International Journal of Forecasting 31 (2015) 575–581

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

News volume information: Beyond earnings forecasting in a global stock selection model

Robert A. Gillam [a], John B. Guerard Jr. [a,*], Rochester Cahan [b,a]

[a] McKinley Capital Management, LLC, 3301 C Street, Suite 500, Anchorage, AK 99503, United States
[b] Empirical Research Partners, New York, NY, United States

ARTICLE INFO

keywords:
Earnings forecasting
I/B/E/S
Portfolio optimization
Active returns

ABSTRACT

Earnings forecasting models produce highly statistically significant asset selection, active equity, and total active returns. We propose a measure of abnormal news volume that controls for the size of the firm and the analyst attention that it receives, and demonstrate that news volume information can enhance returns relative to using only an earnings forecasting model. Furthermore, we show that this measure enhances the predictive power of a global stock selection model using information coefficients, Boolean signals, and efficient frontiers.

1. Introduction

Expected returns on assets are a key input to the mean–variance portfolio selection process. One can estimate global composite models of expected returns using earnings expectations data, price momentum variables, and reported financial data (Guerard, Rachev, & Shao, 2013; Sivaramakrishnan & Stubbs, 2013). In a follow-up analysis to the global composite model analysis, Guerard, Markowitz, and Xu (forthcoming) reported the importance of including an earnings forecasting variable in a global portfolio. In this analysis, we boldly go beyond earnings forecasting and construct portfolios using variations on news volume data for the period from January 2004 to December 2013.

2. Literature review

One can estimate global composite expected returns models using earnings expectations data, price momentum variables, and reported financial data. For examples, the reader is referred to studies by Bloch, Guerard, Markowitz, Todd, and Xu (1993), Guerard (2012), and Guerard et al. (2013). We test a variable created from a new data source described by Cahan, Chen, Luo, Alvarez, Jussa and Wang (2012) in the context of these stock selection models.

Our study connects two strands of literature. First, our focus on the volume of news a company garners is closely related to a body of literature studying the impact of public news flows, or a lack thereof, on the pricing of securities and investor behavior. Second, our work draws on an emerging body of literature that studies the use of computer algorithms for interpreting and quantifying the information contained in textual data, such as news articles from the financial press or the internet.

Barber and Odean (2007) and Fang and Peress (2009) both examine the impact of media coverage on stock returns. Barber and Odean show that individual investors are more likely to buy stocks that have featured prominently in the news. They argue that this is tied to the so-called limited attention hypothesis, which posits that investors' attention is a scarce resource, and therefore the prohibitive cognitive cost of searching among thousands of stocks leads them to focus on those that have

* Corresponding author. Tel.: +1 907 563 4488.
E-mail addresses: jguerard@mckinleycapital.com, jbguerard@gmail.com (J.B. Guerard Jr.).

http://dx.doi.org/10.1016/j.ijforecast.2014.12.007
0169-2070/

received media coverage recently. Barber and Odean report that no excess (or superior, as we report) returns are earned by focusing on attention-buying patterns. Fang and Peress reported that, in the cross-section, stocks with no media attention outperform those with high media attention. Moreover, the chief difference in returns was found for financial stocks that had no media attention. Fang and Peress (2009) reported that the negative returns associated with news stories persisted regardless of whether one modeled excess returns using the CAPM or the Fama and French (1992, 1995, 2008) three-factor model. Fang and Peress reported that stocks with high levels of media coverage earned lower returns, which was consistent with the stocks having higher levels of analyst forecast dispersion and idiosyncratic volatility. Also closely related to our work is the study by Chan (2003), who investigates the tendency of stock prices to exhibit momentum and reversal patterns in the presence of news coverage. Chan finds that momentum effects are stronger when a firm is in the news, while reversal patterns are more prevalent when there is no news about the company. This suggests the use of a trading strategy that over-weights stocks with strong momentum in the presence of news flows. We use this idea in the development of our abnormal news volume variable.

Our work is also tied to a burgeoning body of literature that focuses on the use of sophisticated computer algorithms for quantifying the information content of textual data. Early work in this field, such as that of Tetlock (2010), assessed the sentiment expressed in financial news stories via a simple ratio of the numbers of positive and negative words used. Tetlock's study produced four findings: (1) ten-day reversals were 38% lower on news days; (2) ten-day momentum returns existed only on news days for most stocks; (3) the correlation between abnormal returns and volumes was 35% higher on news days; and (4) the price impact of order flow was 3.3% lower on news days. For small and illiquid stocks, Tetlock (2010) reported that volume-induced momentum occurred only on news days, and reversals occurred on other days. Since then, more sophisticated statistical algorithms have been deployed to enable a better matching of the computer's interpretation of a news story to that of a human reader. For example, Groß-Klußmann and Hautsch (2011) use the same database that we use from Thomson Reuters, where every news story released by the Reuters news agency is processed in real-time, with advanced statistical algorithms being used to quantify the companies mentioned in each story, the positivity or negativity of the sentiment expressed in the story, and the novelty of the story compared to other stories published in the recent past. They find that high frequency returns can be predicted to some extent by combinations of these story-specific data points.

In this analysis, we define the news volume as the number of news articles relating to a certain company over a fixed trailing time period. The data source is a real-time feed from Thomson Reuters News Analytics, as well as a historical archive of news data. We use the following data from Thomson Reuters:

(1) News amount—we calculate the number of news articles published by the Reuters newswire over the past 12 months that mention the company in question.

(2) Relevance scores—how relevant the news story is to the companies mentioned within it. For example, if the article is a story about a specific company's quarterly earnings announcement, it will be highly relevant to that company. On the other hand, if the company is mentioned in passing as part of a general market summary story, the relevance score will be much lower. These scores are determined via proprietary natural language processing algorithms deployed by Thomson Reuters.

The news volume is correlated heavily with the firm size, analyst coverage and liquidity of a company, because larger companies tend to attract the most media coverage and analyst attention. We therefore compute the following robust regression (Beaton & Tukey, 1974) to remove these effects:

$$\begin{aligned} Log(NewsVol)_t &= a_0 + a_1\, Log(MCap)_{t-13} \\ &\quad + a_2\, Log(NOA)_{t-13} \\ &\quad + a_3\, Log(DolVol)_{t-13} + e_t, \end{aligned} \tag{1}$$

where:

$$\begin{aligned} NewsVol &= \text{number of news articles about a company in the past 12 months, weighted by the relevance of the news;} \\ MCap &= \text{market capitalization;} \\ NOA &= \text{number of analysts;} \\ DolVol &= \text{average 20-day dollar volume;} \\ e &= \text{regression residual.} \end{aligned}$$

The residual from the robust estimation of Eq. (1) is the measure of abnormal news volume used in this analysis. A higher abnormal news volume implies that a company has been featured more prominently in the news, after controlling for market capitalization and analyst coverage. Guerard, Chettiappan, and Xu (2010) noted the McKinley Capital strategy of MQ, a model that consists of CTEF, a variable of forecasted earnings acceleration, calculated using equally-weighted consensus stock analysts' revisions, forecasts, and breadths of forecasts, price momentum, and the risk of the stock, which is its historic standard deviation of returns. We define MQMod as a combination of stocks with a high abnormal news volume and a high MQ variable score. Consistent with Chan (2003), our rationale is that the earnings momentum effects captured by MQ will be more prominent among stocks that are garnering more attention from the financial press. We therefore create a proprietary stock selection variable, MQMod, to reflect a composite model.[1] We also report results for another variable, known as the CTEF. CTEF is a composite earnings variable that is calculated using equally-weighted revisions, forecasts, and breadths of forecasts. This variable was originally put forward by Guerard (1997), building upon the Elton, Gruber, and Gultekin (1981) earnings forecasting analysis, and has maintained its strength as a predictor of returns in recent studies such as that of Guerard et al.

[1] The weights in MQMod are proprietary; however, the vast majority of the weight, and effectiveness, of MQMod is composed of CTEF and PM.