Data and Incentives*

Annie Liang† Erik Madsen‡

November 16, 2020

Abstract

We study the impact of big data on incentives for socially valuable effort. While identification of similar individuals is a familiar feature of many markets, big data has significantly increased the resolution of these groupings. We propose a model in which a market segments consumers based on their observable covariates, while consumers exert costly effort to improve the market's perception of their type. Our main results characterize how the distribution of consumer effort changes as the market gains access to new covariates. When covariates are independent, or are related only through their effect on average outcomes, then the impact on effort is uniform for all consumers, and we characterize when effort increases or decreases. Under more general forms of correlation, observation of a new covariate may lead to disparate impact—increasing effort for some consumer groups and decreasing it for others—but the average impact of an additional covariate can be signed.

1 Introduction

Lenders and insurers routinely use consumers' observable covariates to predict their future behaviors. Some of these covariates are simple demographic characteristics,

*We are grateful to Eduardo Azevedo, Dirk Bergemann, Alessandro Bonatti, Sylvain Chassang, Yash Deshpande, Ben Golub, Matt Jackson, Yizhou Jin, Navin Kartik, Rishabh Kirpalani, Alessandro Lizzeri, Steven Matthews, Xiaosheng Mu, Larry Samuelson, Andrzej Skrzypacz, Juuso Toikka, and Weijie Zhong for useful conversations, and to National Science Foundation Grant SES-1851629 for financial support. We thank Changhwa Lee for valuable research assistance on this project.

†Department of Economics, University of Pennsylvania

‡Department of Economics, New York University

such as age or gender. But "big data" gives lenders and insurers access to unprecedented level of detail about their clients, which they can also use to set terms of service: For example, a credit card holder with American Express saw his credit limit fall from \$10,800 to \$3,800 because "Other customers who have used their American Express card at establishments where you recently shopped have a poor repayment history with American Express"; borrowers with the lender CompuCredit have had their credit limits adjusted by a behavioral scoring model which docks points for transactions with automobile repair shops and massage parlors; and drivers may soon begin to see automobile insurance priced based on traffic density on their usual routes, or the property crime rates of their neighborhood. (See Section 3.2 for a fuller description of these settings and others.) The availability of larger sets of covariates may improve the accuracy of forecasting, but also has other important economic consequences that are less well-understood.

In this paper, we study the impact of big data on incentives for socially valuable effort. To fix ideas, suppose an auto insurer partitions a population of drivers based on car age and miles driven. Since there is residual uncertainty about any given driver's accident risk, each driver has a reputational incentive to exert effort to drive safely—this is an important force mitigating moral hazard. Now suppose the auto insurer can additionally observe the frequency with which each driver visits bars, and can base insurance rates on this covariate. How do the drivers' incentives to exert effort change?

To study this issue, we propose a model in which a market segments consumers based on their observable covariates, and agents have a reputational incentive to exert productive effort (Holmström, 1999). In more detail, each agent has an unknown quality type (e.g. risk level) and additionally encounters an idiosyncratic shock to outcomes (e.g. road conditions). We assume that the agent's quality type and shock can both be partly predicted from a set of covariates available to the market. The agent chooses how much (costly) effort to exert, and this choice—along with his quality type and shock—jointly determine an outcome that the market observes (e.g. a claims rate). The agent then receives a reputational payoff equal to the market's posterior expectation of his type. Under standard distributional assumptions, higher outcome realizations are associated with higher underlying types, so the agent has an incentive to exert effort to improve the market's beliefs about his quality. The optimal level of effort depends on the covariates that the market has access to. Our main results characterize how consumer effort changes as the market gains access to new covariates.

How the agent's covariates are correlated turns out to be a key determinant of their impact on effort. Our first result considers a baseline setting in which all covariates are statistically independent from one another. We show that covariates that inform about the agent's type reduce effort, while covariates that inform about a shock to outcomes increase effort. This is because if the new covariate provides no information about other unobserved characteristics of the consumer, then the finer partitioning necessarily reduces uncertainty in each refined subpopulation. In each new subpopulation, the market has a more precise belief about the agent's type, and thus the value of exerting effort to improve the market's beliefs is reduced. Conversely, a dataset providing information about the shock reduces uncertainty about the size of the shock, and so the value to exerting effort to manipulate the (more informative) outcome realization rises.

We next extend our analysis to accommodate correlation between covariates, where new forces emerge. In this case, the impact of a new covariate will generally depend on its realization. We show by example that observing a new covariate can lead to *disparate impact*, increasing effort for some consumer groups and decreasing it for others. To illustrate this idea, consider our previous car insurance example, in which the insurer newly observes how frequently each driver visits bars. Suppose accident rates are more heterogeneous among the subpopulation of "drivers who frequently visit bars" compared to the overall population. For drivers in this subgroup, the value of maintaining a good record is higher than for drivers identified as infrequent bargoers. If this effect is strong enough, the new covariate may lead frequent bargoers to exert *more* effort, even as infrequent bargoers exert less effort due to decreased uncertainty about their type. This behavior contrasts with the independent-covariate baseline, where each additional covariate had the same directional effect on effort for all consumers.

In general, the impact of new covariates is sensitive to details of the correlation across covariates. Nevertheless, we are able to characterize impact for two important classes of correlated variables: those linked via mean shifts, and those that satisfy a suitable affiliation property. When covariates are correlated only via a "mean shift", so that subpopulations exhibit differing average outcomes but the same outcome variability, then our monotonicity result from the independent-covariates baseline fully generalizes: Each new covariate that informs about type reduces effort across the population, while each new covariate that informs about shock increases effort across the population. When the component of type or quality explained by a covariate is affiliated with the unobserved component, then monotonicity holds on average: Conditioning forecasts on new covariates about type reduces effort on average across a subpopulation, while conditioning forecasts on new covariates about circumstances increases effort on average.

1.1 Related Literature

Our paper contributes to an emerging literature studying the behavioral and welfare consequences of forecasting consumer outcomes from big data. A central concern in the literature is the possibility that consumers may be able to distort the data being used for forecasting, opening the door to strategic interaction between consumers and forecasting mechanisms.

One set of papers in this literature examine incentives for "gaming" forecasts, by misreporting private information or exerting effort to distort signals; see, e.g., Frankel and Kartik (2020); Ball (2020); Hu et al. (2019); Eliaz and Spiegler (2019, 2020); Hennessy and Goodhart (2020). These papers treat gaming as intrinsically inefficient, either because it reduces the precision of forecasts or because effort is costly and signal distortion generates no social value. Another set of papers considers settings in which effort which is potentially productive; see, e.g., Frankel and Kartik (2019); Haghtalab et al. (2020). In these papers effort improves the agent's characteristics or signals a type which the market wishes to forecast. Effort is therefore not intrinsically wasteful, though outcomes may still be inefficient if too little or too much effort is incentivized in equilibrium.

Both sets of papers treat the data environment as fixed, and focus on equilibrium outcomes or design of an optimal forecasting mechanism. Our paper instead considers how outcomes change as the set of covariates available to the forecaster varies. We model applications in which effort is productive, as in the second set of papers, and characterize when observation of additional covariates raises or lowers welfare from effort. These features of our work are also present in Tirole (2020), who similarly considers the consequences of varying information available to forecasters in an environment with productive effort. While we focus on how information impacts effort on a task of interest to the market, Tirole (2020) considers whether a designer might wish to muddle a signal to induce effort on an unrelated task.

As discussed in further detail in Section 3.2, a leading application of our model is to insurance markets. Jin and Vasserman (2020) verify empirically that in the auto insurance market, a short-term monitoring program which generates additional signals of accident risk (for example, frequency of night driving and harsh breaking) incentivizes drivers to substantially change driving behavior in order to reduce their future insurance premiums. Their study indicates that incentives for effort deriving from reputational concerns are of significant practical relevance in this market.

Methodologically, our paper models incentives for effort following the career concerns model of Holmström (1999), the classic framework for analyzing the role of reputation-building in motivating effort. Within this framework, we incorporate covariates as additional signals which are correlated with an agent's type and/or output. A small set of papers, most notably Dewatripont et al. (1999) and Rodina (2018), have extended the career concerns model to allow for additional exogenous signals beyond the baseline output signal.[1] These papers take an ex ante perspective on the agent's information, by assuming that the agent must choose effort before observing the realization of any exogenous signals. By contrast, our model studies an ex post environment in which agents are aware of their covariates when choosing effort. This allows us to ask new questions regarding how observation of a new covariate impacts effort across agents with different covariate values.

2 Preliminaries: Reputational Incentives for Effort

Our model of reputational incentives for effort builds on the classic career concerns model of Holmström (1999). We provide here a brief review of this model, which we subsequently augment with a framework of predictive covariates. Readers familiar with the model may skip this section and proceed directly to Section 3.

[1] There is additionally a literature on relative performance comparisons, e.g. Meyer and Vickers (1997), in which the market may additionally observe the output of another agent with correlated unobservables. This literature focuses on the design of incentive contracts, and so is less closely related to our work.

An agent participates in a market across two periods $t = 1, 2$. His productivity in the market is characterized by a quality type $\theta \sim F_\theta$, which is persistent across time and initially unknown to himself and the market. In **period 1**, the agent generates an observable outcome variable

$$Y = e + \theta + \varepsilon,$$

which is determined by the agent's quality θ, a transient shock $\varepsilon \sim F_\varepsilon$ independent of the agent's type, and an effort level $e \in \mathbb{R}_+$ privately chosen by the agent.

The agent must incur a cost to exert effort, which for our main results we take to be $C(e) = \frac{1}{2}e^2$. (We extend our results to general cost functions in Section 5.2.) We will take any payment to the agent to enter the market as sunk, so that the agent's period-1 payoff from participation is just his total cost of effort:

$$U_1 = -C(e) = -\frac{1}{2}e^2$$

In **period 2**, the agent receives a reputational payoff standing in for returns from future participation in the market. We take this payoff to be equal to the market's expectation of his quality conditional on the outcome variable Y. Since the agent's effort choice is private, the market's forecast is based on a conjectured level of effort $\hat{e}$. Letting $Y^{\hat{e}} \equiv \hat{e} + \theta + \varepsilon$ be output supposing the market's effort conjecture is correct, the agent's second-period payoff conditional on realization $Y = y$ is

$$U_2 = \mathbb{E}^{\hat{e}}\left[\theta \mid Y = y\right] \equiv \int \theta dF_\theta(\theta \mid Y^{\hat{e}} = y) \tag{1}$$

where $\mathbb{E}^{\hat{e}}[\theta \mid Y = y]$ denotes the market's (potentially misspecified) expectation of θ, updating to the realization $Y = y$ assuming that $Y \overset{d}{=} Y^{\hat{e}}$. In equilibrium, the market's conjectured level of effort is equal to the effort level that the agent chooses, in which case the assumption that $Y \overset{d}{=} Y^{\hat{e}}$ is correct.

The agent's ex post payoff from participating in the market is the sum of period payoffs, $U = U_1 + U_2$. The agent's expected payoff under effort level e is therefore

$$\mathbb{E}^e[U] = \mathbb{E}^e[\mathbb{E}^{\hat{e}}[\theta \mid Y]] - \frac{1}{2}e^2$$

where $\mathbb{E}^e$ denotes the agent's expectation over output given knowledge of the true effort level e.

As is well known, an equilibrium effort level e^* must satisfy the first-order condition for optimality that the marginal value of effort at the equilibrium level e^* equals its marginal cost:

$$\left. \frac{\partial \mathbb{E}^e[\mathbb{E}^{e^*}[\theta \mid Y]]}{\partial e} \right|_{e=e^*} = e^* \tag{2}$$

The left-hand side of this equation is independent of e^* due to the additive impact of effort on output, so there exists a unique solution to the first-order condition. Throughout this paper, we will assume that the first-order approach is valid, so that this solution constitutes the unique equilibrium effort level. This effort level depends critically on the joint distribution of the variables (θ, ε), which we micro-found in the following section.

3 Data and Forecasts

3.1 Model

We model an individual's quality type θ as a function of *attributes* $\boldsymbol{a} = (a_1, a_2, \ldots, a_J)$, and the shock ε as a function of *circumstances* $\boldsymbol{c} = (c_1, c_2, \ldots, c_K)$, where the sets of attributes and circumstances are taken to be primitives of the model. We assume that each a_j is drawn from a convex set $A_j \subseteq \mathbb{R}$, and similarly that each circumstance c_k is drawn from a convex set $C_K \subseteq \mathbb{R}$. We refer to attributes and circumstances collectively as *covariates*. We treat the agent's attributes and circumstances as random variables, with $\boldsymbol{a} \perp\!\!\!\perp \boldsymbol{c}$ and each vector drawn according to joint distributions to be specified in more detail later.

The agent's attributes and circumstances partly determine his type and shock, via the deterministic effect size functions $\Psi^j : A_j \to \mathbb{R}$ and $\Lambda^k : C_k \to \mathbb{R}$. Specifically, we assume that θ and ε are decomposable as

$$\theta = \mu + \sum_{j=1}^{J} \Psi^j(a_j) + \theta^\perp$$

$$\varepsilon = \sum_{k=1}^{K} \Lambda^k(c_k) + \varepsilon^\perp,$$

where μ is a deterministic scalar and $\theta^\perp$ and $\varepsilon^\perp$ are random variables independent of each other and all attributes and circumstances. (See Section 5.1 for an extension to a more general model that allows for interaction terms between covariates.) Without loss let $\mathbb{E}[\Psi^j(a_j)] = \mathbb{E}[\Lambda^k(c_k)] = \mathbb{E}[\theta^\perp] = \mathbb{E}[\varepsilon^\perp] = 0$ for every j and k, taking $\Psi^j(a_j)$ and $\Lambda^k(c_k)$ to represent the de-meaned impacts of the covariates on outcomes, with $\mu \in \mathbb{R}$ capturing the mean outcome. The residual terms $\theta^\perp$ and $\varepsilon^\perp$ represent the unlearnable components of the agent's type and shock.

We assume that the agent and market commonly observe the distributions of qualities and shocks within certain *subpopulations* of agents, as identified by a set of observed covariates.

Definition 1. *A* subpopulation $\mathcal{S}$ *is a quadruple* $(\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$, *where:*

- $\mathcal{J} \subset \{1, \ldots, J\}$ *is a set of observed attributes and* $\mathcal{K} \subset \{1, \ldots, K\}$ *is a set of observed circumstances,*

- $\boldsymbol{\alpha} \in \prod_{j \in \mathcal{J}} A_j$ *and* $\boldsymbol{\gamma} \in \prod_{k \in \mathcal{K}} C_k$ *are realizations of the covariate vectors* $\boldsymbol{a}_{\mathcal{J}}$ *and* $\boldsymbol{c}_{\mathcal{K}}$.

Given a set of covariates $(\mathcal{J}, \mathcal{K})$, *a* $(\mathcal{J}, \mathcal{K})$-subpopulation *is any subpopulation whose observed covariates are* $(\mathcal{J}, \mathcal{K})$.

The set of observed covariates $(\mathcal{J}, \mathcal{K})$ is a primitive of the model, and we assume that the distributions of θ and ε conditional on the realizations of the observed covariates in each $(\mathcal{J}, \mathcal{K})$-subpopulation are common knowledge. When $\mathcal{J} = \mathcal{K} = \emptyset$, this assumption amounts to the marginal distributions of θ and ε being known, and corresponds to the standard informational environment used in Holmström (1999). As more covariates are observed, and the population is partitioned more finely, the assumption becomes more stringent; we discuss its interpretation in the subsequent Section 3.3.

A simple but important observation is that knowledge of the conditional distributions of types and shocks in a particular $(\mathcal{J}, \mathcal{K})$-subpopulation provides the market the same information for forecasting the agent's type as knowledge of the joint distributions of $\boldsymbol{a}$ and $\boldsymbol{c}$, the functional forms $(\Psi^j)_{j=1}^J$ and $(\Lambda^k)_{k=1}^K$, and the realizations of $(a_j)_{j \in \mathcal{J}}$ and $(c_k)_{k \in \mathcal{K}}$. Indeed, once the conditional distributions of types and shocks are computed from the joint distributions of covariates and the effect size functions, no further information from the model is needed to form conditional expectations of types given outcomes. For ease of exposition, we will cast our analysis in terms of the latter set of knowledge assumptions.

Interactions proceed according to the model described in Section 2:

> $t = 0$: The agent and market observe the $(\mathcal{J}, \mathcal{K})$-subpopulation to which the agent belongs, which induces a common belief about the joint distribution of (θ, ε, Y).
>
> $t = 1$: The agent chooses effort $e \in \mathbb{R}_+$ and incurs the cost of effort. The agent's outcome Y is realized.
>
> $t = 2$: The agent receives the market's forecast of his type.

Each $(\mathcal{J}, \mathcal{K})$-subpopulation is associated with a unique equilibrium effort level, and our main results will describe how this equilibrium effort varies depending on the set of covariates $(\mathcal{J}, \mathcal{K})$ that are used for forecasting.

We impose the following regularity conditions on the distributions of latent variables, which are maintained throughout the paper. The first ensures that the agent's utility function is differentiable, while the second ensures that types and shocks have full support on $\mathbb{R}$.[2]

Assumption 1 (Differentiability)**.** *For every set of observed covariates $(\mathcal{J}, \mathcal{K})$, each $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$, and every effort level e, $\frac{\partial}{\partial Y}\mathbb{E}^e [\theta \mid Y, \mathcal{S}]$ exists and is uniformly bounded across all realizations of Y.*

Assumption 2 (Residual Noise)**.** *The random variables $\theta^\perp$ and $\varepsilon^\perp$ have full support on $\mathbb{R}$.*

At this time, we do not impose any restrictions on the functional forms of Ψ^j and Λ^k, allowing for a completely flexible specification of the relationship between attribute values and effect sizes. The restriction that covariates contribute linearly to the outcome simplifies exposition in the main text, but we show in Section 5.1 that all of our main results extend to a general non-additive model as well. We will, however, subsequently impose certain log-concavity assumptions on the distributions of the type and shock components $\Psi^j(a_j)$ and $\Lambda^k(c_k)$, which constitute joint parametric restrictions on the functional forms of Ψ^j and Λ^k as well as the distributions of the underlying covariates. Discussion of these restrictions is deferred to the statement of our main results.

[2]We expect that our results continue to hold in a more general setting without full support. We impose this assumption to simplify proofs which establish affiliation of sets of random variables.

3.2 Leading Examples

The functioning of many insurance and lending markets relies on firms' ability to identify and pool individuals with comparable risk characteristics. Below we discuss adoption of new predictive covariates in four leading settings.

Health Insurance. Traditionally, risk classification in health insurance is based on basic demographic variables, such as the individual's age, gender and location; certain lifestyle factors such as tobacco use; and past medical and drugs claims information. New sources of big data relevant to insurers include measurements from wearable devices, social media activity, entertainment choices, online shopping, and public records (e.g. criminal and property records). These data sources make it possible for health care premiums to depend on novel kinds of personal characteristics, such as having recently purchased plus-size clothing.

Automobile Insurance. Automobile insurance companies similarly forecast accident risk for the purpose of determining whether to insure an individual, and what insurance premium to charge. Drivers are presently classified into groups based on variables including the age and gender of the driver, the driver's vehicle type, and miles driven in the previous claims period. Geolocation data from mobile devices make it possible for insurers to correlate a driver's routes and parking locations with environmental statistics such as weather patterns, property crime rates, population density, traffic density, and census information. Moreover, vehicle tracking devices provide reports of variables such as frequency of hard braking and average speed. This information may allow insurers to put together a more detailed picture of the context of an accident (e.g. by using speed and braking data to determine which driver was at fault), and may also open the door to insurance pricing based on specific routes driven.

Lending. Assessment of creditworthiness is presently based on a sparse set of covariates, including payment history, credit utilization, length of credit history, new credit, and credit mix as the main determinants. New "alternative credit scores" predict creditworthiness on the basis of a large set of personal covariates, including everything from where the individual shops to the size of their social network to how quickly the loan applicant scrolls through an online terms-and-conditions disclosure.

Law Enforcement. Another use of novel predictive covariates is for forecasting crime. For example, the Los Angeles Police Department has experimented with use of computer-forecasting for prediction of areas where gun violence is likely to occur, and "hot spots" with a high likelihood of property-related crimes, while the Chicago Police Department has experimented with use of an algorithmically-derived "heat list" that predicts the people most likely to commit gun violence or to be a victim of it.

In each of these settings, adoption of novel predictive covariates has so far been limited, and evidence of their effectiveness is limited. Nonetheless, the sustained interest in using big data for prediction suggests that policymakers may eventually have to decide how to regulate use of novel predictive datasets.

3.3 Discussion

We discuss below the interpretation of several key modeling choices.

Knowledge assumptions. In our model, the agent and the market both know the covariates $(\mathcal{J}, \mathcal{K})$ used to define subpopulations, as well as the agent's values for those covariates (i.e. the agent's sub-population). The assumption that agents know the covariates used to define their subpopulation can be relaxed to uncertainty over possible sets of covariates, provided that the necessary statistical assumptions are satisfied by each set of covariates. But it is worth noting that regulations are constantly moving in the direction of enforcing greater transparency about usage of covariates. The assumption that agents know their covariate values can also be relaxed, although it too is one we consider realistic in many cases. Whether agents know their covariates at the time of effort choice distinguishes our results from a related exercise in Dewatripont et al. (1999), as we discuss further in Section 4.

While in our model the agent is commonly informed about the distribution of types and shocks within their subpopulation, we also assume that the agent does not possess any private information about their type beyond what the market possesses. Note that this assumption does not require that the agent is unaware of covariate values unobserved by the market, merely that he cannot forecast how those covariates impact outcomes. The knowledge environment of our model is a good fit for applications to big data in which the impact of novel covariates on outcomes can be estimated only through fitting a model to a large dataset. While the agent may be able to become informed about how a firm's current statistical model classifies their subpopulation, especially under data transparency laws, he is much less likely to know anything about how that model would change in response to augmentation with additional covariates.

Informational environment. We assume that the distribution of qualities and shocks within the agent's subpopulation is known. We interpret this assumption in either of two ways. First, many insurers and banks have records of past outcomes for a large number of agents, allowing them to infer the aggregate distribution of qualities and shocks within observable subpopulations. Our model can be viewed as applying to a possible future in which widespread data collection means that not only are large numbers of covariates collected, but additionally insurers have access to many observations per subpopulation. This interpretation may be less realistic in the short-run, if the number of covariates collected outstrips the number of observed outcomes and the population becomes finely partitioned into groups with small memberships. A second interpretation of our model applicable to such settings is that the distributions of types and shocks for each subpopulation reflect the market's subjective beliefs about that group (as informed by the available data). Under this interpretation, small samples sizes would be captured by more diffuse type and shock distributions. Our results do not depend on whether distributions are interpreted as objective or subjective.

The agent's reputational payoff. As in Holmström (1999), we suppose that the agent's second-period payoff is the market's expectation of his type conditional on his period-1 outcome. A direct microfoundation of this specification is that the agent participates in the market and generates productive output twice, the agent's type is persistent across periods while the agent's shock is drawn anew, and competition between firms to serve the agent results in his being paid his expected output in each period.

Strategic manipulation of covariate values. While some covariates are intrinsic (e.g. height), many of the covariates we discussed in Section 3.2 may be interpreted as choice variables of the agent. Use of large sets of personal covariates to determine lending and insurance policies may cause individuals to distort their behavior—for example, an agent may manipulate the online purchasing decisions they make, or engage in different hobbies than they would otherwise. See, e.g., Haghtalab et al. (2020) for an analysis of classification when agents may strategically manipulate their covariates. We abstract from the possibility of these distortions in the present paper.

4 Main Results

Our main results concern how equilibrium effort changes as the market gains access to new covariates. We establish several versions of our result for settings with different assumptions about correlation across covariates.

Our first result applies in a baseline setting in which covariate realizations are independent from one another, so that beliefs about unobserved covariates do not depend on the values of the observed covariates. Here we demonstrate a stark monotonicity result: For each additional attribute that the market observes, equilibrium effort decreases regardless of the agent's realized covariate values. The effect is the opposite in the case of circumstances: For each additional circumstance that the market observes, equilibrium effort increases.

In a more general setting where an individual's covariates may be correlated, we show that our monotonicity continues to hold so long as covariates are correlated *in mean only*, in a sense we make precise. Beyond that case, we establish that so long as an individual's covariates satisfy an affiliation property—so that higher realizations of one covariate imply higher realizations of the others—monotonicity continues to hold *on average* across all possible covariate realizations. That is, equilibrium effort decreases on average when an additional attribute is observed, and increases on average when an additional circumstance is observed. However, when the residual uncertainty about unobserved covariates depends on the realization of observed covariates, it is no longer guaranteed that observation of an additional covariate has a monotone effect on effort for all realizations of that covariate's value. We show by example that in such settings, observation of an additional covariate can have disparate impact across subpopulations, increasing effort for some individuals but decreasing effort for others.

4.1 Preliminary Definitions and Notation

We derive our monotonicity results from an ex post perspective, holding fixed both a baseline set of observed covariates $(\mathcal{J}, \mathcal{K})$ and a set of realized values $(\boldsymbol{a}_{\mathcal{J}}, \boldsymbol{c}_{\mathcal{K}}) = (\boldsymbol{\alpha}, \boldsymbol{\gamma}) \in \prod_{j \in \mathcal{J}} A_j \times \prod_{k \in \mathcal{K}} C_k$ for those covariates. All conditions we impose on the correlation structure of covariates to obtain monotonicity therefore relate to the conditional distributions of $(\boldsymbol{a}_{-\mathcal{J}}, \boldsymbol{c}_{-\mathcal{K}})$ given $(\boldsymbol{a}_{\mathcal{J}}, \boldsymbol{c}_{\mathcal{K}}) = (\boldsymbol{\alpha}, \boldsymbol{\gamma})$. These conditions can therefore be interpreted as properties of the distribution of unobserved covariates within a subpopulation which ensure monotonicity within the subpopulation when an additional covariate is observed. When we wish to discuss distributional assumptions on random variables which hold conditional on the realized covariate values within a subpopulation $\mathcal{S}$, we will say that that an assumption holds "on $\mathcal{S}$".

To economize on notation, we define the following notion of *type and shock components*, capturing the impact of a given covariate or set of covariates on an agent's type or shock.

Definition 2. *Given an attribute j, define $\theta_j \equiv \Psi^j(a_j)$ to be the jth* type component*. Given a set of attributes $\mathcal{J}$, define*

$$\theta^{-\mathcal{J}} \equiv \sum_{j \notin \mathcal{J}} \theta_j + \theta^{\perp}$$

to be the residual type component*. Similarly, given a circumstance k, define $\varepsilon_k \equiv \Lambda^k(c_k)$ to be the kth* shock component*. Given a set of circumstances $\mathcal{K}$, define*

$$\varepsilon^{-\mathcal{K}} \equiv \sum_{k \notin \mathcal{K}} \varepsilon_k + \varepsilon^{\perp}$$

to be the residual shock component*.*

When the observed covariates are $(\mathcal{J}, \mathcal{K})$, the agent's type θ and the shock ε to his outcome can be decomposed as

$$\theta = \mu + \sum_{j \in \mathcal{J}} \theta_j + \theta^{-\mathcal{J}}$$
$$\varepsilon = \sum_{k \in \mathcal{K}} \varepsilon_k + \varepsilon^{-\mathcal{K}}$$

where $\mu + \sum_{j \in \mathcal{J}} \theta_j$ and $\sum_{k \in \mathcal{K}} \varepsilon_k$ are observed by the agent and market while $\theta^{-\mathcal{J}}$ and $\varepsilon^{-\mathcal{K}}$ are not.

Finally, in our main results we will impose regularity conditions on the distributions of type and shock components relating to log-concavity of their density functions. Log-concavity of latent variables in an additive model is a standard assumption ensuring that higher outputs lead to larger inferred types and shocks, so that higher effort leads to better perceptions of quality. Our results will require that regularity holds both under the baseline set of covariates, but also once an additional covariate is added to the observed set.

Definition 3. *A random variable X is* log-concave *if the distribution function of X admits a log-concave density function. A conditional random variable $X|Y$ is* log-concave *if conditional on any realization of Y, the distribution function of X admits a log-concave density function.*

Definition 4 (Regularity)**.** *Fix a set of covariates $(\mathcal{J}, \mathcal{K})$ and a $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. Say that $\mathcal{S}$ is* regular *if $\theta^{-\mathcal{J}}$ and $\varepsilon^{-\mathcal{K}}$ are log-concave on $\mathcal{S}$.*

Definition 5 ($\mathcal{S}$-Regularity)**.** *Fix a set of covariates $(\mathcal{J}, \mathcal{K})$ and a $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. An attribute $j' \notin \mathcal{J}$ is* $\mathcal{S}$-regular *if $\theta^{-\mathcal{J}\cup\{j'\}} \mid a_{j'}$ is log-concave on $\mathcal{S}$. Similarly, a circumstance $k' \notin \mathcal{K}$ is* $\mathcal{S}$-regular *if $\varepsilon^{-\mathcal{K}\cup\{k'\}} \mid c_{k'}$ is log-concave on $\mathcal{S}$.*

Definition 6 (Global Regularity)**.** *Say that a setting is* globally regular *if for every set of covariates $(\mathcal{J}, \mathcal{K})$, every $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$, and every pair of covariates $j' \notin \mathcal{J}$ and $k' \notin \mathcal{K}$, $\mathcal{S}$ is regular and j' and k' are $\mathcal{S}$-regular.*

We provide several examples below of covariates that satisfy the required log-concavity properties (see Appendix D.1 for details).

Example 1. All covariates are jointly normal, and the functions Ψ^j and Λ^k are affine.[3]

Example 2. The attribute a is a one-dimensional location variable which is uniformly distributed on an interval $[c, d]$, while $\Psi(a) = a - x$ is (signed) distance from a fixed point $x \in [c, d]$.

Example 3. The attribute a is the expected number of friends that one can borrow money from, and it is exponentially distributed, while $\Psi(a) = \sqrt{a}$.

Example 4. The attribute a is days between social media posts, and it has a gamma distribution, while $\Psi(a) = \log(a)$.

[3]This special case of our model closely corresponds to models studied in Meyer and Vickers (1997), Bergemann et al. (2020), and Acemoglu et al. (2019).

Example 5. c is the number of inches of precipitation last month, and it has an exponential distribution, while $\Lambda(c) = -\log(c)$.

4.2 Independent Covariates

Our first result applies to environments in which all covariates are independent. In this case, observation of an additional attribute leads deterministically to a decrease in equilibrium effort, while observation of an additional circumstance leads deterministically to an increase in effort.

Proposition 1. *Suppose that the covariates $(\mathcal{J}, \mathcal{K})$ are observable, and that the agent belongs to the regular $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. Fix any $\mathcal{S}$-regular attribute $j' \notin \mathcal{J}$ and $\mathcal{S}$-regular circumstance $k' \notin \mathcal{K}$. If all covariates are mutually independent, then:*

- *Further observing attribute j' reduces the agent's effort.*
- *Further observing circumstance k' increases the agent's effort.*

Corollary 1. *If the setting is globally regular, then observation of each new attribute reduces effort, while observation of each new circumstance increases effort.*

The key to this result is that—under the assumption that covariates are independent from one another—observation of an additional covariate reduces the market's uncertainty about the corresponding component of output, *regardless of the realized value of the covariate*. Suppose first that a new attribute is acquired. This reduces the market's uncertainty about θ, and so there is "less to learn" about the agent's quality from the realization of the agent's outcome, reducing the marginal value of exerting effort to improve the realization of that outcome. By contrast, acquisition of an additional circumstance reduces uncertainty about the shock ε. This makes the outcome a more informative signal of the agent's type, increasing the marginal value to improving its realization. Recalling the first-order condition (2) from Section 2, these changes in the marginal value of effort are directly inherited by the level of equilibrium effort.

Proposition 1 is closely related to Proposition 5.1 in Dewatripont et al. (1999), with the key difference that our result compares *ex-post* effort conditional on the realization of a set of covariates, while Dewatripont et al. (1999) compares *ex-ante* effort. This difference between the ex-post and the ex-ante analysis is not merely a technical difference, as the ex-ante perspective implies that observation of an additional covariate has the same impact for all consumers, while under an ex-post perspective, consumers in different groups may react differently to acquisition of the same covariate.[4]

In the present setting of mutually independent covariates, the two perspectives coincide: Proposition 1 says that equilibrium effort increases uniformly across realizations when an additional attribute is acquired, and decreases uniformly across realizations when an additional circumstance covariate is acquired, so the impact of effort from an ex-ante perspective holds realization by realization. In fact, not only is the direction of change in effort the same across realizations, but the size of the change in effort is the same. Fixing any set of observed covariates $(\mathcal{J}, \mathcal{K})$, equilibrium effort is the same in every $(\mathcal{J}, \mathcal{K})$-subpopulation. Thus, observing an additional covariate changes effort by the same amount for all consumers.

4.3 Correlation in Mean

In the baseline setting of independent covariates, the addition of a new covariate affects effort through its impact on residual uncertainty about the agent's type or shock. This basic force extends beyond independent covariates. In particular, the logic underlying Proposition 1 suggests that if covariates are correlated in such a way that residual uncertainty does not depend on the realization of the covariate, then the directional impact of new covariates will be the same. We now define a natural class of correlated-covariate models exhibiting this property, and extend Proposition 1 to cover this class.

Definition 7. *Given a set of attributes* $\mathcal{J}$, *the* de-meaned residual type component $\tilde{\theta}^{-\mathcal{J}}$ *is defined to be*

$$\tilde{\theta}^{-\mathcal{J}} \equiv \theta^{-\mathcal{J}} - \mathbb{E}[\theta^{-\mathcal{J}} \mid \boldsymbol{a}_{\mathcal{J}}].$$

Similarly, given any set of circumstances $\mathcal{K}$, *the* de-meaned residual shock component $\tilde{\varepsilon}^{-\mathcal{K}}$ *is defined to be*

$$\tilde{\varepsilon}^{-\mathcal{K}} \equiv \varepsilon^{-\mathcal{K}} - \mathbb{E}[\varepsilon^{-\mathcal{K}} \mid \boldsymbol{c}_{\mathcal{K}}].$$

[4] The ex-ante analysis corresponds to an agent who does not know their covariate values when choosing effort, while the ex-post analysis corresponds to one who does. For many of the covariates that have been discussed (see Section 3.2), we expect the latter to be the case.

Definition 8 (Mean shifters). *Fix a set of covariates* $(\mathcal{J},\mathcal{K})$ *and a* $(\mathcal{J},\mathcal{K})$*-subpopulation* $\mathcal{S}$. *An attribute* $j' \notin \mathcal{J}$ *is an* $\mathcal{S}$-mean shifter *if* $\tilde{\theta}^{-\mathcal{J}\cup\{j'\}}$ *is independent of* $a_{j'}$ *on* $\mathcal{S}$. *Similarly, a circumstance* $k' \notin \mathcal{K}$ *is an* $\mathcal{S}$-mean shifter *if* $\tilde{\varepsilon}^{-\mathcal{K}\cup\{k'\}}$ *is independent of* $c_{k'}$ *on* $\mathcal{S}$.

The assumption that a particular covariate is a mean shifter is a substantive restriction on permitted correlation structures. Note that by construction, the random variables $\tilde{\theta}^{-\mathcal{J}\cup\{j'\}}$ and $\tilde{\varepsilon}^{-\mathcal{K}\cup\{k'\}}$ have mean zero conditional on $\boldsymbol{a}_{\mathcal{J}\cup\{j'\}}$ and $\boldsymbol{c}_{\mathcal{K}\cup\{k'\}}$, regardless of the correlation structure and realization of the agent's observed covariates. In general, however, the higher moments of $\tilde{\theta}^{-\mathcal{J}'}$ and $\tilde{\varepsilon}^{-\mathcal{K}'}$ may depend on the realizations of $a_{j'}$ and $c_{k'}$. The mean shifter property amounts to assuming away all variation in these higher moments.

One prototypical setting in which all covariates are mean shifters is the family of jointly normal covariates.

Example 6 (Jointly Normal Covariates). Suppose the vector of attribute components and circumstance components is multivariate normal. Then the mean-shifter property is satisfied *globally*—that is, for any set of observed covariates $(\mathcal{J},\mathcal{K})$ and each $(\mathcal{J},\mathcal{K})$-subpopulation $\mathcal{S}$, every attribute $j' \notin \mathcal{J}$ is an $\mathcal{S}$-mean shifter and every circumstance $k' \notin \mathcal{K}$ is an $\mathcal{S}$-mean shifter. (See Appendix D.2 for details.)

The following theorem extends the monotonicity result of Proposition 1 to settings in which the markets observes an additional mean shifter.

Theorem 1. *Suppose that the covariates* $(\mathcal{J},\mathcal{K})$ *are observable, and that the agent belongs to the* $(\mathcal{J},\mathcal{K})$*-subpopulation* $\mathcal{S}$. *Fix any* $\mathcal{S}$*-regular attribute* $j' \notin \mathcal{J}$ *and* $\mathcal{S}$*-regular circumstance* $k' \notin \mathcal{K}$. *Then:*

- *If* j' *is an* $\mathcal{S}$*-mean shifter, further observing attribute* j' *reduces the agent's effort.*

- *If* k' *is a* $\mathcal{S}$*-mean shifter, further observing circumstance* k' *increases the agent's effort.*

We prove this result by "de-meaning" the model's covariates, transforming the setting back into one of independent covariates. Fix a set of observed covariates $(\mathcal{J},\mathcal{K})$ and an additional attribute j'. Then output can be decomposed into observed and unobserved components as

$$Y = e + \underbrace{\mu + \textstyle\sum_{j\in\mathcal{J}} \theta_j + \sum_{k\in\mathcal{K}} \varepsilon_k}_{\text{observed}} + \underbrace{\theta_{j'}}_{\text{new covariate}} + \underbrace{\theta^{-\mathcal{J}'} + \varepsilon^{-\mathcal{K}}}_{\text{unobserved}},$$

where $\mathcal{J}' = \mathcal{J} \cup \{j'\}$. The key to the proof is to further decompose $\theta^{-\mathcal{J}'}$ into a sum of the average component $\mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}'}]$, which is observed under $(\mathcal{J}', \mathcal{K})$, and the residual unobserved variation $\tilde{\theta}^{-\mathcal{J}'}$. Under this transformation the previous decomposition of Y may be alternatively written

$$Y = e + \underbrace{\mu + \textstyle\sum_{j\in\mathcal{J}} \theta_j + \sum_{k\in\mathcal{K}} \varepsilon_k}_{\text{observed}} + \underbrace{\theta_{j'} + \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}'}]}_{\text{new covariate}} + \underbrace{\tilde{\theta}^{-\mathcal{J}'} + \varepsilon^{-\mathcal{K}}}_{\text{unobserved}}.$$

When covariates $(\mathcal{J}, \mathcal{K})$ are observed by the market, all variation in output is attributed to variation in the unobserved random variables $\theta_{j'} + \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}'}]$, $\tilde{\theta}^{-\mathcal{J}'}$, and $\varepsilon^{-\mathcal{K}}$. If j' is an $\mathcal{S}$-mean shifter, these three variables are mutually independent on $\mathcal{S}$. Further observing the additional covariate j' reveals $\theta_{j'} + \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}'}]$ and is equivalent to observing an additional type component in an independent-covariates setting. The logic underlying Proposition 1 therefore continues to apply, implying monotonicity.

4.4 General Correlation

In the settings we have studied so far—independent covariates and mean shifters—the impact of observing an additional covariate is deterministic, in the sense that the agent's effort under the expanded set of covariates is independent of the value of the newly observed covariate. In general, outside the class of mean shifters, this need not be true. In fact, not only the magnitude but even the *directional* effect on effort of observing a new covariate may differ across all realizations of that covariate. We demonstrate now in an example the possibility that an additional covariate may have disparate impact across subpopulations.

Example 7 (Disparate Impact). Suppose Y is the negative of a driver's automobile claims amount, which is modeled as

$$Y = e + \Psi^1(a_1) + \Psi^2(a_2) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1)$$

where $a_1 \in [0, 10]$ is the age of the car, while $a_2 \in [0, 1]$ is the driver's education percentile. The effect size functions are $\Psi^1(a_1) = -a_1$ and $\Psi^2(a_2) = -0.01(1 - a_2)$, so risk levels are higher for drivers of older cars, and drivers with lower levels of education.

The joint distribution of (a_1, a_2) is as follows: The education percentile a_2 is uniformly distributed on $[0, 1]$ while the conditional distribution of a_2 is:

$$a_2|a_1 \sim \begin{cases} U([0,1]) & \forall a_1 \geq 0.05 \\ U([0,10]) & \forall a_1 < 0.05 \end{cases}$$

Under this distribution individuals in the bottom 5% of educational attainment, i.e. those without a high-school diploma, drive cars whose ages are both older on average as well as significantly more variable.

If the market observes neither attribute, then all individuals exert the same amount of effort, which is numerically approximated to $e^* \simeq 0.16$. If educational attainment is observed, equilibrium effort for individuals with a high-school diploma declines to $e^{**} \simeq 0.077$ while equilibrium effort for individuals without a high-school diploma rises to $\tilde{e}^{**} \simeq 0.82$. The collection of data on educational attainment thus affects individuals in the population unequally.

The general force driving disparate impact is the possibility that some realizations of a covariate may be associated with small residual uncertainty across remaining covariates, while other realizations imply much larger residual uncertainty.[5] In the example above, although effort reduces for some agents and increases for others, the *average* impact of the new covariate is to reduce effort. This turns out to be a general phenomenon across a large class of correlated covariates.

Definition 9 (Affiliation)**.** *Fix a set of covariates $(\mathcal{J}, \mathcal{K})$ and a $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. The attribute $j' \notin \mathcal{J}$ is $\mathcal{S}$-*affiliated *if $\Psi^{j'}$ is one-to-one and $(\theta_{j'}, \theta^{-\mathcal{J} \cup \{j'\}})$ are affiliated on $\mathcal{S}$. Similarly, the circumstance $k' \notin \mathcal{K}$ is $\mathcal{S}$-*affiliated *if $\Lambda^{k'}$ is one-to-one and $(\varepsilon_{k'}, \varepsilon^{-\mathcal{K} \cup \{k'\}})$ are affiliated on $\mathcal{S}$.*

Affiliation is a well-known strengthening of the notion of positive correlation.[6] If a covariate satisfies $\mathcal{S}$-affiliation, then "good news" about that covariate's contribution

[5] The possibility of disparate impact, and its implications for effort under observation of an additional signal, is a distinctive prediction of an ex post environment in which agents are aware of their covariate values when making effort choices. Ex ante models in which agents choose effort before observing auxiliary signals, as in Dewatripont et al. (1999), effectively abstract from this force.

[6] An n-vector of random variables $\boldsymbol{Z}$ with joint density function $f(\boldsymbol{z})$ is *affiliated* if for every

to the outcome is also good news about the contribution of all unobserved covariates. For example, suppose the shock is decomposed as $\varepsilon = \Lambda^1(c_1) + \Lambda^2(c_2) + \Lambda^3(c_3) + \varepsilon^{\perp}$, where c_1 is precipitation, c_2 is the amount of ice on the ground, and c_3 is the stress level of the driver, with each Λ^k an increasing function. Then precipitation satisfies affiliation if a higher realization of $\Lambda^1(c_1)$ is associated with a higher belief about the sum of $\Lambda^2(c_2) + \Lambda^3(c_3)$.

$\mathcal{S}$-affiliation is useful for disciplining how news about output is attributed across underlying covariates. It ensures that when a type or shock component is revealed, changes in output which are no longer attributed to that component are not offset by larger inferences about movement of unobserved components. As a result, the total portion of an output change attributed to the type or shock is assured to decrease when a component of that random variable becomes observable. The additional requirement that the associated effect size functions be one-to-one simplifies statement of the property, by ensuring that type and shock components θ_j and ε_k are a sufficient statistic for the underlying covariate values a_j and c_k.

Example 8. Suppose that all effect size functions Ψ^j are one-to-one and type components are independent exponentially distributed random variables conditional on a common rate parameter, i.e. $\theta_j|\lambda \sim_{iid} \text{Exp}(\lambda)$, where the rate parameter is distributed as $\lambda \sim \text{Gamma}(\alpha, \beta)$ with $\alpha \geq 1$. Then, for any set of covariates $(\mathcal{J}, \mathcal{K})$ and $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$, each attribute $j' \notin \mathcal{J}$ is $\mathcal{S}$-affiliated. (See Appendix D.3 for a proof.) An analogous statement holds for circumstance variables that are correlated in this way.

The following result establishes that our previous results characterizing the directional effect of new covariates extend on average to settings with affiliated covariates.

Theorem 2. *Suppose that the covariates $(\mathcal{J}, \mathcal{K})$ are observable, and that the agent belongs to the $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. Fix any $\mathcal{S}$-regular attribute $j' \notin \mathcal{J}$ and $\mathcal{S}$-regular circumstance $k' \notin \mathcal{K}$. Then:*

- *If j' is $\mathcal{S}$-affiliated, further observing attribute j' reduces the agent's effort on average.*

$\boldsymbol{z}, \boldsymbol{z'} \in \mathbb{R}^n$, $f(\boldsymbol{z} \wedge \boldsymbol{z'})f(\boldsymbol{z} \vee \boldsymbol{z'}) \geq f(\boldsymbol{z})f(\boldsymbol{z'})$, where $\boldsymbol{z} \wedge \boldsymbol{z'}$ and $\boldsymbol{z} \vee \boldsymbol{z'}$ are the componentwise minimum and maximum of $\boldsymbol{z}$ and $\boldsymbol{z'}$. When f is strictly positive and twice-differentiable everywhere, this condition is equivalent to $\partial^2 \log f / \partial z_i z_j \geq 0$ everywhere for every pair of components i and $j \neq i$.

- *If k' is $\mathcal{S}$-affiliated, further observing circumstance k' increases the agent's effort on average.*

Theorem 2 tells us that the forces identified in our previous results extend in expectation to settings with general affiliated covariates: revealing an additional attribute has the average effect of reducing effort, while revealing an additional circumstance has the average effect of increasing effort. On the other hand, the possibility of disparate impact, as demonstrated in a previous example, tells us that this average effect may mask considerable heterogeneity across different agents. Whenever residual uncertainty is heterogeneous across covariate realizations, agents with small residual uncertainty will experience an enhanced impact on effort relative to the average effect, while agents with large residual uncertainty will experience an attenuated effect. This heterogeneity can be sufficiently extreme so as to overwhelm the direction of the average effect for some agents.

5 Extensions

5.1 Nonlinear Models

Our results so far have been developed in the context of an additive model, where covariates may be correlated but contribute linearly to the outcome. All of our results have natural analogues in a more general model, which we briefly outline.

Suppose that covariates impact outcomes via (potentially nonlinear) effect size functions $\Psi : \prod_{j=1}^{J} A_j \to \mathbb{R}$ and $\Lambda : \prod_{k=1}^{K} C_k \to \mathbb{R}$. Specifically, θ and ε decompose as

$$\begin{aligned}\theta &= \mu + \Psi(\boldsymbol{a}) + \theta^{\perp} \\ \varepsilon &= \Lambda(\boldsymbol{c}) + \varepsilon^{\perp}.\end{aligned}$$

Without loss, let $\mathbb{E}[\Psi(\boldsymbol{a})] = \mathbb{E}[\Lambda(\boldsymbol{c})] = 0$, taking $\Psi(\boldsymbol{a})$ and $\Lambda(\boldsymbol{c})$ to represent the de-meaned impact of the covariates on outcomes, with $\mu \in \mathbb{R}$ capturing the mean outcome. We generalize the notion of type and shock components as follows:

Definition 10. *Given a set of observed attributes $\mathcal{J}$, define*

$$\theta^{\mathcal{J}} \equiv \mathbb{E}[\Psi(\boldsymbol{a}) \mid \boldsymbol{a}_{\mathcal{J}}], \quad \theta^{-\mathcal{J}} \equiv \Psi(\boldsymbol{a}) - \theta^{\mathcal{J}} + \theta^{\perp}$$

to be the associated observed type component *and* residual type component. *Similarly, given a set of observed circumstances* $\mathcal{K}$, *define*

$$\varepsilon^{\mathcal{K}} \equiv \mathbb{E}[\Lambda(\boldsymbol{c}) \mid \boldsymbol{c}_{\mathcal{K}}], \quad \varepsilon^{-\mathcal{K}} \equiv \Lambda(\boldsymbol{c}) - \varepsilon^{\mathcal{K}} + \varepsilon^{\perp}$$

to be the associated observed shock component *and* residual shock component.

Given a set of observed covariates $(\mathcal{J}, \mathcal{K})$, the agent's type and the shock to the agent's outcome can be decomposed as

$$\begin{aligned} \theta &= \mu + \theta^{\mathcal{J}} + \theta^{-\mathcal{J}} \\ \varepsilon &= \varepsilon^{\mathcal{K}} + \varepsilon^{-\mathcal{K}}, \end{aligned}$$

where $\mu + \theta^{\mathcal{J}}$ and $\varepsilon^{\mathcal{K}}$ are observed while $\theta^{-\mathcal{J}}$ and $\varepsilon^{-\mathcal{K}}$ are not. Note that in this general nonlinear framework, residual type and shock components are automatically "de-meaned", in a manner similar to the model with mean shifters.

For any new covariates $j' \notin \mathcal{J}$ and $k' \notin \mathcal{K}$, define

$$\theta^{\mathcal{J}}_{j'} \equiv \theta^{\mathcal{J} \cup \{j'\}} - \theta^{\mathcal{J}}$$

and

$$\varepsilon^{\mathcal{K}}_{k'} \equiv \varepsilon^{\mathcal{K} \cup \{k'\}} - \varepsilon^{\mathcal{K}}$$

to be the market's updates to the agent's expected type and shock when the additional attribute j' and circumstance k' are observed. Then θ and ε may be further decomposed as

$$\begin{aligned} \theta &= \mu + \theta^{\mathcal{J}} + \theta^{\mathcal{J}}_{j'} + \theta^{-\mathcal{J} \cup \{j'\}} \\ \varepsilon &= \varepsilon^{\mathcal{K}} + \varepsilon^{\mathcal{K}}_{k'} + \varepsilon^{-\mathcal{K} \cup \{k'\}}. \end{aligned}$$

Fix a $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. In the general nonlinear model we may define notions of regularity and $\mathcal{S}$-regularity exactly as in the additive environment, with type and shock components in those definitions interpreted according to the definitions just given. We may further define an attribute $j' \notin \mathcal{J}$ to be $\mathcal{S}$-*affiliated* if $\theta^{\mathcal{J}}_{j'}$ is a one-to-one-function of $a_{j'}$ (holding $\boldsymbol{a}_{\mathcal{J}}$ fixed) and $(\theta^{\mathcal{J}}_{j'}, \theta^{-\mathcal{J} \cup \{j'\}})$ are affiliated on $\mathcal{S}$. The notion of $\mathcal{S}$-affiliation for a circumstance is defined analogously. With these definitions in hand, the following result holds:

Theorem 3. *Suppose that the covariates $(\mathcal{J}, \mathcal{K})$ are observable, and that the agent belongs to the $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. Fix any $\mathcal{S}$-regular attribute $j' \notin \mathcal{J}$ and $\mathcal{S}$-regular circumstance $k' \notin \mathcal{K}$. Then:*

- *If j' is $\mathcal{S}$-affiliated, further observing attribute j' reduces the agent's effort on average.*
- *If k' is $\mathcal{S}$-affiliated, further observing circumstance k' increases the agent's effort on average.*

The proof of this theorem follows exactly the same lines as the proof of Theorem 2. Despite its similarity to that theorem, however, Theorem 3 is in fact a generalization of Theorem 1, because the affiliation condition is applied to residual type and shock components which are de-meaned. This parallel can be sharpened by considering a related result under a stronger condition on covariates: say that an attribute $j' \notin \mathcal{J}$ is $\mathcal{S}$-*independent* if $\theta^{-\mathcal{J} \cup \{j'\}}$ is independent of $a_{j'}$ on $\mathcal{S}$, with an analogous definition of $\mathcal{S}$-independence for circumstances. Then the following result holds:

Theorem 4. *Suppose that the covariates $(\mathcal{J}, \mathcal{K})$ are observable, and that the agent belongs to the $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. Fix any $\mathcal{S}$-regular attribute $j' \notin \mathcal{J}$ and $\mathcal{S}$-regular circumstance $k' \notin \mathcal{K}$. Then:*

- *If j' is $\mathcal{S}$-independent, further observing attribute j' reduces the agent's effort.*
- *If k' is $\mathcal{S}$-independent, further observing circumstance k' increases the agent's effort.*

Under $\mathcal{S}$-independence, mononotonicity holds not just on average but uniformly across realizations of the additional covariate, for the same reasons as in Theorem 1. Further, if Ψ and Λ are taken to be additively separable, then Theorems 4 and 1 exactly coincide.

5.2 General Convex Cost Functions

In the main text, we maintained the assumption that effort costs were quadratic: $C(e) = \frac{1}{2}e^2$. Under this cost structure, equilibrium effort is identical to the marginal value of effort, allowing us to characterize the former by analyzing the latter. More generally, when C is a strictly convex cost function, equilibrium effort is a uniquely determined, strictly increasing function of the marginal value of effort:

$$e^* = (C')^{-1}(MV),$$

where MV is the marginal value of effort (which is independent of e^*). As a result, under any strictly convex cost function, a deterministic shift in the marginal value of effort implies a change in effort in the same direction. This implies in particular that the results of sections 4.2 and 4.3, applying to covariates which are independent or mean shifters, extend immediately.

Our results for more general correlation structures become slightly more complex for general convex cost functions. The new force which arises is that average effort may respond to mean-preserving spreads of the marginal value of effort. To illustrate the idea, consider any cost function $C(e) \propto e^k$, where $k > 2$. Under such a cost function the marginal cost of effort is convex, so that equilibrium effort is a concave function of the marginal value of effort. Hence any mean-preserving spread of the marginal value of effort reduces average effort. Conversely, if $2 > k > 1$, effort is a convex function of the marginal value of effort, and a mean-preserving spread of the marginal value of effort increases average effort.

In the case of general correlation (under the conditions of Theorem 2), observing an additional attribute has two effects: it lowers the *average* marginal value of effort, and additionally introduces a *spread* in the distribution of marginal values (relative to its baseline value in the subpopulation). If the marginal cost of effort is convex, these two forces work together to lower average effort; on the other hand, if the marginal cost of effort is sufficiently concave, average effort could increase. Analogous results hold when an additional circumstance is observed.

We view our results for the quadratic-cost case as a natural baseline for analyzing the impact of novel covariates which are "small" contributors to an agent's overall type or shock, in the sense that they don't change the marginal value of effort too much. In that limit, every cost function is approximately quadratic, and the directional effects of adding a covariate identified by our main results will hold.

6 Conclusion

As firms and governments move towards collecting large datasets of consumer transactions and behavior as inputs to decision-making, the question of whether and how to regulate the usage of consumer data has emerged as an important policy question. Recent regulations, such as the European Union's General Data Protection Regulation (GDPR), have focused on protecting consumers' privacy and improving transparency regarding what kind of data is being collected. An important complementary consideration when designing regulations is how data impacts social and economic behaviors. In the present paper, we analyze one such impact—the effect that a market's access to novel covariates has on consumer incentives for socially valuable effort. We find that the economic consequences depend crucially on whether the new covariates are primarily useful for inferring quality or denoising observations, as well as on how the covariates are correlated with one another.

Appendix

A Characterization of MV

Given a subpopulation $\mathcal{S}$, let e^* be equilibrium effort, and define

$$MV(\mathcal{S}) \equiv \frac{d}{de}\mathbb{E}^e\left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right]\Big|_{e=e^*}$$

to be the agent's equilibrium marginal value of effort.

Lemma A.1. *Fix a set of observed covariates $(\mathcal{J}, \mathcal{K})$ and any $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. Then the equilibrium marginal value of effort on $\mathcal{S}$ is*

$$MV(\mathcal{S}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \mid \mathcal{S}\right],$$

where

$$Y^0 \equiv \mu + \sum_{j=1}^{J}\theta_j + \theta^{\perp} + \sum_{k=1}^{K}\varepsilon_k + \varepsilon^{\perp}.$$

Proof. Fix a subpopulation $\mathcal{S} = (\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$. Let $h(y \mid t, e)$ be the conditional density of $Y \mid \theta^{-\mathcal{J}}, e$ on $\mathcal{S}$ and $h(y \mid e)$ be the conditional density of $Y \mid e$ on $\mathcal{S}$. Because effort affects output as an additive shift, $h(y \mid t, e) = h(y-e \mid t, 0)$ and $h(y \mid e) = h(y-e \mid 0)$ for every (y, t, e). So let $f(t \mid y, e)$ be the conditional distribution of $\theta^{-\mathcal{J}} \mid Y, e$ on $\mathcal{S}$, and let $f^0(t)$ be the conditional distribution of $\theta^{-\mathcal{J}}$ on $\mathcal{S}$. Then by Bayes' rule,

$$f(t \mid y, e) = \frac{h(y \mid t, e)f^0(t)}{h(y \mid e)} = \frac{h(y-e \mid t, 0)f^0(t)}{h(y-e \mid 0)} = f(t \mid y-e, 0).$$

Hence

$$\begin{aligned}
\mathbb{E}^{e^*}[\theta \mid Y = y, \mathcal{S}] &= \sum_{j\in\mathcal{J}}\alpha_j + \mathbb{E}^{e^*}[\theta^{-\mathcal{J}} \mid Y = y, \mathcal{S}] \\
&= \sum_{j\in\mathcal{J}}\alpha_j + \int t\, h(t \mid y, e^*)\, dt \\
&= \sum_{j\in\mathcal{J}}\alpha_j + \int t\, h(t \mid y-e^*, 0)\, dt \\
&= \sum_{j\in\mathcal{J}}\alpha_j + \mathbb{E}^0[\theta^{-\mathcal{J}} \mid Y = y-e^*, \mathcal{S}].
\end{aligned}$$

Now, conditional on $e = 0$, $Y \overset{d}{=} Y^0$, and $\theta^{-\mathcal{J}}$ is independent of Y conditional on Y^0. Therefore

$$\mathbb{E}^0[\theta^{-\mathcal{J}} \mid Y = y - e^*, \mathcal{S}] = \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y - e^*, \mathcal{S}].$$

Now,

$$\begin{aligned}
\mathbb{E}^e\left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right] &= \int dy\, h(y \mid e)\mathbb{E}^{e^*}[\theta \mid Y = y, \mathcal{S}] \\
&= \int dy\, h(y \mid e)\left(\sum_{j \in \mathcal{J}} \alpha_j + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y - e^*, \mathcal{S}]\right) \\
&= \int dy\, h(y - e \mid 0)\left(\sum_{j \in \mathcal{J}} \alpha_j + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y - e^*, \mathcal{S}]\right),
\end{aligned}$$

and so by making the variable substitution $y' = y - e$ we may write

$$\mathbb{E}^e\left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right] = \int dy'\, h(y' \mid 0)\left(\sum_{j \in \mathcal{J}} \alpha_j + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y' - e^* + e, \mathcal{S}]\right).$$

Differentiating wrt e and invoking Assumption 1 so that the dominated convergence theorem may be applied yields

$$\frac{\partial}{\partial e}\mathbb{E}^e\left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right]\Big|_{e=e^*} = \int dy'\, h(y' \mid 0)\, \frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}]\Big|_{Y^0 = y'}.$$

Recall that $Y \overset{d}{=} Y^0$ conditional on $e = 0$, so that $h(y' \mid 0)$ is the conditional density of Y^0 on $\mathcal{S}$. The rhs of the previous expression may therefore be written

$$\frac{\partial}{\partial e}\mathbb{E}^e\left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right]\Big|_{e=e^*} = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \;\Big|\; \mathcal{S}\right],$$

as desired. $\square$

B Proof of Theorem 1

B.1 Proof Idea

Fix a set of observed covariates $(\mathcal{J}, \mathcal{K})$ and a subpopulation $\mathcal{S} = (\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$. On $\mathcal{S}$, Y can be decomposed as

$$
\begin{aligned}
Y_{|\mathcal{S}} &\overset{d}{=} e + \mu_{\mathcal{S}} + \theta_{j'|\boldsymbol{a}_{\mathcal{J}}=\boldsymbol{\alpha}} + \theta^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}}=\boldsymbol{\alpha}} + \varepsilon^{-\mathcal{K}}_{|\boldsymbol{c}_{\mathcal{K}}=\boldsymbol{\gamma}} \\
&\overset{d}{=} e + \mu_{\mathcal{S}} + \theta_{j'|\boldsymbol{a}_{\mathcal{J}}=\boldsymbol{\alpha}} + \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}} = \boldsymbol{\alpha}, a_{j'}] + \theta^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}}=\boldsymbol{\alpha}} - \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}} = \boldsymbol{\alpha}, a_{j'}] + \varepsilon^{-\mathcal{K}}_{|\boldsymbol{c}_{\mathcal{K}}=\boldsymbol{\gamma}} \\
&\overset{d}{=} e + \mu_{\mathcal{S}} + \tilde{\theta}_{j'} + \tilde{\theta}^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}'}} + \varepsilon^{-\mathcal{K}}_{|\boldsymbol{c}_{\mathcal{K}}=\boldsymbol{\gamma}} \qquad \text{(B.1)}
\end{aligned}
$$

where $\mu_{\mathcal{S}} = \sum_{j\in\mathcal{J}} \alpha_j + \sum_{k\in\mathcal{K}} \gamma_k$ is a constant, and

$$
\begin{aligned}
\tilde{\theta}_{j'} &\equiv \theta_{j'|\boldsymbol{a}_{\mathcal{J}}=\boldsymbol{\alpha}} + \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}} = \boldsymbol{\alpha}, a_{j'}] \\
\tilde{\theta}^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}'}} &\equiv \theta^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}}=\boldsymbol{\alpha}} - \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}} = \boldsymbol{\alpha}, a_{j'}]
\end{aligned}
$$

are de-meaned type components, as defined in the main text.

When the attribute $a_{j'}$ is additionally observed, with realization $a_{j'} = \alpha'$, then the outcome signal can instead be decomposed as

$$
\begin{aligned}
Y_{|\mathcal{S}, a'_j=\alpha'} &\overset{d}{=} e + \mu_{\mathcal{S}} + \Psi^{j'}(\alpha') + \theta^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}'}=(\boldsymbol{\alpha},\alpha')} + \varepsilon^{-\mathcal{K}}_{|\boldsymbol{c}_{\mathcal{K}}=\boldsymbol{\gamma}} \\
&\overset{d}{=} e + \mu_{\mathcal{S}} + \Psi^{j'}(\alpha') + \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}} = \boldsymbol{\alpha}, a_{j'}] + \\
&\qquad\qquad \theta^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}'}=(\boldsymbol{\alpha},\alpha')} - \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}} = \boldsymbol{\alpha}, a_{j'}] + \varepsilon^{-\mathcal{K}}_{|\boldsymbol{c}_{\mathcal{K}}=\boldsymbol{\gamma}} \\
&\overset{d}{=} e + \mu_{\mathcal{S}} + \kappa + \tilde{\theta}^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}'}=(\boldsymbol{\alpha},\alpha')} + \varepsilon^{-\mathcal{K}}_{|\boldsymbol{c}_{\mathcal{K}}=\boldsymbol{\gamma}}
\end{aligned}
$$

where $\kappa = \Psi^{j'}(\alpha') + \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}} = \boldsymbol{\alpha}, a_{j'}]$ is a constant, and

$$
\tilde{\theta}^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}'}=(\boldsymbol{\alpha},\alpha')} \equiv \theta^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}'}=(\boldsymbol{\alpha},\alpha')} - \mathbb{E}[\theta^{-\mathcal{J}'} \mid \boldsymbol{a}_{\mathcal{J}} = \boldsymbol{\alpha}, a_{j'}]
$$

By assumption that j' is an $\mathcal{S}$-mean shifter, the variables $\tilde{\theta}^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}'}=(\boldsymbol{\alpha},\alpha')}$ and $\tilde{\theta}^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}}=\boldsymbol{\alpha}}$ are identical in distribution, and so

$$
Y_{|\mathcal{S}, a'_j=\alpha'} = e + \mu_{\mathcal{S}} + \kappa + \tilde{\theta}^{-\mathcal{J}'}_{|\boldsymbol{a}_{\mathcal{J}}=\boldsymbol{\alpha}} + \varepsilon^{-\mathcal{K}}_{|\boldsymbol{c}_{\mathcal{K}}=\boldsymbol{\gamma}} \qquad \text{(B.2)}
$$

Comparing (B.1) with (B.2), and observing that κ is a realization of $\tilde{\theta}_{j'}$, the consequence of revealing $a_{j'}$ is to reveal the variable $\tilde{\theta}_{j'}$ which, by assumption that j' is an $\mathcal{S}$-mean shifter, is independent from the other unknowns. The proof below makes this argument precise, and additionally verifies that the de-meaned variables $\tilde{\theta}_{j'}$ and $\tilde{\theta}^{-\mathcal{J}'}_{\boldsymbol{a}_{\mathcal{J}}=\boldsymbol{\alpha}}$ inherit the regularity properties imposed on the true effect sizes necessary for the independent-covariates result to hold.

B.2 Detailed Proof

Fix a set of observed covariates $(\mathcal{J}, \mathcal{K})$ and a $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S} = (\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$. As established in Lemma A.1, the marginal value of effort in subpopulation $\mathcal{S}$ is

$$MV(\mathcal{S}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \;\middle|\; \mathcal{S}\right],$$

where

$$Y^0 \equiv \mu + \sum_{j=1}^{J} \theta_j + \theta^{\perp} + \sum_{k=1}^{K} \varepsilon_k + \varepsilon^{\perp}$$

is the baseline value of output after subtracting out the agent's effort.

Now suppose the market additional observes the additional attribute $j' \notin \mathcal{J}$, and let $\mathcal{J}' = \mathcal{J} \cup \{j'\}$. Under the expanded set of observed covariates, the marginal value of effort becomes

$$MV(\mathcal{S}, a_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, a_{j'}, \mathcal{S}] \;\middle|\; a_{j'}, \mathcal{S}\right],$$

where, on $\mathcal{S}$, $MV(\mathcal{S}, a_{j'})$ is a random variable whose value is a function of the realization of $a_{j'}$.

Now, define functions $\xi(\beta)$ and $\zeta(\beta)$ by

$$\xi(\beta) \equiv \mathbb{E}[\theta^{-\mathcal{J}'} \mid a_{j'} = \beta, \mathcal{S}]$$

and

$$\zeta(\beta) = \Psi^j(\beta) + \xi(\beta).$$

Let $\widetilde{\theta}_{j'} \equiv \zeta(a_{j'})$, and let $\widetilde{\Theta} \equiv \{t \ : \ \zeta(\beta) = t \text{ for some } \beta \in A_{j'}\}$ denote the support of $\widetilde{\theta}_{j'}$. Note that on $\mathcal{S}$, Y^0 may be decomposed as

$$Y^0 = \mu_{\mathcal{S}} + \widetilde{\theta}_{j'} + \widetilde{\theta}^{-\mathcal{J}'} + \varepsilon^{-\mathcal{K}},$$

where $\mu_{\mathcal{S}} \equiv \sum_{j \in \mathcal{J}} a_j + \sum_{k \in \mathcal{K}} \gamma_k$. If j' is a $\mathcal{S}$-mean shifter, then $\widetilde{\theta}^{-\mathcal{J}'}$ is independent of $a_{j'}$ on $\mathcal{S}$, and so the distribution of Y^0 depends on $a_{j'}$ only through $\widetilde{\theta}_{j'}$. Thus

$$\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, a_{j'}, \mathcal{S}] = \xi(a_{j'}) + \mathbb{E}[\widetilde{\theta}^{-\mathcal{J}'} \mid Y^0, a_{j'}, \mathcal{S}] = \xi(a_{j'}) + \mathbb{E}[\widetilde{\theta}^{-\mathcal{J}'} \mid Y^0, \widetilde{\theta}_{j'}, \mathcal{S}].$$

Hence $MV(\mathcal{S}, a_{j'})$ may be written

$$MV(\mathcal{S}, a_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\widetilde{\theta}^{-\mathcal{J}'} \mid Y^0, \widetilde{\theta}_{j'}, \mathcal{S}] \;\middle|\; a_{j'}, \mathcal{S}\right].$$

Note that the random variable inside the outer expectation depends on $a_{j'}$ only through $\widetilde{\theta}_{j'}$. Thus the marginal value of effort after observing j' depends on the realization of $a_{j'}$ only through $\widetilde{\theta}_{j'}$, and so we may denote this marginal value of effort

$$MV(\mathcal{S}, \widetilde{\theta}_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\widetilde{\theta}^{-\mathcal{J}'} \mid Y^0, \widetilde{\theta}_{j'}, \mathcal{S}] \;\middle|\; \widetilde{\theta}_{j'}, \mathcal{S}\right].$$

$MV(\mathcal{S}, \widetilde{\theta}_{j'})$ can be compared to $MV(\mathcal{S})$ as follows. Define

$$\widetilde{F}(t \mid y) \equiv \Pr(\widetilde{\theta}_{j'} \leq t \mid Y^0 = y, \mathcal{S})$$

and

$$\widetilde{\phi}(y, t) \equiv \mathbb{E}[\widetilde{\theta}^{-\mathcal{J}'} \mid Y^0 = y, \widetilde{\theta}_{j'} = t, \mathcal{S}].$$

The random variable $\theta^{-\mathcal{J}}$ may be decomposed as

$$\theta^{-\mathcal{J}} = \widetilde{\theta}_{j'} + \widetilde{\theta}^{-\mathcal{J}'},$$

and so by the law of total probability

$$\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] = \int d\widetilde{F}(t \mid y)\left(t + \widetilde{\phi}(y, t)\right).$$

Lemma B.1. *If j' is an $\mathcal{S}$-mean shifter, then for every y, $\int d\widetilde{F}(t \mid y')\left(t + \widetilde{\phi}(y, t)\right)$ is increasing in y'.*

Proof. We first show that $t + \widetilde{\phi}(y, t)$ is increasing in t. This is equivalent to showing that $\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \widetilde{\theta}_{j'}, \mathcal{S}]$ is increasing in $\widetilde{\theta}_{j'}$ everywhere on $\mathcal{S}$. We prove the sufficient condition for this result that $(\theta^{-\mathcal{J}}, \widetilde{\theta}_{j'}, Y^0)$ are jointly affiliated on $\mathcal{S}$. Let $\tilde{f}(u, t, y)$ be the conditional joint density of $(\theta^{-\mathcal{J}}, \widetilde{\theta}_{j'}, Y^0)$ on $\mathcal{S}$. We will show that $\tilde{f}$ is log-supermodular.

Let $\tilde{f}_{j'}(t)$ be the conditional density of $\widetilde{\theta}_{j'}$ on $\mathcal{S}$, $\tilde{f}_{-\mathcal{J}|j'}(u \mid t)$ be the conditional density of $\theta^{-\mathcal{J}} \mid \widetilde{\theta}_{j'}$ on $\mathcal{S}$, and $h_{Y|-\mathcal{J}}(y \mid u)$ be the conditional density of $Y^0 \mid \theta^{-\mathcal{J}}$ on $\mathcal{S}$. Note that Y^0 is independent of $\widetilde{\theta}_{j'}$ conditional on $\theta^{-\mathcal{J}}$ on $\mathcal{S}$. So $\tilde{f}$ may be decomposed as

$$\tilde{f}(u, t, y) = \tilde{f}_{j'}(t)\tilde{f}_{-\mathcal{J}|j'}(u \mid t)h_{Y|-\mathcal{J}}(y \mid u).$$

It is therefore sufficient to show that $h_{Y|-\mathcal{J}}$ and $\tilde{f}_{-\mathcal{J}|j'}$ are log-supermodular.

First consider $h_{Y|-\mathcal{J}}$. Recall that on $\mathcal{S}$, Y^0 may be written

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \varepsilon^{-\mathcal{K}}.$$

Let $g_{-\mathcal{K}}(z)$ be the conditional density of $\varepsilon^{-\mathcal{K}}$ on $\mathcal{S}$. Then

$$h_{Y|-\mathcal{J}}(y \mid u) = g_{-\mathcal{K}}(y - \mu_{\mathcal{S}} - u).$$

As attribute j' is $\mathcal{S}$-regular, $g_{-\mathcal{K}}$ is log-concave, therefore $h_{Y|-\mathcal{J}}$ is log-supermodular.

As for $\tilde{f}_{-\mathcal{J}|j'}$, let $\tilde{f}_{-\mathcal{J}'}(w)$ be the conditional density of $\widetilde{\theta}^{-\mathcal{J}'}$ on $\mathcal{S}$. Decompose $\theta^{-\mathcal{J}}$ as $\theta^{-\mathcal{J}} = \widetilde{\theta}_{j'} + \widetilde{\theta}^{-\mathcal{J}'}$, and recall that if j' is an $\mathcal{S}$-mean shifter then $\widetilde{\theta}^{-\mathcal{J}'}$ is independent of $a_{j'}$ and hence $\widetilde{\theta}_{j'}$ on $\mathcal{S}$. It follows that

$$\tilde{f}_{-\mathcal{J}|j'}(u \mid t) = \tilde{f}_{-\mathcal{J}'}(u - t),$$

and hence

$$\frac{\partial^2}{\partial u \partial t} \log \tilde{f}_{-\mathcal{J}|j'}(u \mid t) = -\left.\frac{\partial^2}{\partial w^2} \log \tilde{f}_{-\mathcal{J}'}(w)\right|_{w=u-t} = -\frac{\partial^2}{\partial u^2} \log \tilde{f}_{-\mathcal{J}|j'}(u \mid t).$$

Now, let $f^0_{-\mathcal{J}|j'}(u \mid \beta)$ denote the conditional density of $\theta^{-\mathcal{J}} \mid a_{j'}$ on $\mathcal{S}$. Recall that

$$\theta^{-\mathcal{J}} = \zeta(a_{j'}) + \widetilde{\theta}^{-\mathcal{J}'},$$

and so if j' is an $\mathcal{S}$-mean shifter,

$$f^0_{-\mathcal{J}|j'}(u \mid \beta) = \tilde{f}_{-\mathcal{J}'}(u - \zeta(\beta)) = \tilde{f}_{-\mathcal{J}|j'}(u \mid \zeta(\beta)).$$

So fix any $t \in \widetilde{\Theta}$, and let $\beta \in A_{j'}$ be such that $\zeta(\beta) = t$. Then for all u,

$$\frac{\partial^2}{\partial u^2} \log \tilde{f}_{-\mathcal{J}|j'}(u \mid t) = \frac{\partial^2}{\partial u^2} \log f^0_{-\mathcal{J}|j'}(u \mid \beta).$$

As attribute j' is $\mathcal{S}$-regular, this final expression is non-positive, meaning

$$\frac{\partial^2}{\partial u \partial t} \log \tilde{f}_{-\mathcal{J}|j'}(u \mid t) \geq 0$$

for every u and $t \in \widetilde{\Theta}$. Hence $\tilde{f}_{-\mathcal{J}|j'}$ is log-supermodular, as desired.

We complete the proof by arguing that $\mathbb{E}[\widetilde{\theta}_{j'} + \widetilde{\phi}(y, \widetilde{\theta}_{j'}) \mid Y^0, \mathcal{S}]$ is increasing in Y^0, which is equivalent to the lemma statement. Since $(\theta^{-\mathcal{J}}, \widetilde{\theta}_{j'}, Y^0)$ are affiliated on $\mathcal{S}$, so are $(\widetilde{\theta}_{j'}, Y^0)$. Hence for any increasing function $\pi(t)$, $\mathbb{E}[\pi(\widetilde{\theta}_{j'}) \mid Y^0, \mathcal{S}]$ is increasing in Y^0. In particular, letting $\pi(t) = t + \widetilde{\phi}(y, t)$ yields the desired result. $\square$

In light of the previous lemma, for every y and $y' > y$,

$$\begin{aligned}
&\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y', \mathcal{S}] - \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] \\
= &\int d\widetilde{F}(t \mid y') \left(t + \widetilde{\phi}(y', t)\right) - \int d\widetilde{F}(t \mid y) \left(t + \widetilde{\phi}(y, t)\right) \\
\geq &\int d\widetilde{F}(t \mid y) \left(\widetilde{\phi}(y', t) - \widetilde{\phi}(y, t)\right).
\end{aligned}$$

Therefore

$$\frac{\partial}{\partial y}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] \geq \int d\widetilde{F}(t \mid y) \frac{\partial \widetilde{\phi}}{\partial y}(y, t),$$

which may be equivalently written

$$\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \geq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\widetilde{\theta}^{-\mathcal{J}'} \mid Y^0, \widetilde{\theta}_{j'}, \mathcal{S}] \;\middle|\; Y^0, \mathcal{S}\right].$$

Taking the expectation of each side conditional on $\mathcal{S}$ yields

$$MV(\mathcal{S}) \geq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\widetilde{\theta}^{-\mathcal{J}'} \mid Y^0, \widetilde{\theta}_{j'}, \mathcal{S}] \;\middle|\; \mathcal{S}\right].$$

By the law of iterated expectations, the rhs may be expanded as

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\widetilde{\theta}^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; \mathcal{S}\right] \\
= &\mathbb{E}\left[\mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\widetilde{\theta}^{-\mathcal{J}'} \mid Y^0, \widetilde{\theta}_{j'}, \mathcal{S}] \;\middle|\; \widetilde{\theta}_{j'}, \mathcal{S}\right] \;\middle|\; \mathcal{S}\right] \\
= &\mathbb{E}\left[MV(\mathcal{S}, \widetilde{\theta}_{j'}) \mid \mathcal{S}\right].
\end{aligned}$$

Therefore

$$MV(\mathcal{S}) \geq \mathbb{E}[MV(\mathcal{S}, \widetilde{\theta}_{j'}) \mid \mathcal{S}].$$

Thus the marginal value of effort in subpopulation $\mathcal{S}$ is higher than the expected marginal value once attribute j' is additionally observed.

Suppose instead that the market observes the additional circumstance $k' \notin \mathcal{K}$. Now under the expanded set of observed covariates, the marginal value of effort becomes

$$MV(\mathcal{S}, c_{k'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, c_{k'}, \mathcal{S}] \;\middle|\; c_{k'}, \mathcal{S}\right].$$

Let $\mathcal{K}' = \mathcal{K} \cup \{k'\}$, and define functions $\eta(\delta)$ and $\rho(\delta)$ by

$$\eta(\delta) \equiv \mathbb{E}[\varepsilon^{-\mathcal{K}'} \mid c_{k'} = \delta, \mathcal{S}]$$

and

$$\rho(\delta) = \Lambda^j(\delta) + \eta(\delta).$$

Let $\widetilde{\varepsilon}_{k'} \equiv \rho(c_{k'})$, and let $\widetilde{E} \equiv \{z \ : \ \rho(\delta) = z \text{ for some } \delta \in C_{k'}\}$ denote the support of $\widetilde{\varepsilon}_{k'}$. On $\mathcal{S}$, Y^0 may be decomposed as

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \widetilde{\varepsilon}_{k'} + \widetilde{\varepsilon}^{-\mathcal{K}'}.$$

If k' is a $\mathcal{S}$-mean shifter, then $\widetilde{\varepsilon}^{-\mathcal{K}'}$ is independent of $c_{k'}$ on $\mathcal{S}$, and so the distribution of Y^0 depends on $c_{k'}$ only through $\widetilde{\varepsilon}_{k'}$. Thus

$$\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, c_{k'}, \mathcal{S}] = \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \widetilde{\varepsilon}_{k'}, \mathcal{S}].$$

Therefore, in a manner analogous to the attribute case, the marginal value of effort after observing j' depends on $c_{k'}$ only through $\widetilde{\varepsilon}_{k'}$ and may be written

$$MV(\mathcal{S}, \widetilde{\varepsilon}_{k'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \widetilde{\varepsilon}_{k'}, \mathcal{S}] \;\middle|\; \widetilde{\varepsilon}_{k'}, \mathcal{S}\right].$$

We compare $MV(\mathcal{S}, \widetilde{\varepsilon}_{k'})$ and $MV(\mathcal{S})$ in a manner very similar to the attribute case. Define

$$\widetilde{G}(z \mid y) \equiv \Pr(\widetilde{\varepsilon}_{k'} \le z \mid Y^0 = y, \mathcal{S})$$

and

$$\widetilde{\psi}(y, t) \equiv \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \widetilde{\varepsilon}_{k'} = z, \mathcal{S}].$$

Then we may write

$$\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] = \int d\widetilde{G}(z \mid y)\widetilde{\psi}(y, z)$$

and invoke the following lemma:

Lemma B.2. *If circumstance k' is an $\mathcal{S}$-mean shifter, then for every y, $\int d\widetilde{G}(z \mid y')\widetilde{\psi}(y, z)$ is decreasing in y'.*

Proof. On $\mathcal{S}$, Y^0 may be written

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \varepsilon^{-\mathcal{K}}.$$

Taking expectations of each side conditional on $(Y^0, \widetilde{\varepsilon}_{k'}, \mathcal{S})$ yields

$$Y^0 = \mu_{\mathcal{S}} + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \widetilde{\varepsilon}_{k'}, \mathcal{S}] + \mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0, \widetilde{\varepsilon}_{k'}, \mathcal{S}].$$

Hence

$$\int d\widetilde{G}(z \mid y')\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \widetilde{\varepsilon}_{k'} = z, \mathcal{S}]$$
$$= y - \mu_{\mathcal{S}} - \int d\widetilde{G}(z \mid y')\mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0 = y, \widetilde{\varepsilon}_{k'} = z, \mathcal{S}].$$

The desired decreasing monotonicity is therefore equivalent to increasing monotonicity of

$$\int dG(z \mid y')\mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0 = y, \varepsilon_{k'} = z, \mathcal{S}],$$

which may be established along nearly identical lines to the proof of Lemma B.1. Let $f_{-\mathcal{J}}(u)$ be the conditional density of $\theta^{-\mathcal{J}}$ on $\mathcal{S}$ and $g^0_{-\mathcal{K}'|k'}(x \mid z)$ be the conditional density of $\varepsilon^{-\mathcal{K}'} \mid c_{k'}$ on $\mathcal{S}$. The conditions required for the steps of that proof to go through are that $f_{-\mathcal{J}}(u)$ is log-concave, $g^0_{-\mathcal{K}'|k'}(x \mid z)$ is log-concave in x for all z, and $\widetilde{\varepsilon}^{-\mathcal{K}'}$ is independent of $\widetilde{\varepsilon}_{k'}$ on $\mathcal{S}$. The first two properties hold by $\mathcal{S}$-regularity of circumstance k', while the final property holds because k' is an $\mathcal{S}$-mean shifter. □

Hence

$$\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \leq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \widetilde{\varepsilon}_{k'}, \mathcal{S}] \;\middle|\; Y^0, \mathcal{S}\right],$$

implying

$$MV(\mathcal{S}) \leq \mathbb{E}[MV(\mathcal{S}, \widetilde{\varepsilon}_{k'}) \mid \mathcal{S}].$$

Thus the marginal value of effort in subpopulation $\mathcal{S}$ is lower than the expected marginal value of effort when the circumstance k' is additionally observed.

The final step in the proof is establishing that monotonicity holds uniformly across realizations of the additional covariate, and not just on average. This follows immediately once we establish that $MV(\mathcal{S}, a_{j'})$ and $MV(\mathcal{S}, c_{k'})$ are independent of the realizations of $a_{j'}$ and $c_{k'}$. We prove the result for the attribute case, with the circumstance case following from nearly identical work. On $\mathcal{S}$, when attribute j' is additionally observed and is found to have value $a_{j'} = \alpha_{j'}$, Y may be decomposed as

$$Y = e + \mu_{\mathcal{S}} + \Psi^{j'}(\alpha_{j'}) + \xi(\alpha_{j'}) + \widetilde{\theta}^{-\mathcal{J}'} + \varepsilon^{-\mathcal{K}}.$$

Because $\widetilde{\theta}^{-\mathcal{J}'}$ is independent of $a_{j'}$ on $\mathcal{S}$, $\alpha_{j'}$ enters the market's inference problem as a known additive shift to the agent's type distribution, and therefore its value does not impact incentives for effort. So incentives for effort must be independent of $\alpha_{j'}$, as claimed.

C Proof of Theorem 2

Proof. Fix a set of observed covariates $(\mathcal{J}, \mathcal{K})$ and a $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S} = (\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$. As established in Lemma A.1, the marginal value of effort in subpopulation $\mathcal{S}$ is

$$MV(\mathcal{S}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \;\middle|\; \mathcal{S}\right],$$

where

$$Y^0 \equiv \mu + \sum_{j=1}^{J} \theta_j + \theta^{\perp} + \sum_{k=1}^{K} \varepsilon_k + \varepsilon^{\perp}$$

is the baseline value of output after subtracting out the agent's effort.

Now suppose the market additional observes the additional attribute $j' \notin \mathcal{J}$, and let $\mathcal{J}' = \mathcal{J} \cup \{j'\}$. Under the expanded set of observed covariates, the marginal value of effort becomes

$$MV(\mathcal{S}, a_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, a_{j'}, \mathcal{S}] \;\middle|\; a_{j'}, \mathcal{S}\right],$$

where, on $\mathcal{S}$, $MV(\mathcal{S}, a_{j'})$ is a random variable whose value is a function of the realization of $a_{j'}$. Because $\Psi^{j'}$ is a one-to-one mapping, $\theta_{j'}$ is a sufficient statistic for the dependence of the distribution of $\theta^{-\mathcal{J}'}$ on $a_{j'}$. And by construction Y^0 also depends on $a_{j'}$ only through $\theta_{j'}$. So we may equivalently write the agent's marginal value of effort under the expanded set of covariates as

$$MV(\mathcal{S}, \theta_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; \theta_{j'}, \mathcal{S}\right],$$

where this expression depends on $a_{j'}$ via the effect size $\theta_{j'}$.

$MV(\mathcal{S}, \theta_{j'})$ can be compared to $MV(\mathcal{S})$ as follows. Define

$$F(t \mid y) \equiv \Pr(\theta_{j'} \leq t \mid Y^0 = y, \mathcal{S})$$

and

$$\phi(y, t) \equiv \mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0 = y, \theta_{j'} = t, \mathcal{S}].$$

By the law of total probability

$$\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] = \int dF(t \mid y)\,(t + \phi(y, t))\,.$$

Lemma C.1. *If j' is $\mathcal{S}$-affiliated, then for every y, $\int dF(t \mid y')\,(t + \phi(y, t))$ is increasing in y'.*

Proof. We first show that $t + \phi(y, t)$ is increasing in t. This is equivalent to showing that $\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \theta_{j'}, \mathcal{S}]$ is increasing in $\theta_{j'}$ everywhere on $\mathcal{S}$. We prove the sufficient condition for this result that $(\theta^{-\mathcal{J}}, \theta_{j'}, Y^0)$ are jointly affiliated on $\mathcal{S}$. Let $f(u, t, y)$ be the conditional joint density of $(\theta^{-\mathcal{J}}, \theta_{j'}, Y^0)$ on $\mathcal{S}$. We will show that f is log-supermodular.

Let $f_{j'}(t)$ be the conditional density of $\theta_{j'}$ on $\mathcal{S}$, $f_{-\mathcal{J}|j'}(u \mid t)$ be the conditional density of $\theta^{-\mathcal{J}} \mid \theta_{j'}$ on $\mathcal{S}$, and $h_{Y|-\mathcal{J}}(y \mid u)$ be the conditional density of $Y^0 \mid \theta^{-\mathcal{J}}$ on $\mathcal{S}$. Note that Y^0 is independent of $\theta_{j'}$ conditional on $\theta^{-\mathcal{J}}$ on $\mathcal{S}$, and so

$$f(u, t, y) = f_{j'}(t) f_{-\mathcal{J}|j'}(u \mid t) h_{Y|-\mathcal{J}}(y \mid u).$$

It is therefore sufficient to show that $h_{Y|-\mathcal{J}}$ and $f_{-\mathcal{J}|j'}$ are log-supermodular.

Consider $h_{Y|-\mathcal{J}}$. Let $\mu_{\mathcal{S}} \equiv \sum_{j \in \mathcal{J}} \alpha_j + \sum_{k \in \mathcal{K}} \gamma_k + \theta^{-\mathcal{J}}$. On $\mathcal{S}$, Y^0 may be written

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \varepsilon^{-\mathcal{K}}.$$

Let $g_{-\mathcal{K}}(z)$ be the conditional density of $\varepsilon^{-\mathcal{K}}$ on $\mathcal{S}$. Then

$$h_{Y|-\mathcal{J}}(y \mid u) = g_{-\mathcal{K}}(y - \mu_{\mathcal{S}} - u).$$

As attribute j' is $\mathcal{S}$-regular, $g_{-\mathcal{K}}$ is log-concave, therefore $h_{Y|-\mathcal{J}}$ is log-supermodular.

As for $f_{-\mathcal{J}|j'}$, let $f_{-\mathcal{J}'|j'}(w \mid t)$ be the conditional density of $\theta^{-\mathcal{J}'} \mid \theta_{j'}$ on $\mathcal{S}$. As $\theta^{-\mathcal{J}} = \theta_{j'} + \theta^{-\mathcal{J}'}$, it follows that

$$f_{-\mathcal{J}|j'}(u \mid t) = f_{-\mathcal{J}'|j'}(u - t \mid t).$$

Hence by the chain rule,

$$\frac{\partial^2}{\partial u \partial t} \log f_{-\mathcal{J}|j'}(u \mid t) = \left[\frac{\partial^2}{\partial w \partial t} \log f_{-\mathcal{J}'|j'}(w \mid t) - \frac{\partial^2}{\partial w^2} \log f_{-\mathcal{J}'|j'}(w \mid t) \right]_{w=u-t}.$$

As attribute j' is $\mathcal{S}$-regular, the second term is non-negative. Meanwhile since j' is $\mathcal{S}$-affiliated, $(\theta^{-\mathcal{J}'}, \theta_{j'})$ are affiliated on $\mathcal{S}$ and so the first term is also non-negative. Hence

$$\frac{\partial^2}{\partial u \partial t} \log f_{-\mathcal{J}|j'}(u \mid t) \geq 0,$$

establishing the desired log-supermodularity.

To complete the proof we argue that $\mathbb{E}[\theta_{j'} + \phi(y, \theta_{j'}) \mid Y^0, \mathcal{S}]$ is increasing in Y^0, which is equivalent to the lemma statement. Since $(\theta^{-\mathcal{J}}, \theta_{j'}, Y^0)$ are affiliated on $\mathcal{S}$, so are $(\theta_{j'}, Y^0)$. Hence for any increasing function $\pi(t)$, $\mathbb{E}[\pi(\theta_{j'}) \mid Y^0, \mathcal{S}]$ is increasing in Y^0. In particular, letting $\pi(t) = t + \phi(y, t)$ yields the desired result. □

In light of the previous lemma,

$$\frac{\partial}{\partial y}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] \geq \int dF(t \mid y)\frac{\partial \phi}{\partial y}(t, y),$$

which may be equivalently written

$$\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \geq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; Y^0, \mathcal{S}\right].$$

Taking the expectation of each side conditional on $\mathcal{S}$ yields

$$MV(\mathcal{S}) \geq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; \mathcal{S}\right].$$

By the law of iterated expectations, the rhs may be expanded as

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; \mathcal{S}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; \theta_{j'}, \mathcal{S}\right] \;\middle|\; \mathcal{S}\right] \\
&= \mathbb{E}\left[MV(\mathcal{S}, \theta_{j'}) \mid \mathcal{S}\right].
\end{aligned}$$

Therefore

$$MV(\mathcal{S}) \geq \mathbb{E}[MV(\mathcal{S}, \theta_{j'}) \mid \mathcal{S}].$$

Thus the marginal value of effort in subpopulation $\mathcal{S}$ is higher than the expected marginal value once attribute j' is additionally observed.

Suppose instead that the market observes the additional circumstance $k' \notin \mathcal{K}$. Under the expanded set of observed covariates, the marginal value of effort becomes

$$MV(\mathcal{S}, c_{k'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, c_{k'}, \mathcal{S}] \;\middle|\; c_{k'}, \mathcal{S}\right].$$

Because $\Lambda^{k'}$ is a one-to-one-mapping, $\varepsilon_{k'}$ is a sufficient statistic for the dependence of the distribution of $\varepsilon^{-\mathcal{K}'}$ on $c_{k'}$. We may therefore equivalently write the agent's marginal value of effort under the expanded set of covariates as

$$MV(\mathcal{S}, \varepsilon_{k'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \varepsilon_{k'}, \mathcal{S}] \;\middle|\; \varepsilon_{k'}, \mathcal{S}\right].$$

We compare $MV(\mathcal{S}, \varepsilon_{k'})$ with $MV(\mathcal{S})$ in a manner very similar to the case of an additional attribute. Let $\mathcal{K}' = \mathcal{K} \cup \{k'\}$, and define

$$G(z \mid y) \equiv \Pr(\varepsilon_{k'} \leq z \mid Y^0 = y, \mathcal{S})$$

and

$$\psi(y,z) \equiv \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \varepsilon_{k'} = z, \mathcal{S}].$$

Then we may write

$$\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] = \int dG(z \mid y)\psi(y,z)$$

and invoke the following lemma:

Lemma C.2. *If k' is $\mathcal{S}$-affiliated, then for every y, $\int dG(z \mid y')\psi(y,z)$ is decreasing in y'.*

Proof. On $\mathcal{S}$, Y^0 may be written

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \varepsilon^{-\mathcal{K}}.$$

Taking expectations of each side conditional on $(Y^0, \varepsilon_{k'}, \mathcal{S})$ yields

$$Y^0 = \mu_{\mathcal{S}} + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \varepsilon_{k'}, \mathcal{S}] + \mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0, \varepsilon_{k'}, \mathcal{S}].$$

Hence

$$\begin{aligned} &\int dG(z \mid y')\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \varepsilon_{k'} = z, \mathcal{S}] \\ &= y - \mu_{\mathcal{S}} - \int dG(z \mid y')\mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0 = y, \varepsilon_{k'} = z, \mathcal{S}]. \end{aligned}$$

The desired decreasing monotonicity is therefore equivalent to increasing monotonicity of

$$\int dG(z \mid y')\mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0 = y, \varepsilon_{k'} = z, \mathcal{S}],$$

which may be established along nearly identical lines to the proof of Lemma C.1. Let $f_{-\mathcal{J}}(u)$ be the conditional density of $\theta^{-\mathcal{J}}$ on $\mathcal{S}$ and $g_{-\mathcal{K}'|k'}(x \mid z)$ be the conditional density of $\varepsilon^{-\mathcal{K}'} \mid \varepsilon_{k'}$ on $\mathcal{S}$. The conditions required for the steps of that proof to go through are that $f_{-\mathcal{J}}(u)$ is log-concave, $g_{-\mathcal{K}'|k'}(x \mid z)$ is log-concave in x for all z, and $(\varepsilon^{-\mathcal{K}'}, \varepsilon_{k'})$ are affiliated on $\mathcal{S}$. The first two properties hold by $\mathcal{S}$-regularity of circumstance k', while the final property holds by $\mathcal{S}$-affiliation of circumstance k'. $\square$

Hence

$$\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \leq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \varepsilon_{k'}, \mathcal{S}] \;\middle|\; Y^0, \mathcal{S}\right],$$

implying

$$MV(\mathcal{S}) \leq \mathbb{E}[MV(\mathcal{S}, \varepsilon_{k'}) \mid \mathcal{S}].$$

Thus the marginal value of effort in subpopulation $\mathcal{S}$ is lower than the expected marginal value of effort when the circumstance k' is additionally observed. $\square$

D Additional Material

D.1 Supplementary Material to Section 4.1

We show that the examples given in Section 4.1 satisfy regularity. Throughout, let $Z = \Psi(a)$ for notational ease, and denote its distribution by p_Z.

Example 1. Suppose $a \sim \mathcal{N}(\mu, \sigma^2)$. Then $Z = \Psi(a) \sim \mathcal{N}(\mu, \sigma^2)$, and normal distributions are log-concave for all parameter values.

Example 2. $a \sim U([c, d])$ has a log-concave density, and log-concavity of random variables is preserved under affine transformations.

Example 3. The density of Z is

$$\begin{aligned} p_Z(z) &= p_A(\Psi^{-1}(z)) \cdot \left| \frac{d}{dz} \Psi^{-1}(z) \right| \\ &= \lambda e^{-\lambda z^2} \cdot 2z \end{aligned}$$

using in the second equality that Z is supported on $\mathbb{R}_+$. Since $\log(\lambda e^{-\lambda z^2}) = \log(\lambda) - \lambda z^2$ is concave, and $\log(2z)$ is concave, the density $p_Z(z)$ is the product of two log-concave functions, and is hence itself log-concave.

Example 4. a has a log-concave distribution and is supported on the positive reals. Log-concavity of $\log(a)$ follows from Theorem 5 in Borzadaran and Borzadaran (2011).

Example 5. The density of Z is $p_Z(z) = p_A(\Psi^{-1}(z)) \cdot \left|\frac{d}{dz}\Psi^{-1}(z)\right| = \left(\lambda e^{-\lambda e^{-z}}\right) \cdot (e^{-z})$. Since $\log(\lambda e^{-\lambda e^{-z}}) = \log(\lambda) - \lambda e^{-z}$ is concave, and $\log(e^{-z}) = -z$ is weakly concave, the density $p_Z(z)$ is the product of two log-concave functions, and is itself log-concave.

D.2 Supplementary Material to Example 6

We show below that when all covariates are jointly normal, then mean-shifter property is satisfied globally. First consider the two-attribute model

$$Y = e + \theta_1 + \theta_2 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

with

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$$

Using standard formulas for Bayesian updating to Gaussian signals, the conditional distribution of θ_1 given θ_2 is

$$\theta_1 \mid \theta_2 \sim \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(\theta_2 - \mu_2) + \mathcal{N}\left(0, \sigma_1^2(1-\rho^2)\right), \tag{D.1}$$

which depends on θ_2 in mean only. Since the family of normal variables is closed under conditioning and summation, the above model is without loss: that is, for any $\mathcal{J}$ and $j' \notin \mathcal{J}$, we may set θ_1 equal to $\theta^{-\mathcal{J}'}$ and θ_2 equal to $\theta_{j'}$, so that (D.1) implies that attribute j' is an $\mathcal{S}$-mean shifter. (The argument applies identically for circumstance variables.)

D.3 Supplementary Material to Example 8

A standard result about sums of iid exponential random variables is that they are Gamma distributed. In particular, if there are J total attributes, then $\theta^{-\mathcal{J}'}|\lambda \sim \text{Gamma}(J - |\mathcal{J}'|, \lambda)$, and $\theta^{-\mathcal{J}'}$ and $\theta_{j'}$ are independent conditional on λ.

Define $k \equiv \lambda^{-1}$. The joint density of $(\theta_{j'}, \theta^{-\mathcal{J}'})$ conditional on $\boldsymbol{\theta}_\mathcal{J}$ may be written

$$\eta(\theta_{j'}, \theta^{-\mathcal{J}'} \mid \boldsymbol{\theta}_\mathcal{J}) = \int dk\, \eta(k \mid \boldsymbol{\theta}_\mathcal{J})\eta(\theta_{j'} \mid k)\eta(\theta^{-\mathcal{J}'} \mid k).$$

As marginalization preserves affiliation, $(\theta_{j'}, \theta^{-\mathcal{J}'})$ are affiliated conditional on $\boldsymbol{\theta}_\mathcal{J}$ if $(\theta_{j'}, k)$ and $(\theta^{-\mathcal{J}'}, k)$ are each affiliated. Note that

$$\log \eta(\theta_{j'} \mid k) = -\log k - \frac{\theta_{j'}}{k},$$

so that

$$\frac{\partial^2}{\partial\theta_{j'}\partial k} \log \eta(\theta_{j'} \mid k) = \frac{1}{k^2} > 0,$$

while

$$\log \eta(\theta^{-\mathcal{J}'} \mid k) = -N \log k - \log \Gamma(N) + (N-1) \log \theta^{-\mathcal{J}'} - \frac{\theta^{-\mathcal{J}'}}{k}$$

for $N = J - |\mathcal{J}'|$, so that similarly

$$\frac{\partial^2}{\partial \theta^{-\mathcal{J}'} \partial k} \log \eta(\theta^{-\mathcal{J}'} \mid k) = \frac{1}{k^2} > 0.$$

Hence $(\theta_{j'}, k)$ and $(\theta^{-\mathcal{J}'}, k)$ are affiliated, as desired.

Meanwhile, the conditional density of $\theta^{-\mathcal{J}'}$ may be written

$$\eta(\theta^{-\mathcal{J}'} \mid \boldsymbol{\theta}_{\mathcal{J}'}) = \int d\lambda\, \eta(\lambda \mid \boldsymbol{\theta}_{\mathcal{J}'}) \eta(\theta^{-\mathcal{J}'} \mid \lambda).$$

Log-concavity is also preserved by marginalization, so $\eta(\theta^{-\mathcal{J}'} \mid \boldsymbol{\theta}_{\mathcal{J}'})$ is log-concave wrt $\theta^{-\mathcal{J}'}$ if $\eta(\theta^{-\mathcal{J}'} \mid \lambda)$ is log-concave wrt $(\theta^{-\mathcal{J}'}, \lambda)$ and $\eta(\lambda \mid \theta_{\mathcal{J}'})$ is log-concave wrt λ. We have

$$\log \eta(\theta^{-\mathcal{J}'} \mid \lambda) = N \log \lambda - \log \Gamma(N) + (N-1) \log \theta^{-\mathcal{J}'} - \lambda \theta^{-\mathcal{J}'},$$

which is a sum of concave functions of $(\theta^{-\mathcal{J}'}, \lambda)$, hence itself concave. Meanwhile, note that the Gamma function is a conjugate prior for the exponential likelihood function, and so conditional on $\boldsymbol{\theta}_{\mathcal{J}'}$, $\lambda \sim \text{Gamma}\left(\alpha + |\mathcal{J}'|, \beta + \sum_{j \in \mathcal{J}'} \frac{1}{\theta_j}\right)$. The Gamma distribution is log-concave whenever its shape parameter is at least 1, and as $\alpha \geq 1$ it follows that $\eta(\lambda \mid \theta_{\mathcal{J}'})$ is log-concave wrt λ.

This work establishes that this system of attributes satisfies the conditions of Theorem 2 for the attribute case whenever the conditional distribution of $\varepsilon^{-\mathcal{K}}$ is log-concave.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2019): "Too Much Data: Prices and Inefficiencies in Data Markets," Working Paper.

Ball, I. (2020): "Scoring Strategic Agents," Working Paper.

Bergemann, D., A. Bonatti, and T. Gan (2020): "The Economics of Social Data," Working Paper.

Borzadaran, G. R. M. and H. A. M. Borzadaran (2011): "Log-Concavity Property for Some Well-Known Distributions," *Surveys in Mathematics and its Applications*, 6, 203–219.

Dewatripont, M., I. Jewitt, and J. Tirole (1999): "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, 66, 183–198.

Eliaz, K. and R. Spiegler (2019): "The Model Selection Curse," *American Economic Review: Insights*, 1, 127–140.

——— (2020): "On Incentive-Compatible Estimators," Working Paper.

Frankel, A. and N. Kartik (2019): "Muddled Information," *Journal of Political Economy*, 127, 1739–1776.

——— (2020): "Improving Information from Manipulable Data," Working Paper.

Haghtalab, N., N. Immorlica, B. Lucier, and J. Wang (2020): "Maximizing Welfare with Incentive-Aware Evaluation Mechanisms," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 160–166.

Hennessy, C. A. and C. A. E. Goodhart (2020): "Goodhart's Law and Machine Learning," Working Paper.

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

Hu, L., N. Immorlica, and J. W. Vaughan (2019): "The Disparate Effects of Strategic Manipulation," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 259–268.

Jin, Y. and S. Vasserman (2020): "Buying Data from Consumers: The Impact of Monitoring Programs in U.S. Auto Insurance," Working Paper.

Meyer, M. A. and J. Vickers (1997): "Performance Comparisons and Dynamic Incentives," *Journal of Political Economy*, 105, 547–581.

Rodina, D. (2018): "Information Design and Career Concerns," Working Paper.

Tirole, J. (2020): "Digital Dystopia," Working Paper.