New Products*

Abhiroop Mukherjee **Tomas Thörnquist** **Alminas Žaldokas**

Research Proposal for NBER Conference on Big Data and High-Performance Computing for Financial Economics

Abstract

Applying machine learning techniques to unstructured data on corporate press releases on product announcements and merging them with the data on stock price reactions, we construct a systematic database of new product introductions across different industries. In this project, we plan to use this data on inventions to provide insights on across-industry variation in productivity growth and study how it is affected by different economic forces such as product market concentration.

*Mukherjee and Žaldokas: Hong Kong University of Science and Technology; Thörnquist: Shell Street Labs. Emails: amukherjee@ust.hk; tomas@thornquist.me; alminas@ust.hk.

Productivity growth in the US, and in the rest of the world, seems to be slowing. A growing perception is that this could be related to a secular decline. As productivity-enhancing innovations from the decade before, like computing technology, reach maturity, their contribution to growth is tapering off while new inventions have less potential to improve growth.[1] However, measuring inventions in a comprehensive way across industrial sectors is inherently challenging. While physical products in supermarkets might have product codes and each new drug needs to go through FDA approval, other industries might not have a systematic way to track the inventions. This limits our ability to understand what economic forces contribute to variation of inventions across industries.

We believe that we have created a measure that can help systematically value inventions in a forward-looking way. In particular, we apply machine learning tools to unstructured data and combine it with stock market performance. Our approach follows the following steps. We first extract all media articles from Factiva database that involve firms traded on US stock exchanges (NYSE, NASDAQ, AMEX), and that Factiva has filed under *New Products/Services* category over 1989-2010. We then parse out which firms are mentioned in each article and estimate abnormal returns on the mentioned firm's stock on the day of the release of the article. In our further analysis we only consider the firms with a positive stock price reaction on the five days (-2, 2) surrounding the publication of the articles.[2]

In order to classify the articles we employ a state of the art convolutional neural network approach. The labeled training data was sourced from undergraduate students employing a custom built visual interface in the form of mobile phone app. 31 students were asked to classify $2,000$ articles each in a binary fashion indicating whether each article presented to them discusses a major new product introduction. The students were asked not to consider such cases as a minor update of an existing product (especially, software), a repeated presentation of the product at a trade show, and a discussion of future introduction of new

[1] As Peter Thiel famously put it, "We wanted flying cars, instead we got 140 characters".

[2] These first steps are similar to the methodology in Mukherjee et al. (2017) where we use a smaller sample and bag-of-words technique to sort out new product introductions rather than a more comprehensive machine learning approach applied to a much broader sample as in this proposal.

products without any precise schedules.

We randomly assigned each article to two students. Keeping only the articles where both students agreed on their classifications we ended up with a final training set containing 15,160 labeled articles. The remaining articles in the sample were classified using Google's pre-trained Word2Vec word embeddings. The final k-fold out-of-sample results give us a precision (rate of true positives) of 93% and a recall (ratio of positive articles found) of 86% for an F1 score of 89%.

In our final step, we estimate the dollar value added of each new product by taking the product of stock market reaction to the announcement and the firm's market value before it. This approach is similar to the one taken by Kogan et al. (2017) to value patents.

The advantages that our measure brings in are as follows. First, focusing on new product launches leads us to the final good in the innovation process, rather than intermediate steps like R&D investments or patents, that have been the focus of the innovation literature. Second, rather than valuing these products using some statistical or accounting technique, we rely on the collective wisdom expressed through financial markets. This automatically lends a forward-looking aspect to our measure. Under rational expectations, the announcement premium of a new product on the innovating firm's share price will capture the importance of this innovation in a way that statistical or accounting measures cannot. Third, instead of studying a particular industry to collect the data on new products, we can construct a consistent across-industry measure that would allow us to explore economy-wide implications.

In our project, we aim to provide evidence how economic forces such as competition shape productivity growth. The preliminary results indicate that 74% of new product observations occur in high markup industries[3] and while there is strong firm-level persistence in new product creation, no such persistence exists in low markup sectors. We plan to further link the distribution of firms in terms of new product creation to future performance, e.g., whether it gives the rise of the "superstar" firm phenomenon.

[3]Defined as 6-digit NAICS industry with the median industry's gross profit margin larger than the across-industry median in a particular year (based on publicly-listed firm data).

References

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, 2017, Technological innovation, resource allocation, and growth, *The Quarterly Journal of Economics* 132, 665–712.

Mukherjee, Abhiroop, Manpreet Singh, and Alminas Žaldokas, 2017, Do corporate taxes hinder innovation?, *Journal of Financial Economics* 124, 195 – 221.