

Rotman School of Management
University of Toronto
105 St. George Street
Toronto, ON, Canada, M5S 3E6

Office: (416) 946-0723
Mobile: (647) 720-2810
Homepage: www.barbosu.com
Email: sandra.barbosu10@rotman.utoronto.ca

NBER Postdoc Proposal - Sandra Barbosu

At a broad level, my research focuses on the implications of digitization and the proliferation of “big data,” or voluminous amounts of data that require new ways to be analyzed, for firm performance and strategy. The empirical setting of my dissertation is the US motion picture industry. Using data from Amazon Instant Video about users’ movie rental patterns, I uncover implicit similarity between movies based on revealed user preferences, which were previously very difficult, if not impossible, to observe. I identify the effect of a movie’s implicit similarity to other movies on its box office performance, and I explore the formation and latent characteristics of implicit clusters of similar movies. My work has implications for both movie performance and for studios’ movie development strategies.

“Big Data” Goes to the Movies: Revealing latent similarity among movies and its effect on box office performance

The first chapter of my dissertation explores how the rise of big data presents movie studios with the opportunity to gain insights that can influence movie performance and help shape studio movie development processes. It addresses the question: what is the value of big data for studios? Specifically, the paper focuses on Amazon Instant Video’s “Customers who rented this [focal movie] also rented this...” lists to (1) evaluate similarity between pairs of movies based on users’ rental patterns, (2) show that movies that are implicitly similar to others, having many co-rentals in common with other movies, have better box office performance than movies that are far from others, and (3) argue why this would be the case.

I present a novel way to determine implicit similarity between movies from users’ online rental behavior, based on an adaptation of a measure originally developed by Zuckerman (2004) in the stock market setting. A benefit of employing this measure is that it emerges directly from user rental patterns, without having to rely on third-party genre classifications to infer similarity. Using this measure of implicit similarity, I explore the relationship between implicit similarity and box office performance.

This analysis is motivated by the theoretical question of how implicit movie similarity, as determined from revealed consumer rental patterns, matters for perfor-

mance. On one hand, movies that are similar to others tend to form implicit clusters of formulaic movies that appeal to niche groups of consumers. These movies may be high-performing because they are well aligned with niche audience segments, which ensures viewers. On the other hand, movies that are distant from others combine attributes liked by different market segments, giving them the potential to command a larger audience. Results show that movies that are similar to others (having many common co-rentals with other movies) have better box office performance than movies that are far from others, indicating that consumers receive a greater utility from formulaic movies than from movies that combine disparate elements.

Concerns about the causality of this relationship arise in this context for two reasons. First, there is the possibility that an unobservable factor, like movie quality, can impact both movie box office performance and its implicit similarity (i.e. high quality movies earn high revenues and tend to be rented together with other high quality movies online). Second, since implicit similarity is based on online co-rentals, which occur after a movie has played in theaters, it is possible that the relationship is actually due to reverse causality (i.e. high-performing movies may be rented together even if they are not otherwise similar). I disentangle the direction of causality in this relationship by employing a multistep approach combining control function and instrumental variable methods, as well as a test to evaluate robustness to omitted variable bias. Results provide evidence against the two proposed alternative explanations, and lend support to the hypothesized direction, that implicit similarity is a feature that affects movie box office performance.

The paper also explores the formation of implicit clusters of movies, made up of movies with many common co-rentals, and shows that these clusters are different from common genre classifications. Observable characteristics like genre, actor/director star power, studio type and MPAA rating have modest power in explaining why clusters form, which raises the question of what latent similarity dimensions the movies within clusters share. An initial exploration shows that one latent similarity dimension is movie theme, which is not captured through coarse classifications. Exploring other possible latent common features can provide insights to studios, because understanding these latent commonalities could help them make movies that better target different audience segments.

The second chapter of my dissertation is motivated by the observation that different movie studios have different strategies for the types of movies they produce. Some studios consistently make movies with high implicit similarity to other movies, other studios consistently make movies that are far from others, while a third group of studios alternate between making both types of movies. Given this observation,

what factors could account for these differences? Is it possible that some studios are better at understanding latent dimensions than others? In this chapter, I am theoretically analyzing the reasons for heterogeneity in firm-level selection of movie portfolios.

In the third chapter, I am conducting an in-depth empirical analysis of the factors that could be responsible for studio-level heterogeneity, as a test of the theory developed in the second chapter. As part of this, I also want to further explore the potential latent similarity dimensions that movies within clusters share, going beyond theme to explore features such as narrative style, rhythm, visual and sound effects, which studios could be aware of to different extents. An interesting way to explore these types of latent features is to collect movie reviews from critics and users and conduct text analysis, looking for information specific to these features.

Other research

The Puzzle of Crowdsourcing Platforms (with Joshua Gans)

This paper complements my other projects by exploring a different aspect of digitization. We address the question of why crowdsourcing websites, such as Wikipedia or Zooniverse, are sustainable and continue to receive user contributions. These sites face a free-riding problem, since they are essentially public goods: it takes time and effort for a contributor to add content to the site, but the content becomes instantly available to others for free. What allows these sites to remain sustainable despite this issue? This paper presents a theoretical model to argue that the technology of these websites, which allows edits to be divisible so that multiple users can make edits of varying sizes, alleviates the public goods problem, because their design encourages contributions of different magnitudes from people with wide arrays of knowledge.

This theory is currently being tested on the citizen science website Zooniverse, which relies on the efforts of volunteers to parse through and annotate millions of images, with the goal of helping scientists make sense of large volumes of data. Zooniverse has over 30 projects on different topics to which volunteers can contribute. One of the projects on Zooniverse underwent a format change that required users to contribute a larger amount of edit - compared to the initial format - in order to contribute to the project. This project provides an opportune setting in which to test the model, comparing the change in the contribution rate before and after the change in format, to the contribution rates of similar projects that did not undergo format changes. Understanding the role of technology in maintaining sus-

tainable user participation on websites like Zooniverse and Wikipedia is important, as it could inform crowdsourcing platforms about optimal formats to encourage the highest amounts of participation.