Khaled Obaid

Assistant Professor in Finance, Accounting and Finance Department, California State University–East Bay

- Education:

  PhD, Finance, University of Missouri, 2019
  MS, Accountancy, Wake Forest University, 2014
  BS, Finance, Wake Forest University, 2012

- Research Interests:

  empirical asset pricing, behavioral finance, mutual funds, and big data

- Publications:

  Obaid, K., and Pukthuanthong, K., 2021, “Informativeness of mutual fund advertisements: Does advertising communicate fund quality to investors?”, Financial Management 50(1), 203–236. doi.org/10.1111/fima.12323

Kuntara Pukthuanthong

Professor of Finance and the Robert J. Trulaske, Jr. Professor of Finance at the Robert J. Trulaske, Sr. College of Business, University of Missouri.

- Education:

  PhD, Finance, University of California;

  MBA, Finance, Washington University;

  BA, Economics, Chulalongkorn University

- Research Interests:

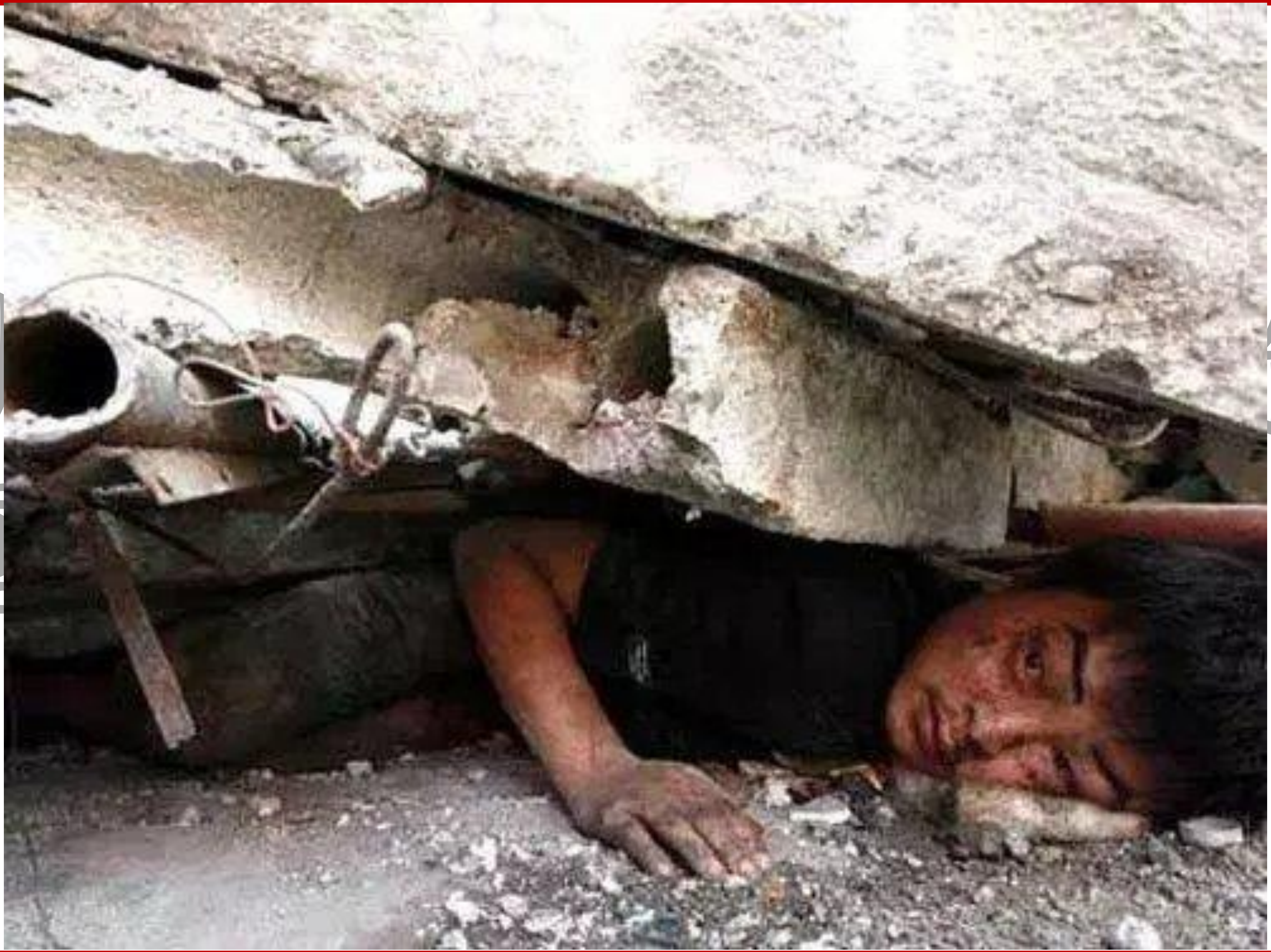
  international finance and asset pricing

- Publications:

  Obaid, K., and Pukthuanthong, K., 2021, "Informativeness of mutual fund advertisements: Does advertising communicate fund quality to investors?", Financial Management 50(1), 203–236. doi.org/10.1111/fima.12323

  Empirical Tests of Asset Pricing Models with Individual Assets: Resolving the Errors-in-Variables Bias in Risk, forthcoming Journal of Financial Economics.

四 告
成



Harried Traders Gird for a Frantic Friday

（焦躁的交易员为疯狂的星期五做好准备）

Analysts Slash GDP Estimates as Coronavirus Ripples Through Economy

（新冠病毒殃及经济，分析师下调CDP预期）

A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news

JFE2022.04

主讲人：王 航

1. 主要研究内容

（1）构建图片悲观指数与文本悲观指数的技术细节

数据选取：

《华尔街日报》2008—2020年涵盖商业、经济、市场、政治及观点五个部分的14余万篇文章中的图片、标题及摘要。

模型选取：

卷积神经网络技术(Convolutional Neural Networks)中的Google Inception(v3)模型

变量构建：

图片悲观指数（PhotoPes）：每天新闻图片中被模型分为悲观情绪组的比例

文本悲观指数（TextPes）：通过文本分析方法对新闻标题与摘要的悲观情绪进行评估得到

（2）图片悲观指数（PhotoPes）预测金融市场活动的能力

收益率逆转：

图片悲观指数上升一个标准差将导致次日美国股指与交易型开放式指数基金下降4.2个基点，在此后第二天与第五天出现共计9.8到10.9个基点的回升。

市场交易量：

纽交所交易量在图片悲观指数波动时增加，表明图片悲观指数影响投资者情绪而非交易成本的增加

套利限制对图片悲观指数的影响幅度：

将资产组合根据特质波动率(Idiosyncratic Volatility)与公司规模进行分类。异质性分析表明图片悲观指数对金融市场回报的影响幅度在资产组合具有更高特质波动率或公司规模更小时更大，表明图片悲观指数对难以套利的资产组合有更大影响。

（3）新闻图片中提取投资者情绪信息的优势

图片在传达恐慌情绪时较文本更为有效：

对本文悲观情绪进行分析（收益逆转）

图片悲观指数影响金融市场回报的系数在恐慌时期约是平时的2.8倍，而文本悲观指数在恐慌时期的系数与平时类似

图片与文本悲观情绪存在相互替代的作用：

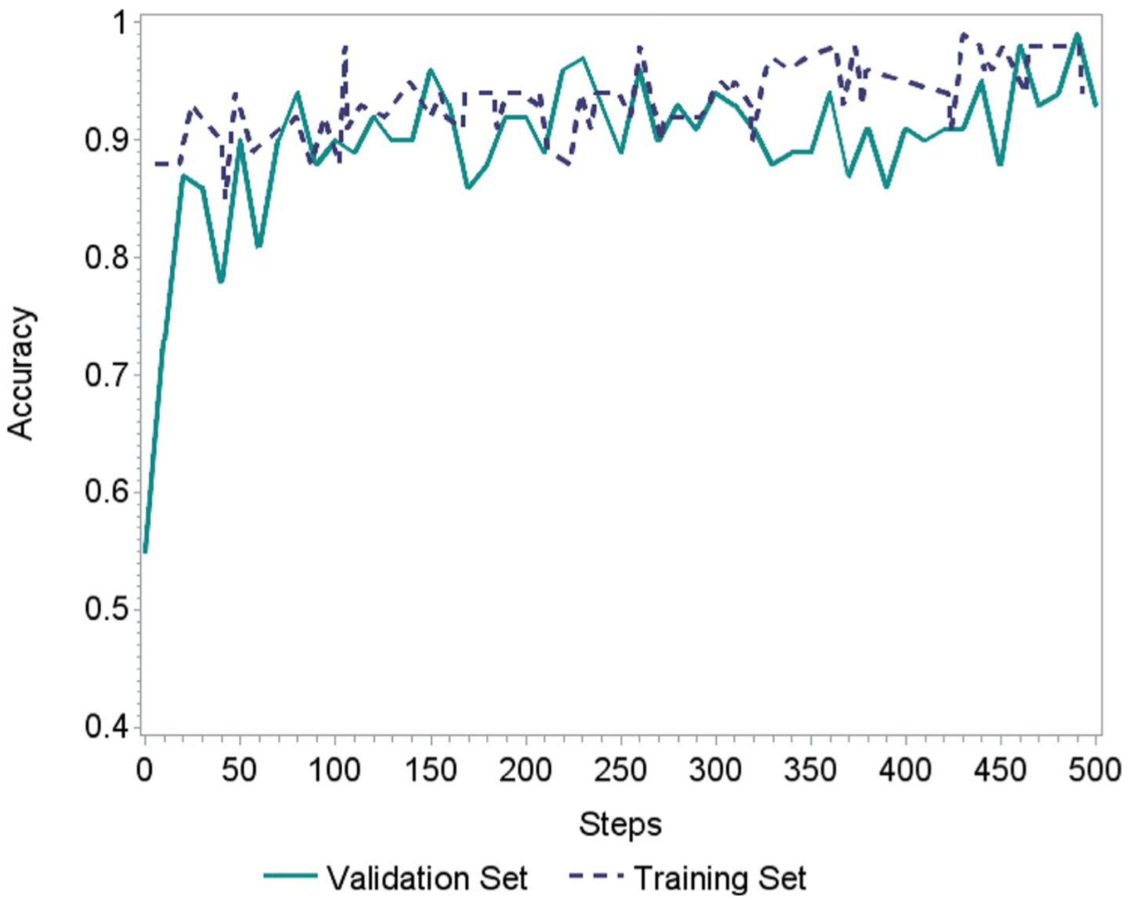
图片悲观指数变动导致的金融市场回报下降幅度在文本悲观指数处于高水平时较小

（4）稳健性分析

2. Data

(1) Photo classification

- For this study, we are interested in a model that is able to predict sentiment that a photo is likely to elicit from investors. We build a photo classification model based on Google Inception (v3).
- We start with a pretrained Google Inception (v3) model and use transfer learning to fine-tune the model for our specific application.
- The output we get from the fine-tuned model is one of two probabilities: the probability that photos have positive sentiment and the probability that photos have negative sentiment.
- To perform transfer learning, we need a training set that consists of photos labeled by sentiment. We use the DeepSent data set for training. For increased reliability, we use photos in which all five MTurk survey participants agree on the photo's sentiment. This restriction reduces the training sample to 882 photos.
- The model is trained using a learning rate of 0.01 and 500 learning steps. We set the train batch size to 100 photos. We reserve 10% of the training photos for the validation sample, and 20% for the test sample (photos are randomly assigned into these sets).



- We achieve 87.1% test accuracy.
- To better test the performance of our model, we also calculate recall (86.2%), precision (94.3%), and F1 (90.1%).
- The model trained with the DeepSent training set might not be able to accurately classify professional photos.
- To address this concern, we randomly select 100 photos from our WSJ sample and classify each photo (classification is preformed in MTurk by five individuals). We ask these individuals to rate the content of each photo based on sentiment. We pass these photos through the model to get predictions and compare the predictions to the responses we collect from MTurk.

		Actual	
		Positive	Negative
Prediction	Positive	64	19
	Negative	5	12

- Based on the confusion matrix above, we calculate the performance of our model for classifying our sample of professional news photos from the WSJ. The accuracy is 76.0%; recall is 92.8%; precision is 77.1%; and F1 is 84.2%.

（2） The Wall Street Journal sample

- Between September 2008 and September 2020, we collect the headline and the summary of each article, any associated photos, and the time stamps of when the article is published from the following WSJ sections: “Business,” “Economy,” “Markets,” “Politics,” and “Opinion.”
- We collect a total of 148,823 articles spanning 3048 trading days. We classify these photos by sentiment using the photo classification model we discussed.

（3） Variable construction

- PhotoPes is calculated as the proportion of photos predicted to be negative on a given date. The formula for PhotoPes on day t is:

$$PhotoPes_t = \frac{\sum_i Neg_{it}}{n_t} \qquad (1)$$

- We use the sentiment tool in Stanford’s CoreNLP software to evaluate the pessimism in each sentence and take the average pessimism score across all sentences in the text as the pessimism score for the article, TextNeg. The formula for TextPes is as follows:

$$TextPes_t = \frac{\sum_i TextNeg_{it}}{n_t} \qquad (2)$$

（4） Descriptive statistics

Panel A: Summary statistics of sentiment variables

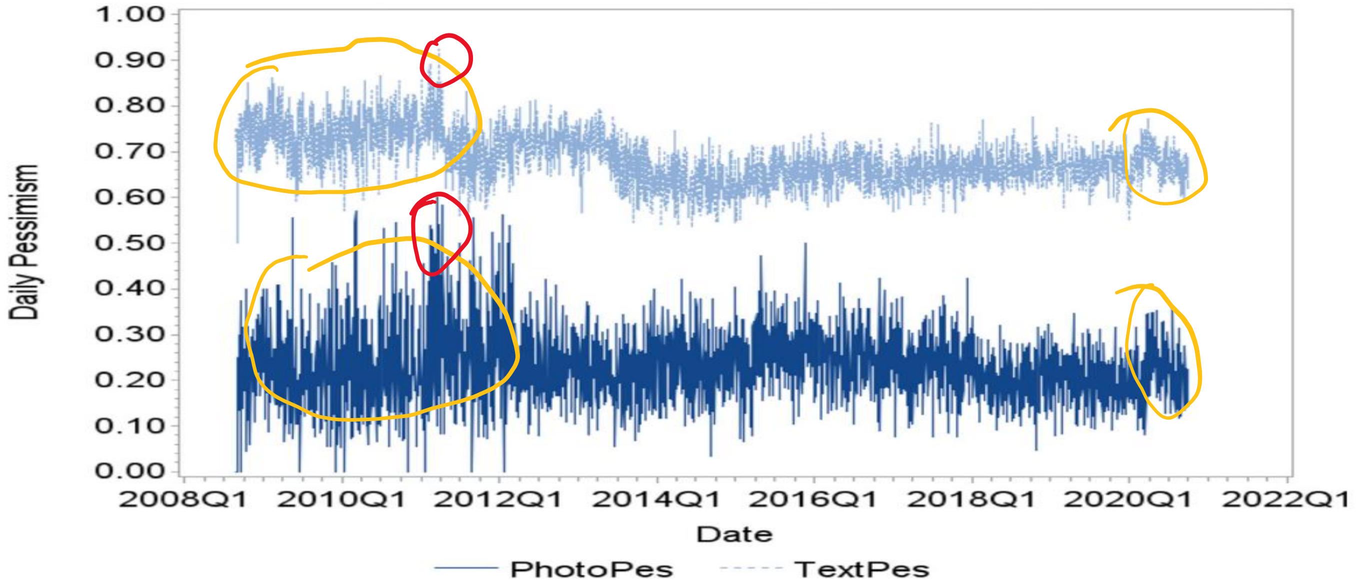
Variable	*N*	*Mean*	*Median*	*P25*	*P75*	*Std dev*
PhotoPes	3048	0.228	0.222	0.180	0.270	0.077
TextPes	3048	0.686	0.681	0.646	0.722	0.056

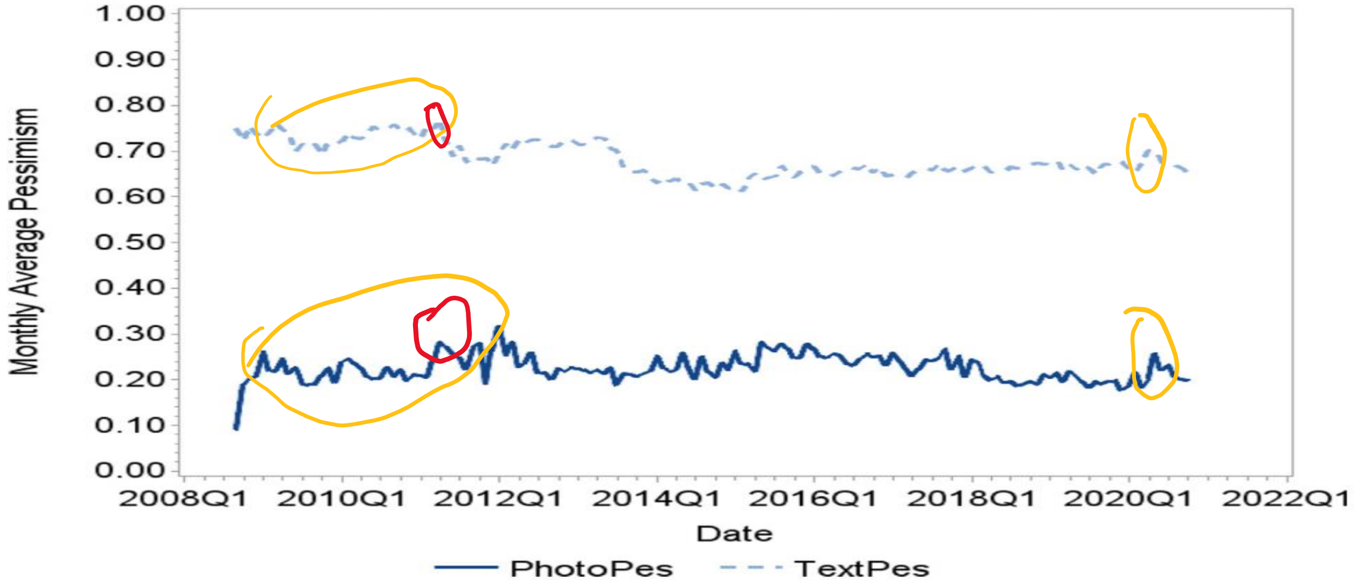
Panel B: Summary statistics of market returns

R_t(%)	*N*	*Mean*	*P50*	*P25*	*P75*	*Std dev*
VWRETD	3048	0.045	0.081	-0.391	0.586	1.332
SPX	3048	0.042	0.070	-0.380	0.570	1.335
SPY	3048	0.049	0.070	-0.370	0.580	1.327
INDU	3048	0.039	0.060	-0.390	0.550	1.283
DIA	3048	0.047	0.070	-0.370	0.550	1.296

Panel C: Correlations between sentiment variables

	PhotoPes
TextPes	0.079***
	$<$0.01


1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Daily Pessimism
2008Q1
2010Q1
2012Q1
2014Q1
2016Q1
2018Q1
2020Q1
2022Q1
Date
PhotoPes
TextPes


1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Monthly Average Pessimism
2008Q1
2010Q1
2012Q1
2014Q1
2016Q1
2018Q1
2020Q1
2022Q1
Date
PhotoPes
TextPes

3. Results

3.1 News sentiment embedded in photos and text

（1） The

Panel A: PhotoPes

	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
$PhotoPes_{t-1}$	-0.042*	-1.837	-0.041*	-1.803	-0.040*	-1.787	-0.046**	-2.182	-0.047**	-2.183
$PhotoPes_{t-2}$	0.055**	2.004	0.051*	1.886	0.046*	1.726	0.043*	1.687	0.038	1.502
$PhotoPes_{t-3}$	-0.033	-1.324	-0.030	-1.213	-0.030	-1.294	-0.024	-1.053	-0.025	-1.142
$PhotoPes_{t-4}$	0.030	1.299	0.024	1.047	0.026	1.143	0.030	1.387	0.033	1.487
$PhotoPes_{t-5}$	0.057**	2.137	0.059**	2.228	0.056**	2.119	0.057**	2.193	0.054**	2.103
Sum t-1 to t-5	0.067		0.063		0.058		0.060		0.053	
Sum t-2 to t-5	0.109		0.104		0.098		0.106		0.100	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	2.272	0.132	2.081	0.149	1.700	0.192	1.979	0.160	1.644	0.200
$\chi^2(1)$[Sum t-2 to t-5 = 0]	6.615**	0.010	6.200**	0.013	5.466**	0.019	6.973***	0.008	6.257**	0.012
Adj. R-squared	0.033		0.038		0.029		0.042		0.040	
N	3044		3044		3044		3044		3044	

Panel B: Predicted Likelihood PhotoPes

	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
$PhotoPes_{t-1}$	-0.052**	-2.240	-0.052**	-2.269	-0.051**	-2.228	-0.055***	-2.581	-0.056**	-2.569
$PhotoPes_{t-2}$	0.039	1.487	0.034	1.345	0.028	1.139	0.029	1.255	0.024	1.031
$PhotoPes_{t-3}$	-0.020	-0.829	-0.017	-0.687	-0.017	-0.716	-0.011	-0.472	-0.011	-0.485
$PhotoPes_{t-4}$	0.019	0.789	0.016	0.676	0.019	0.769	0.026	1.120	0.029	1.242
$PhotoPes_{t-5}$	0.068***	2.596	0.070***	2.686	0.067**	2.576	0.062**	2.516	0.060**	2.390
Sum t-1 to t-5	0.054		0.051		0.046		0.051		0.046	
Sum t-2 to t-5	0.106		0.103		0.097		0.106		0.102	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	1.777	0.182	1.646	0.200	1.311	0.252	1.760	0.185	1.412	0.235
$\chi^2(1)$[Sum t-2 to t-5 = 0]	7.202***	0.007	7.033***	0.008	6.24**	0.012	8.013***	0.005	7.199***	0.007
Adj. R-squared	0.032		0.038		0.029		0.042		0.040	
N	3044		3044		3044		3044		3044	

（2） PhotoPes and sentiment embedded in text

The impact of TextPes on market returns.

$$R_t = \beta_1 L5(TextPes_t) + \beta_2 L5(R_t) + \beta_3 L5(R_t^2) + \beta_4 X_t + \varepsilon_t$$

	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	*t*-stat	β	*t*-stat	β	*t*-stat	β	*t*-stat	β	*t*-stat
$TextPes_{t-1}$	-0.071*	-1.663	-0.083*	-1.904	-0.087**	-1.977	-0.085**	-2.002	-0.086**	-1.998
$TextPes_{t-2}$	-0.056	-1.466	-0.061	-1.590	-0.065*	-1.661	-0.062*	-1.686	-0.061	-1.628
$TextPes_{t-3}$	-0.007	-0.150	0.001	0.018	-0.004	-0.078	0.005	0.108	-0.003	-0.069
$TextPes_{t-4}$	0.021	0.527	0.017	0.436	0.021	0.540	0.027	0.709	0.029	0.756
$TextPes_{t-5}$	0.107**	2.280	0.116**	2.448	0.124***	2.612	0.110**	2.408	0.116**	2.521
Sum t-1 to t-5	-0.006		-0.010		-0.011		-0.005		-0.005	
Sum t-2 to t-5	0.065		0.073		0.076		0.080		0.081	
	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	0.054	0.816	0.112	0.738	0.120	0.729	0.041	0.840	0.030	0.862
$\chi^2(1)$[Sum t-2 to t-5 = 0]	1.697	0.193	2.171	0.141	2.454	0.117	2.812*	0.094	2.931*	0.087
Adj. *R*-squared	0.028		0.035		0.028		0.038		0.039	
N	3044		3044		3044		3044		3044	

	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	*t*-stat	β	*t*-stat	β	*t*-stat	β	*t*-stat	β	*t*-stat
$PhotoPes_{t-1}$	-0.052**	-2.359	-0.049**	-2.229	-0.049**	-2.220	-0.054***	-2.600	-0.054***	-2.596
$TextPes_{t-1}$	-0.027	-0.816	-0.038	-1.183	-0.041	-1.241	-0.042	-1.390	-0.043	-1.372
$(PhotoPesxTextPes)_{t-1}$	0.038*	1.942	0.033*	1.754	0.034*	1.788	0.034*	1.899	0.034*	1.917
$PhotoPes_{t-2}$	0.056**	2.090	0.052**	1.980	0.048*	1.817	0.045*	1.801	0.040	1.595
$PhotoPes_{t-3}$	-0.027	-1.089	-0.025	-1.012	-0.025	-1.076	-0.020	-0.878	-0.021	-0.949
$PhotoPes_{t-4}$	0.032	1.394	0.026	1.125	0.028	1.228	0.031	1.454	0.034	1.581
$PhotoPes_{t-5}$	0.051*	1.936	0.053**	2.018	0.049*	1.896	0.051**	2.003	0.049*	1.897
$TextPes_{t-2}$	-0.040	-1.162	-0.045	-1.277	-0.042	-1.197	-0.047	-1.444	-0.042	-1.295
$TextPes_{t-3}$	-0.024	-0.640	-0.014	-0.387	-0.016	-0.432	-0.003	-0.078	-0.004	-0.119
$TextPes_{t-4}$	-0.016	-0.494	-0.019	-0.578	-0.020	-0.629	-0.014	-0.451	-0.021	-0.671
$TextPes_{t-5}$	0.090**	2.488	0.093***	2.603	0.095***	2.652	0.083**	2.497	0.089***	2.701
Sum t-1 to t-5 *PhotoPes*	0.060		0.057		0.051		0.053		0.048	
Sum t-2 to t-5 *PhotoPes*	0.112		0.106		0.100		0.107		0.102	
Sum t-1 to t-5 *TextPes*	-0.017		-0.023		-0.024		-0.023		-0.021	
Sum t-2 to t-5 *TextPes*	0.010		0.015		0.017		0.019		0.022	
	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value
$\chi^2(1)$[Sum t-1 to t-5 *PhotoPes*=0]	1.866	0.172	1.740	0.187	1.376	0.241	1.633	0.201	1.318	0.251
$\chi^2(1)$[Sum t-2 to t-5 *PhotoPes*=0]	6.983***	0.008	6.412**	0.011	5.693**	0.017	7.187***	0.007	6.483**	0.011
$\chi^2(1)$[Sum t-1 to t-5 *TextPes*=0]	0.332	0.565	0.562	0.453	0.629	0.428	0.658	0.417	0.553	0.457
$\chi^2(1)$[Sum t-2 to t-5 *TextPes*=0]	0.050	0.823	0.146	0.702	0.181	0.671	0.262	0.609	0.328	0.567
Adj. *R*-squared	0.037		0.043		0.034		0.046		0.044	
N	3044		3044		3044		3044		3044	

(3) Attenti

To test the attentio ... n to differentiate between the effect ... photos are salient:

$R_t = (E_t)[\beta_1 L5($ ... $_{-1} + \beta_4 L5(R_t) +$
$\beta_5 L5(R_t^2)] + (1$ ... $\times TextPes)_{t-1} +$

Variables	(1) $VWRETD_t$, E_t=Salient photo period: β	t-stat	γ	t-stat	(2) SPX_t, E_t= Salient photo period: β	t-stat	γ	t-stat	(3) SPY_t, E_t= Salient photo period: β	t-stat	γ	t-stat
$PhotoPes_{t-1}$	-0.070**	-2.479	-0.015	-0.332	-0.064**	-2.295	-0.016	-0.365	-0.063**	-2.260	-0.015	-0.342
$TextPes_{t-1}$	0.047	0.900	-0.070*	-1.883	0.031	0.606	-0.081**	-2.220	0.030	0.585	-0.080**	-2.137
$(PhotoPes \times TextPes)_{t-1}$	0.034	1.524	0.070	1.450	0.029	1.312	0.065	1.362	0.030	1.403	0.060	1.270
$PhotoPes_{t-2}$	0.100***	3.282	-0.034	-0.813	0.099***	3.316	-0.041	-0.978	0.094***	3.173	-0.044	-1.079
$PhotoPes_{t-3}$	-0.020	-0.659	-0.017	-0.411	-0.017	-0.565	-0.020	-0.498	-0.015	-0.526	-0.019	-0.489
$PhotoPes_{t-4}$	0.046*	1.710	0.009	0.216	0.042	1.571	-0.001	-0.033	0.042	1.632	0.003	0.066
$PhotoPes_{t-5}$	0.047	1.469	-0.009	-0.225	0.049	1.538	-0.004	-0.098	0.045	1.383	-0.006	-0.166
$TextPes_{t-2}$	0.044	0.684	-0.066*	-1.795	0.043	0.664	-0.071*	-1.912	0.051	0.761	-0.072*	-1.931
$TextPes_{t-3}$	-0.128**	-2.373	0.013	0.316	-0.113**	-2.174	0.022	0.538	-0.114**	-2.146	0.020	0.486
$TextPes_{t-4}$	-0.079	-1.412	0.002	0.057	-0.083	-1.488	-0.001	-0.018	-0.079	-1.488	-0.002	-0.050
$TextPes_{t-5}$	0.182***	3.133	0.059	1.520	0.186***	3.255	0.064	1.638	0.184***	3.135	0.065*	1.694
Sum t-1 to t-5 *PhotoPes*	0.103		-0.066		0.109		-0.082		0.103		-0.081	
Sum t-2 to t-5 *PhotoPes*	0.173		-0.051		0.173		-0.066		0.166		-0.066	
Sum t-1 to t-5 *TextPes*	0.066		-0.062		0.064		-0.067		0.072		-0.069	
Sum t-2 to t-5 *TextPes*	0.019		0.008		0.033		0.014		0.042		0.011	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 *PhotoPes*=0]	2.905*	0.088	0.602	0.438	3.224*	0.073	0.958	0.328	2.851*	0.091	0.941	0.332
$\chi^2(1)$[Sum t-2 to t-5 *PhotoPes*=0]	9.777***	0.002	0.480	0.489	9.696***	0.002	0.823	0.364	8.943***	0.003	0.833	0.362
$\chi^2(1)$[Sum t-1 to t-5 *TextPes*=0]	0.437	0.509	1.842	0.175	0.405	0.525	2.247	0.134	0.525	0.469	2.443	0.118
$\chi^2(1)$[Sum t-2 to t-5 *TextPes*=0]	0.041	0.840	0.028	0.867	0.115	0.734	0.079	0.778	0.187	0.666	0.052	0.820
Adj. *R*-squared	0.075				0.086				0.071			
N	3044				3044				3044			

Variables	(4) $INDU_t$, E_t= Salient photo period: β	t-stat	γ	t-stat	(5) DIA_t, E_t= Salient photo period: β	t-stat	γ	t-stat
$PhotoPes_{t-1}$	-0.066**	-2.537	-0.034	-0.797	-0.066**	-2.505	-0.034	-0.806
$TextPes_{t-1}$	0.016	0.338	-0.079**	-2.285	0.016	0.345	-0.075**	-2.124
$(PhotoPes \times TextPes)_{t-1}$	0.030	1.471	0.060	1.358	0.031	1.536	0.056	1.253
$PhotoPes_{t-2}$	0.083***	2.901	-0.039	-0.999	0.077***	2.696	-0.044	-1.134
$PhotoPes_{t-3}$	-0.011	-0.411	-0.017	-0.444	-0.010	-0.354	-0.021	-0.556
$PhotoPes_{t-4}$	0.041	1.638	0.016	0.418	0.040	1.603	0.025	0.626
$PhotoPes_{t-5}$	0.041	1.320	0.003	0.080	0.036	1.111	0.003	0.070
$TextPes_{t-2}$	0.049	0.808	-0.076**	-2.169	0.064	1.055	-0.079**	-2.188
$TextPes_{t-3}$	-0.089*	-1.844	0.032	0.814	-0.088*	-1.801	0.032	0.819
$TextPes_{t-4}$	-0.068	-1.308	0.002	0.046	-0.073	-1.394	-0.008	-0.219
$TextPes_{t-5}$	0.159***	2.987	0.058	1.629	0.165***	2.929	0.062*	1.774
Sum t-1 to t-5 *PhotoPes*	0.088		-0.071		0.077		-0.071	
Sum t-2 to t-5 *PhotoPes*	0.154		-0.037		0.143		-0.037	
Sum t-1 to t-5 *TextPes*	0.067		-0.063		0.084		-0.068	
Sum t-2 to t-5 *TextPes*	0.051		0.016		0.068		0.007	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 *PhotoPes*=0]	2.256	0.133	0.772	0.380	1.777	0.183	0.770	0.380
$\chi^2(1)$[Sum t-2 to t-5 *PhotoPes*=0]	8.235***	0.004	0.279	0.598	7.128***	0.008	0.281	0.596
$\chi^2(1)$[Sum t-1 to t-5 *TextPes*=0]	0.529	0.467	2.309	0.129	0.860	0.354	2.590	0.108
$\chi^2(1)$[Sum t-2 to t-5 *TextPes*=0]	0.338	0.561	0.107	0.744	0.598	0.439	0.027	0.869
Adj. *R*-squared	0.096				0.093			
N	3044				3044			

(4) Which ... photos?

We attempt to answ... ...est whether the relation between m... ...during periods of elevated fear.

To examine how th... ...tes to market returns by fear leve...

$R_t = (F_t)[\beta_1 L5(\ldots {}_{-1} + \beta_4 L5(R_t) + \beta_5 L5(R_t^2)] + (1 \ldots TextPes)_{t-1} +$

	(1)				(2)				(3)			
	$VWRETD_t$				SPX_t				SPY_t			
	F_t=Fear period				F_t=Fear period				F_t=Fear period			
Variables	β	t-stat	γ	t-stat	β	t-stat	γ	t-stat	β	t-stat	γ	t-stat
$PhotoPes_{t-1}$	-0.103**	-2.006	-0.037*	-1.893	-0.093*	-1.839	-0.036*	-1.822	-0.092*	-1.793	-0.036*	-1.848
$TextPes_{t-1}$	-0.018	-0.291	-0.019	-0.720	-0.031	-0.517	-0.025	-0.956	-0.035	-0.563	-0.026	-0.987
$(PhotoPes \times TextPes)_{t-1}$	0.087**	2.063	0.009	0.525	0.079*	1.920	0.005	0.294	0.076*	1.830	0.006	0.334
$PhotoPes_{t-2}$	0.065	1.100	0.050**	2.284	0.056	0.956	0.049**	2.272	0.053	0.879	0.049**	2.257
$PhotoPes_{t-3}$	0.018	0.357	-0.054**	-2.423	0.023	0.470	-0.054**	-2.455	0.018	0.374	-0.056**	-2.550
$PhotoPes_{t-4}$	0.093*	1.914	0.001	0.059	0.077	1.587	0.001	0.059	0.078*	1.646	0.004	0.194
$PhotoPes_{t-5}$	0.080	1.431	0.027	1.214	0.082	1.482	0.029	1.307	0.075	1.374	0.030	1.349
$TextPes_{t-2}$	-0.063	-0.919	-0.041	-1.510	-0.071	-1.031	-0.044*	-1.649	-0.059	-0.862	-0.046*	-1.697
$TextPes_{t-3}$	-0.079	-1.094	0.013	0.517	-0.062	-0.885	0.018	0.721	-0.065	-0.920	0.019	0.756
$TextPes_{t-4}$	-0.062	-0.937	0.007	0.263	-0.069	-1.046	0.006	0.235	-0.066	-1.022	0.007	0.267
$TextPes_{t-5}$	0.134*	1.916	0.061**	2.122	0.138**	2.006	0.062**	2.161	0.138**	2.025	0.061**	2.150
Sum t-1 to t-5 *PhotoPes*	0.153		-0.013		0.145		-0.011		0.132		-0.009	
Sum t-2 to t-5 *PhotoPes*	0.256		0.024		0.238		0.025		0.224		0.027	
Sum t-1 to t-5 *TextPes*	-0.088		0.021		-0.095		0.017		-0.087		0.015	
Sum t-2 to t-5 *TextPes*	-0.070		0.040		-0.064		0.042		-0.052		0.041	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 *PhotoPes*=0]	2.259	0.133	0.104	0.747	2.033	0.154	0.065	0.799	1.724	0.189	0.048	0.827
$\chi^2(1)$[Sum t-2 to t-5 *PhotoPes*=0]	7.078***	0.008	0.452	0.502	6.283**	0.012	0.496	0.481	5.629**	0.018	0.567	0.451
$\chi^2(1)$[Sum t-1 to t-5 *TextPes*=0]	1.306	0.253	0.618	0.432	1.562	0.211	0.406	0.524	1.332	0.249	0.360	0.548
$\chi^2(1)$[Sum t-2 to t-5 *TextPes*=0]	0.665	0.415	1.307	0.253	0.580	0.446	1.438	0.231	0.372	0.542	1.432	0.231
Adj. R-squared	0.046				0.053				0.044			
N	3044				3044				3044			

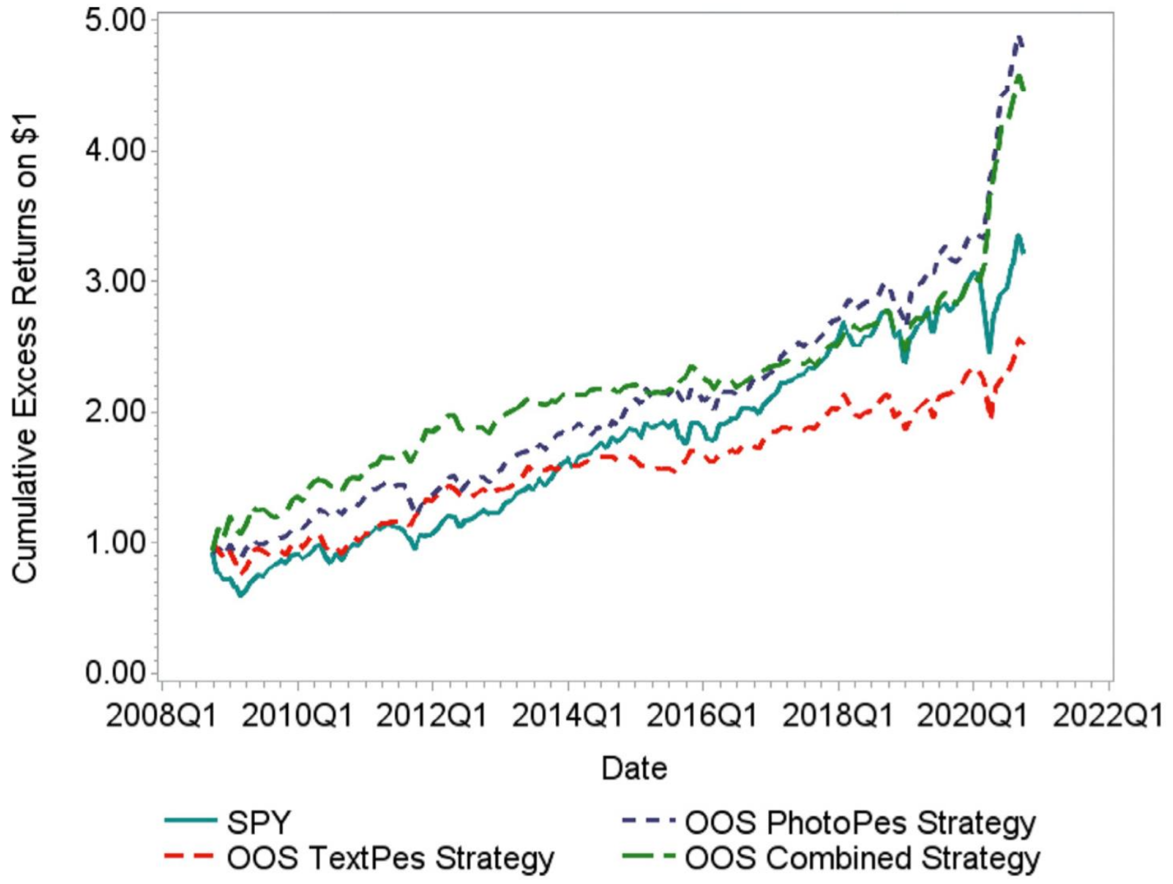
	(4)				(5)			
	$INDU_t$				DIA_t			
	F_t=Fear period				F_t=Fear period			
Variables	β	t-stat	γ	t-stat	β	t-stat	γ	t-stat
$PhotoPes_{t-1}$	-0.101**	-2.102	-0.043**	-2.252	-0.102**	-2.090	-0.043**	-2.260
$TextPes_{t-1}$	-0.031	-0.563	-0.028	-1.134	-0.030	-0.505	-0.027	-1.099
$(PhotoPes \times TextPes)_{t-1}$	0.084**	2.145	0.003	0.215	0.083**	2.107	0.003	0.217
$PhotoPes_{t-2}$	0.039	0.701	0.050**	2.410	0.029	0.511	0.051**	2.473
$PhotoPes_{t-3}$	0.026	0.555	-0.049**	-2.368	0.022	0.487	-0.050**	-2.424
$PhotoPes_{t-4}$	0.070	1.541	0.013	0.622	0.074	1.626	0.015	0.733
$PhotoPes_{t-5}$	0.071	1.322	0.034	1.624	0.068	1.255	0.033	1.614
$TextPes_{t-2}$	-0.079	-1.245	-0.044*	-1.682	-0.072	-1.118	-0.042	-1.637
$TextPes_{t-3}$	-0.047	-0.714	0.027	1.101	-0.048	-0.728	0.025	1.050
$TextPes_{t-4}$	-0.050	-0.815	0.005	0.186	-0.061	-0.978	0.002	0.067
$TextPes_{t-5}$	0.127**	2.017	0.051*	1.913	0.135**	2.185	0.054**	1.995
Sum t-1 to t-5 *PhotoPes*	0.105		0.005		0.091		0.006	
Sum t-2 to t-5 *PhotoPes*	0.206		0.048		0.193		0.049	
Sum t-1 to t-5 *TextPes*	-0.080		0.011		-0.076		0.012	
Sum t-2 to t-5 *TextPes*	-0.049		0.039		-0.046		0.039	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 *PhotoPes*=0]	1.198	0.274	0.019	0.891	0.916	0.338	0.036	0.850
$\chi^2(1)$[Sum t-2 to t-5 *PhotoPes*=0]	5.23**	0.022	1.905	0.168	4.583**	0.032	2.065	0.151
$\chi^2(1)$[Sum t-1 to t-5 *TextPes*=0]	1.307	0.253	0.169	0.681	1.086	0.297	0.169	0.681
$\chi^2(1)$[Sum t-2 to t-5 *TextPes*=0]	0.396	0.529	1.350	0.245	0.309	0.578	1.307	0.253
Adj. R-squared	0.058				0.058			
N	3044				3044			

（5） Applications

- The first strategy is based on the pessimism embedded in news photos: following days in which PhotoPes⊥ is above its historical mean (expanding), we invest in the SPY at the market close of day t + 3 and sell on the market close two days later (t+5).
- The second strategy is based on pessimism embedded in text: following days in which TextPes⊥ is above its historical mean (expanding), we invest in the SPY at the market close of day t + 3 and sell on the market close two days later (t + 5).
- The third strategy involves pessimism from both text and photos: following days in which PhotoPes⊥ and TextPes⊥ are above their historical means, we invest in the SPY at the market close of day t + 3 and sell on the market close two days later (t + 5).

Panel A: Summary statistics of trading strategies					
Strategy	*N*	*Mean*	*t-stat*	*Std dev*	*SR*
PhotoPes	1992	0.058	3.251	1.119	0.052
TextPes	1891	0.037	2.085	1.166	0.032
Combined	1221	0.054	3.547	0.980	0.055
Index	3034	0.047	2.246	1.325	0.036

Panel B: Time series regression				
	(1)		(2)	
	Combined strategy$_t$		*PhotoPesstrategy$_t$*	
Variables	*β*	*t*-stat	*β*	*t*-stat
Alpha	0.021*	1.742	0.014	1.302
Mkt_Rf_t	51.0***	13.347	69.4***	22.912
SMB_t	-1.830	-0.317	-6.612	-1.418
HML_t	-15.6***	-3.517	-13.9***	-3.799
UMD_t	-4.947*	-1.743	-5.489**	-2.121
ST_Rev_t	6.754*	1.708	3.252	1.002
Adj. *R*-squared	0.545		0.706	
N	3034		3034	



3. Results

3.2 Validation of PhotoPes

（1） Limits to arbitrage

To test whether PhotoPes relates to stock returns differently depending on limits to arbitrage, we run the following regression:

$$R_t = \beta_1 L5(PhotoPes_t) + \beta_2 L5(R_t) + \beta_3 L5(R_t^2) + \beta_4 X_t + \varepsilon_t \qquad (7)$$

where R t denotes the value-weighted daily returns on the highest and lowest quintile portfolios and the spread between the two portfolios (High-low) sorted on idiosyncratic volatility and size.

Panel A: VW Idvol (CAPM)$_t$

	(1)		(2)		(3)	
	High		*Low*		*High-low*	
Variables	β	*t*-stat	β	*t*-stat	β	*t*-stat
$PhotoPes_{t-1}$	-0.071**	-2.182	-0.035*	-1.852	-0.040**	-1.986
$PhotoPes_{t-2}$	0.107***	2.833	0.042*	1.902	0.061***	2.810
$PhotoPes_{t-3}$	-0.044	-1.241	-0.030	-1.481	-0.008	-0.381
$PhotoPes_{t-4}$	0.061*	1.745	0.029	1.518	0.028	1.292
$PhotoPes_{t-5}$	0.064*	1.778	0.047**	2.043	0.019	0.867
Sum t-1 to t-5	0.117		0.053		0.060	
Sum t-2 to t-5	0.188		0.088		0.100	
	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	3.431*	0.064	2.054	0.152	2.454	0.117
$\chi^2(1)$[Sum t-2 to t-5 = 0]	9.441***	0.002	6.304**	0.012	7.121***	0.008
Adj. *R*-squared	0.026		0.030		0.028	
N	3044		3044		3044	

Panel B: VW Idvol (FF+UMD)$_t$

	(1)		(2)		(3)	
	High		*Low*		*High-low*	
Variables	β	*t*-stat	β	*t*-stat	β	*t*-stat
$PhotoPes_{t-1}$	-0.066**	-2.073	-0.037*	-1.828	-0.035*	-1.854
$PhotoPes_{t-2}$	0.102***	2.759	0.045*	1.942	0.052**	2.512
$PhotoPes_{t-3}$	-0.049	-1.387	-0.027	-1.300	-0.014	-0.722
$PhotoPes_{t-4}$	0.063*	1.871	0.032	1.633	0.028	1.344
$PhotoPes_{t-5}$	0.062*	1.791	0.049**	2.051	0.015	0.736
Sum t-1 to t-5	0.112		0.062		0.046	
Sum t-2 to t-5	0.178		0.099		0.081	
	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	3.412*	0.065	2.691	0.101	1.560	0.212
$\chi^2(1)$[Sum t-2 to t-5 = 0]	9.026***	0.003	7.362***	0.007	5.133**	0.023
Adj. *R*-squared	0.028		0.032		0.027	
N	3044		3044		3044	

Panel C: VW size$_t$

	(1)		(2)		(3)	
	Large		*Small*		*Large-small*	
Variables	β	*t*-stat	β	*t*-stat	β	*t*-stat
$PhotoPes_{t-1}$	-0.041*	-1.825	-0.073**	-2.452	0.033**	1.984
$PhotoPes_{t-2}$	0.046*	1.776	0.063*	1.914	-0.011	-0.704
$PhotoPes_{t-3}$	-0.028	-1.163	-0.026	-0.810	-0.005	-0.296
$PhotoPes_{t-4}$	0.025	1.110	0.023	0.779	0.000	0.026
$PhotoPes_{t-5}$	0.055**	2.080	0.075**	2.325	-0.016	-1.088
Sum t-1 to t-5	0.057		0.062		0.001	
Sum t-2 to t-5	0.098		0.135		-0.032	
	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	1.805	0.179	1.311	0.252	0.001	0.970
$\chi^2(1)$[Sum t-2 to t-5 = 0]	5.796**	0.016	6.648***	0.010	1.482	0.224
Adj. *R*-squared	0.036		0.043		0.013	
N	3044		3044		3044	

（2） Out-of-sample analysis

Out-of-sample tests are more appropriate to avoid the in-sample overfitting issue. Hence, we examine the out-of-sample predictive performance of PhotoPes .

We evaluate the out-of-sample predictive performance based on the widely used R^2_{OOS} statistic and MSPE-adjusted statistic.

Return	R^2_{OOS} (%)	*CER gain (%)*	*Sharpe ratio (PhotoPes)*	*Sharpe ratio (historical)*
VWRETD	0.251***	1.439	0.470	0.380
SPX	0.162***	1.150	0.441	0.304
SPY	0.302***	1.290	0.490	0.400
INDU	0.123***	1.042	0.396	0.224
DIA	0.158***	1.420	0.490	0.380

3. Results

3.3 The impact of PhotoPes on trading volume

Anothe … annel
helps u

To rem

To test

	$\bar{V}_t$	
Variables	β	t-stat
$PhotoPes_{t-1}$	-0.003	-0.157
$PhotoPes_{t-2}$	-0.006	-0.278
$PhotoPes_{t-3}$	0.014	0.686
$PhotoPes_{t-4}$	-0.015	-0.786
$PhotoPes_{t-5}$	-0.011	-0.582
$\lvert PhotoPes_{t-1}\rvert$	0.080***	2.912
$\lvert PhotoPes_{t-2}\rvert$	0.013	0.453
$\lvert PhotoPes_{t-3}\rvert$	0.008	0.237
$\lvert PhotoPes_{t-4}\rvert$	0.059**	2.059
$\lvert PhotoPes_{t-5}\rvert$	0.015	0.511
Adj. R-squared	0.023	
N	3044	

3. Results

3.4 Robustness

Panel A: Cutoff = {0.55,0.45}

	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
$PhotoPes_{t-1}$	-0.040*	-1.705	-0.038*	-1.652	-0.038*	-1.680	-0.044**	-2.059	-0.046**	-2.140
$PhotoPes_{t-2}$	0.043*	1.685	0.038	1.515	0.036	1.463	0.032	1.355	0.029	1.223
$PhotoPes_{t-3}$	-0.024	-0.977	-0.022	-0.885	-0.025	-1.040	-0.017	-0.727	-0.019	-0.830
$PhotoPes_{t-4}$	0.038	1.605	0.033	1.391	0.035	1.520	0.039*	1.784	0.043*	1.932
$PhotoPes_{t-5}$	0.049*	1.783	0.051*	1.874	0.049*	1.837	0.047*	1.818	0.045*	1.759
Sum t-1 to t-5	0.066	0.062	0.057	0.057	0.052					
Sum t-2 to t-5	0.106	0.100	0.095	0.101	0.098					
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	2.116	0.146	1.900	0.168	1.675	0.196	1.803	0.179	1.520	0.218
$\chi^2(1)$[Sum t-2 to t-5 = 0]	6.118**	0.013	5.591**	0.018	5.223**	0.022	6.366**	0.012	5.969**	0.015
Adj. R-squared	0.029		0.035		0.028		0.039		0.038	
N	3044		3044		3044		3044		3044	

Panel B: No winsorization

	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
$PhotoPes_{t-1}$	-0.044*	-1.819	-0.043*	-1.792	-0.043*	-1.797	-0.049**	-2.194	-0.049**	-2.188
$PhotoPes_{t-2}$	0.061**	2.126	0.057**	2.025	0.055*	1.915	0.049*	1.842	0.044*	1.661
$PhotoPes_{t-3}$	-0.025	-0.911	-0.021	-0.781	-0.025	-0.971	-0.014	-0.545	-0.017	-0.646
$PhotoPes_{t-4}$	0.025	1.039	0.018	0.764	0.022	0.965	0.023	0.995	0.026	1.146
$PhotoPes_{t-5}$	0.053*	1.880	0.056**	2.031	0.052*	1.903	0.054**	2.029	0.051*	1.910
Sum t-1 to t-5	0.070		0.067		0.061		0.063		0.055	
Sum t-2 to t-5	0.114		0.110		0.104		0.112		0.104	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	2.195	0.138	2.094	0.148	1.750	0.186	2.014	0.156	1.684	0.194
$\chi^2(1)$[Sum t-2 to t-5 = 0]	6.408**	0.011	6.178**	0.013	5.525**	0.019	7.02***	0.008	6.245**	0.012
Adj. R-squared	0.030		0.036		0.028		0.039		0.038	
N	3044		3044		3044		3044		3044	

Panel C: GARCH-adjusted returns

	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	t-stat	β	t-stat	β	t-stat	β	t-stat	β	t-stat
$PhotoPes_{t-1}$	-0.036*	-1.948	-0.036*	-1.959	-0.035*	-1.935	-0.042**	-2.348	-0.043**	-2.372
$PhotoPes_{t-2}$	0.042**	2.203	0.041**	2.172	0.041**	2.162	0.042**	2.205	0.044**	2.273
$PhotoPes_{t-3}$	-0.042**	-2.219	-0.042**	-2.178	-0.043**	-2.267	-0.042**	-2.187	-0.043**	-2.261
$PhotoPes_{t-4}$	0.028	1.542	0.024	1.354	0.026	1.465	0.032*	1.763	0.033*	1.841
$PhotoPes_{t-5}$	0.037*	1.936	0.037**	1.963	0.036*	1.886	0.031	1.643	0.029	1.525
Sum t-1 to t-5	0.029		0.024		0.025		0.021		0.020	
Sum t-2 to t-5	0.065		0.060		0.060		0.063		0.063	
	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value	$\chi^2(1)$	p-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	0.673	0.412	0.528	0.468	0.498	0.480	0.370	0.543	0.330	0.566
$\chi^2(1)$[Sum t-2 to t-5 = 0]	3.884**	0.049	3.539*	0.060	3.392*	0.066	3.906**	0.048	3.778*	0.052
Adj. R-squared	0.009		0.009		0.009		0.010		0.010	
N	3044		3044		3044		3044		3044	

(Continued in next page)

Panel D: Trim 1% of the most extreme returns

	(1)		(2)		(3)		(4)		(5)	
	$VWRETD_t$		SPX_t		SPY_t		$INDU_t$		DIA_t	
Variables	β	*t*-stat	β	*t*-stat	β	*t*-stat	β	*t*-stat	β	*t*-stat
$PhotoPes_{t-1}$	-0.047**	-2.238	-0.046**	-2.221	-0.045**	-2.192	-0.047**	-2.384	-0.043**	-2.201
$PhotoPes_{t-2}$	0.040*	1.837	0.040*	1.844	0.040*	1.860	0.028	1.372	0.032	1.625
$PhotoPes_{t-3}$	-0.044*	-1.909	-0.038*	-1.660	-0.037*	-1.685	-0.034	-1.615	-0.035*	-1.703
$PhotoPes_{t-4}$	0.041**	2.062	0.037*	1.902	0.043**	2.217	0.037**	1.979	0.043**	2.326
$PhotoPes_{t-5}$	0.035	1.547	0.042*	1.883	0.032	1.480	0.035*	1.685	0.018	0.898
Sum t-1 to t-5	0.025		0.035		0.033		0.019		0.015	
Sum t-2 to t-5	0.072		0.081		0.078		0.066		0.058	
	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value	$\chi^2(1)$	*p*-value
$\chi^2(1)$[Sum t-1 to t-5 = 0]	0.436	0.509	0.808	0.369	0.722	0.396	0.266	0.606	0.192	0.661
$\chi^2(1)$[Sum t-2 to t-5 = 0]	3.982**	0.046	5.019**	0.025	4.698**	0.030	3.804*	0.051	3.047*	0.081
Adj. *R*-squared	0.018		0.023		0.023		0.024		0.025	
N	3012		3012		3012		3013		3012	

4. Conclusion

- First, we document that PhotoPes predicts market return reversals. This return reversal pattern is consistent with sentiment-induced transient mispricing. This relation is strongest for difficult-to-arbitrage stocks. Moreover, we show that PhotoPes can predict market trading volume.
- Second, we show that the pessimism embedded in photos and the pessimism embedded in news text act as substitutes for each other in predicting returns. Moreover, the pessimism embedded in news photos serves to grab attention away from text during periods when photos are salient. Our evidence shows that PhotoPes is especially useful for predicting market returns during periods of elevated fear.
- Third, we demonstrate the benefit of using cutting-edge photo classification techniques to study how the information obtained from a large sample of news photos is relevant to the context of financial markets.

5. Key contributions to the literature

- First, we demonstrate the importance of visual content in making predictions about market returns.
- Second, we demonstrate how to overcome key hurdles of studying the importance of visual content in financial markets by employing machine learning techniques for large-scale photo classification.
- Third, we compare the predictive ability of PhotoPes and pessimism embedded in news text.

Thank You!