The colour of finance words[*]

Diego García[†]   Xiaowen Hu[‡]   Maximilian Rohrer[§]

May 29, 2020

Abstract

We study a standard machine learning algorithm, the multinomial inverse regression (MNIR, Taddy, 2013) to measure sentiment in accounting finance related text. We show how dictionaries constructed from such an algorithm perform much better than existing techniques in predicting stock return movements. Most of the improvement is driven by using bi-grams instead of single words, consistent with much of the natural language literature on sentiment. Our paper focuses on the transcripts from earnings calls, but we also study the release of 10-K statements (Loughran and McDonald, 2011) and general market financial news (Tetlock, 2007).

JEL classification: D82, G14.

Keywords: measuring sentiment, machine learning.

[†]Diego García, University of Colorado Boulder, Email: `diego.garcia@colorado.edu`; Webpage: `http://leeds-faculty.colorado.edu/garcia/`

[‡]Xiaowen Hu, University of Colorado Boulder, Email: `xiaowen.hu@colorado.edu`; Webpage: `https://www.colorado.edu/business/xiaowen-hu`

[§]Maximilian Rohrer, Norwegian School of Economics, Email: `Maximilian.Rohrer@nhh.no`; Webpage: `https://www.nhh.no/en/employees/faculty/maximilian-rohrer/`