



return rose or fell by at least 2.5%. The graph shows that for the March “jump” days the U.S. common stocks had both highly scattered returns and very volatile average returns.

Against this enormously volatile overall market and individual stock reactions to COVID-19 news, Prof Davis and his co-authors extract information on risk factors driving stock returns and thus investors’ view on future resource allocation. In a nutshell, they used firms’ discussions of Risk Factors in 2010-2016 filings to characterize firm-level risk exposures, used these exposures to explain equity returns on jump days, and then interpreted the drivers of these returns.

The authors first used a conventional dictionary method that uses 36 sets of terms curated by experts to capture a particular type of risk exposure. Using machine counting, they quantified each firm’s exposure to those 36 categories of risk. They regressed firm-level abnormal returns on these risk factors, controlling for standard industry fixed effects and firm-level market capitalization and leverage. They ran the regressions for each of the 17 “jump days” and also pooled them.

The second method is supervised machine learning that uses the multinomial inverse regression (MNIR) method introduced by Taddy (2013). Under this approach, an algorithm selects terms in a very large feature space useful in explaining an outcome of interest. It has a vector of 19,000 different terms, compared the dictionary method’s 200 terms.

Both approaches showed remarkable congruence in predicted firm-level returns. But, the MNIR model explained half of the firm-level abnormal return variations on pandemic fallout days, compared to the dictionary method which explained only a third of such variations.

However, it is hard to extract clear insights from the raw MNIR results. Prof Davis and his co-authors thus devised a hybrid approach. They identified seed terms that MNIR weights highly in capturing firm-level risk exposures, then used those seeds to automatically generate a set of related terms based on how similar they were in linguistic context and MNIR weights. These terms were then pruned down using human domain expertise into sets defining new risk exposure categories.

This hybrid approach outperforms the dictionary approach in terms of fit. It does not perform as well as the MNIR approach statistically but yields results with richer characterizations of the forces driving nuanced firm-level returns. “It is how we uncover the role of exposures to social distancing restrictions, drug trials, e-commerce and more. It is also how we uncover the role of downstream demand linkages.”

For instance, downstream exposure to aircraft production predicts negative firm-level returns in reaction to bad COVID-19 news. The same news predicts positive returns for firms with high exposures to specific metals, such as tantalum, that are used in semiconductors, lasers, integrated circuits as well as for cloud computing and telecommunications.

Prof Davis sees their hybrid approach as flexible and adaptable enough to be useful in many other applications. For instance, applying the method to characterize returns the day after the 2020 Super Tuesday elections – when the market rose 4 per cent – they found that election results news drove negative returns for firms with high exposure to hotels, gambling, fracking and financial management, and positive returns for firms with high exposure to healthcare, health insurance, REITs, property rentals, communications and construction.

Responding to a question at the ABFER keynote, Prof Davis shared one observation he had picked up – that the MNIR surpassed human-constructed expert dictionaries when it came to highlighting risks to do with where a firm sits in the input-output structure, and detecting aspects of the global supply

chain. He said a hybrid approach could be used to draw out the input-output structure in a far more granular manner than previously done using industry-level input-output matrices.

Regulatory filings are but one source of text that could be used to predict, understand and interpret firm-level outcomes beyond the standard measures. “There’s just many mounds of raw text sitting out there waiting to be turned into useful information. We have the computing speed to do it now. We have the digital archives in many cases that provide the text. We have these methods that can be applied. I see it as a gold mine for empirical researchers,” said Prof Davis.

In their own application of the hybrid approach, the findings are significant. The pandemic-induced return reactions they uncovered foretell future corporate earnings surprises. They also foreshadowed other shifts in the real economy. For instance, the traditional retail sector’s weak returns were followed by major job losses, while online shopping and delivery firms hired more. Similar patterns unfolded with the devastation of airlines and job cuts in aircraft production, oil and gas companies going bankrupt, and plunging print media ad revenues. Meanwhile, cloud computing and digital services surged on the back of higher demand from the work-from-home boom.

If, as the researchers note, the social and economic fallout of these economic shifts will present major policy changes for years to come, early interpretation of company-level risk factors could at least point out where the subtler reallocative shifts are already underfoot.

---