Online Appendix for

"A picture is worth a thousand words:

Measuring investor sentiment by combining machine learning and photos from news"

PhotoPes and market returns over long horizon

In this section, we examine the relation between *PhotoPes* and market returns using a sample of photos from *Getty Images* between 1926 and 2018. The main advantage of the *Getty Images* sample over the *WSJ* sample that we use in the paper is the longer-time horizon. We first discuss the *Getty Images* sample and document the baseline results using this sample. Second, we discuss the limitations of this sample.

We select photos daily between January 1926 and June 2018 from *Getty Images* Editorial News Section. Their website describes photos in this section as follows: "*show real-world people, places, events and things and are intended to be used only in connection with events that are newsworthy or of general interest*."[1] Focusing on the Editorial News Section helps us focus on the most widely distributed news that can play an important role in financial markets. The *Getty Images* news database contains photos that are taken by photographers working for *Getty Images* in addition to photos taken by photojournalists working for various news outlets. News outlets such as *Bloomberg*, *The Los Angeles Times*, and *The Washington Post* license their photos through *Getty Images*. *Getty Images* makes its news photos database available for licensing to news organization and other interested parties.

Each day, we sort photos by popularity, and select the 10 most popular photos. In addition to collecting the photos, we gather information on the popularity ranking of the photo, date the photo is taken, photo description, and photo identification number. The popularity rank considers purchase history and number of views. *Getty Images* does not make the actual popularity scores available. Moreover, we do not have access to the actual number of purchases and views for each image.[2]

To help filter out irrelevant photos or photos related to unimportant or frivolous events, we apply two filters to the sample of *Getty Image* photos we collect. First, we require that a given day has at least 15 total

[1] Source: https://www.gettyimages.com/faq/workingfiles

[2] We have made several attempts to get this information but it is strictly proprietary by Getty.

photos available. This rule is important for excluding days earlier in our sample with sparse photos that do not have apparent link to financial markets or important events. Second, we require that the photo description contains at least one negative or positive word according to the Loughran and McDonald (2011) dictionaries (LM). This filter removes photos with descriptions not in English and helps increase the chance that photo is related to a news story with relevance to financial markets. In total, we have 74,044 qualifying photos that pass the two filters and are used to construct our main variable.

We process all qualifying photos we select from *Getty Images* into the photo classification model we train in the paper to get predictions on likelihood of positive or negative sentiment in the photo. Our main variable, *PhotoPes*, is calculated as the proportion of photos predicted to be negative on a given date. We weight photos from *Getty Images* based on their popularity (i.e., $\frac{1}{Popularity\ Rank_{it}}$).[3] The purpose behind weighting photos by popularity is to give more popular photos higher weights compared to less popular photos since those photos are more likely to capture events that attract more market attention. The formula for *PhotoPes* on day *t* for the *Getty Images* sample is the following:

$$PhotoPes_t = \frac{\sum_i (Neg_{it} \times \frac{1}{Popularity\ Rank_{it}})}{\sum_i \frac{1}{Popularity\ Rank_{it}}}, \tag{1}$$

where Neg_{it} is an indicator variable for whether photo *i* at time *t* is predicted to have negative sentiment.[4] The denominator is the sum of the weights.

To test the relation between *PhotoPes* using the *Getty Images* sample, we run the following, with Newey and West (1987) t-statistics:

$$R_t = \beta_1 PhotoPes_t + \beta_2 L4(PhotoPes_t) + \beta_3 L5(R_t) + \beta_4 L5(R_t^2) + \beta_5 X_t + \varepsilon_t \tag{2}$$

where R_t denotes daily log-returns on VWRETD and $PhotoPes_t$ is the proportion of the most popular photos that are predicted to be negative on time *t*. $L4$ and $L5$ transform a variable into a row vector consisting of four

[3] The most popular photo on a given day will have a popularity rank of 1, and the 10th most popular photo in a given day will have a popularity rank of 10.

[4] The probability cutoff for Neg_{it} is 50%.

and five lags of that variable, respectively. X_t is a set of exogenous variables that include an intercept, day-of-the-week indicators (except for Monday), and an indicator variable for whether time *t* is in a recession period.

In Table B1, $PhotoPes_t$ is negatively related to VWRETD at the 5% level. We examine the lags of *PhotoPes* to determine whether a reversal of the initial effect occurs in the following four days. We note that the coefficient on $PhotoPes_{t-3}$ is positive and significant at the 10% level suggesting some of the initial effect is reversed. In contrast to the main results in the paper and the results in Tetlock (2007) and Garcia (2013), the Chi-Square test shows that the sum of the *PhotoPes* coefficients between lags one and four is not significantly different from zero

Overall, we find that *PhotoPes* helps explain and predict returns after extending the sample period substantially. However, the *Getty Images* sample has four key flaws that make us cautious about interpreting the above results. First, although *Getty Images* provides a date with each photo, the time the photo is published is not clear. This lack of clarity regarding timing makes it difficult to know whether the significant negative relation between *PhotoPes* and returns is contemporaneous or predictive. We err on the side of caution and interpret it as a contemporaneous relation. Second, photos in the *Getty Images* sample are not finance-focused and thus we are not sure about the audience of these *Getty Images* photos.[5] Third, the popularity rank that we use to determine which photos are included concerns readers because we do not know the popularity rank with certainty at the time the photo is taken and the measure is proprietary to *Getty Images*.

Finally, one of our main goals in this paper is to compare pessimism embedded in news photos and text. We are better able to compare photo and text using the *WSJ* sample because photo descriptions from the *Getty Images* sample often fail to capture the overall tone of the article. For example, we might have a photo of a politician making a speech. The photo can convey emotions expressed by the politician and crowd. However, the photo description might simply state the name of the politician and the fact that the politician was giving a

[5] Our results are consistent with studies documenting the importance of general news on financial markets. For instance, Calomiris and Mamaysky (2019) show how news articles in general (i.e. not strictly financial news) have a large impact on stock markets in 51 developed and emerging economies.

speech. To alleviate this concern, we use the *WSJ* sample and capture pessimism embedded in news text of articles and compare directly with the pessimism embedded in the photo associated with that article.

Table B1: The impact of *PhotoPes* on market returns using *Getty Images* sample

This table reports β_1 and β_2 from the following time-series regression:

$$R_t = \beta_1 PhotoPes_t + \beta_2 L4(PhotoPes_t) + \beta_3 L5(R_t) + \beta_4 L5(R_t^2) + \beta_5 X_t + \varepsilon_t,$$

where $PhotoPes_t$ is calculated as the proportion of the 10 most popular photos from *Getty Images* predicted to be negative on time *t*. $L4$ and $L5$ transform a variable into a row vector consisting of four and five lags of that variable, respectively. X_t contains a set of exogenous variables including a constant term, day of the week dummies (except for Monday), and a recession dummy. *PhotoPes* is winsorized at 1% and standardized to have zero mean and unit variance. R_t is log daily return on CRSP value-weight ($VWRETD_t$) index. The sample period is between January 1926 and June 2018. Newey and West (1987) standard errors are applied to compute t-stat for all specifications. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	(1)	
	$VWRETD_t$	
Variables	β	t-stat
$PhotoPes_t$	-0.019**	-2.256
$PhotoPes_{t-1}$	0.008	0.958
$PhotoPes_{t-2}$	-0.003	-0.356
$PhotoPes_{t-3}$	0.015*	1.765
$PhotoPes_{t-4}$	-0.013	-1.505
Sum t to t-4	-0.012	
Sum t-1 to t-4	0.007	
	$\chi^2(1)$	*p*-value
$\chi^2(1)$[Sum t to t-4=0]	0.569	0.451
$\chi^2(1)$[Sum t-1 to t-4=0]	0.214	0.644
Adj R-sq	0.015	
N	16,935	