Desk rejection of submissions to academic journals: an efficient screening process?

Menghan Li, PhD
Georgetown University

Abstract

This paper studies the desk reject decisions using the submissions in one leading journal and matches the citations and publication records via Microsoft Academic. The result shows that the desk reject decisions are not random. First, this paper finds that desk-reject submissions are less likely to be published in other journals compared with papers that pass the desk review round but end up being rejected. Second, this paper develops a control function model and shows that non-desk-reject papers tend to have significantly higher citation counts than desk-reject papers when controlling the predicted probability of avoiding desk rejections.

Moreover, by using natural language processing (NLP) and machine learning (ML) tools, this paper manages to construct and select important features from the abstract of the submissions, which are overlooked by previous literature. The machine learning models attempt to predict the desk reject decisions and achieve an accuracy of 72%, which is 10% higher than the baseline model which assumes rejecting every paper. The NLP models show that papers with more concrete and descriptive words are more likely to be published. The way a paper is written has proven to be very important in editorial decisions.

Introduction

Sometimes, papers with high scientific contributions are at the risk of being rejected by leading journals in the field. The American Economic Review, for example, rejected Nobel Laureate Akerlof's seminal paper on the market for lemons, stating that "AER did not publish such trivial study." The paper was rejected by the Journal of Political Economy because "it was too general to be true." It was rejected by the Review of Economic Studies because, again, "it was too trivial."(Shugan, 2007)[1] As a result, it's critical to understand why papers are desk rejected and whether or not the desk rejection is an efficient screening process.

The literature has given rise to many theories on the factors that may be crucial to editorial decisions.(Angrist et al.,2017; Card et al.,2013)[2,3] The text itself is overlooked, nevertheless. First, I generate variables using NLP techniques to capture the traits of a particular paper. The presenting style, keywords, and research topics make up the three factors. Then, I create several ML algorithms to discover the editors' preferences. The feature selection model's outcome confirms that author attributes like submission history and coauthorship are essential to editors in the desk review round. More importantly, the feature selection model demonstrates that desk rejection judgments are heavily influenced by the presenting style, including word choice and the length of the abstract.

The third step is to incorporate the selected variables into the econometric model.(Card and DellaVigna,2020)[4] The empirical result shows that non-desk-rejected papers have higher citation counts than desk-rejected papers when controlling the predicted probability of avoiding desk rejection. Moreover, the model indicates that editors favor submissions that use concrete and descriptive language over those that use terminology and abstract ideas. Additionally, editors favor lengthy abstracts containing keywords relevant to trending subjects.

Model

In addition to the characteristics indicated in the earlier literature, this paper succeeds in extracting information from the paper's abstract. Assume that the editor's observable information x_{1i} about paper i consists of:

$$x_{1i} = f(P_i, W_i, F_i, A_i)$$

where $x_{1i} = f(P_i, W_i, F_i, A_i)$ is a function of presentation style (P_i), keywords (W_i), research topic (F_i), author information (A_i).

The features are selected based on their ability to predict editor's desk-reject decisions using numerous machine learning methods. More specifically, the feature is more significant the more variety in editors' decisions it can account for. I require a baseline model to determine whether the machine learning algorithm is capable of learning editorial decisions. Desk rejecting all submissions to journal is one potential baseline model. The baseline model's accuracy is 62% since desk rejections account for 62% of the submissions.

The effectiveness of each machine learning algorithm is displayed in the Table 1. The accuracy of every model is greater than 62%, proving that the algorithms perform better than just rejecting every piece of writing. In other words, artificial intelligence is somewhat capable of picking up editorial decisions.

Figure 1 illustrates that various features generated by NLP are essential to desk rejection. a significant portion of the variation in the prediction result can be explained by the topic of the submissions, which is obtained from the LDA algorithm.

The parameters of the desk reject decision model are identified using the control function method. I first fit a Probit model with the instrument variable and the observables for the non-desk reject decision. The generalized residual from the Probit model is then estimated. Running OLS for citations including the observables, the estimated generalized residual, and the non-desk reject decision is the final step. The type of variables that should be included in the observables relies on the outcome of the machine learning algorithm. Based on the feature importance, I incorporate regressors into the control function model such as coauthorship, author submission history, paper topics, unique keywords, average sentence length, paper descriptiveness, and concreteness.

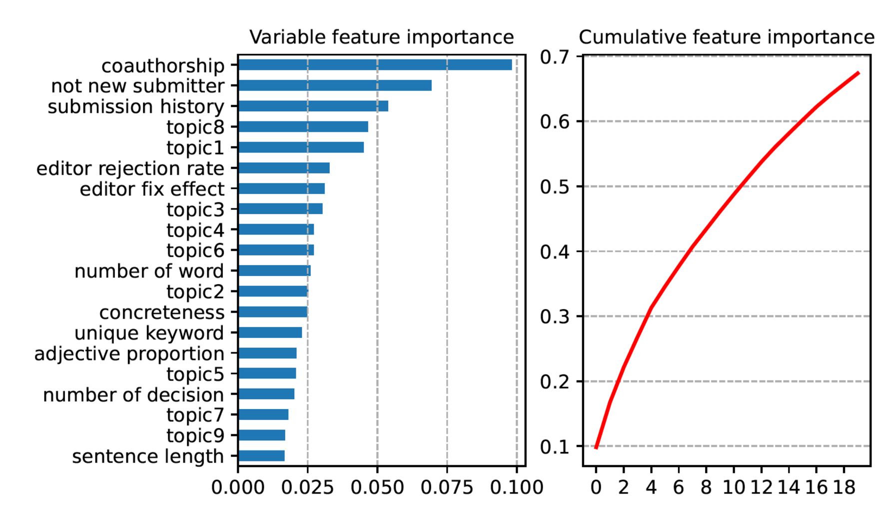

Figure 1. Gradient boosted tree feature importance

Methods	Accuracy
Logit	0.7075
Lasso	0.7047
Ridge	0.7086
Elastic net	0.7069
Support vector machine	0.6762
Neural net(ANN)	0.6891
Random forest	0.7157
Gradient boosted tree	0.7212

Table 1. Model evaluation

Result

Figure 2 reveals that the desk rejection is not that random by contrasting the percentage of submissions published in other journals between the desk-rejected manuscript and the rejected papers that make it through the desk review round.

Figure 3 illustrates how the editor's desk rejection decisions and realized citations relate to one another. Based on the Probit regression from the control function approach, the average predicted probability of avoiding desk rejection is calculated. It demonstrates that there is a sizable difference in the average predicted probability of avoiding desk rejection for papers that receive desk rejection and those that do not. It implies that the desk reject decision is not made at random if the editor maximizes the expected citations of the published papers to raise the journal's impact factor.

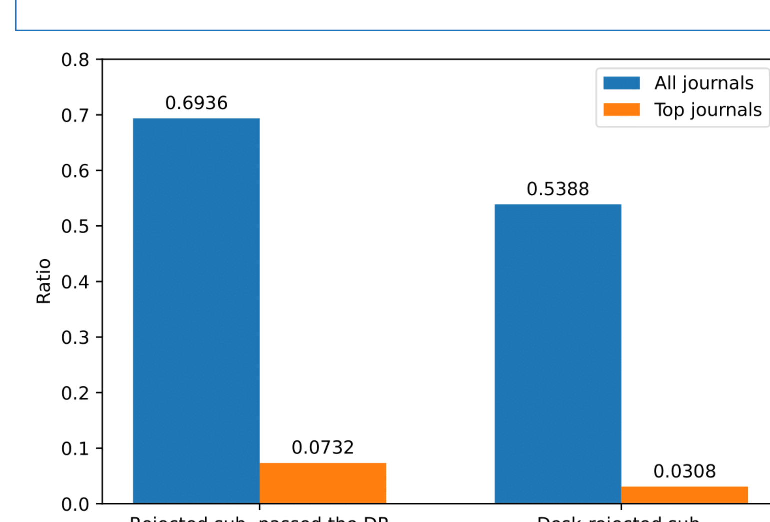

Figure 2. Rejected submissions published in other journals

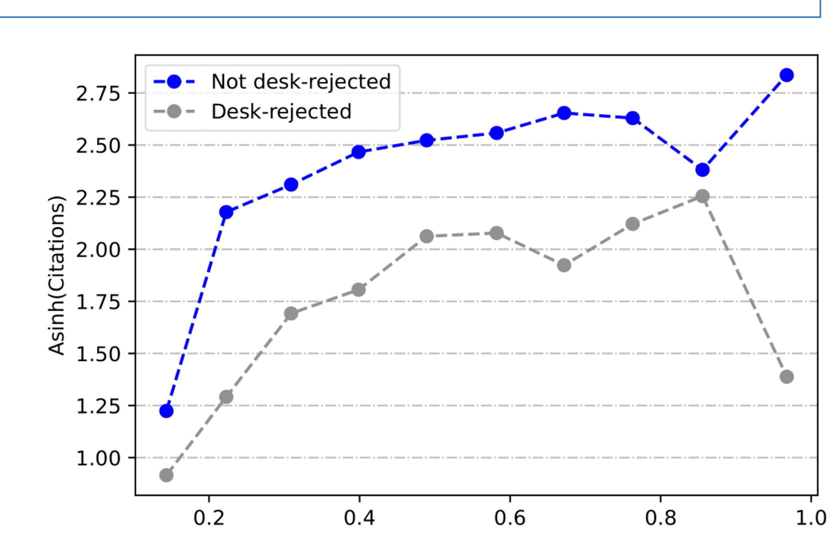

Figure 3. Informativeness of the editor's private signal

Conclusions

The paper examines whether the desk rejection of submissions to academic journals is an efficient screening process by answering the following two questions. Will desk-reject submissions be less likely to get published in other journals compared with papers that pass the desk review round but end up being rejected? And do non-desk-reject papers have higher citation counts than desk-reject papers when controlling the predicted probability of avoiding desk rejections? By tracing the publication record of the rejected paper, I find that desk-reject submissions are less likely to get published in other journals. Additionally, the editor has the ability to select the outstanding paper from all the submissions. Desk rejection of submissions to academic journals is efficient, or at the very least, not a random screening procedure.

Contact

Menghan Li
Georgetown University
Email: ml1629@georgetown.edu
Website: https://gufaculty360.georgetown.edu/s/contact/00336000014Se5jAAC/menghan-li
Phone: (202)-705-4033

References

1. Steven M Shugan. The editor's secrets. *Marketing Science*, 26(5):589-595, 2007.
2. Joshua Angrist, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu. Economic research evolves: Fields and styles. *American Economic Review*, 107(5):293-97, 2017.
3. David Card and Stefano DellaVigna. Nine facts about top journals in economics. *Journal of Economic Literature*, 51(1):144-61, 2013.
4. David Card and Stefano DellaVigna. What do editors maximize? evidence from four economics journals. *Review of Economics and Statistics*, 102(1):195-217, 2020.