to the literature on the commonality of mispricing. We show that deep learning models can capture the common mispricing. Stambaugh and Yuan (2017a) propose a set of mispricing factors by averaging several characteristics documented in the previous research. Daniel et al. (2020) provide theoretical motivation for constructing mispricing factors by behavioral biases and empirically show the model works. One of the advantages of deep learning methods is that those models do not require prior human knowledge. Therefore, with minimum knowledge about cross-sectional or time-series properties regarding asset pricing, deep learning models for asset pricing discover the commonality of behavioral mispricing. Our result supports the validity of the literature.

This paper is organized as follows. Section 2 develops our hypotheses. Section 3 describes how we construct the variable of capturing the distinctive feature of deep learning models, the NML signal. Section 4 describes data. Section 5 provides our empirical results. Finally, section 6 concludes our paper with a summary and areas of future research.

2. Hypothesis Development

Uncertain and complex and information has an important implication on firms' valuation (Zhang (2006)). Based on psychological evidence, individuals have a behavioral bias that is to pay less attention to uncertain and complex information (Tversky and Kahneman (1974)). The behavioral bias is also applicable to the investors in financial markets: investors are inattentive to uncertain and complex information so that firms with more complex and uncertain information are more likely to be mispriced (e.g., Hirshleifer (2001), Hirshleifer et al. (2018)). Thus, uncertain and complex information amplifies behavioral biases such as investor inattention and induces mispricing. Many studies support the argument. For example, in the viewpoint of considering momentum strategy as an anomaly by behavioral bias, Zhang (2006) and Jiang et al. (2005) provide the evidence. Several studies also provide similar evidence in more specific settings of complex and uncertain information such as valuation of complicated conglomerates (Cohen and Lou (2012)), technological originality (Hirshleifer et al. (2018)), and technological similarity between firms (Lee et al. (2019)).

This paper argues that deep learning models' performance is mainly originated from capturing such behavioral mispricing, which is particularly concentrated in firms with uncertain and complex information. We start by focusing on the consensus of deep learning models that their nonlinear structure is good at discovering useful latent features, particularly from the complex and uncertain types of information (LeCun et al. (2015)). In empirical asset pricing as well, the nonlinearity of deep learning models is good at capturing latent features of asset prices from a set of complex and high-dimensional firm characteristics associated with firms' value (Gu et al. (2020)). Although understanding the economic meaning of those latent features is important, due to the black-box-like nature of deep learning models, their economic meaning is unclear (Karolyi and Van Nieuwerburgh (2020)). By relying on the information uncertainty literature discussed above, the latent features in uncertain and complex information might be behavioral mispricing, and we hypothesize that deep learning models might capture it.

To test the hypotheses empirically, we need two things. First, we need a variable representing the distinctive feature of the nonlinear structure of deep learning models. Because the nonlinear structure is the main advantage of deep learning models (LeCun et al. (2015), Gu et al. (2020), Gu et al. (2021)). To construct the variable, we use the Conditional Autoencoder (CA) model of Gu et al. (2021) because of its several advantages.[2] Using the model, we simply take the difference between the predicted returns from the nonlinear CA model (deep learning model) and the linear CA model (linear machine learning model). We call the variable as NML, *nonlinear-minus-linear*.[3] Second, we need proxies of information uncertainty. The literature on information uncertainty is well-established. Zhang (2006) refers to information uncertainty as high ambiguity on new information for the valuation of firms or poor information quality. He suggests proxies of high information uncertainty as high volatility in stock returns or cash flow, high analyst dispersion, or young age.

Now, with the NML signal and proxies of information uncertainty, our hypotheses can be rephrased more specifically. In cross-section, if the nonlinear structure of deep learning models

[2] First, making both nonlinear machine learning model (deep learning) and linear machine learning model is straightforward in the CA model. Second, the output of the CA model has economic interpretation by their factor structure. Third, the loadings on factors can be directly written as a function of firm characteristics: it makes cross-sectional comparison easier.

[3] More details on the NML signal will be discussed in the next section.

captures behavioral mispricing, the NML signal should predict future returns. Furthermore, if the mispricing is especially related to information uncertainty, the return predictability of the NML signal should be stronger in firms with high information uncertainty.

In time-series, the trading strategy based on the NML signal of buying high NML stocks and shorting low NML stocks (NML strategy) should earn significant risk-adjusted returns because they would exploit behavioral mispricing. Moreover, as behavioral biases like investor inattention can induce common mispricing, the NML strategy should have positive exposures to the behavioral factors of capturing such investor inattention, and the exposure should be higher in the short leg (Daniel et al. (2020)). Similar arguments also can be applied to the double-sorted portfolios by information uncertainty and NML signal. The NML strategy using firms with higher information uncertainty should generate higher profits than the strategy using firms with lower information uncertainty. Furthermore, the exposures to behavioral factors should be stronger for the NML strategy using firms with higher information uncertainty but weaker or no exposures for the strategy using firms with lower information uncertainty.

3. Capturing Nonlinearity of Deep Learning Models

The variable capturing the distinctive feature in the nonlinear structure of deep learning models is constructed by taking the difference between a deep learning model's signal and a linear machine learning model's signal. We call the variable as NML (*nonlinear-minus-liner*). In this section, we describe how we construct the variable.

The nonlinear and linear signals are generated by the Conditional Autoencoder (CA) model by Gu et al. (2021). Our nonlinear deep learning model is a CA model with three layers of nonlinear structure, and the linear machine learning model is a CA model with no layers of nonlinear structure. CA models have the following specification:

$$r_{i,t} = \alpha(z_{i,t-1}) + \beta(z_{i,t-1})' f_t + \epsilon_{i,t}.$$

The excess returns $r_{i,t}$ possess a latent K-factor structure by f_t. Therefore, the CA model estimates the systematic part, $\beta(z_{i,t-1})'f_t$, and purely idiosyncratic mispricing part, $\alpha(z_{i,t-1})$. The systematic part $\beta(z_{i,t-1})'f_t$, and the pure idiosyncratic mispricing part $\alpha(z_{i,t-1})$ are functions of lagged firm characteristics $z_{i,t-1}$. Each of the latent K-factors, f_t, is constructed by a linear combination of contemporaneous characteristics-managed portfolios by following Gu et al. (2021). The weights in the linear combination are also estimated by the CA model using the lagged firm characteristics.

The machine learning models produce predicted returns. For each firm-month observations in our test sample, trained CA models produce machine learning signals as follows:

$$\hat{r}_{i,t} = \hat{\alpha}(z_{i,t-1}) + \hat{\beta}(z_{i,t-1})'\lambda_{t-1}$$

where λ_{t-1} is the historical average of latent factors up to the start of the year. The machine learning signals $\hat{r}_{i,t}$ are predicted returns composed of the systematic part and the idiosyncratic mispricing part. The systematic part $\hat{\beta}(z_{i,t-1})'\lambda_{t-1}$ represents the covariation between individual stock returns with the latent factors that are mainly explained by firm characteristics.[4] For example, in the viewpoint of the conventional risk-based framework, the systematic part may represent the higher risk premium on the small firms by the size factor risk (Fama and Kenneth (1993)). In contrast, in the perspective of the commonality of mispricing by behavioral biases, the systematic part can represent the stronger mispricing in speculative stocks by market-wide investor sentiment (Baker and Wurgler (2006)).

We estimate the CA models with 6 latent factors. The machine learning literature in the empirical asset pricing assumes the no-arbitrage condition (e.g., Gu et al. (2021), Kelly et al. (2019), Chen et al. (2019)). By following the literature, our baseline specification also does not have the idiosyncratic mispricing part.

[4]Note that the machine learning models estimate not only the factor loading ($\hat{\beta}^L$ or $\hat{\beta}^N$) but also the latent factors (λ^L or λ^N) as linear combinations of the characteristics-managed portfolios. The weights on the linear combinations are estimated using the lagged firm characteristics as well: the latent factors of each of the models can be affected by the nonlinearity of each model. This form of predicted returns is used to examine the profitability of machine learning strategies in Gu et al. (2020) and Avramov et al. (2021)

We construct a *linear* machine learning signal $CA0$ using a linear CA model having no layers of nonlinear activation functions in the systematic part:

$$CA0_{i,t-1} = \hat{\beta}^L(z_{i,t-1})'\lambda^L_{t-1}$$

where $\hat{\beta}^L$ is a linear function of the lagged firm characteristics $z_{i,t-1}$.

We also construct a *nonlinear* machine learning signal $CA3$ using a nonlinear CA model having three layers of nonlinear activation functions in the systematic part:

$$CA3_{i,t-1} = \hat{\beta}^N(z_{i,t-1})'\lambda^N_{t-1}$$

where $\hat{\beta}^N$ is a nonlinear function of the lagged firm characteristics $z_{i,t-1}$.

How are the linear and nonlinear signals different? Figuring out the reasons for the difference is the key to understanding the economic mechanism of deep machine learning models. To study the topic, we introduce our key variable of capturing the distinctive feature of nonlinearity in deep learning models. The variable is constructed by simply taking the difference between the nonlinear and the linear signal. We call the variable as NML (*nonlinear-minus-linear*):

$$NML_{i,t-1} = CA3_{i,t-1} - CA0_{i,t-1}.$$

It represents the predicted returns mainly captured by the nonlinearity of deep learning models. There are several important reasons why we construct NML by taking the difference between $\hat{\beta}^N(z_{i,t-1})'\lambda^N_{t-1}$ and $\hat{\beta}^L(z_{i,t-1})'\lambda^L_{t-1}$ rather than comparing directly $\hat{\beta}^N(z_{i,t-1})$ and $\hat{\beta}^L(z_{i,t-1})$. Although $\hat{\beta}^N(z_{i,t-1})$ and $\hat{\beta}^L(z_{i,t-1})$ are the explicit nonlinear and linear function of firm characteristics $z_{i,t-1}$ respectively, they are the "vector" of exposures to the 6 latent factors, so that direct comparison is meaningless as long as we cannot specify the latent factors. Furthermore, the nonlinearity of the CA3 model affects not only the estimation of the beta part but also the estimation of the latent factors. Therefore, we construct the NML signals using the predicted returns rather than each beta.

Table 1 reports the summary of the machine learning signals generated by CA models including $CA0$, $CA3$, and NML. The mean and standard deviation of the $CA0$ and $CA3$ signals are similar because they are fitted to reproduce the returns by the autoencoder structure (Gu et al. (2021)). However, their performance is different. Their descriptive performance is measured by the predictive R^2. The predictive R^2 represents how well the signal explains the variation of the out-of-sample realized returns.[5] $CA3$ shows 0.63% of predictive R^2, while $CA0$ shows a lower performance that is 0.36% of predictive R^2. The result is consistent with with Gu et al. (2021) that $CA3$ shows better performance than $CA0$ because of $CA3$'s more flexible form than $CA0$. More importantly, $CA3$ is the systematic part, i.e., covariance. Therefore, the results implies that the nonlinear CA model captures the common variation of stock returns with latent factors better than the linear CA model by estimating the latent factors and exposures to the factors more accurately than the linear CA model.

The performance of NML is also notable. The predictive R^2 by NML is significant as 0.25%. It represents almost all of the performance difference between $CA3$ and $CA0$. The significant predictive R^2 by NML represents the covariance of stock returns and systematic factors that is well-explained by using the intricate combinations between firm characteristics captured by the nonlinear CA model but is not easily captured using only the linear CA model.

It is worth discussing why we use the CA model to construct the NML signal to study the machine learning model's economic interpretability. To examine the economic interpretability, cross-sectional comparisons might be crucial. Among many machine learning models in empirical asset pricing such as the neural network model of Gu et al. (2020), or SDF model of Chen et al. (2019), we believe the CA model of Gu et al. (2021) is the most suitable model to construct NML for our purpose. At first, a predicted return from the CA model is written

[5] The predictive R^2 is $1 - \sum_{(i,t)\in OOS}(r_{i,t} - CA0_{i,t-1})^2 / \sum_{(i,t)\in OOS}(r_{i,t}^2)$ where OOS is firm-month observations in the out-of-sample test years by following Gu et al. (2021).

directly as a function of the firm characteristics. It makes the cross-sectional comparisons easier. In comparison, the SDF models by Chen et al. (2019), or Bryzgalova et al. (2020) require further estimation of the exposures to SDF of each stock again. Thus, it might incur additional estimation noise. Second, the CA model has a direct interpretation as a factor model. For example, their systematic part can be interpreted as the covariance captured by the nonlinear structure of the CA model. However, the neural network models of Gu et al. (2020) are pure prediction models, so that it is hard to interpret them from a perspective of the asset pricing factor model. Therefore, using the CA model might be the most simple and effective way of conducting a cross-sectional study on the economic interpretability of machine learning models.

4. Data

4.1 Machine Learning Signals

In the training of the CA models, we use the NYSE, AMEX, and NASDAQ stocks that are ordinary commons shares incorporated in the US. The sample begins in August 1962 and ends in December 2017, covering 56 years. The stock return data is obtained from the Center for Research in Security Prices (CRSP) monthly stock file. We employ the full sample of NYSE, AMEX, and NASDAQ without any filters (Gu et al. (2020), Gu et al. (2021), Avramov et al. (2021)). The unique number of firms in our sample is around 17,000, and the monthly average number of firms is around 4,600.

The lagged firm characteristics are 94 variables constructed by following Green et al. (2017).[6] These firm characteristics are used on the latent factors models as inputs for the systematic parts and latent factors, $z_{i,t-1}$, are The updating frequencies of the firm-level

[6]The set of firm characteristics is widely used in the machine learning literature (Gu et al. (2020), Gu et al. (2021), Avramov et al. (2021)). Details are listed on the Appendix of Green et al. (2017), and the Internet Appendix of Gu et al. (2020). We thank Jeremy Green for building the SAS code to produce the 94 characteristics and sharing the code via his website: https://sites.google.com/site/jeremiahrgreenacctg/home

characteristics are annual (e.g., firm age, gross profitability), quarterly (e.g., returns on equity, revenue surprise), or monthly (e.g., beta, stock return volatility). The variables are rank-transformed into a unit interval for each month (e.g., see Kelly et al. (2019), Gu et al. (2020)). To minimize the forward-looking biases, our firm characteristics are all lagged concerning their updating frequencies.

The initial training sample is 12 years (1962-1973), the validation sample is 12 years (1974-1985), and the out-of-sample test sample is the remaining 32 years (1986-2017). By following Gu et al. (2021), CA models are re-trained for every year in the test sample by increasing the size of the training sample by 1 year while maintaining the size of the validation sample as 12 years by rolling it forward.

4.2 Information Uncertainty

We construct several proxies of information uncertainty by following the literature (e.g., Zhang (2006), Jiang et al. (2005)). Panel B of Table 1 reports 4 proxies of information uncertainty, including stock return volatility ($SIGMA$), firm age (AGE), cash flow volatility ($CFVOL$), and analyst dispersion ($DISP$). The column labeled by High Information Uncertainty indicates the leg of each proxy that is associated with high information uncertainty.

Firms with volatile stock returns have higher information uncertainty (Zhang (2006), Jiang et al. (2005)). Therefore, firms with volatile stock return volatility might have more mispricing by information uncertainty. Stock return volatility ($SIGMA$) is the monthly standard deviation of daily returns.

Young age firms are hard-to-value so that investors are prone to more behavioral biases on young firms (Kumar (2009)). In a similar vein, Hirshleifer et al. (2018) also perceive young age firms as having higher valuation uncertainty. Therefore, younger firms would be more prone to be mispriced. Firm age (AGE) is the number of years since a firm appeared in CRSP.

Volatile cash flows represent higher ambiguity of information in firms' financial reports to a firm's valuation. Therefore, firms with higher volatility in cash flow might have more mispricing (Zhang (2006), Lam and Wei (2011)). Cash flow volatility ($CFVOL$) is the standard deviation of cash flow scaled by assets.

Higher dispersion of analyst forecast refers to higher uncertainty on the firms' prospects (Lang and Lundholm (1996)). Therefore, firms with higher dispersion should have more mispricing. Analyst dispersion ($DISP$) is the standard deviation of analyst forecasts of earnings-per-share from Institutional Brokers' Estimate System (I/B/E/S) (Diether et al. (2002)).

5. Empirical Results

In this section, we discuss our empirical results. Our main focus is on the NML signal. It represents the covariance, instrumented by firm characteristics, mainly captured by the nonlinear structure of deep learning models. We first investigate the return predictability of our NML signal and the cross-sectional difference of the return predictability by the level of information uncertainty. We next examine the trading strategy by the NML signal by portfolio sorts. Finally, we check the exposures of the trading strategy on behavioral factors.

5.1 Return Predictability

5.1.1. Return Predictability of *NML*

This section reports the cross-sectional return predictability of machine learning signals by using Fama-Macbeth regression (Fama and MacBeth (1973)). Each of the machine learning signals is an aggregation of anomaly variables (Avramov et al. (2021)). Therefore, regressing returns on the machine learning signals can be considered as regressing returns on various anomalies' composite signals. Kelly et al. (2021) run a predictive regression of returns on the signals generated by their machine learning model. We also take a similar approach of testing the return predictability of our machine learning signals by using the Fama-Macbeth regression.

Table 2 reports the Fama-Macbeth regression results of investigating the return predictability of our machine learning signals including $CA3$, $CA0$, and NML. Column (1) reports the result using the $CA3$ signal. $CA3$ signal has significant and positive return predictability. A 1% increase of $CA3$ signal predicts 0.94% higher returns.[7] Column (2) reports the result using the $CA0$ signal. It also shows that the $CA0$ signal has statistically significant return predictability that a 1% increase of $CA0$ signal predicts 0.71% higher returns. In Column (3), the NML signal also has significant return predictability that a 1% increase of NML predicts 1.20% higher returns. As $CA3$ is decomposed into $CA0$ and NML, we examine the return predictability of the components together in Column (4). The result shows that $CA0$ and NML have separate return predictability. The magnitude is similar to the univariate cases.

There could be several omitted variables in our regression. Firm size, book-to-market, and momentum are the widely-used determinants of future returns (e.g., Daniel et al. (1997)). We need to control them to see more clear return predictability of our machine learning signals. Column (5) shows a consistent result with the empirical asset pricing literature that small, value and high momentum stocks earn higher returns. However, in Column (6), $CA3$ subsumes almost all of the return predictability of those variables. It is consistent with the result of Kelly et al. (2021) that their machine learning signals subsume predictability of other conventional return predictors such as momentum.[8] In Column (7), using $CA0$ shows similar results. The magnitude of coefficients on $CA3$ also remains the same with the univariate case. In Column (8), we observe that although the NML signal does not subsume the predictability of the ordinary firm characteristics, it still maintains its predictability. It represents that NML has separate return predictability than the common firm characteristics. Lastly, Column (9) shows that controlling the linear signal, $CA0$ effectively subsumes the predictability of common firm characteristics while does not affect NML's return predictability.

In summary, the NML signal shows significant return predictability. Then, what is the source of the return predictability? The following section examines the cross-sectional dif-

[7] Note that the standard deviation of $CA0$, $CA3$, and NML signals are around 1% from Table 1 so that 1% increase of those signals is almost interpretable as economic significance.

[8] The result once again validates the common argument about the performance of machine learning models that they efficiently aggregate multiple anomalies into one signal.

ference of return predictability of the NML signal to discover the origin of the return predictability.

5.1.2. Return Predictability Conditional on Information Uncertainty

In this section, we test whether the return predictability of the NML signal is higher in firms with high information uncertainty. There are empirical evidence that several studies document that the return predictability of anomaly variables is stronger on firms with high information uncertainty (e.g., Zhang (2006), Cohen and Lou (2012), Hirshleifer et al. (2018)). The argument is also supported by the machine learning literature that deep learning models (i.e., nonlinear machine learning) are good at discovering hidden but valuable patterns in complex and uncertain forms of information (LeCun et al. (2015)).

To tackle the empirical question, we run Fama-Macbeth regression focusing on the interaction of NML with a dummy variable of representing firms with high information uncertainty (IU):[9]

$$\begin{aligned} r_{i,t} =& \alpha + \beta_{NML} NML_{i,t-1} + \beta_{IU} HighIU_{i,t-1} + \gamma_{IU} HighIU_{i,t-1} \times NML_{i,t-1} \\ &+ \beta_{CA0} CA0_{i,t-1} + X_{i,t-1} + \epsilon_{i,t} \end{aligned}$$

where $HighIU_{i,t-1}$ is a dummy variable that is 1 when the proxy of information uncertainty $IU_{i,t-1}$ is greater than its monthly cross-sectional median, and $X_{i,t-1}$ represents the lagged control variables including size, book-to-market, and momentum. The proxies of IU are stock return volatility ($SIGMA$), (inverse) firm age ($1/AGE$), cash flow volatility ($CFVOL$), and analyst dispersion ($DISP$). The main variable of interest is the interaction term of NML and $HighIU$. If NML has stronger return predictability in firms with higher information uncertainty, then the coefficient γ_{IU} should be significantly positive.

[9] This empirical strategy is widely used in the literature to study the differential effect of information uncertainty on the anomaly profits (e.g., Cohen and Lou (2012), Hirshleifer et al. (2018), and Lee et al. (2019)).

Table 3 shows the panel regression results of the following month's excess stock returns on the interaction of NML and proxies of information uncertainty. Panel A reports the results using stock return volatility ($SGIMA$) as a proxy of information uncertainty. The results show that the return predictability of NML is higher on firms with higher information uncertainty. Although $CA0$ signal interacts with $HighIU$ in Column (1), the regression model having both $CA0$ and NML signals in Column (4) shows that the interaction of $CA0$ with $HighIU$ becomes insignificant while NML shows significant interaction. Column (4) shows that in the firms with higher information uncertainty proxied by stock return volatility, a 1% increase of NML predicts 0.80% higher returns than the lower uncertainty firms.

Panels B, C, and D use (inverse) firm age ($1/AGE$), cash flow volatility ($CFVOL$), and analyst dispersion ($DISP$) as proxies of information uncertainty (IU). The results are qualitatively similar to Panel A, which uses stock return volatility as the proxy. In Panel B, Column (4) shows that in firms with higher information uncertainty proxied by younger age, a 1% increase of NML predicts 0.23% significantly higher returns than the old firms. However, the stronger return predictability of the $CA0$ signal in firms with high information uncertainty is not statistically different from lower information uncertainty firms. Using cash flow volatility or analyst dispersion as proxies of information uncertainty also shows similar results. Columns (4) in each of Panel C and D show that the NML signal has higher return predictability on firms with higher information uncertainty than lower information uncertainty firms. In contrast, the return predictability of the $CA0$ signal is insignificant or even lower in higher uncertainty firms.

In Daniel et al. (2020), they argue that the interaction of the return predictability associated with their $PEAD$ factor and proxies of limits-to-arbitrage represents that the $PEAD$ factor captures mispricing by behavioral biases, which is not easily explained by the conventional risk-based framework. Our results that NML interacts with proxies of information uncertainty, which also hinders arbitrage opportunities (e.g., Lam and Wei (2011)), also support that NML captures behavioral mispricing.

However, an important point to address is that NML represents the systematic part of the latent factor model. It seems contradictory with our findings that the NML signal captures behavioral mispricing, it seems contradictory. Therefore, we suggest two possible explanations. First, as we do not incorporate the idiosyncratic mispricing part in the latent factor model, the results could be driven by the misspecified term. Second, on the other hand, as NML is a systematic part capturing behavioral mispricing, it might represent the commonality of mispricing by behavioral biases (e.g., Stambaugh and Yuan (2017b), Kozak et al. (2018), Daniel et al. (2020)). Distinguishing the explanations is worth investigating to improve the economic interpretability of machine learning models.

5.1.3. Misspecified Idiosyncratic Mispricing Part

This section discusses the first potential explanation that the misspecified idiosyncratic part derive our results. In the latent K-factor structure of CA model of Gu et al. (2021), predicted returns with idiosyncratic part are written as follows:

$$\hat{r}_{i,t} = \alpha(z_{i,t-1}) + \beta(z_{i,t-1})\lambda_{t-1}.$$

The idiosyncratic part is also a function of firm characteristics $z_{i,t-1}$. It captures the pure idiosyncratic mispricing predicted by firm characteristics that are not covarying with the latent factors. We generate two machine learning signals $CA0^{\alpha+\beta}$ and $CA3^{\alpha+\beta}$. The linear signal, $CA0^{\alpha+\beta}$, is generated by a linear CA model having an idiosyncratic mispricing part: $CA0^{\alpha+\beta}_{i,t-1} = \alpha^L(z_{i,t-1}) + \beta^L(z_{i,t-1})'\lambda^L_{t-1}$ where α^L and β^L are linear functions of $z_{i,t-1}$. The nonlinear signal, $CA3^{\alpha+\beta}$, is generated by a nonlinear CA model having an idiosyncratic mispricing part: $CA3^{\alpha+\beta}_{i,t-1} = \alpha^N(z_{i,t-1}) + \beta^N(z_{i,t-1})'\lambda^L_{t-1}$ where α^N and β^N is nonlinear functions of $z_{i,t-1}$. $NML^{\alpha}_{i,t-1}$ is the difference between the alpha parts of the signals that is $\alpha^N_{i,t-1} - \alpha^L_{i,t-1}$. It represents the pure idiosyncratic mispricing that is easily captured by the nonlinear structure, but not easily by using only the linear structure. $NML^{\beta}_{i,t-1}$ is the difference between the systematic parts of the signals that is $\beta^N(z_{i,t-1})'\lambda^N_{t-1} - \beta^L(z_{i,t-1})'\lambda^L_{t-1}$. It represents the covariance instrumented by firm characteristics, that are captured by the nonlinear structure, but not easily captured by using only the linear structure.

Table 4 summarizes the construction of the machine learning signals and reports summary statistics. The comparison of the predictive R^2 of each machine learning signal is notable. The predictive R^2 of $CA0^{\alpha+\beta}$ is 0.36%, while that of $CA3^{\alpha+\beta}$ is 0.64%. The magnitude of performance and outperformance of the nonlinear signal is similar to the non-idiosyncratic case. It implies that the inclusion of the idiosyncratic part does not have material effects on the performance of the CA models. It is consistent with Kelly et al. (2019), and Gu et al. (2021) who conclude that the pure idiosyncratic part is not necessary to explain the cross-section of asset returns once the latent factor models have enough number of factors, for example, 6. The predictive R^2 of NML^{α} and NML^{β} further validate our arguments. The performance of NML^{α} is 0.02%, while that of NML^{β} is around 0.23%. It means that NML^{α} does not have enough explanatory power for future returns, while NML^{β} has significant explanatory power.

The Fama-Macbeth regression results are also consistent with the descriptive statistics. Table 5 shows the return predictability of machine learning signals with idiosyncratic mispricing parts. In Columns (1) and (2), the return predictability of $CA3^{\alpha+\beta}$ and $CA0^{\alpha+\beta}$ are significantly positive. A 1% increase of each of $CA3^{\alpha+\beta}$ and $CA0^{\alpha+\beta}$ signals predict 0.94% and 0.71% of higher future returns, respectively.[10] In Column (3), NML^{α} also shows significant return predictability in the univariate regression. Since NML^{α} has a smaller standard deviation compared to other machine learning signals as 0.17%, we mention the economic significance formally: a 1 standard deviation increase of NML^{α} predicts around 1.02% of higher future returns (0.173*5.922=1.02). In Column (4), NML^{β} also predicts future returns significantly. A 1 standard deviation increase of NML^{β} predicts 0.86% higher future returns. Then, how does the predictability become different when we control all signals together? Importantly, in Column (5), NML^{α} is no longer significant, while NML^{β} is still significant. Columns (6) to

[10] As both signals have around 1% of standard deviations, the estimated coefficients are almost read by the economic significance.

(11) show that the results are also consistent after controlling firm size, book-to-market, and momentum.

Table 6 further supports that the inclusion of idiosyncratic parts does not have material effects on the higher return predictability of the "systematic" NML in firms with higher information uncertainty (IU).[11] For brevity we report the regression model using $CA0^{\alpha+\beta}$, NML^{α}, and NML^{β}, which is the model in Column (4) of Table 3. In Column (1), the proxy of IU is $SIGMA$. In the firms with high information uncertainty ($HighIU$), NML^{α} has no stronger return predictability. However, NML^{β} has significant and stronger return predictability in firms with higher information uncertainty than the firms with lower information uncertainty. In Columns (2), (3), and (4), by using different proxies of information uncertainty as (inverse) firm age, cash flow volatility, and analyst dispersion, the results are similar that NML^{α}, the pure idiosyncratic part, has no additional return predictability in higher information uncertainty firms. In contrast, NML^{β}, the covariance, has significantly stronger return predictability in firms with higher information uncertainty than the firms with lower information uncertainty.

In summary, we examine the return predictability of NML, the machine learning signal mainly attributable to the nonlinear structure in the systematic part of the latent factors. NML has significant return predictability, and it is stronger for firms with higher information uncertainty. Based on Daniel et al. (2020), the higher return predictability of the NML signal in firms with higher information uncertainty is not clearly explained by the conventional risk-based framework. Rather it represents that NML captures mispricing incurred by behavioral biases; in our case, investor inattention strongly affects the firms with higher information uncertainty. To rule out the potential effect of the misspecified idiosyncratic part to derive our results, we explicitly include the idiosyncratic mispricing part and show that the part of NML does not interact with the proxies information uncertainty.

In contrast, the systematic part of NML still interacts with information uncertainty.

[11] There are 4 proxies of information uncertainty that are stock return volatility ($SIGMA$), (inverse) firm age ($1/AGE$), cash flow volatility ($CFVOL$), and analyst dispersion ($DISP$).

Therefore, a plausible explanation for our findings is that the "systematic" NML signal captures the commonality of mispricing by behavioral biases (e.g., for theory, see Kozak et al. (2018), for empirical factors models see Stambaugh and Yuan (2017a), Daniel et al. (2020)). We test the idea by portfolio-level analysis.

5.2 Portfolios Analysis

5.2.1. Performance of Portfolios Sorted by *NML*

If the NML signal captures the commonality of behavioral mispricing, then a trading strategy that we call the NML strategy, of buying high NML stocks and shorting low NML stocks, might exploit such behavioral mispricing. So the strategy should be profitable. Furthermore, the NML strategy should have high exposure to certain systematic factors that capture such behavioral biases. This section investigates those hypotheses using NML-sorted portfolio. First, we test whether NML-sorted portfolios are profitable, and second, test whether those portfolios have higher exposures on the behavioral factors of Daniel et al. (2020).

Primarily, we sort stocks by their machine learning signals that are $CA0$, $CA1$, $CA2$, $CA3$, and NML. The portfolios are constructed by their NYSE breakpoints, and all returns are valued-weighted.[12] The risk-adjusted returns are estimated by using the five-factor model of Fama and French (2015) augmented by the momentum factor of Carhart (1997).

The construction of $CA0$ portfolios are straightforward. In each month, stocks are sorted into a decile by following NYSE breakpoints of the $CA0$ signal. $CA1$, $CA2$, and $CA3$ portfolios are constructed similarly. NML portfolios are constructed by following the procedure of Ang et al. (2006) to control the $CA0$ signal.[13] Each month, stocks are sorted into one of five portfolios by following NYSE breakpoints of $CA0$ signals. Within each of the five portfolios, stocks are further sorted into a decile using NYSE breakpoints of the NML signal. Then, for each NML decile assignment, five $CA0$ portfolios have the same assignment. The five

[12] We follow the literature to mitigate the effects of microcaps (e.g., Hou et al. (2020))

[13] It rules out $CA0$'s effect on explaining our results.

$CA0$ portfolios are averaged over each of the ten portfolios having the same NML decile assignment. As a result, the final ten portfolios are NML decile portfolios controlling for the $CA0$ signal.

Table 7 reports the risk-adjusted returns of portfolios sorted by machine learning signals. The risk-adjusted returns of long-short portfolio (D10−D1) from the $CA0$ signal is 0.41% per month ($t = 1.82$). On the other hand, the performance of the $CA1$ portfolio with one layer of nonlinear activation function shows better performance as 0.56% per month ($t = 2.33$). As $CA2$ also has two layers of nonlinear activation functions, it also shows higher performance than the $CA0$ portfolio. $CA3$ portfolio that has three layers of nonlinear activation functions shows 0.66% per month ($t = 2.82$). NML portfolio also shows robust performance as 0.68% per month ($t = 3.92$). The portfolio sorting results suggest an important aspect of the nonlinearity in the machine learning models. By adding nonlinearity to the model, performance is improved, and the effect is material that the portfolio sorted by the NML signal also shows significant performance.

5.2.2. Behavioral Factors and Portfolios Sorted by *NML*

As the systematic part of the machine learning model, NML captures mispricing signals by information uncertainty in the cross-section. Then, in the time-series, it might be related to the commonality of mispricing by behavioral biases such as investor inattention. The proxy of investor inattention is adopted from the behavioral factor model (Daniel et al. (2020)). In the model, $PEAD$ captures short-term behavioral mispricing by investor inattention, and FIN captures long-term behavioral mispricing by overconfidence.

Table 8 shows the time-series regression results of the returns of NML portfolios on the behavioral factor model. Columns (1) uses the long-short portfolio returns of NML sorted portfolios as the main dependent variable. The result shows that the returns of NML sorted portfolios have significant covariation with the behavioral factors $PEAD$ and FIN, while it has no exposures on the market factor. By a 1% contemporaneous increase of $PEAD$ (FIN)

factor, NML-sorted portfolio has statistically significant covariation as 0.45% (0.23%). In addition, the magnitude of coefficients is greater for $PEAD$ than FIN. It represents that our NML signal mainly captures mispricing mainly originated from investor inattention proxied by $PEAD$.

As Daniel et al. (2020) document that the mispricing by behavioral biases is related to the short leg, we also check the exposure of the short leg and long leg of the NML sorted portfolios to the behavioral factors. In Column (2), the returns of the NML portfolio are significantly covarying with $PEAD$ and FIN. By a 1% contemporaneous increase of $PEAD$ (FIN) factor, NML-sorted portfolio has statistically significant covariation as -0.52% (-0.31%). On the other hand, in Column (5), the long leg of the NML portfolio does not have covariation with $PEAD$ or FIN.

Collectively, those results explain how the NML signal that is the "systematic" part can capture behavioral mispricing in firms with high information uncertainty. That is because the NML signal captures the commonality of behavioral mispricing by investor inattention, which might be more severe in firms with high information uncertainty. We find that the well-diversified NML-sorted portfolio has significant exposure to behavioral factors. Furthermore, the covariation with the behavioral factor is more concentrated in the short leg, which is also consistent with many empirical findings that short-sale constraints exacerbate mispricing (e.g., Daniel et al. (2020)). In summary, the NML signal captures the commonality of behavioral mispricing that could be stronger in firms with higher information uncertainty.

5.2.3. Performance of Portfolios Double-Sorted by Information Uncertainty and *NML*

The trading strategy of buying high NML stocks and shorting low NML stocks should be more profitable in the group of firms with higher information uncertainty because the behavioral mispricing is stronger on those firms (e.g., Hirshleifer (2001), Hirshleifer et al. (2018)). To test the idea, we construct the double-sorted portfolios by information uncertainty

(IU) and NML.

The IU-NML double-sorted portfolios are constructed as follows. As a first step, each month, by using a proxy of IU, stocks are sorted into one of the IU quintiles by following NYSE breakpoints of the IU proxy. Next, within each IU quintile, NML decile portfolios are constructed by following the procedure of Ang et al. (2006) to control $CA0$ signals. For a IU quintile, stocks are sorted into five portfolios by following NYSE breakpoints of $CA0$ signals. Within each of the five portfolios, stocks are further sorted into a decile using NYSE breakpoints of the NML signal. Then, for each NML decile assignment, five $CA0$ portfolios have the same assignment. The five $CA0$ portfolios are averaged over each of the ten portfolios having the same NML decile assignment. As a result, the final ten portfolios are NML decile portfolios controlling for the $CA0$ signal. Repeating the procedure for each of the IU quintiles produces a total of fifty IU-NML double-sorted portfolios by controlling the $CA0$ signal. In each IU quintile, the long-short portfolio is constructed by buying the highest NML decile and shorting the lowest NML decile.

Table 9 reports the risk-adjusted performance of portfolios doubled sorted by IU-NML signal. The proxies of information uncertainty (IU) are stock return volatility ($SIGMA$), (inverse) firm age ($1/AGE$), cash flow volatility ($CFVOL$), and analyst dispersion ($DISP$). Panel A reports the results using $SIGMA$ as a proxy of IU. Each row represents IU quintiles. For brevity, we report the lowest, fifth, and highest decile of NML portfolios as columns labeled by D1, D5, and D10, respectively. Column D10−D1 shows the risk-adjusted long-short portfolio returns. In the highest IU quintile, the NML long-short portfolio earns a significant 2.54% per month ($t = 7.36$). The short leg shows -1.40% per month ($t = -5.69$), while the long leg shows slightly lower performance in the absolute value of the magnitude, 1.14% per month ($t = 4.70$). As IU decreases, the performance of the long-short portfolio decreases dramatically. In the fourth quintile Q4, the NML long-short portfolio earns just 0.33% per month ($t = 1.49$), it is substantially lower performance than the highest IU quintile. These results imply that in the highest IU quintile, NML effectively exploits behavioral mispricing so that the NML long-short strategy is very profitable. However, the behavioral mispricing captured by NML is disproportionately concentrated on the highest IU quintile proxied by $SIGMA$. Therefore, we explore more proxies of IU to get a more clear view.

In Panel B, $1/AGE$ is the proxy of IU. Similarly, in the highest quintile of IU (young firms), the NML long-short portfolio yields 1.16% per month ($t = 5.33$). As IU decreases (old firms), the long-short risk-adjusted returns decrease gradually. In the lowest quintile, the NML long-short portfolio does not show any significant return that is 0.26% per month ($t = 1.20$). By using $CFVOL$ and $DISP$ in Panels C and D, we still observe similar results. In the highest IU quintile, the NML long-short portfolio risk-adjusted returns are very significant. As IU decreases, the profitability decreases, and in the lowest IU quintile, there is virtually no profit from the NML strategy.

Those results collectively show that the NML strategy exploits behavioral mispricing obscured information uncertainty. It is consistent with the cross-sectional regression that the return predictability of the NML signal becomes stronger in firms with higher information uncertainty.

5.2.4. Behavioral Factors and Portfolios Double-Sorted by Information Uncertainty and *NML*

Lastly, we test whether the good performance of doubled-sorted portfolios by information uncertainty and NML in the group of firms with higher information uncertainty are explained by their exposures on the behavioral factors.

At first, among the IU-NML double-sorted portfolios, we focus on the highest IU quintile's NML long-short portfolio. As a "systematic" part, if the NML strategy harvests significant profits in the firms with high information uncertainty that are more prone to behavioral biases, it should have significant positive exposure to the behavioral factors capturing the commonality of behavioral mispricing.

Table 10 reports the time-series regression results to test the argument. In Panel A, the main dependent variables are the highest IU quintile's NML long-short portfolio returns. For all proxies of IU, the NML strategy shows significant exposure on the $PEAD$ factor. Except for the $SIGMA$, the NML strategy also shows significant exposure on the FIN factor.

On the other hand, we might expect that in firms with low information uncertainty, the NML strategy does not necessarily generate good returns because, in those firms, the behavioral biases would be less important.

Panel B reports the results using the lowest IU quintile's NML long-short portfolio returns as the main dependent variables. The results show that regardless of using any of the proxies of IU, the NML long-short strategy has virtually no exposure to behavioral factors.

Collectively, NML exploits mispricing by behavioral biases in firms with high information uncertainty. Therefore, the performance of the NML long-short strategy using the group of firms with high information uncertainty is largely explained by behavioral factors capturing investor inattention. However, in the group of firms with low information uncertainty, the NML strategy does not produce significant returns.

6. Conclusion

In this paper, we study the economic mechanism of machine learning models. The nonlinear structure of deep learning models captures behavioral mispricing by investor inattention, particularly concentrated on firms with high information uncertainty. Using the NML signal, representing the distinctive feature of the nonlinear structure, we show that the NML signal has return predictability, and the return predictability becomes stronger in firms with high information uncertainty. The cross-sectional findings are consistent with the behavioral asset pricing literature (e.g., Zhang (2006), Daniel et al. (2020)). The idiosyncratic mispricing part does not drive our findings. Therefore, the finding leads to a test on the time-series property of whether the NML signal is related to the common mispricing. In time-series, the NML-sorted portfolio has strong exposure to behavioral factors of Daniel et al. (2020). Combining the cross-section and time-series results, the economic mechanism of deep learning models is to capture the commonality of behavioral mispricing and loadings to factors proxying the commonality.

We contribute to the machine learning literature by investigating the economic meaning of deep learning models. The nonlinear structure of deep learning models capture behavioral mispricing. We also contribute to the common mispricing literature by showing that deep learning models can capture the commonality of behavioral mispricing without prior knowledge.

We suggest several areas of future research. First, it can be possible to explore more specific channels of explaining the economic mechanism of deep learning models. For example, further specifying the categories of information uncertainty would improve our understanding of the nonlinearity of deep learning models. Second, as information uncertainty matters, developing machine learning models that perform better in a more complex information environment is promising. Third, our findings are built on the deep learning model of latent factors, which we believe the most suitable models for studying the cross-sectional comparison of deep learning models' performance. However, there are many other deep learning models with different meanings. Exploring economic meanings in those models would also be valuable. These areas of future research would promise a more successful adoption of machine learning models in empirical asset pricing.

References

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61:259–299.

Avramov, D., S. Cheng, and L. Metzker. 2021. Machine learning versus economic restrictions: Evidence from stock return predictability. *American Financial Association Annual Meeting, Virtual Conference, 3-5 January 2021* .

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

Bryzgalova, S., M. Pelger, and J. Zhu. 2020. Forest through the trees: Building cross-sections of stock returns. *Available at SSRN 3493458* .

Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82.

Chen, L., M. Pelger, and J. Zhu. 2019. Deep learning in asset pricing. *Available at SSRN 3350138* .

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

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52:1035–1058.

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

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

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

Fama, E. F., and R. Kenneth. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3–56.

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

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

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

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

Hirshleifer, D. 2001. Investor psychology and asset pricing. *Journal of Finance* 56:1533–1597.

Hirshleifer, D., P.-H. Hsu, and D. Li. 2018. Innovative originality, profitability, and stock returns. *Review of Financial Studies* 31:2553–2605.

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

Hou, K., C. Xue, and L. Zhang. 2020. Replicating anomalies. *Review of Financial Studies* 33:2019–2133.

Israel, R., and T. J. Moskowitz. 2013. The role of shorting, firm size, and time on market anomalies. *Journal of Financial Economics* 108:275–301.

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

Karolyi, G. A., and S. Van Nieuwerburgh. 2020. New methods for the cross-section of returns. *Review of Financial Studies* 33:1879–1890.

Kelly, B. T., T. J. Moskowitz, and S. Pruitt. 2021. Understanding momentum and reversal. *Journal of Financial Economics* .

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

Kozak, S., S. Nagel, and S. Santosh. 2018. Interpreting factor models. *Journal of Finance* 73:1183–1223.

Kumar, A. 2009. Hard-to-value stocks, behavioral biases, and informed trading. *Journal of Financial and Quantitative Analysis* pp. 1375–1401.

Lam, F. E. C., and K. J. Wei. 2011. Limits-to-arbitrage, investment frictions, and the asset growth anomaly. *Journal of Financial Economics* 102:127–149.

Lang, M. H., and R. J. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* pp. 467–492.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. *Nature* 521:436–444.

Lee, C. M., S. T. Sun, R. Wang, and R. Zhang. 2019. Technological links and predictable returns. *Journal of Financial Economics* 132:76–96.

Stambaugh, R. F., J. Yu, and Y. Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104:288–302.

Stambaugh, R. F., and Y. Yuan. 2017a. Mispricing factors. *Review of Financial Studies* 30:1270–1315.

Stambaugh, R. F., and Y. Yuan. 2017b. Mispricing factors. *Review of Financial Studies* 30:1270–1315.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–1131.

Zhang, X. F. 2006. Information uncertainty and stock returns. *Journal of Finance* 61:105–137.

Table 1: Variable Definitions and Summary Statistics

This table reports definitions and summary statistics of our main variables, including machine learning signals and proxies of information uncertainty. Panel A reports definitions and summary statistics of the machine learning signals generated by CA models. The machine learning signals $CA0_{i,t-1}$, $CA3_{i,t-1}$ are generated by the Conditional Autoencoder (CA) model (Gu et al. (2021)) having the following specification: $r_{i,t} = \beta(z_{i,t-1})' f_t + \epsilon_{i,t}$ where $r_{i,t}$ is the excess returns, $z_{i,t-1}$ is firm characteristics, $\beta(z_{i,t-1})$ is the systematic parts as a linear (or nonlinear) function of $z_{i,t-1}$, and f_t is a K latent factors. For each firm-month observations of firm characteristics $z_{i,t-1}$ in a particular year of the test sample, trained CA models produce machine learning signals as $\hat{\beta}(z_{i,t-1})'\lambda_{t-1}$ where λ_{t-1} is the historical average of the latent factors upto time $t-1$. $CA0_{i,t-1}$ is linear machine learning signals that is the linear systematic part $\hat{\beta}^L(z_{i,t-1})\lambda^L_{t-1}$. $CA3_{i,t-1}$ is nonlinear machine learning signals that is the nonlinear systematic part $\hat{\beta}^N(z_{i,t-1})\lambda^N_{t-1}$. $NML_{i,t-1}$ is the difference between $CA3_{i,t-1}$ and $CA0_{i,t-1}$ that is $CA3_{i,t-1} - CA0_{i,t-1}$. The CA models have 6 latent factors. The CA models do not have idiosyncratic mispricing α part. By following Gu et al. (2021), the initial training sample is 12 years (1962-1973), the validation sample is 12 years (1974-1985), and the out-of-sample test sample is the remaining 32 years (1986-2017). CA models are re-trained for every year in the test sample by increasing the size of the training sample by 1 year while maintaining the size of the validation sample as 12 years by rolling it forward. The monthly stock returns are obtained from the Center for Research in Security Prices (CRSP). Firm characteristics $z_{i,t-1}$ are the 94 variables constructed by Green et al. (2017). The column labeled by Linearity indicates the linearity of systematic part of CA models. The column labeled by α Part shows whether the CA model for each signal uses an idiosyncratic mispricing part. The column labeled by Construction shows the expression for each machine learning signal. Summary statistics, including mean and standard deviation, are reported. The predictive R^2 of each signal is calculated by following Gu et al. (2021). Panel B reports the summary statistics of proxies of information uncertainty, including stock return volatility ($SIGMA$), (inverse) firm age ($1/AGE$), cash flow volatility ($CFVOL$), and analyst dispersion ($DISP$). $SIGMA$ is the monthly standard deviation of daily returns. Firm age (AGE) is the number of years since a firm appeared in CRSP. Firm age is an inverse proxy of information uncertainty, so that $1/AGE$ is a proxy for information uncertainty. $CFVOL$ is the standard deviation of cash flow scaled by assets (Zhang (2006)). Analyst dispersion ($DISP$) is the standard deviation of analyst forecasts of earnings-per-share from Institutional Brokers' Estimate System (I/B/E/S) (Diether et al. (2002)). The column labeled by High Information Uncertainty indicates which characteristics are associated with high information uncertainty.

Panel A: Machine Learning Signals

ML Signal	Linearity	α Part	Construction	Mean (%)	Std. Dev. (%)	Predictive R^2 (%)
$CA0_{i,t-1}$	Linear	No	$\hat{\beta}^L(z_{i,t-1})'\lambda^L_{t-1}$	0.802	1.026	0.356
$CA3_{i,t-1}$	Nonlinear	No	$\hat{\beta}^N(z_{i,t-1})'\lambda^N_{t-1}$	0.831	1.245	0.631
$NML_{i,t-1}$	Nonlinear	No	$CA3_{i,t-1} - CA0_{i,t-1}$	0.029	0.79	0.254

Panel B: Proxies of Information Uncertainty

Variable	Explanation	High Information Uncertainty	Mean	Std. Dev.
$SIGMA$	Stock Return Volatility	Volatile	0.035	0.028
$1/AGE$	(Inverse) Firm Age	Young	14.011	11.852
$CFVOL$	Cash Flow Volatility	Volatile	0.05	0.041
$DISP$	Analyst Dispersion	High Dispersion	0.153	0.407

Table 2: Return Predictability of *NML* Signal

This table reports the Fama-Macbeth regression (Fama and MacBeth (1973)) results of the excess stock returns $r_{i,t}$ on the machine learning signals including $CA3_{i,t-1}$, $CA0_{i,t-1}$, and $NML_{i,t-1}$. The machine learning signals are generated by Conditional Autoencoder (CA) model using the 94 firm characteristics. $CA3_{i,t-1}$ is machine learning signals from a nonlinear CA model having 3 layers of nonlinear activation functions in the systematic part and having no idiosyncratic mispricing part. $CA0_{i,t-1}$ is machine learning signals from a linear CA model having no layers of nonlinear activation functions in the systematic part and having no idiosyncratic mispricing part. $NML_{i,t-1}$ is the difference between $CA3_{i,t-1}$ and $CA0_{i,t-1}$. All machine learning signals are calculated in the out-of-sample. Columns (1) to (4) include $CA3_{i,t-1}$, $CA0_{i,t-1}$ and $NML_{i,t-1}$. Columns (5) to (9) additionally include control variables as firm size ($SIZE_{i,t-1}$), book-to-market ratio ($BM_{i,t-1}$), and momentum of stock returns ($r_{i,t-12,t-2}$). $SIZE_{i,t-1}$, $BM_{i,t-1}$, and $r_{i,t-12,t-2}$ is rank-transformed into a unit interval in month $t-1$. Standard errors adjusted by Newey-West adjustment. t-statistics are reported in the parentheses. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variable					$r_{i,t}$				
$CA3_{i,t-1}$	0.938***					0.978***			
	(11.04)					(11.58)			
$CA0_{i,t-1}$		0.710***		0.782***			0.742***		0.789***
		(8.86)		(9.40)			(8.87)		(9.36)
$NML_{i,t-1}$			1.199***	1.290***				1.276***	1.321***
			(11.25)	(11.42)				(11.63)	(11.72)
$SIZE_{i,t-1}$					-1.033**	0.387	0.117	-1.182***	0.027
					(-2.57)	(1.05)	(0.31)	(-2.87)	(0.07)
$BM_{i,t-1}$					0.662**	0.444	0.394	0.803***	0.532*
					(2.13)	(1.41)	(1.25)	(2.67)	(1.75)
$r_{i,t-12,t-2}$					0.771**	0.206	0.139	1.087***	0.437
					(1.99)	(0.48)	(0.31)	(2.81)	(0.98)
Constant	0.164	0.369	0.890***	0.279	0.723	-0.387	0.018	0.534	-0.226
	(0.52)	(1.18)	(2.74)	(0.89)	(1.13)	(-0.62)	(0.03)	(0.84)	(-0.37)
#Month	384	384	384	384	384	384	384	384	384
Adjusted R^2	0.010	0.007	0.005	0.013	0.022	0.029	0.027	0.027	0.032

Table 3: Return Predictability of *NML* Signal and Information Uncertainty

This table reports the Fama-Macbeth regression results of the excess stock returns $r_{i,t}$ on the machine learning signals including $CA0_{i,t-1}$ and $NML_{i,t-1}$, and their interactions with dummy variables constructed by proxies of information uncertainty (IU). The machine learning signals are generated by Conditional Autoencoder (CA) model using the 94 firm characteristics. $CA3_{i,t-1}$ is machine learning signals from a nonlinear CA model having 3 layers of nonlinear activation functions and having no idiosyncratic mispricing part. $CA0_{i,t-1}$ is machine learning signals from a linear CA model having no layers of nonlinear activation functions and having no idiosyncratic mispricing part. $NML_{i,t-1}$ is the difference between $CA3_{i,t-1}$ and $CA0_{i,t-1}$. All machine learning signals are calculated in the out-of-sample. There are 4 proxies of information uncertainty (IU) including stock return volatility ($SIGMA$), (inverse) firm age ($1/AGE$), analyst dispersion ($DISP$), and cash flow volatility ($CFVOL$). The dummy variable $HighIU_{i,t-1}$ is 1 when $IU_{i,t-1}$ is greater than its monthly median. Control variables are monthly rank-transformed firm size, book-to-market-ratio, and momentum of stock returns. Panel A reports the results by using $SIGMA$ as a proxy of IU. Column (1) reports the result by using $HighIU_{i,t-1}$ as the main independent variable. Column (2) reports the result by using $HighIU_{i,t-1}$ and its interaction with $CA0_{i,t-1}$. Column (3) reports the result by using $HighIU_{i,t-1}$ and its interaction with $NML_{i,t-1}$. Column (4) reports the result by using $HighIU_{i,t-1}$ and its interaction with $CA0_{i,t-1}$ and $NML_{i,t-1}$. Panel B reports the result by using $1/AGE$, Panel C reports the result by using $CFVOL$, and Panel D reports the result by using $DISP$. Standard errors are estimated by Newey-West adjustment. t-statistics are reported in the parentheses. ***, **, * denotes 1%, 5%, and 10% statistical significance.

Panel A: IU Proxied by $SIGMA$

Variable	(1)	(2)	(3)	(4)
	$r_{i,t}$			
$CA0_{i,t-1}$	0.806***	0.983***	0.692***	0.676***
	(9.84)	(11.43)	(8.53)	(7.52)
$NML_{i,t-1}$	1.313***	1.380***	0.685***	0.692***
	(11.81)	(12.27)	(6.17)	(5.84)
$HighIU_{i,t-1}$	-0.290	-0.074	-0.283	-0.306
	(-1.58)	(-0.36)	(-1.56)	(-1.39)
$\times\ CA0_{i,t-1}$		-0.260***		0.029
		(-3.99)		(0.36)
$\times\ NML_{i,t-1}$			0.810***	0.803***
			(9.19)	(7.08)
Constant	0.051	-0.101	0.097	0.125
	(0.11)	(-0.21)	(0.20)	(0.27)
Controls	Yes	Yes	Yes	Yes
#Month	384	384	384	384
Adjusted R^2	0.037	0.037	0.037	0.038

Panel B: IU Proxied by $1/AGE$

Variable	(5)	(6)	(7)	(8)
	$r_{i,t}$			
$CA0_{i,t-1}$	0.791***	0.813***	0.778***	0.781***
	(9.32)	(10.10)	(9.19)	(9.56)
$NML_{i,t-1}$	1.318***	1.321***	1.179***	1.167***
	(11.74)	(11.83)	(9.64)	(9.49)
$HighIU_{i,t-1}$	-0.106	-0.067	-0.116	-0.102
	(-1.49)	(-0.84)	(-1.64)	(-1.25)
$\times\ CA0_{i,t-1}$		-0.042		-0.011
		(-1.01)		(-0.23)
$\times\ NML_{i,t-1}$			0.217***	0.233***
			(2.92)	(2.94)
Constant	-0.125	-0.148	-0.107	-0.114
	(-0.22)	(-0.26)	(-0.18)	(-0.20)
Controls	Yes	Yes	Yes	Yes
#Month	384	384	384	384
Adjusted R^2	0.033	0.033	0.033	0.034

Table 3 Continues

Panel C: *IU* Proxied by *CFVOL*

Variable	(1)	(2)	(3)	(4)
	$r_{i,t}$			
$CA0_{i,t-1}$	0.794***	0.825***	0.781***	0.805***
	(9.54)	(9.22)	(9.48)	(9.02)
$NML_{i,t-1}$	1.269***	1.276***	1.146***	1.159***
	(11.79)	(11.75)	(10.52)	(10.26)
$HighIU_{i,t-1}$	-0.198***	-0.155*	-0.199***	-0.167**
	(-2.77)	(-1.92)	(-2.74)	(-2.00)
$\times\ CA0_{i,t-1}$		-0.056		-0.043
		(-1.24)		(-0.88)
$\times\ NML_{i,t-1}$			0.186**	0.173**
			(2.22)	(1.97)
Constant	0.148	0.128	0.155	0.141
	(0.26)	(0.23)	(0.28)	(0.25)
Controls	Yes	Yes	Yes	Yes
#Month	384	384	384	384
Adjusted R^2	0.032	0.032	0.032	0.033

Panel D: *IU* Proxied by *DISP*

	(5)	(6)	(7)	(8)
	$r_{i,t}$			
$CA0_{i,t-1}$	0.785***	0.974***	0.748***	0.873***
	(8.39)	(9.91)	(7.94)	(8.24)
$NML_{i,t-1}$	1.025***	1.076***	0.725***	0.848***
	(8.54)	(8.77)	(5.79)	(5.80)
$HighIU_{i,t-1}$	-0.137	0.043	-0.180	-0.050
	(-1.17)	(0.36)	(-1.53)	(-0.38)
$\times\ CA0_{i,t-1}$		-0.300***		-0.185***
		(-7.15)		(-3.03)
$\times\ NML_{i,t-1}$			0.453***	0.310**
			(4.85)	(2.52)
Constant	0.281	0.182	0.357	0.266
	(0.42)	(0.27)	(0.54)	(0.41)
Controls	Yes	Yes	Yes	Yes
#Month	348	348	348	348
Adjusted R^2	0.048	0.048	0.048	0.049

Table 4: Definitions of Machine Learning Signals Having Idiosyncratic Mispricing Part

This table reports definitions and summary statistics of machine learning signals generated by Conditional Autoencoder (CA) models having idiosyncratic mispricing parts. The model specification of the CA models is as follows: $r_{i,t} = \alpha(z_{i,t-1}) + \beta(z_{i,t-1})' f_t + \epsilon_{i,t}$. The excess returns $r_{i,t}$ possess a latent K-latent factor structure by f_t. The exposure to the K-latent factors are characterized by the systematic part $\beta(z_{i,t-1})$. The model also has an idiosyncratic mispricing part $\alpha(z_{i,t-1})$. Both of and the systematic and the idiosyncratic mispricing parts are instrumented by firm characteristics $z_{i,t-1}$. The monthly stock returns are obtained from Center for Research in Security Prices (CRSP). The firm characteristics are 94 variables constructed by following Green et al. (2017). The initial training sample is 12 years (1962-1973), the validation sample is 12 years (1974-1985), and the out-of-sample test sample is the remaining 32 years (1986-2017). By following Gu et al. (2021), CA models are re-trained for every year in the test sample by increasing the size of the training sample by 1 year while maintaining the size of the validation sample as 12 years by rolling it forward. For each firm-month observations in a particular year of the test sample, trained CA models produce machine learning signals as $\hat{\alpha}(z_{i,t-1}) + \hat{\beta}(z_{i,t-1})'\lambda_{t-1}$ where λ_{t-1} is the historical average of latent factors upto the start of the year. All CA models have 6 latent factors and the idiosyncratic mispricing part. $CA0^{\alpha+\beta}_{i,t-1}$ is machine learning signals from a linear CA model having no layers of nonlinear activation functions in the idiosyncratic mispricing and the systematic part: $CA0^{\alpha+\beta}_{i,t-1} = \hat{\alpha}^L(z_{i,t-1}) + \hat{\beta}^L(z_{i,t-1})'\lambda^L_{t-1}$ where $\hat{\alpha}^L$ and $\hat{\beta}^L$ are linear functions of $z_{i,t-1}$. The column labeled by Linearity indicates whether a CA model to generate each signal has a linear or nonlinear structure in its idiosyncratic mispricing part and systematic part. The column labeled by α Part shows whether the CA model for each signal uses an idiosyncratic mispricing part α or not. The column labeled by Construction shows the expression for each machine learning signal. Summary statistics, including mean and standard deviation, are reported. The predictive R^2 of each signal is also reported. The predictive R^2 is $1 - \sum_{(i,t)\in OOS}(r_{i,t} - CA0^{\alpha+\beta}_{i,t-1})^2 / \sum_{(i,t)\in OOS}(r^2_{i,t})$ where OOS is firm-month observations in the out-of-sample test years by following Gu et al. (2021). $CA3^{\alpha+\beta}_{i,t-1}$ is machine learning signals from a nonlinear CA model having no layers of nonlinear activation functions in the idiosyncratic mispricing and the systematic part: $CA3^{\alpha+\beta}_{i,t-1} = \hat{\alpha}^N(z_{i,t-1}) + \hat{\beta}^N(z_{i,t-1})'\lambda^L_{t-1}$ where $\hat{\alpha}^N$ and $\hat{\beta}^N$ is nonlinear functions of $z_{i,t-1}$. $NML^{\alpha}_{i,t-1}$ is the difference between the alpha parts of the signals that is $\hat{\alpha}^N_{i,t-1} - \hat{\alpha}^L_{i,t-1}$. $NML^{\beta}_{i,t-1}$ is the difference between the systematic parts of the signals that is $\hat{\beta}^N(z_{i,t-1})'\lambda^N_{t-1} - \hat{\beta}^L(z_{i,t-1})'\lambda^L_{t-1}$.

ML Signal	Linearity	α Part	Construction	Mean (%)	Std. Dev. (%)	Predictive R^2 (%)
$CA0^{\alpha+\beta}_{i,t-1}$	Linear	Yes	$\hat{\alpha}^L(z_{i,t-1}) + \hat{\beta}^L(z_{i,t-1})'\lambda^L_{t-1}$	0.805	1.029	0.359
$CA3^{\alpha+\beta}_{i,t-1}$	Nonlinear	Yes	$\hat{\alpha}^N(z_{i,t-1}) + \hat{\beta}^N(z_{i,t-1})'\lambda^N_{t-1}$	0.801	1.258	0.637
$NML^{\alpha}_{i,t-1}$	Nonlinear	Yes	$\hat{\alpha}^N(z_{i,t-1}) - \hat{\alpha}^L(z_{i,t-1})$	-0.095	0.173	0.018
$NML^{\beta}_{i,t-1}$	Nonlinear	No	$\hat{\beta}^N(z_{i,t-1})'\lambda^N_{t-1} - \hat{\beta}^L(z_{i,t-1})'\lambda^L_{t-1}$	0.091	0.757	0.233

Table 5: Return Predictability of *NML* Signal Having Idiosyncratic Mispricing Parts

This table reports the Fama-Macbeth regression results of the excess stock returns $r_{i,t}$ on the machine learning signals including $CA3^{\alpha+\beta}_{i,t-1}$, $CA0^{\alpha+\beta}_{i,t-1}$, $NML^{\alpha}_{i,t-1}$, and $NML^{\beta}_{i,t-1}$. The machine learning signals are generated by Conditional Autoencoder (CA) models using the 94 firm characteristics $z_{i,t-1}$. $CA3^{\alpha+\beta}_{i,t-1}$ is machine learning signals from a nonlinear CA model having 3 layers of nonlinear activation functions in both of the systematic part $\hat{\beta}^N(z_{i,t-1})\lambda^N_{t-1}$ and the idiosyncratic mispricing part $\hat{\alpha}^N(z_{i,t-1})$ where λ^N_{t-1} is the historical average of latent factors. $CA0^{\alpha+\beta}_{i,t-1}$ is machine learning signals from a linear CA model having no layers of nonlinear activation function in both of the systematic part $\hat{\beta}^L(z_{i,t-1})\lambda^L_{t-1}$ and the idiosyncratic mispricing part $\hat{\alpha}^L(z_{i,t-1})$ where λ^L_{t-1} is the historical average of latent factors. $NML^{\alpha}_{i,t-1}$ is the difference between the idiosyncratic mispricing parts that is $\hat{\alpha}^N(z_{i,t-1}) - \hat{\alpha}^L(z_{i,t-1})$. $NML^{\beta}_{i,t-1}$ is the difference between the systematic parts that is $\hat{\beta}^N(z_{i,t-1})'\lambda^N_{t-1} - \hat{\beta}^L(z_{i,t-1})'\lambda^L_{t-1}$. All machine learning signals are calculated in the out-of-sample. Columns (1) to (5) include $CA3^{\alpha+\beta}_{i,t-1}$, $CA0^{\alpha+\beta}_{i,t-1}$, $NML^{\alpha}_{i,t-1}$ and $NML^{\beta}_{i,t-1}$. Columns (6) to (11) additionally include control variables as firm size ($SIZE_{i,t-1}$), book-to-market ratio ($BM_{i,t-1}$), and momentum of stock returns ($r_{i,t-12,t-2}$). $SIZE_{i,t-1}$, $BM_{i,t-1}$, and $r_{i,t-12,t-2}$ is rank-transformed into a unit interval in month $t-1$. Standard errors are estimated by Newey-West adjustment. t-statistics are reported in the parentheses. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Variable						$r_{i,t}$					
$CA3^{\alpha+\beta}_{i,t-1}$	0.940***						0.976***				
	(11.64)						(12.06)				
$CA0^{\alpha+\beta}_{i,t-1}$		0.712***			0.819***			0.744***			0.850***
		(8.87)			(5.92)			(8.86)			(5.86)
$NML^{\alpha}_{i,t-1}$			5.922***		1.415				7.854***		1.263
			(2.80)		(0.52)				(3.97)		(0.49)
$NML^{\beta}_{i,t-1}$				1.130***	1.285***					1.248***	1.351***
				(9.55)	(6.74)					(10.90)	(7.13)
$SIZE_{i,t-1}$						-1.033**	0.353	0.128	-1.972***	-1.002**	0.177
						(-2.57)	(0.94)	(0.34)	(-5.70)	(-2.48)	(0.75)
$BM_{i,t-1}$						0.662**	0.399	0.392	-0.177	0.885***	0.346*
						(2.13)	(1.26)	(1.25)	(-0.73)	(2.93)	(1.69)
$r_{i,t-12,t-2}$						0.771**	0.131	0.126	-0.451	1.227***	0.327
						(1.99)	(0.30)	(0.28)	(-1.08)	(3.18)	(0.88)
Constant	0.190	0.366	1.381***	0.820**	0.318	0.723	-0.279	0.018	2.852***	0.251	-0.169
	(0.60)	(1.17)	(5.00)	(2.52)	(1.40)	(1.13)	(-0.44)	(0.03)	(5.15)	(0.40)	(-0.50)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
#Month	384	384	384	384	384	384	384	384	384	384	384
Adjusted R^2	0.010	0.007	0.015	0.005	0.032	0.022	0.029	0.027	0.030	0.026	0.043

Table 6: Return Predictability of *NML* Signal Having Idiosyncratic Mispricing Parts and Information Uncertainty

This table reports the Fama-Macbeth regression results of the excess stock returns $r_{i,t}$ on the machine learning signals including $CA0^{\alpha+\beta}_{i,t-1}$, $NML^{\alpha}_{i,t-1}$, and $NML^{\beta}_{i,t-1}$, and their interactions with dummy variables constructed by using proxies of information uncertainty (IU). The machine learning signals are generated by Conditional Autoencoder (CA) models using the 94 firm characteristics $z_{i,t-1}$. $CA3^{\alpha+\beta}_{i,t-1}$ is machine learning signals from a nonlinear CA model having 3 layers of nonlinear activation functions in the systematic part $\hat{\beta}^N(z_{i,t-1})\lambda^N_{t-1}$ and having an idiosyncratic mispricing part $\hat{\alpha}^N(z_{i,t-1})$ where λ^N_{t-1} is the historical average of latent factors. $CA0^{\alpha+\beta}_{i,t-1}$ is machine learning signals from a linear CA model having no layers of nonlinear activation function in the systematic part $\hat{\beta}^L(z_{i,t-1})\lambda^N_{t-1}$ and having an idiosyncratic mispricing part $\hat{\alpha}^L(z_{i,t-1})$ where λ^L_{t-1} is the historical average of latent factors. $NML^{\alpha}_{i,t-1}$ is the difference between the idiosyncratic mispricing parts that is $\hat{\alpha}^N(z_{i,t-1}) - \hat{\alpha}^L(z_{i,t-1})$. $NML^{\beta}_{i,t-1}$ is the difference between the systematic parts that is $\hat{\beta}^N(z_{i,t-1})'\lambda^N_{t-1} - \hat{\beta}^L(z_{i,t-1})'\lambda^L_{t-1}$. All machine learning signals are calculated in the out-of-sample. There are 4 proxies of information uncertainty (IU) including stock return volatility ($SIGMA$), (inverse) firm age ($1/AGE$), cash flow volatility ($CFVOL$), and analyst dispersion ($DISP$). The dummy variable $HighIU_{i,t-1}$ is 1 when $IU_{i,t-1}$ is greater than its monthly median. Control variables are firm size, book-to-market-ratio, and momentum of stock returns that are rank-transformed in each month. For each Column labeled by one of IU reports the results by using $HighIU_{i,t-1}$ and its interactions with $CA0^{\alpha+\beta}_{i,t-1}$, $NML^{\alpha}_{i,t-1}$, and $NML^{\beta}_{i,t-1}$ as main independent variables. Standard errors are estimated by Newey-West adjustment. t-statistics are reported in the parentheses. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)	(2)	(3)	(4)
IU Proxied by	$SIGMA$	$1/AGE$	$CFVOL$	$DISP$
Variable	$r_{i,t}$			
$CA0^{\alpha+\beta}_{i,t-1}$	0.657***	0.816***	0.821***	0.899***
	(4.22)	(5.55)	(5.23)	(4.75)
$NML^{\alpha}_{i,t-1}$	1.960	1.254	0.994	1.656
	(0.63)	(0.48)	(0.36)	(0.53)
$NML^{\beta}_{i,t-1}$	0.605***	1.170***	1.108***	0.828***
	(3.33)	(5.81)	(5.67)	(3.49)
$HighIU_{i,t-1}$	-0.498***	-0.146	-0.129	-0.016
	(-2.70)	(-1.40)	(-1.21)	(-0.11)
$\times$ $CA0^{\alpha+\beta}_{i,t-1}$	0.129	0.042	-0.058	-0.180*
	(1.32)	(0.61)	(-0.87)	(-1.93)
$\times$ $NML^{\alpha}_{i,t-1}$	-1.267	-0.084	0.757	-0.291
	(-0.95)	(-0.15)	(0.86)	(-0.28)
$\times$ $NML^{\beta}_{i,t-1}$	1.015***	0.288***	0.260**	0.284*
	(6.70)	(2.81)	(2.41)	(1.78)
Constant	0.220	-0.049	0.325	0.346
	(0.72)	(-0.15)	(1.06)	(0.83)
Controls	Yes	Yes	Yes	Yes
#Month	384	384	384	348
Adjusted R^2	0.046	0.045	0.044	0.062

Table 7: Risk-Adjusted Returns of Portfolios Sorted by Machine Learning Signals

This table reports the the value-weighted risk-adjusted returns of portfolios sorted by machine learning signals estimated by Conditional Autoencoder (CA) models. The machine learning signals are $CA0$, $CA1$, $CA2$, $CA3$, and NML. $CA0$ is machine learning signals from a linear CA model having no layers of nonlinear activation functions. $CA1$ ($CA2$, $CA3$) is machine learning signals from a nonlinear CA model with 1 (2, 3) layers of nonlinear activation functions. All the CA models do not have idiosyncratic mispricing parts. NML is the difference between $CA3$ and $CA0$. To construct $CA0$ portfolios, in each month, stocks are sorted into a decile by following NYSE breakpoints of the $CA0$ signal. $CA1$, $CA2$, and $CA3$ portfolios are constructed similarly. NML portfolios are constructed by following the procedure of Ang et al. (2006) to control the $CA0$ signal. Each month, stocks are sorted into one of five portfolios by following NYSE breakpoints of $CA0$ signals. Within each of the five portfolios, stocks are further sorted into a decile using NYSE breakpoints of the NML signal. Then, for each NML decile assignment, five $CA0$ portfolios have the same assignment. The five $CA0$ portfolios are averaged over each of the ten portfolios having the same NML decile assignment. As a result, the final ten portfolios are NML decile portfolios controlling for the $CA0$ signal. The excess returns are value-weighted, and their risk-adjusted returns are estimated by the five-factor model of Fama and French (2015) augmented by momentum factor of Carhart (1997). The standard errors are estimated by Newey-West adjustment. t-statistics are reported in the parentheses. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	ML Signal Decile										
ML Signal	D1 (Low)	D2	D3	D4	D5	D6	D7	D8	D9	D10 (High)	D10−D1
$CA0$	-0.172	0.006	0.070	0.114	-0.121	0.062	0.146	-0.076	0.147	0.238	0.410*
	(-1.56)	(0.08)	(0.83)	(1.52)	(-1.41)	(0.85)	(1.41)	(-0.79)	(0.97)	(1.51)	(1.82)
$CA1$	-0.158	0.041	-0.069	-0.041	0.015	-0.060	0.209**	0.020	0.184	0.398**	0.556**
	(-1.10)	(0.44)	(-0.81)	(-0.53)	(0.16)	(-0.82)	(2.15)	(0.23)	(1.38)	(2.37)	(2.33)
$CA2$	-0.221	0.021	0.133	-0.057	0.033	-0.095	0.099	0.162	0.141	0.312*	0.532*
	(-1.48)	(0.21)	(1.50)	(-0.76)	(0.46)	(-1.19)	(0.94)	(1.62)	(1.30)	(1.65)	(1.92)
$CA3$	-0.214	0.032	0.013	-0.158*	0.048	-0.036	0.265***	0.157	0.041	0.446***	0.660***
	(-1.47)	(0.31)	(0.18)	(-1.89)	(0.64)	(-0.49)	(3.10)	(1.63)	(0.30)	(2.87)	(2.82)
NML	-0.166	0.067	0.054	-0.061	0.115*	0.022	-0.024	0.054	0.113	0.511***	0.677***
	(-1.15)	(0.74)	(0.72)	(-0.86)	(1.70)	(0.31)	(-0.36)	(0.63)	(1.35)	(4.51)	(3.92)

Table 8: Portfolios Sorted by *NML* Signal, and Their Relationship with Behavioral Factors

This table reports the time-series regression results of regressing the returns of value-weighted portfolios sorted by NML signals on the short- and long-horizon behavioral factors of Daniel et al. (2020). The NML signals are obtained by taking the difference between the $CA3$ signal and $CA0$ signal, $NML = CA3 - CA0$. The $CA3$ ($CA0$) signals are machine learning signals from a nonlinear (linear) CA model having 3 (no) layers of nonlinear activation functions. All the CA models do not have idiosyncratic mispricing parts. Each month, $CA0$ and $CA3$ portfolios are constructed by sorting stocks into a decile by following NYSE breakpoints using the $CA0$ or $CA3$ signals. NML portfolios are constructed by following the procedure of Ang et al. (2006) to control $CA0$ signals. Each month, stocks are sorted into one of five portfolios by following NYSE breakpoints of $CA0$ signals. Within each of the five portfolios, stocks are further sorted into a decile using NYSE breakpoints of the NML signal. Then, for each NML decile assignment, five $CA0$ portfolios have the same assignment. The five $CA0$ portfolios are averaged over each of the ten portfolios having the same NML decile assignment. As a result, the final ten portfolios are NML decile portfolios controlling for the $CA0$ signal. Column (1) reports the results using the returns of long-short (D10−D1) NML portfolios as the main dependent variable. The long-short (D10−D1) portfolio is constructed by buying the highest decile (D10) and selling the lowest decile (D1). The main independent variables are the short- and long-horizon behavioral factors of Daniel et al. (2020) including market factor ($MKTRF_t$), short-horizon behavioral factor ($PEAD_t$), and long-horizon behavioral factor (FIN_t). Columns (2) reports the results using the lowest decile (D1) NML portfolio returns as the main dependent variable, and Columns (3) reports the results using the highest decile (D10) NML portfolio returns. The standard errors are estimated by Newey-West adjustment. t-statistics are reported in the parentheses. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)	(2)	(3)
NML Portfolio	D10−D1	D1	D10
Variable		r_t	
$MKTRF_t$	-0.029	1.148***	1.119***
	(-0.51)	(20.04)	(35.29)
$PEAD_t$	0.452***	-0.522***	-0.070
	(2.99)	(-3.46)	(-0.85)
FIN_t	0.234**	-0.313***	-0.079
	(2.57)	(-6.46)	(-1.19)
Constant	0.574***	0.102	0.677***
	(2.77)	(0.56)	(5.14)
Observations	384	384	384
Adjusted R^2	0.129	0.828	0.844

Table 9: Performance of Portfolios Double-Sorted by Information Uncertainty and *NML* Signal

This table reports the value-weighted risk-adjusted returns of portfolios double-sorted by information uncertainty (IU) and NML signals. The proxies of information uncertainty (IU) are stock return volatility ($SIGMA$), (inverse) firm age ($1/AGE$), cash flow volatility ($CFVOL$), and analyst dispersion ($DISP$). The NML signals are obtained by taking the difference between the $CA3$ signal and $CA0$ signal, $NML = CA3 - CA0$. The $CA3$ ($CA0$) signals are generated by a nonlinear (linear) Conditional Autoencoder model having 3 (no) layers of nonlinear activation functions in the systematic part. The IU-NML double-sorted portfolios are constructed as follows. As a first step, each month, by using a proxy of IU, stocks are sorted into one of the IU quintiles by following NYSE breakpoints of the IU proxy. Next, within each IU quintile, NML decile portfolios are constructed by following the procedure of Ang et al. (2006) to control $CA0$ signals. For a IU quintile, stocks are sorted into one of five portfolios by following NYSE breakpoints of $CA0$ signals. Within each of the five portfolios, stocks are further sorted into a decile using NYSE breakpoints of the NML signal. Then, for each NML decile assignment, five $CA0$ portfolios have the same assignment. The five $CA0$ portfolios are averaged over each of the ten portfolios having the same NML decile assignment. As a result, the final ten portfolios are NML decile portfolios controlling for the $CA0$ signal. Repeating the procedure for each of the IU quintiles produces a total of fifty IU-NML double-sorted portfolios by controlling the $CA0$ signal. In each IU quintile, the long-short portfolio is constructed by buying the highest NML decile and selling the lowest NML decile. Panel A reports the result by using the proxy of IU as $SIGMA$. For the IU quintiles, Q1 is the lowest IU quintile, while Q5 is the highest IU quintile. For NML deciles, the risk-adjusted returns of lowest decile (D1 (Low)), D5, and the highest decile (D10 (High)), and the long-short portfolio (D10−D1) are only reported for brevity. Similarly, Panel B uses $1/AGE$, Panel C uses $CFVOL$, and Panel D uses $DISP$ as the proxy of IU. The risk-adjusted returns are estimated by the five-factor model of Fama and French (2015) augmented by momentum factor of Carhart (1997). The standard errors are estimated by Newey-West adjustment. t-statistics are reported in the parentheses. ***, **, * denotes 1%, 5%, and 10% statistical significance.

Panel A: IU Proxied by $SIGMA$

	NML Decile			
IU Quintile	D1 (Low)	D5	D10 (High)	D10−D1
Q1 (Low)	0.114	-0.121	0.191	0.078
	(1.09)	(-1.08)	(1.55)	(0.44)
Q2	0.032	0.071	0.043	0.010
	(0.26)	(0.73)	(0.33)	(0.06)
Q3	-0.037	-0.099	-0.011	0.026
	(-0.30)	(-0.81)	(-0.08)	(0.15)
Q4	-0.014	-0.286**	0.317*	0.331
	(-0.08)	(-2.30)	(1.84)	(1.49)
Q5 (High)	-1.398***	-0.082	1.144***	2.542***
	(-5.69)	(-0.51)	(4.70)	(7.36)

Panel B: IU Proxied by $1/AGE$

	NML Decile			
IU Quintile	D1 (Low)	D5	D10 (High)	D10−D1
Q1 (Low)	-0.262	-0.031	-0.000	0.262
	(-1.59)	(-0.30)	(-0.00)	(1.20)
Q2	-0.511***	-0.104	0.205	0.716***
	(-2.66)	(-0.96)	(1.53)	(3.22)
Q3	-0.309*	0.047	0.392**	0.701***
	(-1.83)	(0.37)	(2.21)	(3.05)
Q4	-0.411***	0.322***	0.542***	0.953***
	(-2.61)	(3.07)	(3.82)	(4.50)
Q5 (High)	-0.566***	0.170	0.593***	1.159***
	(-3.28)	(1.26)	(3.73)	(5.33)

Table 9 Continues

Panel C: *IU* Proxied by *CFVOL*

IU Quintile	*NML* Decile D1 (Low)	D5	D10 (High)	D10−D1
Q1 (Low)	0.168	-0.038	0.172	0.004
	(0.82)	(-0.29)	(1.18)	(0.02)
Q2	-0.269	-0.034	0.144	0.413
	(-1.34)	(-0.29)	(0.79)	(1.55)
Q3	-0.207	0.001	0.293**	0.500**
	(-1.14)	(0.00)	(2.22)	(2.28)
Q4	-0.586***	-0.094	0.693***	1.279***
	(-2.99)	(-0.71)	(4.15)	(5.04)
Q5 (High)	-0.885***	-0.052	0.596***	1.481***
	(-4.98)	(-0.46)	(3.37)	(6.54)

Panel D: *IU* Proxied by *DISP*

IU Quintile	*NML* Decile D1 (Low)	D5	D10 (High)	D10−D1
Q1 (Low)	0.064	0.100	0.256*	0.192
	(0.51)	(0.69)	(1.79)	(1.05)
Q2	-0.021	0.047	-0.108	-0.086
	(-0.15)	(0.41)	(-0.86)	(-0.39)
Q3	-0.188	-0.028	0.144	0.332
	(-1.21)	(-0.25)	(1.00)	(1.47)
Q4	-0.180	0.024	0.608***	0.788***
	(-0.68)	(0.16)	(3.33)	(2.62)
Q5 (High)	-0.825***	-0.373**	0.628***	1.454***
	(-2.93)	(-2.12)	(3.37)	(4.75)

Table 10: Portfolios Double-Sorted by Information Uncertainty and *NML* Signal, and Their Relationship with Behavioral Factors

This table reports the time-series properties of the returns of portfolio double-sorted by information uncertainty (IU) and NML signal. The time-series property is examined by their relationship with behavioral factors and market state variables. The proxies of information uncertainty (IU) are stock return volatility ($SIGMA$), (inverse) firm age ($1/AGE$), analyst dispersion ($DISP$), and cash flow volatility ($CFVOL$). The NML signals are obtained by taking the difference between the $CA3$ signal and $CA0$ signal, $NML = CA3 - CA0$. The $CA3$ ($CA0$) signals are machine learning signals from a nonlinear (linear) CA model having 3 (no) layers of nonlinear activation functions. Then, NML is $CA3 - CA0$. All the CA models do not have idiosyncratic mispricing parts. The IU-NML double-sorted portfolios are constructed as follows. As a first step, each month, by using a proxy of IU, stocks are sorted into one of the IU quintiles by following NYSE breakpoints of the IU proxy. Next, within each IU quintile, NML decile portfolios are constructed by following the procedure of Ang et al. (2006) to control $CA0$ signals. Panel A reports the time-series regression results using the long-short NML portfolio returns in the highest IU quintile as the main dependent variable. The main independent variable is the long- and short-horizon behavioral factors of Daniel et al. (2020). Column (1) uses $SIGMA$ as the proxy of IU. Column (2) uses $1/AGE$, Column (3) uses $CFVOL$, and Column (4) uses $DISP$. Panel B reports the time-series regression results using the long-short NML portfolio returns in the lowest IU quintile as the main dependent variable. The standard errors are estimated by Newey-West adjustment. t-statistics are reported in the parentheses. ***, **, * denotes 1%, 5%, and 10% statistical significance.

Panel A: NML Portfolios in the Highest IU Quintile

	(1)	(2)	(3)	(4)
High IU Proxied by	$SIGMA$	$1/AGE$	$CFVOL$	$DISP$
Variable		r_t		
$MKTRF_t$	-0.176***	-0.185***	-0.123	-0.246**
	(-2.71)	(-2.62)	(-1.62)	(-2.34)
$PEAD_t$	0.669***	0.648***	0.598***	0.768***
	(4.06)	(4.26)	(3.95)	(3.14)
FIN_t	0.099	0.431***	0.250***	0.248**
	(0.97)	(3.78)	(2.87)	(2.29)
Constant	2.284***	1.142***	1.225***	1.250***
	(6.42)	(4.37)	(5.38)	(3.70)
Observations	384	384	384	348
Adjusted R^2	0.093	0.247	0.158	0.139

Panel B: NML Portfolios in the Lowest IU Quintile

	(1)	(2)	(3)	(4)
Low IU Proxied by	$SIGMA$	$1/AGE$	$CFVOL$	$DISP$
Variable		r_t		
$MKTRF_t$	0.123	0.134**	-0.023	0.124**
	(1.52)	(2.12)	(-0.28)	(2.06)
$PEAD_t$	-0.076	-0.102	0.228	0.226**
	(-0.95)	(-0.86)	(1.62)	(2.12)
FIN_t	0.011	0.129*	0.109	-0.014
	(0.16)	(1.86)	(1.42)	(-0.18)
Constant	0.148	0.193	-0.017	0.219
	(0.85)	(0.85)	(-0.07)	(1.36)
Observations	384	384	384	348
Adjusted R^2	0.030	0.014	0.014	0.034