

Tilinpäätöksen negatiivisten sanojen frekvenssi osaketuottojen ennustamisessa

On helppoa olettaa, että nykyteknologialla onnistutaan kaivamaan tilinpäätöksistä piilotettuja totuuksia, joiden perusteella pystyisi ennakoimaan osaketuoton kehitystä tilinpäätöksen julkaisua seuraavina päivinä. Tätä efektiä on tutkittu yhdysvaltojen osakemarkkinoilla analysoimalla SEC:n julkaisemia tilinpäätösdocumentteja. Dokumenteista on parsittu negatiivisten sanojen määriä suhteutettuna dokumentin kokoon, ja näin luokiteltu niitä negatiivisuusasteikolla. Amerikkalaiset tutkijat Loughran and McDonald vuonna 2011 julkaistussa tutkimuksessa liittävät tämän tilinpäätösten negatiivisuusluokittelun, julkaisun jälkeisiin negatiivisiin tuottoihin. Omassa tutkimuksessani näytän, että aiemmat tutkimustulokset eivät yleisty uudempaan dataan sekä avaan ongelmia liittyen taloudellisen tekstin analysointiin nykyisillä keinoilla.

Data ja menetelmät

Seuraan Loughranin and McDonaldin vuonna 2011 julkaisemaa tutkimusta lataamalla ensin kaikki 10-K tilinpäätös documentit SEC:n nettisivulta aikaväliltä 1994-2008. Toistan tällä näytteellä heidän havaintonsa, jonka jälkeen käytän samaa menetelmää uudempaan näytteeseen aikaväliltä 2008-2018.

Kerään tarvittavat taloudelliset muuttujat kuten yrityksen koko, B/P-luku, osakkeen kiertonopeus sekä indikaattorimuuttujan NASDAQ-listaukselle. Testaan negatiivisten sanojen taloudellista merkitystä käyttäen selittävänä muuttujana osakkeen ylituottoja neljän päivän pituisella aikajaksolla, alkaen tilinpäätöksen julkaisusta. Teen Fama-MacBeth (1973) poikittais-regressiot dataan käyttäen Newey-West (1987) keskivirhettä yhdellä viiveellä. Huomioin myös toimialavaihtelun käyttämällä indikaattorimuuttujia perustuen Fama and French (1997) 48-toimialan luokitteluun.

Tulokset

Osoitan, että aikasemmin havaitun suhteen tuottojen ja negatiivisten sanafrekvenssien välillä pystyy toistamaan aikavälillä 1994-2008, mutta sen jälkeen vaikutus on kadonnut. Kun testaan aiemmin pätevänä pidettyä mallia uudemmalla datalla aikaväliltä 2008-2018, ei mikäänlaista kytköstä negatiivisten sanafrekvenssien ja niitä seuraavien osaketuottojen välillä ole havaittavissa.

Päätelmät ja sudenkuopat

Yleinen sanafrekvensseihin perustuva tekstin luokittelumenetelmä ei ole tarpeeksi selitysvomainen arvioimaan tilinpäätöksiä ja niistä seuraavaa osaketuoton kehitystä luotettavasti. Koska vaikutus on kuitenkin havaittavissa vanhemmassa datassa, on mahdollista että ammattisijoittajat ovat suunnitelleet sijoitustrategioita, joiden seurauksena vaikutus on hävinnyt.

Lisäksi pelkkiin sanafrekvensseihin perustuva luokittelu on herkkä virheille sekä jättää kokonaan huomiotta kontekstin, missä kukin sana ilmenee. SEC:n datankirjausstandardit eivät ole täysin vakiintuneita, joten virheet tekstin parsinnassa voivat aiheuttaa vääriä tulkintoja. Kyseessä on vasta alkava tutkimuksen ala, jossa on vielä paljon potentiaalia kehitykselle.