



# Data and Incentives

ANNIE LIANG, University of Pennsylvania

ERIK MADSEN, New York University

CCS Concepts: • **Applied computing** → **Economics**.

Additional Key Words and Phrases: big data, incentives, economics, algorithmic scoring

## ACM Reference Format:

Annie Liang and Erik Madsen. 2020. Data and Incentives. In *Proceedings of the 21st ACM Conference on Economics and Computation (EC '20), July 13–17, 2020, Virtual Event, Hungary*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3391403.3399474>

In many industries, firms provide services whose profitability depends on an unobserved characteristic, or “quality,” of recipients. For example, the profitability of a car insurance policy depends on the insuree’s driving risk, and the profitability of issuing a loan depends on the borrower’s creditworthiness. While a recipient’s quality is not directly observable, service providers can use data to help forecast it. This paper is about the interaction between two kinds of data which inform these forecasts: traditional *past outcome* data—e.g. insurance claims rates or credit card repayment—and novel *consumer segment* data identifying “similar” individuals based on aggregated online activity and other digitally tracked behaviors.

To understand the interaction between these two kinds of data, consider two prototypical consumers, Alice and Bob. If Alice reported an automobile accident last claims cycle, that event suggests a higher future accident risk for Alice. If consumer segment data suggests that outcomes for Alice and Bob are correlated, then Alice’s accident also impact the provider’s beliefs about Bob’s accident risk. We study and contrast two kinds of correlations between Alice and Bob’s observed outcomes:

- *Quality linkages*: Some agents have intrinsic similarities in lifestyle or personal characteristics. For example, if the provider learns that Alice and Bob are both graduate students, then Alice’s accident from the last claims cycle may cause the provider to perceive an increased risk from all graduate students, and hence assess Bob to be a higher-risk individual as well.
- *Circumstance linkages*: Other agents experience common transient shocks to outcomes. For example, if Alice and Bob drove through the same blizzard, the provider may assess Bob to be a *lower-risk* individual on account of Alice’s outcome.

Linkages across agents always improve forecast accuracy. But their use also impacts other aspects of service provision. In many settings in which past outcome data is used for forecasting, agents can strategically exert effort to impact those outcomes: for example, drivers decide how carefully to drive, while borrowers decide how much financial caution to exert. Although effort is costly, agents will exert it if terms of service—e.g. an insurance premium or interest rate—improve following a good outcome. When service providers additionally use another agent’s outcomes to make predictions about a given agent, this changes the impact of the agent’s own outcome on

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

EC '20, July 13–17, 2020, Virtual Event, Hungary

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7975-5/20/07.

<https://doi.org/10.1145/3391403.3399474>

future terms of service, reshaping their incentives for effort. We show that quality linkages and circumstance linkages turn out to have very different implications for consumer effort.

Formally, we consider a multiple-agent version of the [Holmström 1999] career concerns model, in which each agent's outcomes depend on the agent's unknown type, and also on their choice of effort. We suppose that outcomes are correlated across agents through either a common component of agents' types (i.e. a quality linkage) or through a common component of random shocks to outcomes (i.e. a circumstance linkage). This correlation creates an informational externality, as each agent's choice of effort has an effect not only on the market's perception of *his* future outcomes, but also on the market's perception of outcomes for other linked agents.

Our main results demonstrate that quality and circumstance linkages have opposing effects on incentives for effort, with quality linkages depressing effort while circumstance linkages encourage it. This is because under quality linkages, other agents' outcomes *substitute* for a given agent's outcomes, decreasing the marginal informativeness of the agent's own outcome for forecasting. Thus, the value to exerting effort to improve one's outcome is decreased. In contrast, when agents are linked by a circumstance linkage, other agents' outcomes improve the informativeness of the agent's own outcome by helping the market to more effectively separate quality from noise. This improves the agent's incentives to exert costly effort.

We establish these effort comparative statics in a model with general type and noise distributions, imposing only a standard log-concavity condition ensuring that posterior estimates of latent variables are monotone in signal realizations. Reasoning about incentives for effort in such an environment is challenging, because outside of special cases, type forecasts are a complex nonlinear function of signal realizations. A technical contribution of our paper is the development of techniques for establishing how marginal incentives for effort change as the number of correlated signals grows large.

We next use our results to understand when data usage for forecasting helps or hurts welfare. As a benchmark, we consider an environment "without linkages" in which the market is permitted to use only an agent's *own* past data to predict his type. We show that the welfare impact of linkages depends crucially on the agent's discount factor. When the discount factor is large (so that agents sufficiently value the future), effort is inefficiently high without linkages. Thus, quality linkages—which depress effort—improve welfare, while circumstance linkages—which intensify effort—reduce welfare. The direction of these welfare impacts is reversed when the discount factor is small and effort is inefficiently low absent linkages. These results suggest that markets would generally benefit from targeted regulations specifying what kinds of data may be used for forecasting.

Finally, we comment on a current policy debate regarding whether firms should have proprietary ownership of their data, or if this data should be shared across an industry (as for example recently recommended by the European Commission). We extend our model by grouping firms onto several distinct *submarkets* focusing on distinct consumer groups. We contrast two different data regimes—under *proprietary data*, firms observe data linkages only for agents in their submarket, while under *data sharing*, all linkages are observed by firms in every submarket. We show that data sharing may reduce the welfare generated by consumer effort in the presence of a linkage, indicating a potential drawback to proposed data sharing regulations.

Full paper available at: <https://arxiv.org/abs/2006.06543>

## REFERENCES

Bengt Holmström. 1999. Managerial Incentive Problems: A Dynamic Perspective. *Review of Economic Studies* 66, 1 (1999), 169–182.