The Ronald O. Perelman Center for Political Science and Economics (PCPSE)
133 South 36th Street
Philadelphia, PA 19104-6297

pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper
19-006

Selling Data

CARLOS SEGURA-RODRIGUEZ
University of Pennsylvania

April 21, 2019

Selling Data*

Carlos Segura-Rodriguez†

April 21, 2019

Abstract

I study how a monopolist data broker (seller), who wants to maximize profits, should present and sell consumer data to a firm (buyer). The buyer has an interest in forecasting a particular consumer characteristic, but the seller is uncertain about which characteristic the buyer wants to forecast and how much the buyer values information. I assume that the joint distribution of both the unknown characteristics and the data is elliptical. This information environment reduces to a multidimensional, multi-product mechanism design problem in which the buyer's payoffs are nonlinear. Hence, I cannot use the common differential approach to solve for the optimal mechanism. I obtain two main results. First, I show that the seller should optimally offer statistics that are linear combinations of the data and independent noise. Second, by using a direct approach, I show that in the optimal mechanism the seller might want to offer a continuum of different statistics, and these statistics, without containing independent noise, are less correlated than they would be if the seller could perfectly price discriminate. Thus this distortion affects the mimicking type more than the mimicked type.

Keywords: Information Design, Mechanism Design, Multidimensional Screening, Product Design, Elliptical Distribution.

JEL Classification: D42, D82, D83, D86.

---

*I would like to express my gratitude to George Mailath, Annie Liang, Rakesh Vohra, Andrew Postlewaite, and Aislinn Bohren for their insightful comments and support throughout this project. I also thank Dirk Bergemann and Nima Haghpanah for their valuable input. Thanks as well to my fellow classmates, especially Ashwin Kambhampati, Youngsoo Heo, Joao Granja, and Paolo Martellini for listening and providing help during the project. I also gratefully acknowledge all the participants in the UPenn Micro Theory Lunch and UPenn Micro Theory Seminar for their feedback. All errors are my own.

†University of Pennsylvania, Department of Economics, The Ronald O. Perelman Center, 133 South 36th Street; Philadelphia, PA 19104, cseg@sas.upenn.edu.

Contents

1 Introduction

Consumer data is valuable to economic agents that want to forecast an unobservable consumer characteristic. Consider a firm that plans to introduce a new product to the market. The firm can either send advertisements to all consumers or target those who are more likely to buy the new product. A targeted marketing campaign is effective only if the firm can identify accurately which consumers will react positively to its marketing efforts. As another example, consider a politician who wants to win an election. Sending personalized messages to undecided voters, in which the candidate explains how his/her promises can fulfill the voters' expectations, might be effective. To implement this strategy, the politician needs to identify the undecided voters and their expectations.

Due to the ease of gathering information from electronic commerce and online surfing, a new industry that specializes in collecting consumer data has emerged. This is a billion dollar and growing business, controlled by a small number of large data brokers.[1] These data brokers have collected a large amount of information about millions of consumers, and according to the Federal Trade Commission (2014), most of these brokers do not directly sell the data they have collected; instead, they either sell internally produced analyses or provide buyers with consumer "scores." These scores can refer, for example, to how likely a consumer is to make a purchase.

I study how monopolist data brokers (sellers) with access to consumer data can maximize profits by presenting and selling that data in an environment in which they lack information about buyers' motives to buy the data. The key novel issue in this environment is that different presentations of the data, which I called statistics, act as imperfect substitutes. Consider the problem of selling data to a coffee shop and a restaurant, each of which wants to advertise only to those consumers with high willingness to pay for a coffee or a lunch, respectively. Some observable characteristics might be useful for forecasting unknown variables that are relevant to both businesses. For instance, a consumer with a higher income might be willing to pay more for both coffee and a lunch. In that case, the forecast of a consumer's willingness to pay for a lunch might help

[1] According to Marr (2017), the main data brokers are Acxiom, Nielsen, Experian, Equifax and Corelogic.

the coffee shop owner to make an informed decision by using that same information to draw inferences about the consumer's willingness to pay for a coffee. I study how sellers (data brokers) should optimally produce adequate statistics for each product and charge prices for each statistic such that buyers purchase the statistics targeted to the products they want to advertise.

In my model, a random vector θ represents the set of consumer characteristics the buyer might want to forecast. The buyers are different in two dimensions: the unknown consumer characteristic they want to forecast and the value that they place on information. Each buyer wants to forecast only one of the components of θ and faces a quadratic loss function that is weighted by how valuable this information is for her. There is a seller, who does not know the buyer type, but who can access data about some other observable consumer characteristics. I assume that the seller can commit to a mechanism before observing the realization of the data.[2] While the data is observed only by the seller, the seller and the buyer share a common prior about the joint distribution of all consumer characteristics, observable or not. I assume that this joint distribution is elliptical, following the usual practice in applied work. I show that the seller should optimally offer only mechanisms in which the statistics are linear combinations of the data and independent noise. The argument, which I believe is novel in economics, depends crucially on the existence of conflict between the seller and the buyers in the designing environment.[3]

My model leads to a mechanism design problem in which the buyer's preferences are multidimensional, the seller needs to design multiple products, and the ranking of the buyer's payoffs is nonlinear. In the literature there is no general solution for this class of problem. I study a novel example where all of the properties just indicated are inherent to the economic environment, and I characterize the main properties satisfied by the optimal mechanism.

To better understand the problem, in Section 4.1 I characterize the optimal mecha-

[2]The commitment assumption is satisfied if, for example, the seller can publish menus only in discrete periods of time, but new consumer data arrives continuously.

[3]As some examples in the literature show (see, for example, Witsenhausen (1968)), using only linear statistics is not always optimal in the absence of conflict between players, even when the joint distribution of all variables is normal.

nism for when there are only two information types who share a common valuation type. The main property of the optimal mechanism is that the offered statistics are less correlated than they would be if the seller could perfectly price discriminate. Without introducing independent noise, the seller degrades the statistic targeted to the mimicked type by relatively decreasing the weights assigned to the variables that are more informative for the mimicking type, affecting the mimicking type more than the mimicked type. Although these coefficients are reduced, they are generically different from zero; that is, the seller does not create differentiation by selling a set of variables to one type and a different set of variables to another type. Furthermore, the mimicked type always receives a statistic that is informative for her.[4,5]

In section 4.2, I consider the more general problem in which there are two information types, and for each of them there is a continuum of valuation types. When fixing an information type, the problem that the seller faces is the problem studied by Myerson (1981). But the difficulty in my environment comes from the extra IC constraints across information types. Since the buyer's payoffs are nonlinear, I cannot use Myerson's differential approach to deal with them, and instead I use a direct method to find the optimal mechanism.

The solution to the relaxed problem without the IC constraints across-information types is simple. It consists of two take-it-or-leave-it offers for a statistic that for each information type is equivalent to receiving all the data. In the general problem, the solution is more involved. The seller offers to the type paying the highest price in the relaxed problem two statistics: a statistic that is equivalent to receiving all data targeted to high valuation types and another less informative statistic targeted to an intermediate range of valuation types. The novelty is that this less informative statistic can be produced by either introducing independent noise or by reducing the coefficients in the variables that are more informative for the other information type. To the information type paying the lowest price in the relaxed problem, the seller offers either a unique statistic or a

[4] This contrasts with the results of the literature in quality degradation, where it is always possible to find parameters such that a type is excluded from the mechanism. The reason for this is that in my environment the ranking of payoffs is product dependent.

[5] I use masculine pronouns for the seller (data broker) and feminine pronouns for the buyer (firm).

continuum of degraded statistics. In both cases, the statistics are degraded by reducing the coefficients in the variables that are more informative for the first information type.

When considering many information types, I cannot solve completely for the optimal mechanism since I cannot identify ex-ante which IC constraints are relevant. There are two reasons for this. First, reducing the price of the statistic targeted to one information type might provoke an incentive problem in which another type will want to mimic the first one. Second, changing the coefficients of the statistic targeted to an information type might make the modified statistic too informative for another type. I provide examples that show these complications and how they affect the optimal mechanism.

In spite of these issues, I partially characterize the optimal mechanism and show that the properties that I have highlighted are still satisfied with a caveat: even when there is only one valuation type, there might be some types that are excluded from the mechanism and do not receive any useful information. Furthermore, I show that at least one type will receive a statistic that for her is equivalent to receiving all of the data. But this type does not necessarily coincide with the type with the highest willingness to pay for all the data, as would be the case if there were only two information types. This generalizes the nondistortion at the top condition that appears in environments where the payoffs are linear.[6]

1.1 Literature Review

My paper contributes to a growing body of literature that studies the optimal ways in which a monopolist can sell information. The work that comes closest to my approach is Bergemann, Bonatti, and Smolin (2018). They study an environment in which the seller knows perfectly the realization of a random variable that can take a finite number of values, and in which all buyer types want to match the realization of the same unknown random variable. The types differ in their prior, so that each type willingness to pay for a signal depends on how informative the signal is for different realizations of the state. In contrast, I consider the problem in which the monopolist has multidimensional

[6]For the unidimensional case, see for example, Mussa and Rosen (1978), while Rochet and Choné (1998) shows that a similar condition is satisfied in the multidimensional case.

data that is informative, though not necessarily perfectly revealing, about a random vector. In my model, the types share the same prior, but each type wants to reduce the variance associated with only one of the random variables, and therefore they are interested in different features of the data. This new approach addresses some novel questions. I am able to study the optimal way to degrade the quality of the information in each of the dimensions of the data and consider whether or not the monopolist is better off giving only a subset of the data to certain buyers. To answer these questions, I analyze a simplified environment. Instead of studying any discrete distribution and payoff function for the buyer, I restrict the focus to cases in which the joint distribution of the data and states is elliptical and the buyer weights the goodness-of-fit of its forecast according to a quadratic loss function.[7]

There are a few other papers that have studied similar problems in which one agent buys information from another one. In an environment without private information, Bergemann and Bonatti (2015) have studied the problem of selling information when the market for it is competitive and the seller of information can only sell signals that reveal perfectly one of the states. Yang (2018) considers an environment in which the data broker reveals information to the consumers about their valuations for a good. However, the data broker charges the sellers of goods for this information. In this environment the data brokers' optimal strategy is to reveal perfectly the consumers' willingness to pay when the consumers have low valuations and pool all the high valuation consumers. This way the sellers of goods can sell to all of their best clients. Babaioff, Kleinberg, and Paes Leme (2012) consider a problem where both the seller and the buyer have private information about one distinct state, but the buyer's payoff function depends on both of them. Assuming that the outcome of the mechanism can depend on the signal observed by the seller, they study the conditions under which the revelation principle holds, and they find algorithms that approximate the optimal mechanism.

The problem I study is technically related to the literature on multidimensional mechanism design (see, for example, Armstrong (1996); Thanassoulis (2004); Manelli and Vin-

[7]In Appendix A, I show that when the monopolist is interested in forecasting the intercept of a linear demand function, the implied loss is quadratic.

cent (2006); Daskalakis, Deckelbaum, and Tzamos (2017)). While this literature normally assumes that a buyer may want to buy multiple products, and the products are fixed, in my environment each type strictly prefers a unique product, and I allow the monopolist to design the products he will offer. By designing the products, the data broker makes sure that out of the many partially substitutable products, the buyer buys the product specifically targeted to her.

In this sense, my paper is related to the literature that considers the problem of selling substitutable goods. The problem I study reduces to a problem analogous to a discrete version (see Vohra (2011)) of a problem solved by Wilson (1993)—except that the payoff is nonlinear in my environment, while in his is linear— with a completely different solution. In a linear two-goods environment, Pavlov (2011) has shown that the optimal mechanism involves lotteries that either assign one of the two goods with probability 1 or assign none of the goods to the buyer. Balestrieri, Izmalkov, and Leao (2015) show in a Hotelling-type model, in which there are two goods located on the extremes of a line and the consumers face a transportation cost, that even in a unidimensional environment the optimal mechanism may involve lotteries depending on the shape of the transportation cost function.

Finally, when the data broker designs the statistics he wants to offer, he chooses the quality of each type's inference. A similar problem has been studied before in the literature on quality degradation (see, for example, Mussa and Rosen (1978); Maskin and Riley (1984)) and product design (see, for example, Anderson and Celik (2015)). In these branches of literature, it is normally assumed that the preferences are unidimensional and that the ranking of buyer types, in terms of their willingness to pay, is product independent. In my multidimensional environment, the ranking of types according to their valuations is statistic dependent, in the sense that statistics that are considered to be of high quality by some types are considered to be of bad quality by others.

2 Model

Data brokers sell consumer data to firms. I study how a monopolist data broker, who wants to maximize profits, should present and sell the consumer data he can access when he is uncertain about the buyer's motives for purchasing the data. By assuming that each consumer's characteristics come from the same distribution, I focus on the problem of selling data about a representative consumer.

The data broker has access to some observable variables about a consumer $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ that are potentially relevant for the firm to estimate an unknown consumer's characteristic in the set $\theta = (\theta_1, \theta_2, \dots, \theta_n)$. I represent the firm's preferences by a two-dimensional type. The first component is one of n possible information types: $t_1, \dots, t_n$; the second component is a valuation type v that can take any value in an interval $[\underline{v}_i, \bar{v}_i]$. This interval might depend on the first component. Type (t_i, v)'s forecast loss is equal to

$$-vE[(\theta_i - a)^2];$$

that is, type (t_i, v) is interested in making an accurate forecast of the consumer's characteristic θ_i and faces a quadratic loss function that is weighted, according to how valuable the estimation is for her, by v.[8] Since acquiring information allows the firm to make a better forecast, the firm is willing to buy all, a part of, or a summary of the data available to the data broker. I assume that the firm's global payoff is quasilinear in money; it is equal to the forecast loss minus the price she pays for the information.

The data broker faces the issue that he does not know the firm's type, so he cannot first order price discriminate. The data broker assigns probability α_i to the firm being information type t_i and believes that the valuation type is distributed, conditional on the information type being t_i, according to an absolutely continuous distribution $G(v \mid t_i)$ that admits a density $g(v \mid t_i)$. I assume that the data broker, before knowing the data realization, can commit to a selling mechanism and look for the incentive-compatible and ex-ante individually rational direct mechanism that maximizes the broker's profit.

[8] In Appendix A, I prove that when a monopolist is uncertain about the intercept of a linear demand, his optimal decisions lead to a profit loss that is quadratic with respect to the forecast of the intercept that the monopolist uses in his decision-making process.

I endow the data broker with a zero-mean random variable ϵ that is independent of X and θ and that allows him to flexibly lessen the informativeness of the data statistics he offers. In this environment, a direct mechanism is a pair (ψ, p), such that if the firm reports type (t_i, v), she receives a statistic of the data and noise $\psi(t_i, v) : \mathbb{R}^{k+1} \to \mathbb{R}^m$ and pays a price $p(t_i, v) \in \mathbb{R}$.

The firm has a prior over θ, denoted by F_θ, and believes that the random variable X is distributed, conditional on θ, according to $F_{X|\theta}$. This information environment is common knowledge to both the data broker and the firm.

In Figure 1 I summarize the timing in the model. The data broker commits to a direct mechanism before knowing the realization of the data. After this, nature draws the random vector θ and the data X according to the distribution $F_{(X,\theta)}$. When the firm reports her type, the data broker provides her with the data transformation he promised at the price he committed to in the mechanism. Once the firm has observed this extra information, she makes a forecast.

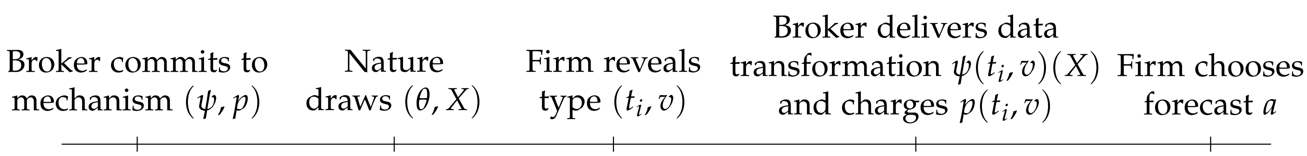


Figure 1: Timing in the model.

2.1 Distribution Restriction

The model that I have specified is too general to generate meaningful conclusions. I restrict the model by assuming that the joint distribution of (X, θ, ϵ) is elliptical and has finite second moments; the covariance matrix of the vector X, $Var(X)$, is positive definite; and $Var(\epsilon) = \sigma^2 > 0$.

Assumption 1 *The joint distribution of (X, θ, ϵ) is elliptical with finite second moments.*

Some multivariate elliptical distributions with finite second moments are the multivariate normal distribution, the multivariate t-student distribution, the multivariate symmetric Laplace distribution, the multivariate logistic distribution, and the multivariate symmetric hyperbolic distribution. While this assumption restricts the set of distributions that I consider, it is general enough to include distributions that are skewed or leptokurtic, and distributions for which zero covariance is different from independence.

The distinct characteristic of an elliptical distribution is that the iso-density plot for a two-dimensional random vector is an ellipse (and a generalization of an ellipse for higher dimensions). An example is presented in Figure 2.

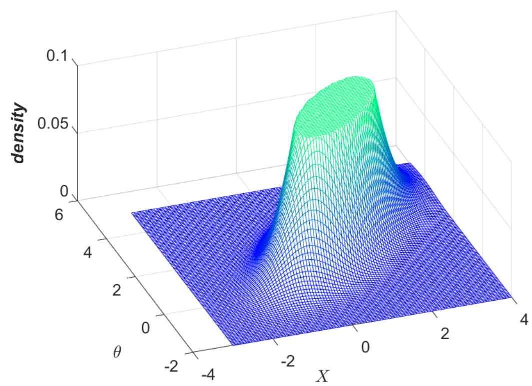


Figure 2: General shape of probability density function for an Elliptical distribution.

This family of distributions satisfies two properties that are important for my purposes. First, for any elliptical distribution, the conditional expectation is linear; second, the linear combination of elliptical distributions is an elliptical distribution. The proof of the following claim can be found in Fang, Kotz, and Ng (1990).

Claim 1 *Suppose the joint distribution of (X, Y) is jointly elliptical. Then*

1. *$E[Y \mid X]$ is linear in X, and*

2. *any linear transformation of (X, Y), $A(X, Y)$, is elliptical.*

These properties of the elliptical distributions are important for my analysis since they imply that when the seller offers a linear statistic of the data, the buyer's conditional expectation after observing such a statistic is linear. I show in Theorem 1 that under this restriction, it is optimal for the seller to only offer linear statistics. This result helps to simplify my analysis.

3 Value of Information

I start the analysis by calculating the maximum price that type (t_i, v) is willing to pay for any statistic ψ. Since the firm utility is quadratic, the optimal forecast for type (t_i, v) is the conditional expectation of θ_i given ψ. With this forecast, type (t_i, v) suffers an expected loss equal to the negative of the expected conditional variance of θ_i given ψ times v.[9]

Claim 2 *Given that the data broker provides the firm with a statistic $\psi(x)$, type (t_i, v)'s optimal forecast is $a^* = \mathbb{E}(\theta_i \mid \psi(x))$, and type (t_i, v) suffers an ex-ante expected loss equal to*

$$-v\mathbb{E}_X[Var(\theta_i \mid \psi(x))].$$

I can use this result to formally define what incentive compatibility and individual rationality mean in my environment. Because types are two-dimensional, I need to consider the possibility that a buyer lies about her information and/or valuation type. I say that the mechanism (ψ, p) is incentive compatible (IC) if

$$-v\mathbb{E}_X[Var(\theta_i \mid \psi(t_i, v))] - p(t_i, v) \geq -v'\mathbb{E}_X[Var(\theta_i \mid \psi(t_j, v'))] - p(t_j, v') \;\forall (t_i, v), (t_j, v'),$$

and it is individually rational (IR) if

$$-v\mathbb{E}_X[Var(\theta_i \mid \psi(t_i, v))] - p(t_i, v) \geq -vVar_{F_\theta}(\theta_i) \;\forall (t_i, v).$$

These constraints have two distinct properties that make them slightly different from those in the problems previously studied in the literature. First, in the IC constraint two possible deviations are embedded: The first deviation is the usual one that the buyer can misreport her type. Furthermore, when the buyer does this, she can use the statistic received to make a forecast about the variable she is interested in, which might not coincide with the variable that the statistic was targeted to. The second property is that the right-hand side of the IR constraint is not the same across types, since when the buyer does not buy any information, her best forecast is her prior expectation, resulting in an expected loss equal to the negative of her prior variance times her valuation type.

[9]The proof of Claim 2 and all others that are not in the main text can be found in Appendix B.

By reordering the IR constraint, it is possible to conclude that the maximum that type (t_i, v) is willing to pay is v times the reduction in the variance that type (t_i, v) can achieve by buying the statistic that is targeted to her. Therefore, in this environment, the buyer's only interest is to reduce her forecast variance.

With this in mind I can show in Theorem 1 that in this environment it is optimal for the seller to only sell linear statistics. When none of the IC constraints bind, the argument is trivial since the seller will want to target to each type the conditional expectation of the variable the firm is interested in, and by Claim 1 this conditional expectation is linear. The significant part of the result, and to some extent a surprising one, is that even when some IC constraints bind, the seller is still better off only offering linear statistics, even though they do not necessarily coincide with the conditional expectation.

The proof of the theorem, which was inspired by the argument in one of the examples in Basar (2008), consists of two steps. In the first step, by an argument similar to the one in Kamenica and Gentzkow (2011), I show that no loss results from considering mechanisms in which the seller offers forecasts that are followed by the buyer. In the second step, I consider a simultaneous zero-sum game between the seller and a fictitious player. In this game the payoff function is equal to the seller's profits in a mechanism, and the conflict arises from the seller choosing statistics to maximize his profits while the fictitious player chooses updating rules to minimize any type's posterior variance when observing the statistic targeted to another type. I show that in the unique saddle-point of this zero-sum game the seller offers linear statistics and the fictitious player uses linear updating rules. The optimal mechanism corresponds to the seller's saddle-point strategy, which completes the argument. The existence of conflict between the two players is crucial for this argument: Witsenhausen (1968) introduced an example without conflict between two players in which linear statistics and updating rules are not always optimal.

Theorem 1 *In the optimal mechanism, the seller only offers statistics that are linear combinations of the data and independent noise; that is, the seller targets to each type a statistic of the form* $\psi(t_i, v) = L(t_i, v)^T x + \ell(t_i, v)\epsilon \in \mathbb{R}$, *where* $L(t_i, v) \in \mathbb{R}^k$ *and* $\ell(t_i, v) \in \mathbb{R}$.

Since Theorem 1 implies that it is optimal for the seller to offer a mechanism in which all statistics are linear combinations of the data and independent noise, I start the analysis in Lemma 1 by calculating the reduction of θ_i's forecast error variance when type (t_i, v) receives a linear statistic.

Lemma 1 *θ_i's forecast error variance reduction when observing a non-null linear combination of the data and noise, $L^T X + \ell\epsilon$, corresponds to*

$$\frac{(L^T Cov(\theta_i, X))^2}{L^T Var(X) L + \ell^2\sigma^2}.$$

Furthermore, Claim 2 implies that type (t_i, v)'s willingness to pay for a statistic is maximized when she observes the conditional expectation of θ_i given X. If the data broker were not facing any feasibility constraints, he would sell these statistics to the firm. The following lemma shows the conditional expectation of θ_i given X and the maximum reduction of θ_i's forecast error variance for any data set.

Lemma 2 *The conditional expectation of θ_i given X is*

$$E(\theta_i \mid X) = E(\theta_i) + Cov(\theta_i, X)^T Var(X)^{-1}(X - E[X]),$$

and the maximum θ_i's forecast error variance reduction that can be achieved with data X equals

$$Cov(\theta_i, X)^T Var(X)^{-1} Cov(\theta_i, X).$$

Therefore, type (t_i, v)'s maximum willingness to pay is $vCov(\theta_i, X)^T Var(X)^{-1} Cov(\theta_i, X)$. This formula is not very helpful for understanding how much type (t_i, v) is willing to pay for a statistic that is not sufficient to learn the conditional expectation of θ_i given X. I therefore introduce new terminology that is easier to work with and to interpret. I define the vector $\gamma_i \equiv Var(X)^{-1/2} Cov(\theta_i, X)$, which measures how valuable any linear statistic is for forecasting θ_i.[10] Consider a linear statistic $L^T X + \ell\epsilon$ and let $\lambda = Var(X)^{1/2} L$ denote the transformed vector of coefficients that lives in the same space as γ_i. Type (t_i, v) is

[10]This vector is well defined since I assumed that $Var(X)$ is positive definite. Formally, I define $Var(x)^{-1/2}$ as $Var(X)^{-1/2} = SX^{1/2}S^T$ where $X^{1/2}$ is the diagonal matrix whose entries are the positive roots of the eigenvalues of $Var(X)^{-1}$ and S is an orthogonal matrix whose columns are the eigenvectors of $Var(X)^{-1}$. All the results that follow are true as long as I choose a square root of $Var(X)$ that is invertible.

willing to pay for this statistic the amount $v\frac{(\gamma_i^T\lambda)^2}{\lambda^T\lambda+\ell^2\sigma^2}$. Since $\gamma_i^T\gamma_i = Var(E[\theta_i \mid X])$, and $\gamma_i^T\lambda$ is the covariance between $E[\theta_i \mid X]$ and $L^TX + \ell\epsilon$, type (t_i, v)'s willingness to pay for this statistic is proportional to the square of the correlation between $E[\theta_i \mid X]$ and the signal. In other words, type (t_i, v) is willing to pay more for statistics that are more correlated with the forecast that she would implement if she could observe all the data.

An important special case is the one in which the statistic does not include any independent noise. In this case the correlation between this statistic and the conditional expectation is completely determined by the angle between γ_i and λ. Figure 3 represents type (t_i, v)'s willingness to pay for the transformed vector of coefficients λ when her preferred vector is γ_i. The distance between the origin and the blue locus represents type (t_i, v)'s willingness to pay for a noiseless linear combination that generates a vector pointing in any particular direction—for example, in the same direction as λ_0. As the angle between these two vectors increases from 0 to $\pi/2$, type (t_i, v)'s willingness to pay changes continuously from $vCov(\theta_i, X)^T Var(X)^{-1} Cov(\theta_i, X)$ to 0.

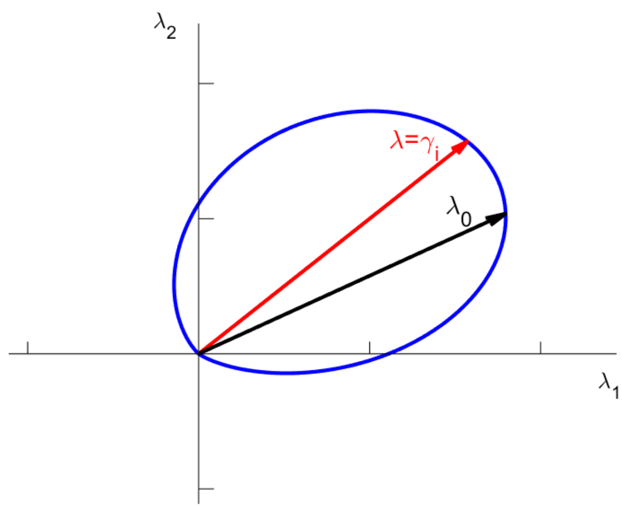


Figure 3: The distance between the origin and the blue locus represents the willingness to pay, by type with optimal vector γ_i, when it receives a noiseless linear combination with coefficients pointing in that direction.

A similar comparison holds if the linear statistic contains independent noise. However, adding independent noise reduces proportionally the willingness to pay for any linear statistic, and the slope with respect to the angle becomes flatter.

Overall, type (t_i, v) prefers statistics that have a higher correlation with $E[\theta_i \mid X]$ and statistics for which uncorrelated noise accounts for a smaller portion of the statistic variance. The objective of the rest of the paper is to understand how the data broker designs these linear statistics to maximize profits when the mechanism must satisfy the IC and IR constraints.

To rule out cases in which two information types want to use the data in exactly the same way, I assume that no two types have more preferred statistics that are parallel. Technically, I assume that no two information types have γ vectors pointing in the same (or opposite) direction.

Assumption 2 *For each $i, j \in \{1, \ldots, n\}$, $i \neq j$, and for any $k \neq 0$, $\gamma_i \neq k\gamma_j$.*

4 Optimal Mechanism with Two Information Types

In this section I provide a complete characterization of the optimal mechanism when there are only two information types. The general intuition from the previous section is that in the optimal mechanism the data broker needs to reduce the correlation between the statistic targeted to any mimicked type and the mimicking type's preferred statistic. I present the optimal way in which the seller modifies the statistics he sells to reach this objective for two distinct cases: a unique common valuation type and a continuum of valuation types for each of the information types.

4.1 A Unique Common Valuation Type

In this subsection I provide a complete characterization of the simplest case in which for each information type there is a unique valuation type that is common across information types, and I normalize it to one. This analysis provides insights about the forces that affect the design of the optimal mechanism in the general case. The main result is that in the optimal mechanism the data broker, without introducing independent noise, will deteriorate the quality of the statistic targeted to the mimicked type by distorting the coefficients assigned to each of the variables in the data in the direction opposite to the one the mimicking type prefers the most. Though this quality degradation affects both types, it is designed in a way that affects the mimicking type more than the mimicked type.

Without loss, I assume that $\|\gamma^1\| > \|\gamma^2\|$;[11] that is, type t_1 is willing to pay more for all the data than type t_2 is, and that $\gamma_1^T \gamma_2 > 0$.[12]

First, I introduce some of the properties that the optimal mechanism must satisfy and that simultaneously simplify the problem. Proposition 1 shows that type t_1 will receive a statistic that from the perspective of that firm type is equivalent to observing all the data, and type t_2 will receive zero information rent. These properties are a result of type t_1 being willing to pay more for all the data than what type t_2 is willing to pay for it, so that the data broker can choose a price high enough for all the data such that type t_2 never wants to report type t_1.

The same proposition presents an easy to check condition that characterizes when the IC constraint for type t_1 reporting type t_2 binds and when it does not. Intuitively, this constraint binds only if type t_1 can make an accurate guess about $E[\theta_1 \mid X]$ from observing $E[\theta_2 \mid X]$. As was pointed out in the previous section, this depends on the correlation between the two conditional expectations or, equivalently, on the angle between γ_1 and γ_2. If they are perfectly correlated, type t_1 can learn the same amount from observing either of the two and will buy the cheapest one. If they are conditionally independent, the IC constraint will never bind since type t_1 cannot learn anything about $E[\theta_1 \mid X]$ from observing $E[\theta_2 \mid X]$. Proposition 1 shows that the IC constraint for type t_1 reporting type t_2 binds if and only if the vectors γ_1 and γ_2 point in a similar direction, while γ_1's magnitude is large relative to γ_2's magnitude. This means that the IC constraint for type t_1 reporting type t_2 binds if and only if both types are interested in similar features of the data, but one type is willing to pay substantially more for all the data than the other one.

[11] I use the symbol $\|\|$ to denote the Euclidean norm of a vector.

[12] If $\gamma_1^T \gamma_2 < 0$, I can always convert the problem to the one in which type t_1 measure of information is given by $-\gamma_1$. Lemma 1 shows that these problems are equivalent since a rational agent learns exactly the same when receiving the linear combination $L^T X$ and when receiving $-L^T X$.

Finally, Proposition 1 states that in the optimal mechanism the data broker offers statistics that do not contain any independent noise. Modifying the direction of the statistic targeted to type t_2 is more efficient than adding independent noise to it; adding independent noise reduces proportionally the willingness to pay of both types, while modifying the direction of the signal targeted to type t_2 in the direction opposite to type t_1's preferred direction affects type t_1 more than type t_2.[13]

Proposition 1 *In the optimal mechanism, when there are two information types and a unique common valuation type, it must be the case that*

1. *type t_1 receives a statistic that reduces the variance of θ_1 the same amount as observing data x;*
2. *type t_2 receives zero information rent;*
3. *the IC constraint for type t_1 reporting type t_2 binds if and only if*
$$\cos(\beta) > \frac{\|\gamma_2\|}{\|\gamma_1\|}, \text{ or equivalently, } \|\gamma_1 - \gamma_2\|^2 < \|\gamma_1\|^2 - \|\gamma_2\|^2,$$
where β is the angle between the vectors γ_1 and γ_2; and
4. *the data broker never adds independent noise to any of the statistics.*

Proposition 1 implies that the data broker's problem can be simplified to:

$$\begin{aligned}
\max_{p_1, L_2} \quad & \alpha_1 p_1 + (1-\alpha_1)\frac{(L_2^T Cov(\theta_2, X))^2}{L_2^T Var(X) L_2} \\
\text{s.t.} \quad & \gamma_1^T \gamma_1 - p_1 \geq \frac{(L_2^T Cov(\theta_1, X))^2}{L_2^T Var(X) L_2} - \frac{(L_2^T Cov(\theta_2, X))^2}{L_2^T Var(X) L_2} \\
& \gamma_1^T \gamma_1 - p_1 \geq 0
\end{aligned}$$

If the IC constraint for type t_1 reporting type t_2 does not bind, the optimal mechanism for the data broker is to offer to each type a distinct statistic that allows them to recover the conditional expectation they are interested in and charge them the most they are

[13]The analysis in Section 4.2 present an instance in which adding independent noise is one of the ways, though not the only one, in which the optimal mechanism can be implemented.

willing to pay for all the data.[14] When the IC constraint for type t_1 reporting type t_2 binds, the data broker cannot simultaneously offer all available information to both types while setting monopolist prices. Since, by Proposition 1, adding independent noise is never optimal, the data broker is left with two possibilities. He can either reduce the price targeted to type t_1 or he can deteriorate the quality of the signal targeted to type t_2 by modifying the coefficients that are used to create the statistic targeted to type t_2. The data broker will optimally weight each of these two possibilities.

If I assume that the IR constraint for type t_1 does not bind and that the IC constraint for type t_1 reporting type t_2 binds, I can plug in for the value of p_1 and obtain a maximization problem that depends only on the coefficients L_2:

$$\max_{L_2} \alpha_1 \left(\gamma_1^T \gamma_1 - \frac{(L_2^T Cov(\theta_1, X))^2}{L_2^T Var(X) L_2} \right) + \frac{(L_2^T Cov(\theta_2, X))^2}{L_2^T Var(X) L_2}$$

Taking the first order condition and solving, I obtain the following equation for the optimal coefficient vector:

$$\tilde{L}_2 = Var(X)^{-1} \left(Cov(\theta_2, X) - c Cov(\theta_1, X) \right),$$

where $c = \alpha_1 \frac{\tilde{L}_2^T Cov(\theta_1, X)}{\tilde{L}_2^T Cov(\theta_2, X)} > 0$. This is an implicit equation in both $\tilde{L}_2$ and c. Since c is the solution to a single variable fixed-point problem, I am able to solve for it in closed form, which allows me to obtain a closed-form solution for $\tilde{L}_2$.[15] The proof presents the details. In the optimal mechanism, the data broker deteriorates the informativeness of the statistic targeted to type t_2 by obfuscating it in the direction opposite to type t_1's preferred direction. This distortion reduces the correlation between type t_1's preferred statistic and the statistic targeted to type t_2, affecting type t_1 more than type t_2.

So far I have assumed that type t_1's IR constraint does not bind. This assumption is satisfied as long as $\frac{(\tilde{L}_2^T Cov(\theta_1, X))^2}{\tilde{L}_2^T Var(X) \tilde{L}_2} > \frac{(\tilde{L}_2^T Cov(\theta_2, X))^2}{\tilde{L}_2^T Var(X) \tilde{L}_2}$. In the proof of Theorem 2 I show that this inequality holds if and only if $\alpha < \tilde{\alpha} = \frac{\gamma_2^T (\gamma_1 - \gamma_2)}{\gamma_1^T (\gamma_1 - \gamma_2)} \in (0, 1)$. When type t_1's IR

[14] Offering only the rough data is never optimal since it allows both types to calculate the conditional expectation, resulting in each type reporting the type that pays the lowest price.

[15] The objective function is not globally concave. One of the main steps in the proof shows that there is a solution to the fixed-point problem for the constant c that is actually a solution to the maximization problem.

constraint binds, the data broker does not have an incentive to distort further the statistic targeted to type t_2, since he cannot increase the price targeted to type t_1. This implies that the quality of the statistic targeted to type t_2 is never degraded until the point at which type t_2 is willing to pay nothing for it. This contrasts with the literature in quality degradation, where there is always a distribution over types such that type t_2 is offered a quality that gives him a null payoff (see, for example, Mussa and Rosen (1978)). This difference comes from the fact that in my environment the ranking of types according to their valuations is statistic-dependent, while in the previous literature, it has been normally assumed that the ranking of types according to their valuations is product-independent. This also implies that the data broker can extract all the surplus created by the mechanism even when some of the IC constraints bind, although the surplus created by the mechanism is smaller than the maximum social surplus.

Theorem 2 provides a closed-form solution for the optimal mechanism.

Theorem 2 *In the case of two information types, the data broker maximizes profits by offering the mechanism with coefficients* $L(t_1) = Var(X)^{-1}Cov(\theta_1, X)$ *and coefficients*

1. $L(t_2) = Var(X)^{-1}Cov(\theta_2, X)$ *if* $\|\gamma_1 - \gamma_2\|^2 \geq \|\gamma_1\|^2 - \|\gamma_2\|^2$*; and*
2. $L(t_2) = Var(X)^{-1}\left(Cov(\theta_2, X) - cCov(\theta_1, X)\right)$ *if* $\|\gamma_1 - \gamma_2\|^2 < \|\gamma_1\|^2 - \|\gamma_2\|^2$*;*

where $c = \frac{\|\gamma_2\|^2+\alpha_1\|\gamma_1\|^2-\sqrt{\left(\|\gamma_2\|^2+\alpha_1\|\gamma_1\|^2\right)^2-4\alpha_1\left(\gamma_1^T\gamma_2\right)^2}}{2\alpha_1\gamma_1^T\gamma_2}$*, if* $\alpha < \tilde{\alpha} = \frac{\gamma_2^T(\gamma_1-\gamma_2)}{\gamma_1^T(\gamma_1-\gamma_2)}$*, and* $c = \tilde{\alpha}$ *otherwise. Furthermore,* $p(t_1) = \gamma_1^T\gamma_1$ *in case 1, and it is the value that makes the IC constraint for type* t_1 *reporting type* t_2 *to hold with equality in case 2, while* $p(t_2)$ *is the value that makes the IR constraint for type* t_2 *to hold with equality.*[16]

The amount of distortion of the statistic targeted to type t_2 depends on the value of α_1. As the probability of facing firm type t_1 increases, the data broker wants to reduce the information rent given to type t_1, and to reach this objective the data broker degrades the quality of the statistic targeted to type t_2. In terms of the notation in Theorem 2, this means that c is nondecreasing with respect to α_1.

[16] Explicit closed-form solutions for these prices can be found in the proof of the theorem.

Corollary 1 *Suppose that the IC constraint for type t_1 reporting type t_2 binds. Then the optimal mechanism satisfies the following properties:*

1. *the degradation of quality is increasing in α_1; that is, c is nondecreasing in α_1, and as α_1 tends to 0 the data broker almost offers the undistorted statistic to type t_2; that is, $\lim_{\alpha_1 \to 0} c = 0$; and*

2. *the information rent given to type t_1 is nonincreasing in α_1, and as α_1 tends to 0 the information rent tends to $\frac{(\gamma_1^T \gamma_2)^2}{\gamma_2^T \gamma_2} - \gamma_2^T \gamma_2$.*

Using the characterization of the optimal mechanism in Theorem 2, I can study which data sets the data broker would like to acquire. . If two data sets are offered to the data broker at the same cost, he would choose the data set that allows vectors γ_1 and γ_2 to be as nearly orthogonal as possible; that is, he would like to acquire data that generates the smallest correlation between the conditional expectations of θ_1 and θ_2 given all the data. The following corollary states precisely the conditions under which the profits of the data broker are increasing in the angle between the vectors γ_1 and γ_2.

Corollary 2 *Let π be the optimal profits for some α, γ_1, and γ_2 and let $0 < \beta < \pi/2$ be the angle between γ_1 and γ_2. For any data set that generates γ_1' and γ_2' with $\|\gamma_1\| = \|\gamma_1'\|$, $\|\gamma_2\| = \|\gamma_2'\|$ and $\beta < \beta' \leq \pi/2$, where β' is the angle between γ_1' and γ_2', the optimal profits are larger than or equal to π.*

Unfortunately, I cannot directly translate the results in the corollary to a direct comparison between the covariance vectors. First, an uncountable number of statistical models (covariance matrices of unknown and know consumer characteristics) generate exactly the same vectors γ_1 and γ_2.[17] Second, even if I fixed a covariance matrix $Var(X)$, it is not true that when two vectors γ_i and γ_j are closer to being orthogonal, the corresponding vectors $Cov(\theta_i, X)$ and $Cov(\theta_j, X)$ are also closer to being orthogonal. Figure 4 presents such an example where the vectors γ_1 and γ_2 are orthogonal ($cos^2(\beta) = 0$) only when the angle between $Cov(\theta_i, X)$ and $Cov(\theta_j, X)$ is strictly smaller than $\pi/2$.

[17] Fix any positive definite matrix A of dimension 2×2 and say that this is the variance of the data X. Then, by picking $Cov(\theta_i, X) = A^{1/2}\gamma_i$, we guarantee that the statistical model exactly generates the vector γ_i.

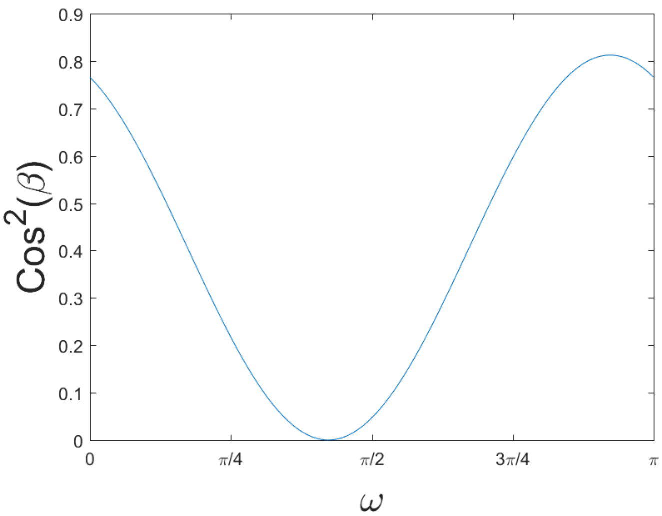


Figure 4: Value of $cos^2(\beta)$, with β being the angle between γ_1 and γ_2, for different values of ω, the angle between $Cov(\theta_1, X)$ and $Cov(\theta_2, X)$, when $Var(X) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ and $Cov(\theta_1, X)^T = (0.866, 0.5)$.

4.2 A Continuum of Valuation Types

I now proceed to a more general environment in which the buyer has both vertical and horizontal private information. I still assume there are two information types $\{t_1, t_2\}$, but for each information type there is a continuum of valuation types that are distributed according to absolutely continuous distributions $G(v \mid t_1)$ and $G(v \mid t_2)$, with densities $g(v \mid t_1)$ and $g(v \mid t_2)$ and supports $[\underline{v}_1, \bar{v}_1]$ and $[\underline{v}_2, \bar{v}_2]$, respectively.

Since I do not assume any special property relating to the support of these distributions, the analysis in Section 3 implies that it is without loss to assume that $\|\gamma_1\| = \|\gamma_2\| = 1$. This assumption means that for $i \in \{1, 2\}$ the maximum reduction of θ_i's forecast error variance is equal to 1.

To facilitate comparison with the classical paradigm, I impose the Regularity Condition that has normally been assumed in the mechanism design literature.

Assumption 3 *I assume that for each t_i, $i \in \{1, 2\}$, the conditional distribution of valuation types, satisfies that*

$$v - \frac{1 - G(v \mid t_i)}{g(v \mid t_i)}$$

is nondecreasing in v.

To simplify the notation, I introduce two new notions. I let q_{iv} be the reduction of information type t_i's forecast error variance when observing the linear combination targeted to type (t_i, v); that is, $q_{iv} = \frac{(L_{iv}^T Cov(\theta_i, X))^2}{L_{iv}^T Var(X) L_{iv} + \ell_{iv}^2 \sigma^2}$, and I let δ_{jiv} be the reduction of information type t_j's forecast error variance when observing the linear combination targeted to type (t_i, v); that is, $\delta_{jiv} = \frac{(L_{iv}^T Cov(\theta_j, X))^2}{L_{iv}^T Var(X) L_{iv} + \ell_{iv}^2 \sigma^2}$.[18] The problem that the seller faces can be rewritten as

$$\max_{\{L_{iv}\},\{\ell_{iv}\},\{p_{iv}\}} \quad \alpha_1 \int_{\underline{v}_1}^{\bar{v}_1} p_{1v} f(v \mid t_1) dv + \alpha_2 \int_{\underline{v}_1}^{\bar{v}_1} p_{2v} f(v \mid t_2) dv$$

$$st. \quad vq_{iv} - p_{iv} \geq vq_{iv'} - p_{iv'} \quad \text{for } i \in \{1,2\},\ \forall v, v' \in [\underline{v}_i, \bar{v}_i] \qquad \text{(IC iv-iv')}$$

$$vq_{iv} - p_{iv} \geq v\delta_{ijv'} - p_{jv'} \quad \text{for } i \neq j,\ \forall v \in [\underline{v}_i, \bar{v}_i], v' \in [\underline{v}_j, \bar{v}_j] \qquad \text{(IC iv-jv')}$$

$$vq_{iv} - p_{iv} \geq 0 \quad \text{for } i \in \{1,2\},\ \forall v \in [\underline{v}_i, \bar{v}_i] \qquad \text{(IR iv).}$$

Except for the constraints *IC iv-jv'*, the problem looks like two copies of the one studied by Myerson (1981). This allows me to draw two conclusions that follow almost directly from his analysis. First, for each $i \in \{1,2\}$ the constraints *IC iv-iv'* and *IR iv* can be summarized by an integral condition and a monotonicity condition that are easier to work with. Lemma 3 presents the formal statement.[19]

Lemma 3 *The constraints IC 1v-1v', IC 2v-2v', IR 1v and IR 2v are equivalent to:*

1. $vq_{iv} - p_{iv} = \underline{v}_i q_{i\underline{v}_i} - p_{i\underline{v}_i} + \int_{\underline{v}_i}^{v} q_{iw}\, dw$ *for* $i \in \{1,2\}$.

2. q_{iv} *is nondecreasing for* $i \in \{1,2\}$.

3. $\underline{v}_i q_{i\underline{v}_i} - p_{i\underline{v}_i} \geq 0$ *for* $i \in \{1,2\}$.

Second, if none of the constraints *IC iv-jv'* bind, the problem has a simple solution. The seller will offer a menu with two packages. Each package is targeted to a specific

[18] Remember that type (t_i, v) wants to forecast the stochastic variable θ_i. Therefore, I can use q_{iv} as a measure of the quality of the statistic that is targeted to type (t_i, v).

[19] I omit the proof of Lemma 3 and Proposition 2 since they follow directly from standard arguments used in Myersonian environments.

information type and contains a take-it-or-leave-it offer for a distinct signal that allows this firm type to recover the conditional expectation in which she is interested.[20] Without loss I assume that the take-it-or-leave-it price for information type t_1, which I denote by v_1^*, is larger than the take-it-or-leave-it price for information type t_2, which I denote by v_2^*. Proposition 2 formally presents the optimal mechanism for this case and introduces a condition that establishes when none of the constraints *IC iv-jv′* binds, a generalization of the condition in Proposition 1.

Proposition 2 *Suppose that none of the constraints IC iv-jv′ bind. There exist values $v_1^* > v_2^*$ such that in the optimal mechanism*

$$q_{iv} = \begin{cases} 1 \text{ if } v \geq v_i^* \\ \\ 0 \text{ if } v < v_i^*, \end{cases}$$

and $p_{iv} = \mathbb{1}_{q_{iv}=1} v_i^*$. Furthermore, none of the constraints IC iv-jv′ bind if and only if $v_1^* cos^2(\beta) \leq v_2^*$, where β is the angle between γ_1 and γ_2.

From now on I consider the interesting case in which at least some of the constraints *IC iv-jv′* bind. The challenge is that *a priori* I do not know which of them actually bind. Furthermore, since the payoffs are nonlinear I cannot apply the results in the literature that deliver necessary and sufficient conditions that summarize the IC constraints.[21] Proceeding constructively, I characterize the optimal mechanism for the relaxed problem without the constraints *IC 2v-1v′* and argue that its solution is actually the solution to the original problem.

The next lemma generalizes part of Proposition 1. It states that, in the relaxed problem, as long as the seller wants to sell an informative signal to type (t_2, v), the signal targeted to this type cannot contain any independent noise. The reason is similar to what we saw in the previous case: by modifying the coefficients in the right direction,

[20] As pointed out before in footnote 14, offering all rough data to each information type will not be optimal even in this case, since the seller, in general, wants to offer different take-it-or-leave-it prices to each information type.

[21] In particular the characterization of optimality in Rochet and Choné (1998) does not apply in my environment since the payoffs are nonlinear.

the seller can create a larger difference effect in the types' willingness to pay than he can by adding independent noise.

Lemma 4 *In the relaxed problem without the constraints IC 2v-1v', if* $q_{2v} \neq 0$, $\ell_{2v} = 0$.

This lemma allows me to rewrite the relaxed problem as a problem in which the seller chooses the angles between the statistic targeted to type t_2 and the vectors γ_1 and γ_2 that represent the preferred statistics of type t_1 and type t_2, respectively. Let β_{2v} be the angle between $Var(X)^{1/2}L_{2v}$ and γ_2 and β be the angle between γ_1 and γ_2. Then $q_{2v} = cos^2(\beta_{2v})$ and $\delta_{12v} = cos^2(\beta + \beta_{2v})$, so that $\delta_{12v} = g(q_{2v})$ with $g(q_{2v}) = cos^2(\beta + cos^{-1}(\sqrt{q_{2v}}))$.[22] Furthermore, in the relaxed problem the seller and the buyer are indifferent between any pair (L_{1v}, ℓ_{1v}) that produces the same quality q_{1v}.[23]

Plugging in the prices that are implied by Lemma 3, using integration by parts in the traditional fashion, and denoting $U(\underline{v}_i, t_i) = \underline{v}_i q_{i\underline{v}_i} - p_{i\underline{v}_i}$, the relaxed problem can be written as

$$(RP) \quad \max_{\{q_{iv}\},\{U(\underline{v}_i,t_i)\}} \quad \int_{\underline{v}_1}^{\bar{v}_1} q_{1v}\left(v - \frac{1-F(v|t_1)}{f(v|t_1)}\right) f(v \mid t_1)dv f(t_1) - U(\underline{v}_1, t_1)f(t_1)$$
$$+ \int_{\underline{v}_2}^{\bar{v}_2} q_{2v}\left(v - \frac{1-F(v|t_2)}{f(v|t_2)}\right) f(v \mid t_2)dv f(t_2) - U(\underline{v}_2, t_2)f(t_2)$$

$$st. \quad U(\underline{v}_1, t_1) + \int_{\underline{v}}^{v} q_{1w}\, dw \geq v g(q_{2v'}) - v' q_{2v'} + U(\underline{v}_2, t_2) + \int_{\underline{v}}^{v'} q_{2w}\, dw \quad \forall v \in [\underline{v}_1, \bar{v}_1], v' \in [\underline{v}_2, \bar{v}_2]$$

q_{iv} non-decreasing for $i \in \{1, 2\}$

$U(\underline{v}_i, t_i) \geq 0$ for $i \in \{1, 2\}$.

Lemma 5 presents two properties that the optimal solution to the relaxed problem must satisfy. First, the seller must target to types (t_1, v) with $v \geq v_1^*$ a statistic that for them is equivalent to observing all the data: the seller has no reason to distort the information targeted to them since I have ignored the constraints *IC 2v-1v'*. Second, the

[22]In principle I need to make sure that the angle inside the cos is smaller than $\pi/2$. However, it is straightforward to conclude that as long as q_{2v} is positive, the seller does not want to distort the statistic targeted to type (t_2, v) beyond the point at which type (t_1, v) does not obtain any information from observing this statistic, which happens when the angle is exactly $\pi/2$.

[23]When I interpret the final solution to the relaxed problem, I will discuss again the multiple mechanisms that implement the solution.

problem of implementing the mechanism presented in Proposition 2 is that type v_1^* has an incentive to mimic type v_2^*. To reduce this incentive, the seller has an extra tool that he did not have in the simplest version of the problem: he can choose the lowest valuation type of information type t_2, to whom he sells a statistic. Lemma 5 shows that the seller always increases the minimum valuation type of information type t_2, to whom he sells above v_2^*, but this adjustment is bounded by v_1^* since type (t_1, v_1^*) places less value on the statistic targeted to information type t_2 than type (t_2, v_1^*) does.

Lemma 5 *In the solution to the relaxed problem, it must be the case that*

1. $q_{1v} = 1$ *for all* $v \geq v_1^*$, *and*

2. $\hat{v}_2 \in (v_2^*, v_1^*)$, *with* $\hat{v}_2 = \inf\{v : q_{2v} > 0\}$.

Lemma 6 completely characterizes the solution for the allocation targeted to type t_1, q_{1v}. It shows that the seller will sell a homogeneous statistic to some valuation types below v_1^*, and the quality of this statistic is only a function of $q_{2\hat{v}_2}$. The reason for this is that all of the types below v_1^* want to mimic type $\hat{v}_2$, and since all of them have the same incentives and a negative virtual value, the seller should treat them equally.

Lemma 6 *In the solution to problem RP, there is a threshold* $\hat{v}_2 < \hat{v}_1 = \frac{\hat{v}_2 q_{2\hat{v}_2}}{g(q_{2\hat{v}_2})} < v_1^*$ *such that* $q_{1v} = 0$ *for* $v < \hat{v}_1$ *and* $q_{1v} = g(q_{2\hat{v}_2})$ *for all* $v \in [\hat{v}_1, v_1^*)$.

To conclude, I must indicate the optimal allocation targeted to information type t_2, q_{2v}. The next lemma shows that the only constraints that really matter, once I have fixed an optimal allocation q_{1v}, are the constraints *IC* $1v_1^*$-$2v'$ for v' larger than $\hat{v}_2$. The reason for this is that for all types above v_1^*, the incentives to report some type (t_2, v') are smaller than v_1^*'s incentives. Given that the constraints *IC* $1v_1^*$-$2v'$ bind, the lemma shows that the optimal allocation q_{2v} has at most two distinct sections: it is at first constant and may increase for high valuation types.

Lemma 7 *Any allocation* (q_1, q_2) *with* q_1 *and* q_2 *nondecreasing, such that* $q_{1v} = 1$ *for all* $v \geq v_1^*$, $q_{1v} = g(q_{2v})$ *for all* $v \in [\hat{v}_1, v_1^*]$*—and* $q_{1v} = 0$ *otherwise; and such that the constraints*

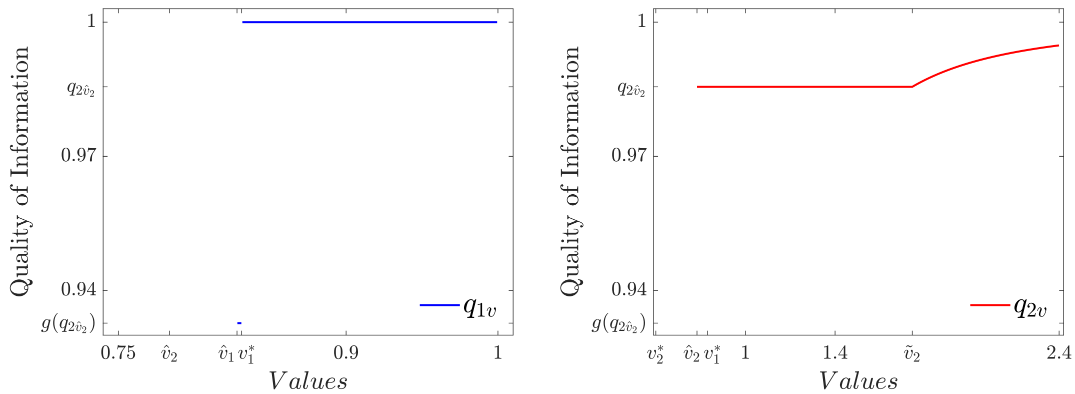
IC $1v_1^* - 2v'$ *for all* $v' \geq \hat{v}_2$ *are satisfied—is a feasible allocation for problem P. In the solution to the relaxed problem, all the constraints IC* $1v_1^*$*-*$2v'$ *for* $v' \in [\hat{v}_2, \bar{v}_2]$ *bind.*

Furthermore, there exists $\tilde{v}_2 \in (\hat{v}_2, \bar{v}_2]$ *such that the optimal solution to problem P is given by* $q_{2v'} = q_{2\hat{v}_2}$ *for* $v' \in (\hat{v}_2, \tilde{v}_2]$*, and for* $v' > \tilde{v}_2$*,* $q_{2v'}$ *is equal to the solution to* $g'(q_{2v'}) = \frac{v'}{v_1^*}$*.*

Theorem 3 summarizes the characterization of the optimal mechanism that I have described throughout the lemmas. Figure 5 presents an instance in which the optimal mechanism exhibits all the properties discussed above. Buyers that want to forecast θ_2 are affected in two ways by the presence of the other information type. First, some extra types will be excluded from the mechanism. Second, none of the valuation types that buy a statistic receive a statistic that is equivalent to observing all of the data for them: intermediate valuation types receive the same degraded statistic and large valuation types receive a more informative statistic, but it is never fully informative. Buyers that want to forecast θ_1 are benefited in two ways. First, types in $[\hat{v}_1, v_1^*]$ will buy an informative signal due to the presence of the other information type, and second, types with high valuations will receive a statistic that for them is equivalent to receiving all data for a price strictly below their willingness to pay.

Theorem 3 *There exist cutoffs* $\hat{v}_1$*,* $\hat{v}_2$*, and* $\tilde{v}_2$ *with* $\hat{v}_2 > v_2^*$*,* $\hat{v}_2 < \hat{v}_1 = \frac{\hat{v}_2 q_{2\hat{v}_2}}{g(q_{\hat{v}_2})} < v_1^*$*, and* $\tilde{v}_2$ *the value that solves* $g'(q_{2\hat{v}}) = \frac{\tilde{v}_2}{v_1^*}$*, such that in the optimal mechanism the seller targets to types* (t_1, v) *with* $v > v_1^*$ *a statistic that reduces* θ_1*'s forecast variance by* 1*, to types* (t_1, v) *with* $v \in [\hat{v}_1, v_1^*]$ *a statistic that reduces* θ_1*'s forecast variance by* $g(q_{2\hat{v}_2})$*, to types* (t_2, v) *with* $v \in [\hat{v}_2, \tilde{v}_2]$ *a statistic that reduces* θ_2*'s forecast variance by* $q_{2\hat{v}_2}$*, and to types* (t_2, v) *with* $v > \tilde{v}_2$ *a statistic that reduces* θ_2*'s forecast variance by* q_{2v}*, where* q_{2v} *solves* $g'(q_{2v}) = \frac{v}{v_1^*}$*.*

The theorem deserves two comments. First, if $\tilde{v}_2 < \bar{v}_2$, the seller will offer a continuum of statistics as in Figure 5. There are other examples in which the seller only targets one statistic to information type t_2. Second, the proposition describes the maximum variance reduction that is targeted to each type, but it does not specify the actual statistics that are offered. As in Section 4.1, to types (t_1, v) with $v > v_1^*$ the data broker provides all the data, or a sufficient statistic for it, and to types (t_2, v) with $v > \hat{v}_2$ the seller targets, without adding independent noise, a linear combination that distorts the



(a) Optimal q_{1v}.

(b) Optimal q_{2v}.

Figure 5: Optimal Mechanism when the angle between γ_1 and γ_2 is equal to 0.14° and $\alpha_1 = 0.85$, and the distributions of valuation types satisfy that $\underline{v}_1 = \underline{v}_2 = 0$, $\bar{v}_1 = 1$ and $\bar{v}_2 = 2.4$ with $g(v \mid t_1) = 15v^{14}$ and $g(v \mid t_2) = 0.8\bar{3}$ for $v < 0.8$ and $g(v \mid t_2) \propto 4(2.4 - v)^3$.

coefficients on the data away from information type t_1's preferred direction. There is a new class of statistics that needs to be offered when dealing with multiple valuation types since types (t_1, v) with $v \in [\hat{v}_1, v_1^*]$ are targeted with an only partially informative statistic. This can be implemented in multiple ways. The data broker can do one of the following: modify the coefficients on the data in the opposite direction of information type t_2's preferred direction, include independent noise to information type t_1's preferred linear combination, or offer a statistic that contains both distortions.

5 Many Information Types

In this section I present the main difficulties that impede me from obtaining a complete characterization of the optimal mechanism in the general environment. In spite of these difficulties, I show that the main properties of the optimal mechanism that I have highlighted previously are still true. To simplify the analysis, I assume that there is a unique common valuation type that I normalize to 1. I define, as before, the vector $\gamma_i = Var(X)^{-\frac{1}{2}} Cov(\theta_i, X)$, and without loss assume that $\gamma_1^T \gamma_1 \geq \gamma_2^T \gamma_2 \geq \ldots \geq \gamma_n^T \gamma_n$, with at least one strict inequality.[24] This means that type t_1 is willing to pay more for all

[24] If all the inequalities are actually equalities, the data broker can reveal all the information to all types and charge a uniform price equal to $\gamma_1^T \gamma_1$.

available data than type t_2 is, and by analogy the same is true for all the other comparisons.

Finding a condition such that none of the IC constraints bind is not a hard task. I need only ensure that all the players want features of the data that are sufficiently distinct. The next proposition presents a sufficient and necessary condition for none of the IC constraints to bind. In such a case, the data broker can target to each type a statistic that for her is equivalent to receiving all the data and then charge all of them them according to their willingness to pay without worrying that some types will mimic others.

Proposition 3 *Let β_{ij} be the angle between the vectors γ_i and γ_j. None of the IC constraints bind if and only if for all $i < j$,*

$$\cos^2(\beta_{ij}) < \frac{\|\gamma_j\|^2}{\|\gamma_i\|^2}.$$

The issue that prevents me from characterizing the optimal mechanism is that when some IC constraints bind, it is not possible to identify *a priori* the relevant IC constraints. The difficulty arises from two different sources. First, the seller may want to charge to one type a price that is below the firm's willingness to pay for the statistic that is targeted to her. This might provoke an incentive problem since another type's willingness to pay for this statistic could now be larger than its price. Second, if the seller changes the coefficients of the linear statistic targeted to one type, the modified statistic might become very informative to a third type. I will use some examples to make these sources more explicit.

The first two examples demonstrate that *a priori* I cannot identify which downward IC constraints bind.[25] The examples show that the condition in Proposition 1 is neither necessary nor sufficient. The first example presents a case in which $\|\gamma_1 - \gamma_2\|^2 > \|\gamma_1\|^2 - \|\gamma_2\|^2$ but the IC constraint for type t_1 reporting type t_2 binds. The reason is that the IC constraint from type t_2 to type t_3 binds, and the seller wants to reduce the price of the statistic targeted to type t_2. But if this price is severely reduced, the type t_1

[25] I designate an IC constraint as downward if it is one in which type t_i considers reporting type t_j with $i < j$. Otherwise it is upward.

firm will receive a positive surplus when buying the statistic targeted to type t_2.

Example 1 *Suppose that* $\gamma_1 = (6,3)$, $\gamma_2 = (2,5)$ *and* $\gamma_3 = (0,4)$. *It can be checked that* $\|\gamma_1 - \gamma_2\|^2 > \|\gamma_1\|^2 - \|\gamma_2\|^2$, $\|\gamma_1 - \gamma_3\|^2 > \|\gamma_1\|^2 - \|\gamma_3\|^2$, *and* $\|\gamma_2 - \gamma_3\|^2 < \|\gamma_2\|^2 - \|\gamma_3\|^2$. *Following the analysis in Section 4.1, in a naive first attempt I assume that only IC 2-3 binds. Under this assumption, the seller's optimal strategy is to target to types* t_1 *and* t_2 *an undistorted statistic and target to type* t_3 *a statistic that is distorted in the direction opposite to type* t_2*'s preferred statistic. When* α_2 *is small enough relative to* α_3*, according to Corollary 1, type* t_2 *receives an information rent close to* $\frac{(\gamma_2^T\gamma_3)^2}{\gamma_3^T\gamma_3} - \gamma_3^T\gamma_3 = 9$, *or equivalently a price close to* 20. *At this price, type* t_1 *can report type* t_2 *and obtain a surplus of* $\frac{(\gamma_1^T\gamma_2)^2}{\gamma_2^T\gamma_2} - 20 = 5.24$, *so that the constraint IC 1-2 is not satisfied.*[26] *It can be shown that in the optimal mechanism the constraints IC 1-2 and IC 2-3 bind. In the optimal mechanism, as* α_2 *is too small, to satisfy the constraint IC 1-2, the data broker offers some information rent to type* t_2 *and modifies the recommendation targeted to type* t_2 *in the opposite direction of type* t_1*'s preferred statistic, so that type* t_1 *does not receive any information rent.*[27]

Example 2 presents a case in which $\|\gamma_1 - \gamma_2\|^2 < \|\gamma_1\| - \|\gamma_2\|$, but the IC constraint for type t_1 reporting type t_2 does not bind. This demonstrates that the condition in Proposition 1 is not sufficient in this general environment. In the example, type t_1 can get a larger surplus by reporting type t_3 than by reporting type t_2. To eliminate the incentive for type t_1 of reporting type t_3, the data broker will want to give a large information rent to type t_1, and this rent may be larger than the surplus type t_1 can obtain by reporting type t_2, so that the solution to the relaxed problem satisfies IC_{12}.

Example 2 *Suppose that* $\gamma_1 = (6,2)$, $\gamma_2 = (3.5,4)$ *and* $\gamma_3 = (5,0.5)$. *It can be checked that* $\|\gamma_1 - \gamma_2\|^2 < \|\gamma_1\|^2 - \|\gamma_2\|^2$ *and* $\|\gamma_1 - \gamma_3\|^2 < \|\gamma_1\|^2 - \|\gamma_3\|^2$. *However,* $12.81 = \frac{(\gamma_1^T\gamma_3)^2}{\|\gamma_3\|^2} - \|\gamma_3\|^2 > 1.52 = \frac{(\gamma_1^T\gamma_2)^2}{\|\gamma_2\|^2} - \|\gamma_2\|^2$. *That is, without any reduction in quality and when charging them their willingness to pay, type* t_1 *can obtain a higher rent by reporting type* t_3 *than by reporting type* t_2. *Consider the relaxed problem with only the constraint IC 1-3. In*

[26] The surplus is positive as long as $\frac{\alpha_2}{\alpha_2+\alpha_3} < 0.279$.

[27] The complete results of the numerical solution for this example and those following are available upon request.

this problem the seller targets to types t_1 and t_2 the undistorted statistics and charges type t_2 according to her willingness to pay. Therefore, by mimicking type t_2, type t_1 can obtain a rent of $1.52 = \frac{(\gamma_1^T \gamma_2)^2}{\gamma_2^T \gamma_2} - \gamma_2^T \gamma_2$. *At the same time, when α_1 is small relative to α_3, by Corollary 1, the information rent given to type t_1 is close to* $12.81 = \frac{(\gamma_1^T \gamma_3)^2}{\gamma_3^T \gamma_3} - \gamma_3^T \gamma_3$. *Then, for α_1 small, the constraint IC 1-2 is satisfied.*[28]

The problem with two information types was simplified significantly when I show that the upward IC constraint never binds. The next two examples present instances in which some upward constraints do bind.

Example 3 presents a case in which the solution to the two-type problem involves modifying the direction of the statistic targeted to one type away from the mimicking type's most preferred statistic. While this avoids a situation one type mimics the other, it also makes the distorted statistic more valuable for a third type. If this distortion is large enough it gives the third type an incentive to report the originally mimicked type.

Example 3 *Suppose that $\gamma_1 = (6, 2)$, $\gamma_2 = (5, 1.8)$, and $\gamma_3 = (0, 5)$. If none of the IC constraints involving type t_3 bind, the data broker will offer to this type a statistic that she considers equivalent to receiving all the data at a price equal to her willingness to pay. In the solution to the relaxed problem with only the constraint* IC 1-2, *the data broker distorts the statistic targeted to type t_2 in the opposite direction of type t_1's preferred statistic—that is, by rotating it towards type t_3's preferred direction.*

Figure 6 presents the optimal recommendations in this relaxed problem for types t_1 and t_2 when $\frac{\alpha_1}{\alpha_1+\alpha_2} = 0.8$, *where the length of the recommendation vectors represents the optimal price. The statistic λ_2 that is targeted to type t_2 is almost in the same direction as γ_3, and it is offered at a low price relative to type t_3's willingness to pay. Therefore, the constraint IC 3-2 is not satisfied.*[29]

It can be shown that constraints IC 1-2 and IC 3-2 are the ones that bind. More interestingly, in the optimal mechanism the data broker only targets to types t_1 and t_3 a statistic that for them is equivalent to observing all the data, and he charges them a price equal to their willingness to

[28] In this example, this happens whenever $\frac{\alpha_1}{\alpha_1+\alpha_3} < 0.58$.
[29] This is true as long as $\frac{\alpha_1}{\alpha_1+\alpha_2} > 0.7$.

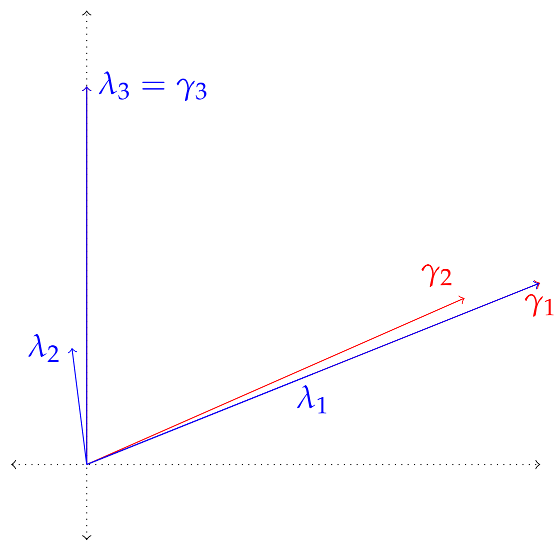


Figure 6: Optimal solution of the relaxed problem in Example 3 when $\frac{\alpha_1}{\alpha_1+\alpha_2} = 0.8$. The length of the blue vectors represents the optimal price.

pay. The data broker does not offer any valuable statistic to type t_2, even though type t_2 is not the one with the lowest willingness to pay.

Finally, Example 4 presents a case in which the reduction in the price charged to one type makes the upward constraint *IC 2-1* bind. This means that in this environment there could be distortions at the top—that is, the type with the highest willingness to pay for all the data may receive a distorted statistic.[30]

Example 4 *Suppose that $\gamma_1 = (4.8, 2)$, $\gamma_2 = (5, 0)$, and $\gamma_3 = (4, 2)$. It is easy to check that $\|\gamma_1 - \gamma_3\|^2 < \|\gamma_1\|^2 - \|\gamma_3\|^2$ and that the other two analogous conditions hold with the opposite sign. Consider the relaxed problem with only constraint* IC 1-3*. Under such an assumption, neither type t_2 nor t_3 receives any surplus under the optimal mechanism, and by Corollary 1, when α_1 is small enough relative to α_3, type t_1 receives an information rent close to $\frac{(\gamma_1^T\gamma_3)^2}{\gamma_3^T\gamma_3} - \gamma_3^T\gamma_3 = 6.91$. Since in this relaxed problem the seller targets an undistorted statistic to type t_1, he charges type t_1 a price equal to $\gamma_1^T\gamma_1 - 6.91 = 20.12$. Given this price for the statistic targeted to type t_1, type t_2 has an incentive to report type t_1 since $\frac{(\gamma_1^T\gamma_2)^2}{\gamma_1^T\gamma_1} - 20.12 = 1.17$.*[31]

It can be shown that the constraints IC 1-3 and IC 2-1 are the ones that bind. In the optimal

[30] In Theorem 4, I show that there is always a type that receives an undistorted statistic. In this example, it is type t_2.

[31] The same is true as long as $\frac{\alpha_1}{\alpha_1+\alpha_3} < 0.5910$.

mechanism the data broker distorts the statistic targeted to type t_1 in the opposite direction of type t_2's preferred statistic and targets to type t_2 an undistorted statistic. This reduces type t_1's willingness to pay for the information targeted to her, but it does not change type t_1's incentive to report type t_3.

The examples show that it is difficult to know *a priori* the relevant IC constraints to consider when solving for the optimal mechanism in this general environment. Despite this difficulty, I am able to show that the optimal mechanism satisfies two important properties. First, at least one type receives zero information rent. If this were not the case, the data broker could easily increase profits by increasing all prices uniformly without affecting incentives. Second, there is at least one type that receives an undistorted statistic, in the sense that receiving this statistic is for that type equivalent to receiving all data. But this type is not necessarily the one with the highest willingness to pay for all the data. Example 4 presented one such a case, where the type with the second highest willingness to pay for all the data is the only one that receives an undistorted statistic. This is a generalization of the "nondistortion at the top" property that holds in unidimensional mechanism design problems with payoffs that satisfy the single crossing condition.[32] The reasoning behind this result is subtler. I show that if there is no type that receives an undistorted statistic, the data broker can offer a new package with all the data at a price high enough such that only one type is willing to pay for it, thus satisfying all constraints.

Theorem 4 *The optimal mechanism satisfies the following properties:*

1. *at least one type receives zero information rent; and*
2. *there is at least one type t_i that receives a nondistorted statistic; that is, from the statistic she can learn $E[\theta_i \mid X]$.*

Although I cannot explicitly find the solution to the seller's problem, I am able to characterize how the statistics that are offered in the optimal mechanism look. These

[32] Rochet and Choné (1998) show that in a multidimensional environment with linear payoffs, a similar property holds. In their environment there is at least one boundary type that receives the optimal quality.

statistics are built to satisfy the main property I found in the case of two information types: when an IC constraint binds, the data broker should ideally modify the statistic given to the mimicked type in the opposite direction of the information preferred by the mimicking type(s). There are two differences. First, there could be multiple types that want to report one single type, so the data broker must weight the direction in which to distort the information. Second, it might be optimal for the data broker not to sell a valuable statistic to certain types. It is worth noting that in this environment, as occurred in Example 3, it is not always true that the types who are willing to pay less for all the data are the ones that are left out of the mechanism. Instead the seller wants to leave out those types that other types have a higher incentive to report.

Theorem 5 *In the optimal mechanism, the vector of coefficients of the statistic targeted to type t_i, to whom the data broker wants to sell, is one of the solutions to the fixed-point problem*

$$L_i = Var(X)^{-1} \left(Cov(\theta_i, X) - \sum_{j \in I, j \neq i} \frac{\lambda_{ji} c_{ji}}{\sum_{j \in I, j \neq i} \lambda_{ij} + \mu_i} Cov(\theta_j, X) \right)$$

where $I \subseteq \{1, \ldots, n\}$ is the set of types to whom the data broker wants to sell, $\lambda_{k\ell}$ is the Lagrange multiplier associated with the constraint IC k-ℓ, *μ_i is the Lagrange multiplier associated with constraint* IR i, *and $c_{k\ell} = \frac{L_k^T Cov(\theta_k, X)}{L_\ell^T Cov(\theta_\ell, X)}$.*

References

Anderson, S. P., and L. Celik (2015): "Product line design," *Journal of Economic Theory*, 157, 517–526.

Armstrong, M. (1996): "Multiproduct nonlinear pricing," *Econometrica: Journal of the Econometric Society*, pp. 51–75.

Babaioff, M., R. Kleinberg, and R. Paes Leme (2012): "Optimal mechanisms for selling information," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 92–109. ACM.

Balestrieri, F., S. Izmalkov, and J. Leao (2015): "The market for surprises: selling substitute goods through lotteries," .

Basar, T. (2008): "Variations on the theme of the Witsenhausen counterexample," in *2008 47th IEEE Conference on Decision and Control*, pp. 1614–1619. IEEE.

Basar, T., and G. J. Olsder (1999): *Dynamic noncooperative game theory*, vol. 23. Siam.

Bergemann, D., and A. Bonatti (2015): "Selling cookies," *American Economic Journal: Microeconomics*, 7(3), 259–94.

Bergemann, D., A. Bonatti, and A. Smolin (2018): "The design and price of information," *American Economic Review*, 108(1), 1–48.

Daskalakis, C., A. Deckelbaum, and C. Tzamos (2017): "Strong Duality for a Multiple-Good Monopolist," *Econometrica*, 85(3), 735–767.

Fang, K., S. Kotz, and K. Ng (1990): *Symmetric multivariate and related distributions*. Chapman and Hall.

Federal Trade Commission (2014): "Data brokers: A call for transparency and accountability," *Washington, DC*.

Kamenica, E., and M. Gentzkow (2011): "Bayesian persuasion," *American Economic Review*, 101(6), 2590–2615.

Manelli, A. M., and D. R. Vincent (2006): "Bundling as an optimal selling mechanism for a multiple-good monopolist," *Journal of Economic Theory*, 127(1), 1–35.

Marr, B. (2017): "Where Can You Buy Big Data? Here Are The Biggest Consumer Data Brokers," .

Maskin, E., and J. Riley (1984): "Monopoly with incomplete information," *The RAND Journal of Economics*, 15(2), 171–196.

Mussa, M., and S. Rosen (1978): "Monopoly and product quality," *Journal of Economic theory*, 18(2), 301–317.

Myerson, R. B. (1981): "Optimal auction design," *Mathematics of operations research*, 6(1), 58–73.

Pavlov, G. (2011): "Optimal mechanism for selling two goods," *The BE Journal of Theoretical Economics*, 11(1).

Rochet, J.-C., and P. Choné (1998): "Ironing, sweeping, and multidimensional screening," *Econometrica*, pp. 783–826.

Thanassoulis, J. (2004): "Haggling over substitutes," *Journal of Economic theory*, 117(2), 217–245.

Vohra, R. V. (2011): *Mechanism design: a linear programming approach*, vol. 47. Cambridge University Press.

Wilson, R. B. (1993): *Nonlinear pricing*. Oxford University Press on Demand.

Witsenhausen, H. S. (1968): "A counterexample in stochastic optimum control," *SIAM Journal on Control*, 6(1), 131–147.

Yang, K. H. (2018): "Non-linear Pricing of Information Structure: An Equivalence Result," .

A A Micro Foundation for Quadratic Loss

Suppose there is a monopolist that faces a linear demand $q(p) = \bar{a} - bp$ and the monopolist knows b but he is uncertain about the value of $\bar{a}$.

If the monopolist knows that the value of the intercept is $\bar{a}$ he would choose to produce a quantity $\bar{q} = \frac{\bar{a}}{2}$ and he would receive a profit of $\bar{\pi} = \frac{\bar{a}^2}{4b}$.

Since the monopolist is not certain about $\bar{a}$ he would choose some quantity q. Let $a = 2q$, that is, the monopolist chooses the quantity q as if he thinks that $\bar{a} = a$. By choosing this quantity the monopolist obtains profits $\pi = q\left(\frac{\bar{a}}{b} - \frac{q}{b}\right) = \frac{a}{2}\left(\frac{\bar{a}}{b} - \frac{a}{2b}\right) = \frac{a\bar{a}}{2b} - \frac{a^2}{4b}$.

Then by choosing q the monopolist is leaving on the table the quantity

$$\bar{\pi} - \pi = \frac{\bar{a}^2}{4b} - \frac{a\bar{a}}{2b} + \frac{a^2}{4b} = \frac{(a-\bar{a})^2}{4b},$$

that is, the monopolist weights the losses of its uncertainty about the parameter $\bar{a}$ according to a quadratic loss function.

B Proofs

Proof Claim 2

For any function $g : \psi(X) \to \mathbb{R}$, where $\psi(X)$ is the set of possible realizations of statistics,

$$\begin{aligned} -\mathbb{E}_{x,\theta}\left[(g(\psi(x)) - \theta)^2 \mid \psi(x)\right] &\leq -\mathbb{E}_{X,\theta}\left[(\mathbb{E}(\theta \mid \psi(x)) - \theta)^2 \mid \psi(x)\right] \\ &= -\mathbb{E}_X\left[\mathbb{E}_\theta\left[(\mathbb{E}(\theta \mid \psi(x)) - \theta)^2 \mid \psi(x)\right]\right] = -\mathbb{E}_X[Var(\theta \mid \psi(x))]. \end{aligned}$$

The inequality follows because $\mathbb{E}\left[(b-\theta)^2 | \psi(x)\right]$ is minimized by setting $b = \mathbb{E}[\theta|\psi(x)]$. The first equality follows from the Law of Iterated Expectations, and the second equality follows from the definition of conditional variance.

Proof Theorem 1

First, I present a detailed argument of why selling linear mechanisms is optimal when there are only two information types that share a common valuation type, and then argue why this argument still applies to the general model.

Suppose there are two information types with a unique common valuation type. Since by Proposition 1 the IC constraint for type t_2 mimicking type t_1 does not bind, the problem that the Data Broker faces when the IC constraint for type t_1 mimicking type t_2 binds, can be rewritten as

$$\begin{aligned} \max_{\psi(t_2)} \quad & \alpha_1\left(\gamma_1^T\gamma_1 - Var(\theta_1) + \mathbb{E}[Var(\theta_1 \mid \psi(t_2))]\right) + Var(\theta_2) - \mathbb{E}[Var(\theta_2 \mid \psi(t_2))] \\ \text{s.t} \quad & Var(\theta_1) - \mathbb{E}[Var(\theta_1 \mid \psi(t_2))] \geq Var(\theta_2) - \mathbb{E}[Var(\theta_2 \mid \psi(t_2))], \end{aligned}$$

where the remaining constraint represents the IR constraint for type t_1.

First, suppose that the remaining constraint does not bind. By an argument similar to the one in Kamenica and Gentzkow (2011), it is without loss to assume that the Data Broker gives to type t_2 a recommendation that this type will follow, that is, $E(\theta_2 \mid \psi(t_2)) = \psi(t_2) \in \mathbb{R}$. This is true because, by Claim 2, type t_2 learns the same by observing the statistic $\psi(t_2)$ or by observing an a statistic the expected value of θ_2 given the original statistic, while type t_1 cannot learn more about θ_2 by observing the modified, probably aggregate, statistic than by observing the original one. Claim 2 implies that $\mathbb{E}[Var(\theta_1 \mid \psi(t_2))]$ is the smallest variance that type t_1 can reached by optimally choosing the updating rule he uses to forecast θ_1 when buying the statistic targeted to type t_2, $a_1(\psi(t_2))$. By allowing type t_1 to choose a_1, I can define the functional

$$G(a_1, \psi(t_2)) = \alpha_1 \mathbb{E}(\theta_1 - a_1(\psi(t_2)(x)))^2 - \mathbb{E}(\theta_2 - \psi(t_2)(x) \mid x)^2.$$

Then the solution to the original problem has to be an equilibrium of the stochastic zero-sum game in which type t_1 wants to minimize G, where type t_1 chooses a_1 to minimize G and the Data Broker chooses $\psi(t_2)$ to maximize G.

First, suppose that $\psi(t_2)$ is a linear combination of x. Claim 1 implies that the random vector $(\theta_1, \psi(t_2)(x))$ is jointly elliptical. By Claim 2, type t_1 chooses $a_1^* = E(\theta_1 \mid \psi(t_2)(x))$, a linear function of $\psi(t_2)$ since $(\theta_1, \psi(t_2)(x))$ is elliptical.

Second, suppose that a_1 is linear, that is, $a_1 = c_1\psi(t_2)(x) + c_0$. In such a case the seller wants to maximize

$$\mathbb{E}\left(\alpha_1(\theta_1 - c_1\psi(t_2)(x) - c0)^2 - (\theta_2 - \psi(t_2)(x))^2 \mid x\right).$$

The First Order Condition implies that the optimal solution for the seller has to satisfy

$$\psi^*(t_2)(x) = \frac{\mathbb{E}[\theta_2 \mid x] - \alpha_1 c_1 \mathbb{E}[\theta_1 \mid x] - \alpha_1 c_0}{1 - \alpha_1 c_1^2},$$

a linear mapping since the conditional expectations are linear. Furthermore, the seller's problem is strictly concave as long as $1 > \alpha_1 c_1^2$, and I show in Theorem 2 that there is a unique fixed point of this linear mapping that satisfies this condition.

Therefore, the pair $(a_1^*, \psi^*(t_2))$ is a linear saddle-point strategy, that is, for any other a_1 and $\psi(t_2)$ I have $G(a_1^*, \psi(t_2)) \leq G(a_1^*, \psi^*(t_2)) \leq G(a_1, \psi^*(t_2))$. To finish the argument

I use a well-known property of zero-sum games called *ordered interchangeability property* (see for example Basar and Olsder (1999)). For completeness I present a general proof of this property.

Property 1 *In a simultaneous two-player zero-sum game, let* (a_1^*, a_2^*) *and* $(\tilde{a}_1, \tilde{a}_a)$ *be two saddle-point strategies. Then* $(a_1^*, \tilde{a}_2)$ *and* $(\tilde{a}_1, a_2^*)$ *are also saddle-point strategies.*

Proof Let u be the payoff function of the zero-sum game, and suppose that player 1 wants to minimize u, while player 2 wants to maximize it. Let (a_1^*, a_2^*) and $(\tilde{a}_1, \tilde{a}_a)$ be two saddle-point strategies, that is, for any a_1 and a_2, $u(a_1, a_2^*) \geq u(a_1^*, a_2^*) \geq u(a_1^*, a_2)$, and analogously for and $(\tilde{a}_1, \tilde{a}_2)$.

Then $\max_{a_2} u(a_1^*, a_2) \leq u(a_1^*, a_2^*) \leq u(a_1, a_2^*) \leq \max_{a_2} u(a_1, a_2) \; \forall a_1$, that is, $a_1^* \in \arg\min_{a_1} \max_{a_2} u(a_1, a_2)$. This means that a_1^* is a security strategy for player 1. Similarly, it can be argued that a_2^* is a security point for player 2, and conclude that $\min_{a_1} \max_{a_2} u(a_1, a_2) = \max_{a_2} \min_{a_1} u(a_1, a_2) = v = u(a_1^*, a_2^*) = u(\tilde{a}_1, \tilde{a}_2)$, where the last inequality follows from an analogous argument.

Then $v = min_{a_1} u(a_1, \tilde{a}_2) \leq u(a_1^*, \tilde{a}_2) \leq max_{a_2} u(a_1^*, a_2) = v$, implying that $v = u(a_1^*, \tilde{a}_2)$. This means that $(a_1^*, \tilde{a}_2)$ is a saddle-point strategy, since by the definition of v for any a_1 and a_2, $u(a_1, \tilde{a}_2) \geq u(a_1^*, \tilde{a}_2) \geq u(a_1^*, a_2)$. By an analogous argument, $(\tilde{a}_1^*, a_2^*)$ is also a saddle-point strategy. ∎

The ordered interchangeability property implies that the linear addle-point strategy $(a_1^*, \psi^*(t_2))$ is the unique saddle point of the zero-sum game between the data broker and type t_1; if $(\hat{a}_1, \hat{\psi}(t_2))$ were another saddle-point strategy then the pair $(a_1^*, \hat{\psi}(t_2))$ is also a saddle-point strategy. If this is the case, by Claim 2 it has to be that $E(\theta_1 \mid \tilde{\psi}(t_2)(x))$ is a linear mapping with the same coefficients as $E(\theta_1 \mid \psi^*(t_2)(x))$, meaning that $\tilde{\psi}(t_2)(x)$ and $\psi^*(t_2)(x)$ generate the same conditioning σ-algebras, that is, they are equal almost everywhere.

If the constraint does bind, the argument has to be slightly modified. Let a_1^* be type t_1's optimal updating rule as a function of the statistic that is offered by the seller. The constraint can be rewritten as $Var(\theta_1) - E[(\theta_1 - a_1^*(\psi^*(t_2)))^2] = Var(\theta_2) - E[Var(\theta_2 - \psi^*(t_2))^2]$. Define $\Psi_{\mathcal{F}} = \{\psi(t_2) : (\psi(t_2), a_1^*(\psi(t_2)))$ satisfies the constraint$\}$. The modified zero-sum game where the seller's available strategies are given by the set $\Psi_{\mathcal{F}}$ has a unique linear saddle-point. The argument is analogous to the previous one. In particular, the seller's best response is linear when type t_1's updating rule is linear since the constraint is quadratic in such a case. The proof of Theorem 2 shows that the mapping created by this linear responses has at least one fixed point and one of them is the solution to the seller's problem.

Now I extend the argument to consider the general case with many information types and many valuation types. I assume that I know which are the constraints that bind. This is without loss since I argue that for any set of binding constraints the optimal mechanism is linear.

It is always possible to either write $p(t_i, v) = v(Var(\theta_i) - \mathbb{E}[(\theta_i - \psi(t_i))^2]$, if there is not IC constraint for type (t_i, v) reporting another type (t_j, v') that binds, or $p(t_i) = -v\mathbb{E}[(\theta_i - \psi(t_i))^2] + v\mathbb{E}[(\theta_i - a_{iv,j}(\psi(\theta_j)))^2] + p(t_j)$, if there is an IC constraint for type (t_i, v) reporting another type (t_j, v') that binds, where $a_{iv,j}$ represents how type (t_i, v) updates his forecast when observing the signal targeted to type (t_j, v'). Plugging in one of these prices in the seller's objective generates a functional analogous to functional G. The critical property of the functional G is that, for any i, j, $\mathbb{E}[(\theta_i - \psi(t_i))^2]$ always appears with a negative sign and $\mathbb{E}[(\theta_i - a_{iv,j}(\psi(t_i)))^2]$ always appears with a positive sign. This keeps the conflict that the seller wants to maximize $\mathbb{E}[(\theta_i - a_{iv,j}(\psi(t_i)))^2]$ and type (t_i, v) wants to minimize it, and the property that the seller wants to minimize $\mathbb{E}[(\theta_i - \psi(t_i))^2]$. If there are multiples ways to write the price for some type (t_i, v), some extra constraints have to be satisfied. Let $\Psi_{\mathcal{F}} = \{\psi : (\psi, a^*(\psi))$ satisfies all the constraints$\}$, where a^* denotes the buyers' optimal decisions when they buy a statistic in ψ that is not targeted to them.

Consider the zero-sum game between the seller that wants to maximize G by choosing any ψ in $\Psi_{\mathcal{F}}$ and a fictitious player that wants to minimize G by choosing the updating rules a, representing the buyers' interests in the mechanism. By the assumption that the joint distribution is elliptical and Claim 2, if the seller offers linear statistics, the fictitious player wants to choose linear forecasting rules. At the same time, if the fictitious player announces linear forecasting rules, the seller's best response is to choose linear statistics. This is a result of the problem being analogous to the one before with some extra constraints that are linear combinations of quadratic terms. Theorem 5 shows that this linear best response mapping has at least one fixed point, and that one of them is a solution to the seller's problem when the buyer uses linear forecasting rules. Therefore, by the same argument as in the two type case, there is a unique saddle-point strategy in which the seller offers linear statistics and the fictitious player uses linear forecasting rules.

Proof Lemma 1

Consider the family of functions $g(L,\ell,k) = \mathbb{E}_{\theta_i}[\theta_i] + k(L^T X + \ell\epsilon - \mathbb{E}_X[L^T X])$ for $k \in \mathbb{R}$. The conditional expectation belongs to this family of functions since the joint distribution of (θ, X) is elliptical. When the buyer uses $g(L,\ell,k)$ as her estimator when observing $L^T X + \ell\epsilon$, the variance of θ_i's forecasting error is equal to

$$\begin{aligned}
\mathbb{E}_{X,\epsilon}[\mathbb{E}_{\theta_i}[(\theta_i - g(L,\ell,k))^2 \mid L^T X + \ell\epsilon]] &= \mathbb{E}_{X,\epsilon}[\mathbb{E}_{\theta_i}[((\theta_i - \mathbb{E}_{\theta_i}[\theta_i] - kL^T(X - \mathbb{E}_X[X]) - k\ell\epsilon)^2) \mid L^T X + \ell\epsilon]] \\
&= \mathbb{E}_{\theta_i}[(\theta_i - E(\theta_i))^2] - 2kL^T\mathbb{E}_{X,\epsilon}(\mathbb{E}_{\theta_i}((\theta_i - \mathbb{E}_{\theta_i}[\theta_i])(X - \mathbb{E}_X[X])) \mid L^T X + \ell\epsilon)) \\
&\quad + k^2 L^T \mathbb{E}_{X,\epsilon}[\mathbb{E}_{\theta_i}[(X - E(X))(X - E(X))^T \mid L^T X + \ell\epsilon]]L + k^2\ell^2\mathbb{E}_\epsilon[\epsilon^2] \\
&= \mathbb{E}_{\theta_i}[(\theta_i - \mathbb{E}_{\theta_i}(\theta_i))^2] - 2kL^T\mathbb{E}_{X,\theta_i}[(\theta_i - E_{\theta_i}[\theta_i])(X - \mathbb{E}_X[X])] \\
&\quad + k^2 L^T \mathbb{E}_X[(X - \mathbb{E}_X[X])(X - \mathbb{E}_X[X])^T]L + k^2\ell^2\sigma^2 \\
&= Var(\theta_i) - 2kL^T Cov(\theta_i, X) + k^2 L^T Var(X) L + k^2\ell^2\sigma^2
\end{aligned}$$

where the the second equality follows from linearity of the expectation and from ϵ being independent of X and θ, the third one from the Law of Iterated Expectations and the last one from the definition of variance and covariance. Claim 2 implies that the conditional expectation has to minimized this forecasting error. The First Order Condition with respect to k is

$$-2L^T Cov(\theta_i, X) + 2kL^T Var(X)L + 2k\ell^2\sigma^2 = 0,$$

that is, $\hat{k} = \frac{L^T Cov(\theta_i, X)}{L^T Var(X) L + \ell^2\sigma^2}$. Plugging in this value of k gives the conditional variance, which is equal to

$$Var(\theta_i) - \frac{(L^T Cov(\theta_i, X))^2}{L^T Var(X) L + \ell^2 \sigma^2}.$$

Proof Lemma 2

Lemma 1 presented the value of the conditional variance when the buyer observes any linear combination of the data. Since the conditional expectation minimizes the conditional variance, it is enough to find the set of coefficients that minimize the expression in Lemma 1. Immediately it can be seen that $\ell = 0$. The first order condition for L gives

$$\frac{-2L^T Cov(\theta_i, X) Cov(\theta_i, X) L^T Var(X) L + 2(L^T Cov(\theta_i, X))^2 L^T Var(X)}{(L^T Var(X) L)^2} = 0,$$

from which we obtain that $\hat{L} = Var(X)^{-1} Cov(\theta_i, X)$. Therefore, the conditional variance of θ_i given X is equal to

$$Var(\theta_i) - Cov(\theta_i, X)^T Var(X)^{-1} Cov(\theta_i, X).$$

Proof Proposition 1

To prove parts 1. and 2. I only need to argue that the IC constraint for type t_2 reporting type t_1 never binds. Suppose that in the optimal mechanism that targets statistics $\psi(t_1)$ and $\psi(t_2)$ to types t_1 and t_2, respectively, this constraint binds. Then it holds with equality, and by Claim 2 it can be expressed as

$$Var_{F_\theta}(\theta_2) - \mathbb{E}_X[Var(\theta_2 \mid \psi(t_1))] - p(t_1) = Var_{F_\theta}(\theta_2) - \mathbb{E}_X[Var(\theta_2 \mid \psi(t_2))] - p(t_2). \quad (1)$$

Lemma 2 implies that $\gamma_2^T \gamma_2 \geq Var_{F_\theta}(\theta_2) - E_x[Var(\theta_2 \mid \psi(t_1))]$. Therefore, $p(t_1) \leq \gamma_2^T \gamma_2 - Var_{F_\theta}(\theta_2) + E_x[Var(\theta_2 \mid \psi(t_2))] + p(t_2)$. Since the IR constraint for type t_2 implies that $Var_{F_\theta}(\theta_2) - E_x[Var(\theta_2 \mid \psi(t_2))] \geq p(t_2)$, it has to be that $p(t_1) \leq \gamma_2^T \gamma_2$. Similarly, $p(t_2) \leq \gamma_2^T \gamma_2$.

If one of these constraints is strict, the monopolist can do better by selling all the data to both types, and charging to them the price $\gamma_2^T \gamma_2$, satisfying trivially all the constraints.

Now suppose that $p(t_1) = p(t_2) = \gamma_2^T \gamma_2$. In such a case to satisfy the IR constraint for type t_2 it has to be that $\psi(t_2)(x) = E(\theta_2 \mid x)$. Consider the alternative mechanism that targets to type t_1 the statistic $\psi'(t_1)$ and charge him the price $p'(t_1) =$

$\gamma_1^T\gamma_1 - Var_{F_\theta}(\theta_2) + E_x[Var(\theta_1 \mid \psi(t_2))] + p(t_2) > p(t_1)$, since by Claim 2 and assumption 2, $\gamma_1^T\gamma_1 > Var_{F_\theta}(\theta_2) - E_x[Var(\theta_1 \mid \psi(t_2))]$. Clearly this mechanism satisfies all the constraints and gives to the monopolist higher profits than the original mechanism, contradicting the assumption that the original mechanism was optimal.

Now I proceed to prove part 3. According to Lemma 2, when the data broker offers to type t_2 the statistic $\psi(t_2) = \mathbb{E}[\theta_2 \mid X]$, the maximum price type t_2 is willing to pay for it is $\gamma_2^T\gamma_2 = \|\gamma_2\|^2$, and the maximum price that type t_1 is willing to pay for the same statistic is

$$\frac{(Cov(\theta_2, X)^T Var(X)^{-1} Cov(\theta_1, X))^2}{Cov(\theta_2, X)^T Var(X)^{-1} Cov(\theta_2, X)} = \frac{(\gamma_1^T\gamma_2)^2}{\|\gamma_2\|^2}.$$

Therefore, when the data broker targets to type t_2 the statistic $\psi(t_2) = \mathbb{E}[\theta_2 \mid X]$ and charges him price $p(t_2) = \|\gamma_2\|^2$, type t_1 will report type t_2 if and only if

$$(Cov(\theta_2, X)^T Var(X)^{-1} Cov(\theta_1, X))^2 > (Cov(\theta_2, X)^T Var(X)^{-1} Cov(\theta_2, X))^2$$

$$\Leftrightarrow \quad (\gamma_1^T\gamma_2)^2 > (\gamma_2^T\gamma_2)^2$$

$$\Leftrightarrow \quad cos^2(\beta)\gamma_2^T\gamma_2\gamma_1^T\gamma_1 > (\gamma_2^T\gamma_2)^2$$

$$\Leftrightarrow \quad cos(\beta) > \tfrac{\|\gamma_2\|}{\|\gamma_1\|},$$

where β is the angle between γ_1 and γ_2, and the second equivalence follows from the fact that the angle between vectors γ_1 and γ_2 satisfies the equality $cos(\beta) = \frac{\gamma_1^T\gamma_2}{\|\gamma_1\|\|\gamma_2\|}$, while the last equivalence from the assumption that $\gamma_1^T\gamma_2 > 0$. This provides the first equivalence in the theorem. Since $\gamma_2^T\gamma_1 > 0$, the expression after the first equivalence can be rewritten as $\gamma_1^T\gamma_2 - \gamma_2^T\gamma_2 > 0$, that is, the dot product between the vectors γ_2 and $\gamma_1 - \gamma_2$ has to be positive. Figure 7 helps to understand how to derive the second condition.

The dot product between two vectors is positive if and only if the angle between them is less than 90°. Therefore, $\gamma_2^T(\gamma_1 - \gamma_2) > 0$ if and only if the angle between this two vectors is smaller than 90°. In the figure this implies that $\alpha + \beta < 90°$. By the properties of the measures of angles between parallel lines, $\alpha = \alpha'$. Therefore, the IC constraints will bind if and only if $\omega > 90°$. By Pythagoras theorem it has to be that the squared of

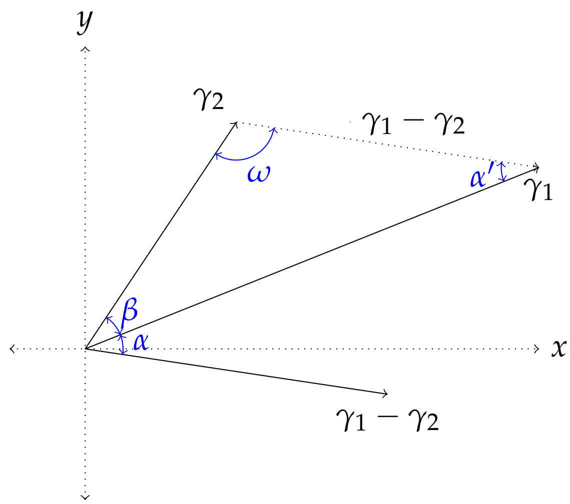


Figure 7: Graphic representation of covariance adjusted vectors.

the magnitude of the vector γ_1 is larger than the sum of the square of the magnitudes of vectors γ_2 and $\gamma_1 - \gamma_2$, that is, $\|\gamma_1\|^2 > \|\gamma_2\|^2 + \|\gamma_1 - \gamma_2\|^2$, which implies the second condition.

Finally, I proceed to prove part 4. If the IC constraint for type t_1 reporting type t_1 does not bind the seller can give the optimal statistics to each type, so it does not include any noise. Now suppose it binds. Since the IR constraint for type t_2 binds and type t_1 receives the optimal amount of information the prices are given by

$$p(t_1) = p(t_2) + \gamma_1^T \gamma_1 - \frac{(L_2^T Cov(\theta_1, X))^2}{L_2^T Var(X) L_2 + \ell_2^2 \sigma^2}, \text{ and}$$

$$p(t_2) = \frac{(L_2^T Cov(\theta_2, X))^2}{L_2^T Var(X) L_2 + \ell_2^2 \sigma^2}.$$

Therefore, the objective of the seller is to solve the problem

$$\pi = \max_{L_2, \ell_2} \alpha_1 \left(\gamma_1^T \gamma_1 - \frac{(L_2^T Cov(\theta_1, X))^2}{L_2^T Var(X) L_2 + \ell_2^2 \sigma^2} \right) + \frac{(L_2^T Cov(\theta_2, X))^2}{L_2^T Var(X) L_2 + \ell_2^2 \sigma^2}$$

Suppose that in the solution of this problem $\ell_2^* > 0$. If we reduce ℓ_2 by an small amount, the change in the seller's profits is equal to

$$2\sigma^2 \frac{(L_2^T Cov(\theta_2, X))^2}{(L_2^T Var(X) L_2 + \ell^{*2}_2 \sigma^2)^2} - 2\sigma^2 \alpha_1 \frac{(L_2^T Cov(\theta_1, X))^2}{(L_2^T Var(X) L_2 + \ell^{*2}_2 \sigma^2)^2}.$$

Suppose that $\alpha_1 (L_2^T Cov(\theta_1, X))^2 \geq (L_2^T Cov(\theta_2, X))^2$. Then $\pi \leq \alpha_1 \gamma_1^T \gamma_1$. Consider the alternative mechanism with coefficients $\tilde{L}_2$, such that $\tilde{L}_2$ is orthogonal to $Cov(\theta_2) - Cov(\theta_1)$, that is, $\tilde{L}_2^T Cov(\theta_1, X) = \tilde{L}_2^T Cov(\theta_1, X)$, and prices $\tilde{p}(t_1) = \gamma_1^T \gamma_1$ and $\tilde{p}(t_2) = \frac{(\tilde{L}_2^T Cov(\theta_2,X))^2}{\tilde{L}_2^T Var(X) \tilde{L}_2}$. $\tilde{p}(t_2)$ is strictly positive since $\tilde{L}_2$ is not orthogonal to $Cov(\theta_2)$ by assumption 2. The new contract satisfies all IC and IR constraints and give a profit $\tilde{\pi} = \alpha_1 \gamma_1^T \gamma_1 + \tilde{p}(t_2) > \pi$. Therefore, $\alpha_1 (L_2^T Cov(\theta_1, X))^2 < (L_2^T Cov(\theta_2, X))^2$, and the net benefit of decreasing ℓ^* is positive. Therefore, $\ell^* > 0$ cannot be the optimal solution, that is, the data broker will never include independent noise into the statistic.

Proof of Theorem 2

Part 1. follows directly from part 3. in Proposition 1 and Lemma 1.

To solve for the optimal mechanism design when the IC constraint for type 1 reporting type 2 binds, I first assume that the IR constraint for type t_1 does not bind. From the analysis in the main text, the seller's maximization problem is:

$$\pi = \max_L \alpha_1 \left(\gamma_1^T \gamma_1 - \frac{(L^T Cov(\theta_1, X))^2}{L^T Var(X) L} \right) + \frac{(L^T Cov(\theta_2, X))^2}{L^T Var(X) L}$$

Letting $a_1 = \frac{L^T Cov(\theta_1,X)}{L^T Var(X) L}$ and $a_2 = \frac{L^T Cov(\theta_2,X)}{L^T Var(X) L}$, the First Order Condition is:

$$-2\alpha_1 (a_1 Cov(\theta_1, X) - a_1^2 \alpha_1 Var(X) L) + 2a_2 Cov(\theta_2, X) - 2a_2^2 Var(X) L = 0,$$

from which

$$L = \frac{Var(X)^{-1} \left(Cov(\theta_2, X) - \frac{a_1}{a_2} \alpha_1 Cov(\theta_1, X) \right)}{a_2 - \alpha_1 \frac{a_1^2}{a_2}}.$$

Rename $c = \frac{a_1}{a_2} = \frac{L^T Cov(\theta_1,X)}{L^T Cov(\theta_2,X)}$ and normalize the denominator to 1. Lemma 1 implies that this normalization is without loss. Therefore, if the IR constraint of type t_1 does not bind the result is true. The IR constraint for type t_1 does not bind if and only if

$$\begin{aligned}
& \frac{(\gamma_2^T \gamma_1 - c\alpha_1 \gamma_1^T \gamma_1)^2}{\|\gamma_2 - \alpha_1 c \gamma_1\|^2} > \frac{(\gamma_2^T \gamma_2 - c\alpha_1 \gamma_1^T \gamma_2)^2}{\|\gamma_2 - \alpha_1 c \gamma_1\|^2} \\
\Leftrightarrow \quad & \frac{(\gamma_2^T \gamma_1 - c\alpha_1 \gamma_1^T \gamma_1)^2}{(\gamma_2^T \gamma_2 - c\alpha_1 \gamma_1^T \gamma_2)^2} > 1 \\
\Leftrightarrow \quad & c^2 > 1.
\end{aligned}$$

Furthermore, from the proof of Theorem 1, the problem is strictly convex if and only if

$1 > \alpha_1 c^2$. Therefore, I need to check that there is a solution of the fixed-point problem with $c^2 < 1/\alpha_1$.

We have

$$c = \frac{\gamma_2^T \gamma_1 - c\alpha_1 \gamma_1^T \gamma_1}{\gamma_2^T \gamma_2 - c\alpha_1 \gamma_1^T \gamma_2},$$

which generates the quadratic equation $c^2 \alpha_1 \gamma_1^T \gamma_2 - c(\gamma_2^T \gamma_2 + \alpha_1 \gamma_1^T \gamma_1) + \gamma_1^T \gamma_2 = 0$. This equation has two solutions given by

$$\frac{\|\gamma_2\|^2 + \alpha_1 \|\gamma_1\|^2 \pm \sqrt{(\|\gamma_2\|^2 + \alpha_1 \|\gamma_1\|^2)^2 - 4\alpha_1 (\gamma_1^T \gamma_2)^2}}{2\alpha_1 \gamma_1^T \gamma_2}.$$

In both solutions the denominator is positive, so that the two solutions have the same sign and the sign coincide with the sign of $\gamma_1^T \gamma_2$.

To make the notation easier let c be the negative solution and d the positive solution. The following lemma shows that only the negative solution of the equation is a solution of the original problem and that the IR constraint binds only when $\alpha_1 > \tilde{\alpha}$ with $\tilde{\alpha} \in (0, 1)$.

Lemma 8 *For all α_1, $d^2 > 1/\alpha_1$, that is, d is never a solution of the problem, and $c^2 < 1/\alpha_1$, that is, c is always a solution of the problem. Furthermore, $c^2 > 1$ iff $\alpha_1 < \tilde{\alpha} = \frac{\gamma_2^T(\gamma_1 - \gamma_2)}{\gamma_1^T(\gamma_1 - \gamma_2)}$, that is, the IR constraint binds only if $\alpha_1 > \tilde{\alpha}$.*

Proof Since by assumption $\gamma_1^T \gamma_2 > 0$, both c and d are positive. I first prove that $d > 1/\alpha_1$ independently of the parameters. This is true if

$$\begin{aligned} d^2 > 1/\alpha_1 \quad &\Leftrightarrow \|\gamma_2\|^2 + \alpha_1 \|\gamma_1\|^2 + \sqrt{(\|\gamma_2\|^2 + \alpha_1 \|\gamma_1\|^2)^2 - 4\alpha_1 (\gamma_1^T \gamma_2)^2} > 2\sqrt{\alpha_1} \gamma_1^T \gamma_2 \\ &\Leftrightarrow \|\gamma_2\|^2 + \alpha_1 \|\gamma_1\|^2 - 2\sqrt{\alpha_1} \gamma_1^T \gamma_2 > -\sqrt{(\|\gamma_2\|^2 + \alpha_1 \|\gamma_1\|^2)^2 - 4\alpha_1 (\gamma_1^T \gamma_2)^2}. \end{aligned}$$

The RHS is always negative and the LHS is always positive since $\|\gamma_2\|^2 - 2\sqrt{\alpha_1} \gamma_1^T \gamma_2 + \alpha_1 \|\gamma_1\|^2 \geq (\|\gamma_2\| - \sqrt{\alpha_1} \|\gamma_1\|)^2 \geq 0$.

Now I prove that for any α_1, $c^2 < 1/\alpha_1$. From the definition of c

$$c^2 > 1/\alpha_1 \quad \Leftrightarrow \|\gamma_2\|^2 + \alpha_1\|\gamma_1\|^2 - \sqrt{(\|\gamma_2\|^2 + \alpha_1\|\gamma_1\|^2)^2 - 4\alpha_1(\gamma_1^T\gamma_2)^2} > 2\sqrt{\alpha_1}\gamma_1^T\gamma_2$$

$$\Leftrightarrow \|\gamma_2\|^2 + \alpha_1\|\gamma_1\|^2 - 2\sqrt{\alpha_1}\gamma_1^T\gamma_2 > \sqrt{(\|\gamma_2\|^2 + \alpha_1\|\gamma_1\|^2)^2 - 4\alpha_1(\gamma_1^T\gamma_2)^2}$$

$$\Leftrightarrow 8\alpha(\gamma_1^T\gamma_2)^2 \leq 4\sqrt{\alpha_1}(\gamma_1^T\gamma_2)(\gamma_2^T\gamma_2 + \alpha_1\gamma_1^T\gamma_1)$$

$$\Leftrightarrow 0 \leq \|\gamma_2\|^2 - 2\sqrt{\alpha_1}\gamma_1^T\gamma_1^2 + \alpha_1\|\gamma_1\|^2,$$

and it was argued before that this inequality is always true.

Finally, I prove that $c > 1$ only for small values of α_1. From the definition of c

$$c > 1 \quad \Leftrightarrow \|\gamma_2\|^2 + \alpha_1\|\gamma_1\|^2 - 2\alpha_1\gamma_1^T\gamma_2 > \sqrt{(\|\gamma_2\|^2 + \alpha_1\|\gamma_1\|^2)^2 - 4\alpha_1(\gamma_1^T\gamma_2)^2}$$

$$\Leftrightarrow -4\alpha_1\gamma_1^T\gamma_2(\|\gamma_2\|^2 + \alpha_1\|\gamma_1\|^2) + 4\alpha_1^2(\gamma_1^T\gamma_2)^2 > -4\alpha_1(\gamma_1^T\gamma_2)^2$$

$$\Leftrightarrow 0 > \gamma_2^T\gamma_2 + \alpha_1\gamma_1^T\gamma_1^2 - \alpha_1\gamma_1^T\gamma_2 - \gamma_1^T\gamma_2$$

$$\Leftrightarrow 0 > -\gamma_2^T(\gamma_1 - \gamma_2) + \alpha_1\gamma_1^T(\gamma_1 - \gamma_2) = (\alpha_1\gamma_1 - \gamma_2)^T(\gamma_1 - \gamma_2)$$

To complete the proof, notice that the vectors $\alpha_1\gamma_1 - \gamma_2$ and $\gamma_1 - \gamma_2$ form a triangle

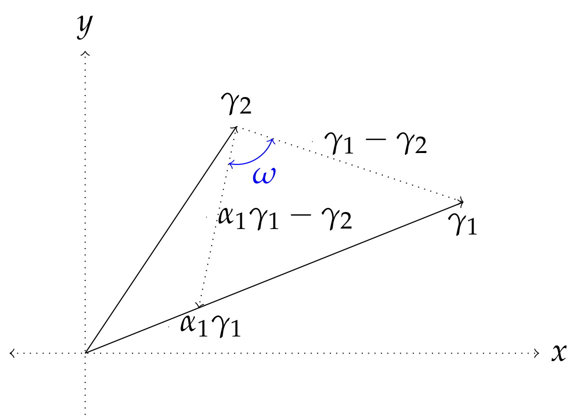


Figure 8: Graphic representation for proof of Lemma 8.

where the third side is given by the vector $(1 - \alpha_1)\gamma_1$ (See figure 8). As α_1 increases, the length of the vector $(1 - \alpha_1)\gamma_1$ decreases and by the sinus law the angle ω becomes smaller. As long as $\omega < 90°$, the condition in the last equation is satisfied, and when ω

becomes larger than 90° the inequality is reversed. To find when the measure of angle ω is exactly 90° notice that

$$-\gamma_2^T(\gamma_1-\gamma_2)+\alpha_1\gamma_1^T(\gamma_1-\gamma_2)=0$$

$$\Leftrightarrow \quad \alpha_1=\tilde{\alpha}=\tfrac{\gamma_2^T(\gamma_1-\gamma_2)}{\gamma_1^T(\gamma_1-\gamma_2)},$$

which is positive by Proposition 1, and by the fact that γ_1 is always between γ_2 and $\gamma_1-\gamma_2$. Besides, $\tilde{\alpha}<1$, since $\gamma_1^T\gamma_1-2\gamma_1^T\gamma_2+\gamma_2^T\gamma_2=\|\gamma_1-\gamma_2\|^2>0$. ■

Then the statistics in optimal mechanism are

$$\begin{aligned}\psi(t_1) &= \mathbb{E}[\theta_1]+Cov(\theta_1,X)^TVar(X)^{-1}(X-\mathbb{E}[X]) \text{ and}\\ \psi(t_2) &= E(\theta_2)+(Cov(\theta_2,X)-c\alpha_1Cov(\theta_1,X))^T\,Var(X)^{-1}(X-\mathbb{E}[X])\end{aligned}$$

with $c(\alpha_1)=\frac{\|\gamma_2\|^2+\alpha_1\|\gamma_1\|^2-\sqrt{(\|\gamma_2\|^2+\alpha_1\|\gamma_1\|^2)^2-4\alpha_1(\gamma_1^T\gamma_2)^2}}{2\alpha_1\gamma_1^T\gamma_2}$ if $\alpha<\tilde{\alpha}$, and $c(\alpha)=\tilde{\alpha}$, otherwise. When $\alpha_1<\tilde{\alpha}$, the optimal prices can be found by plugging in the IR constraint for type t_2 and the IC constraint for type t_1 and are equal to $p(t_1)=\frac{\|\gamma_1\|^2\|\gamma_2\|^2-(\gamma_1^T\gamma_2)^2}{\|\gamma_2-c\alpha_1\gamma_1\|^2}+t(\theta_2)$ and $p(t_2)=\frac{(\|\gamma_2\|^2-c\alpha_1\gamma_1^T\gamma_2)^2}{\|\gamma_2-c\alpha_1\gamma_1\|^2}$. If $\alpha>\tilde{\alpha}$, the optimal prices can be found by plugging in the IR constraints and they are equal to $p(t_1)=\|\gamma_1\|^2$ and $p(t_2)=\frac{(\|\gamma_2\|^2-\tilde{\alpha}\gamma_1^T\gamma_2)^2}{\|\gamma_2-\tilde{\alpha}\gamma_1\|^2}$.

Proof Corollary 1

I first prove 1. By the definition of the function c

$$\frac{\partial c(\alpha_1)}{\partial\alpha_1} = \frac{\gamma_1^T\gamma_1-1/2\left((\gamma_2^T\gamma_2+\alpha_1\gamma_1^T\gamma_1)^2-4\alpha_1(\gamma_1^T\gamma_2)^2\right)^{-0.5}\left(2(\gamma_2^T\gamma_2+\alpha_1\gamma_1^T\gamma_1)\gamma_1^T\gamma_1-4(\gamma_1^T\gamma_2)^2\right)}{2\gamma_1^T\gamma_2}$$

It is enough to argue that the numerator is positive. The numerator is non-negative iff

$$\begin{aligned}&\gamma_1^T\gamma_1\sqrt{(\gamma_2^T\gamma_2+\alpha_1\gamma_1^T\gamma_1)^2-4\alpha_1(\gamma_1^T\gamma_2)^2}\geq(\gamma_2^T\gamma_2+\alpha_1\gamma_1^T\gamma_1)\gamma_1^T\gamma_1-2(\gamma_1^T\gamma_2)^2\\ \Leftarrow\quad &(\gamma_1^T\gamma_1)^2((\gamma_2^T\gamma_2+\alpha_1\gamma_1^T\gamma_1)^2-4\alpha_1(\gamma_1^T\gamma_2)^2)\\ \geq\quad &4(\gamma_1^T\gamma_2)^4-4\gamma_1^T\gamma_1(\gamma_1^T\gamma_2)^2(\gamma_2^T\gamma_2+\alpha_1\gamma_1^T\gamma_1)+(\gamma_2^T\gamma_2+\alpha_1\gamma_1^T\gamma_1)^2(\gamma_1^T\gamma_1)^2\\ \Leftrightarrow\quad &4(\gamma_1^T\gamma_2)^2\gamma_2^T\gamma_2\gamma_1^T\gamma_1\geq4(\gamma_1^T\gamma_2)^4.\end{aligned}$$

The last inequality is always satisfied since $(\gamma_1^T\gamma_1)^2 = \|\gamma_1\|^2\|\gamma_2\|^2 cos^2(\beta) < \|\gamma_1\|^2\|\gamma_2\|^2$, where β is the angle between the vectors γ_1 and γ_2.

Further, see that

$$\lim_{\alpha_1\to 0} c(\alpha_1) = \lim_{\alpha_1\to 0}\frac{\|\gamma_2\|^2+\alpha_1\|\gamma_1\|^2-\sqrt{(\|\gamma_2\|^2+\alpha_1\|\gamma_1\|^2)^2-4\alpha_1(\gamma_1^T\gamma_2)^2}}{2\gamma_1^T\gamma_2} = 0.$$

For part 2., the information rent given to type t_1 is equal to

$$Rent = \frac{((\gamma_2-c(\alpha_1)\gamma_1)^T\gamma_1)^2-((\gamma_2-c(\alpha_1)\gamma_1)^T\gamma_2)^2}{(\gamma_2-c(\alpha_1)\gamma_1)^T(\gamma_2-c(\alpha_1)\gamma_1)},$$

and by taking the derivative with respect to $c(\alpha_1)$ I obtain

$$\begin{aligned}
\frac{\partial Rent}{\partial c(\alpha_1)} &\propto \left(-(\gamma_1^T\gamma_2\gamma_1^T\gamma_1-\gamma_2^T\gamma_2\gamma_1^T\gamma_2)+c(\alpha_1)((\gamma_1^T\gamma_1)^2-(\gamma_1^T\gamma_2)^2)\right)\left(\gamma_2^T\gamma_2-2c(\alpha_1)\gamma_1^T\gamma_2+c^2(\alpha_1)\gamma_1^T\gamma_1\right)\\
&-\left(-\gamma_1^T\gamma_2+c(\alpha_1)\gamma_1^T\gamma_1\right)\left((\gamma_1^T\gamma_2)^2-(\gamma_2^T\gamma_2)^2-2c(\alpha_1)(\gamma_1^T\gamma_2\gamma_1^T\gamma_1-\gamma_2^T\gamma_2\gamma_1^T\gamma_2)+c^2(\alpha_1)((\gamma_1^T\gamma_1)^2-(\gamma_1^T\gamma_2)^2)\right)\\
&=(-\gamma_1^T\gamma_2+c(\alpha_1)(\gamma_1^T\gamma_1+\gamma_2^T\gamma_2)-c^2(\alpha_1)\gamma_1^T\gamma_2)(\gamma_1^T\gamma_1\gamma_2^T\gamma_2-(\gamma_1^T\gamma_2)^2)\\
&\propto -\gamma_1^T\gamma_2+c(\alpha_1)(\gamma_1^T\gamma_1+\gamma_2^T\gamma_2)-c^2(\alpha_1)\gamma_1^T\gamma_2
\end{aligned}$$

The last expression is increasing in $c(\alpha_1)$ since $\gamma_1^T\gamma_1+\gamma_2^T\gamma_2-2c\alpha_1\gamma_1^T\gamma_2 > \gamma_1^T\gamma_1+\gamma_2^T\gamma_2-2\gamma_1^T\gamma_2 = \|\gamma_1-\gamma_2\|^2 > 0$, where the first inequality follows from $c(\tilde{\alpha}) = \tilde{\alpha}$ and from $c(\alpha_1)$ being non-decreasing in α_1. Therefore, I only need to prove that the last expression is negative when evaluated at $\tilde{\alpha}$. This expression becomes

$$\begin{aligned}
&-(1+\tilde{\alpha}^2)\gamma_1^T\gamma_2+\tilde{\alpha}(\gamma_1^T\gamma_1+\gamma_2^T\gamma_2)\\
\propto\ & -\left((\gamma_1^T\gamma_1-\gamma_1^T\gamma_2)^2+(\gamma_1^T\gamma_2-\gamma_2^T\gamma_2)^2\right)\gamma_1^T\gamma_2+(\gamma_1^T\gamma_2-\gamma_2^T\gamma_2)(\gamma_1^T\gamma_1+\gamma_2^T\gamma_2)(\gamma_1^T\gamma_1-\gamma_1^T\gamma_2)\\
=\ & (-\gamma_1^T\gamma_1+2\gamma_1^T\gamma_2-\gamma_2^T\gamma_2)(\gamma_1^T\gamma_1\gamma_2^T\gamma_2-(\gamma_1^T\gamma_2)^2) = -\|\gamma_1-\gamma_2\|^2(\gamma_1^T\gamma_1\gamma_2^T\gamma_2-(\gamma_1^T\gamma_2)^2) < 0
\end{aligned}$$

Therefore, $\frac{\partial Rent}{\partial c(\alpha_1)} < 0$. Since by the first part $\frac{\partial c(\alpha_1)}{\alpha_1} > 0$, $\frac{\partial Rent}{\partial \alpha_1} < 0$ as stated. The limit condition is a direct result the definition of rent and part 1.

Proof of Corollary 2

If the IC constraint does not bind when the vectors are γ_1' and γ_2', the data broker charges the monopolist prices and obtains the highest profits he is able to.

If the IC constraint binds when the vectors are γ_1' and γ_2', then it binds when the vectors are γ_1 and γ_2. Since the angle between γ_1' and γ_2' is larger than the angle between γ_1 and γ_2 and the measure of all angles is smaller than $\pi/2$, $\gamma_1^T\gamma_2 > \gamma_1'^T\gamma_2'$.

Remember that the profits are given by

$$\pi = \frac{\alpha_1\left(\gamma_1^T\gamma_1\gamma_2^T\gamma_2 - (\gamma_1^T\gamma_2)^2\right) + \left(\gamma_2^T\gamma_2 - c(\alpha_1)\gamma_1^T\gamma_2\right)^2}{\gamma_2^T\gamma_2 - 2c(\alpha_1)\gamma_1^T\gamma_2 + c^2(\alpha_1)\gamma_1^T\gamma_1}$$

where c is defined in theorem 2. Then

$$\begin{aligned}\frac{\partial\pi}{\partial\gamma_1^T\gamma_2} \;&\propto\; (-\alpha_1\gamma_1^T\gamma_2 - c(\alpha_1)(\gamma_2^T\gamma_2 - c(\alpha_1)\gamma_1^T\gamma_2))(\gamma_2^T\gamma_2 - 2c(\alpha_1)\gamma_1^T\gamma_2 + c^2(\alpha_1)\gamma_1^T\gamma_1) \\ &\quad + c(\alpha_1)\left(\alpha_1\left(\gamma_1^T\gamma_1\gamma_2^T\gamma_2 - (\gamma_1^T\gamma_2)^2\right) + \left(\gamma_2^T\gamma_2 - c(\alpha_1)\gamma_1^T\gamma_2\right)^2\right) \\ &= (\gamma_1^T\gamma_2 - c(\alpha_1)\gamma_1^T\gamma_1)(\gamma_2^T\gamma_2 - c(\alpha_1)\gamma_1^T\gamma_2)(c^2(\alpha_1) - \alpha_1)\end{aligned}$$

Since $c(\alpha_1) \leq \tilde{\alpha}$ it is easy to check that the first two factors are non-negative. Lemma 8 shows that the third factor is negative. Therefore, the profits decrease when increasing $\gamma_1^T\gamma_2$, which implies the result.

Proof Lemma 4

Suppose that $q_{2v} \neq 0$ and $\ell_{2v} \neq 0$. Let $\lambda_{2v} = Var(X)^{1/2}L_{2v}$ and $\hat{\ell}_{2v} = \frac{\ell_{2v}}{\|\lambda_{2v}\|^2}$. I first show that it is without loss to assume that the angle β_{2v} between λ_{2v} and γ_2 is smaller than the angle β_{12v} between λ_{2v} and γ_1.

Suppose that $\beta_{12v} \leq \beta_{2v}$. Let β be the angle between γ_1 and γ_2. Pick $\tilde{\lambda}_{2v}$ as the vector with the same magnitude as λ_{2v} such that the angle between $\tilde{\lambda}_{2v}$ and γ_1 satisfies $\beta_{\gamma_1,\tilde{\lambda}_{2v}} = \beta_{2v} + \beta$ and the angle between $\tilde{\lambda}_{2v}$ and γ_2 satisfies $\beta_{\gamma_2,\tilde{\lambda}_{2v}} = \beta_{2v}$. After this change the willingness to pay by any type (v', t_1) for the statistic that is targeted to type (v, t_2) has decrease, while it keeps constant the willingness to pay by any type (v', t_2) for this statistic. This uniformly relaxes all constraints *IC 1v'-2v*, while it keeps the rest of constraints unchanged. This construction shows that it is without loss to assume that $\beta_{12v} > \beta_{2v}$.

Remember that $q_{2v} = \frac{cos^2(\beta_{2v})}{1+\hat{\ell}_{2v}\sigma^2}$. I will show that there is another pair $(\beta'_{2v}, \ell'_{2v})$ with $\ell'_{2v} = 0$ such that type (v, t_2) still obtains the same forecast's variance reduction, but such that $\delta_{12v} > \delta'_{12v}$, implying that all constraints *IC 1v'-2v* are relaxed. This completes the argument.

Let $\beta'_{2v} = cos^{-1}\left(\sqrt{\frac{cos^2(\beta_{2v})}{1+\hat{\ell}_{2v}\sigma^2}}\right)$ and $\ell'_{2v} = 0$. It is inmediate that $q'_{2v} = cos^2(\beta'_{2v}) = q_{2v}$. Furthermore,

$$\begin{aligned}
\sqrt{\delta'_{12v}} &= cos(\beta'_{2v} + \beta) \\
&= cos(\beta'_{2v})cos(\beta) - sin(\beta'_{2v})sin(\beta) \\
&= \frac{cos(\beta_{2v})}{\sqrt{1+\hat{\ell}_{2v\sigma 2}}}cos(\beta) - \sqrt{1 - \frac{cos^2(\beta_{2v})}{1+\hat{\ell}_{2v\sigma 2}}}sin(\beta) \\
&< \frac{cos(\beta_{2v})}{\sqrt{1+\hat{\ell}_{2v\sigma 2}}}cos(\beta) - \frac{\sqrt{1-cos^2(\beta_{2v})}}{\sqrt{1+\hat{\ell}_{2v\sigma 2}}}sin(\beta) \\
&= \frac{cos(\beta_{2v})cos(\beta) - sin(\beta_{2v})sin(\beta)}{\sqrt{1+\hat{\ell}_{2v\sigma 2}}} \\
&= \frac{cos(\beta_{2v}+\beta)}{\sqrt{1+\hat{\ell}_{2v\sigma 2}}} = \sqrt{\delta_{12v}},
\end{aligned}$$

where in the third equality I use that $sin(cos^{-1}(x)) = \sqrt{1-x^2}$.

Proof Lemma 5

1. Suppose that in the solution to problem P, $q_{1v} < 1$ for some $v \in [v_1^*, \hat{v}]$. This cannot be a solution since for $v \in [v_1^*, \hat{v}]$ the virtual value is positive and increasing q_{1v} relaxes all constraints *IC 1v'-2w* for $v' > v$ and any w.

2. Suppose that in the solution to problem P $\hat{v}_2 < v_2^*$, that is, $q_{2v} > 0$ for some $v < v_2^*$. Since the virtual value for information type t_2 is negative for $v < v_2^*$, by taking $q_{2v} = 0$ for $v < \hat{v}_2$ the objective function increases and all constraints are still satisfied. Therefore, $\hat{v}_2 \geq v_2^*$. Now, if $\hat{v}_2 = v_2^*$ the seller can increase his profits. Consider slightly increasing $\hat{v}_2$. This directly affects the profits coming from information type t_2 at a quadratic rate, but it reduces the information rent given to types $v > v_1^*$ in a linear way. Therefore, for a small increase of the threshold the profits have to increase, and $\hat{v}_2 > v_2^*$.

   Now suppose that $\hat{v}_2 \geq v_1^*$. The allocation $q_{1v} = q_{2v} = 1$ for all $v \geq v_1^*$ and 0 otherwise, satisfies all the constraints. By continuity, the seller can pick $q_{2v} = 1$ for some neighborhood $(v_1^* - \epsilon, v_1^*)$ and still satisfy the constraints. By picking $q_{2v} = 1$

for $v > v_1^* - \epsilon$ and $q_{1v} = 1$ for $v \geq v_1^*$, the seller can increases his profits. Therefore, $\hat{v}_2 < v_1^*$.

Proof Lemma 6

I show a property that is stronger than the statement. I show that for any allocation q_{2v} for type t_2 with q_{2v} non-decreasing, and $\hat{v}_2 < v_1^*$ such that the constraints $IC\ 1v_1^* - 2v'$ for all $v' \geq \hat{v}_2$ are satisfied, it has to be that in the constrained solution to problem P, there is a threshold $\hat{v}_1 = \frac{\hat{v}_2 q_{2\hat{v}_2}}{g(q_{2\hat{v}_2})}$ such that $q_{1v} = 0$ for $v < \hat{v}_1$ and $q_{1v} = g(q_{2\hat{v}_2})$ for all $v \in [\hat{v}_1, v_1^*)$.

Let $\hat{v}_1 = \frac{\hat{v}_2 q_{2\hat{v}_2}}{g(q_{2\hat{v}_2})}$. That $\hat{v}_2 < \hat{v}_1$ is obvious from the definition. I show that in the optimal solution to problem P, $\hat{v}_1 < v_1^*$. First, $\hat{v}_2$ has to be such that $\frac{\hat{v}_2}{g(1)} \leq v_1^*$. If not, the seller could offer $q_{2v'} = 1$ to the type $v' > v_2^*$ that solves this equation, which satisfies all the constraints and increases his profits because this type's virtual value is positive. Therefore, $\hat{v}_1 \leq v_1^*$. Suppose that $\hat{v}_1 = v_1^*$. Then $q_{2\hat{v}_2} = 1$ and $\hat{v}_2 = g(1)v_1^*$. As in Section 4.1 this cannot be a solution to the seller's profits, reducing $q_{2\hat{v}_2}$ slightly has a negligible impact in type $(t_2, \hat{v}_2)'$ willingness to pay for this statistic, but it significantly reduces type $(t_1, v_1^*)'$ willingness to pay for it. Formally, the derivative of type $(t_2, \hat{v}_2)'$ willingness to pay is zero when evaluated at $q_{2\hat{v}_2} = 1$ since this means that the angle between type $(t_2, \hat{v}_2)$'s preferred vector of coefficients and that statistic targeted to her has zero magnitude, and the derivative of the function *cos* is zero when the angle is zero, but it is non-zero for any angle in $(0, \pi/2)$.

Now I argue that picking the allocation $q_{1v} = 0$ for $v < \hat{v}_1$ and $q_{1v} = g(q_{2\hat{v}_2})$ for $v \in [\hat{v}_2, v_1^*)$ and the values $U(\underline{v}_1, t_1) = U(\underline{v}_2, t_2) = 0$ is feasible in problem P. With this values the constraints *IC 1v-2v'* with $v < v_1^*$ and $v' \geq \hat{v}_2$ are satisfied since

$$\begin{aligned}
\int_{\underline{v}}^{v} q_{1w}\, dw &= \int_{\underline{v}}^{v_1^*} q_{1w}\, dw - \int_{v}^{v_1^*} q_{1w}\, dw \\
&\geq v_1^* g(q_{2v'}) - v' q_{2v'} + \int_{\underline{v}}^{v'} q_{2w}\, dw - g(q_{2\hat{v}_2})(v_1^* - v) \\
&\geq v g(q_{2v'}) - v' q_{2v'} + \int_{\underline{v}}^{v'} q_{2w}\, dw,
\end{aligned}$$

where in the first inequality I use that the constraint $IC\ 1v^* - 2v'$ is satisfied, and in the second inequality I use that $v_1^* > v$.

Choosing $U(\underline{v}_1, t_1) = U(\underline{v}_2, t_2) = 0$ is clearly optimal and picking $q_{1v} = 0$ when possible for $v < v_1^*$ is optimal since the virtual value is negative in this case.

Now suppose that for an small interval to the right of $\hat{v}_1$ the constraints *IC 1v-2*$\hat{v}_2$ do not bind. Since the virtual value for these types is negative, the solution would be to pick $q_{1v} = 0$ for all of them, but these types can report type $(\hat{v}_2, t_2)$ and obtain a positive profit since for them $vg(q_{2\hat{v}_2}) > \hat{v}_1 g(q_{2\hat{v}_2}) = \hat{v}_2 q_{2\hat{v}_2}$. Then the constraints *IC 1v-2*$\hat{v}_2$ have to bind for any $v \in (\hat{v}_1, v_1^*)$. Since virtual value is negative in this interval, the seller wants to pick the lowest value of q_{1v} that satisfies these constraints and this value is $q_{1v} = g(q_{2\hat{v}_2})$.

Proof Lemma 7

I prove the lemma in many steps. First, I fix an allocation (q_1, q_2) with q_1 and q_2 non-decreasing with q_1 satisfying the conditions in Lemma 6, and I show that if the constraints *IC* $1v_1^* - 2v'$ for all $v' > \hat{v}_2$ are satisfied, then the allocation (q_1, q_2) is feasible for problem P. Finally, I show that all these constraints have to bind, and, this implies, that the optimal solution q_{2v} satisfies the conditions in the statement.

Fix an allocation (q_1, q_2) with q_1 and q_2 non-decreasing, and q_1 satisfying the conditions in Lemma 6. I first show that if the constraints *IC* $1v_1^* - 2v'$ for all $v' > \hat{v}_2$ are satisfied, then the allocation (q_1, q_2) is feasible for problem P. For $v' < \hat{v}_2$ and any v, the IC constraints *IC 1v-2v'* are automatically satisfied since $q_{2v'} = g(q_{2v'}) = 0$. In the proof of Lemma 6 I showed that for all $v < v_1^*$ and $v' \geq \hat{v}_2$ the constraints *IC 1v-2v'* are satisfied. Now, I show that for all $v \geq v_1^*$ and $v' \geq \hat{v}_2$ the constraints *IC 1v-2v'* are satisfied since

$$\begin{aligned}
\int_{\underline{v}}^{v} q_{1w}\, dw &= \int_{\underline{v}}^{v_1^*} q_{1w}\, dw + \int_{v_1^*}^{v} q_{1w}\, dw \\
&\geq v_1^* g(q_{2v'}) - v' q_{2v'} + \int_{\underline{v}}^{v'} q_{2w}\, dw + (v - v_1^*) \\
&\geq v g(q_{2v'}) - v' q_{2v'} + \int_{\underline{v}}^{v'} q_{2w}\, dw,
\end{aligned}$$

where the first inequality follows from the constraint *IC 1*v_1^**-2v'* being satisfied and from the fact that $q_{1v} = 1$ for all $v > v_1^*$.

Now, I show that all the constraints *IC* $1v_1^* - 2v'$ for $v' \geq \hat{v}_2$ bind. First, suppose that the constraint *IC* $1v_1^*$-$2\hat{v}_2$ does not bind. Since the virtual value is positive for $v > v_2^*$ the seller maximizes its revenue by setting $\hat{v}_2 = v_2^*$ and $q_{2\hat{v}_2} = 1$ and $\hat{v}_1 = v_1^*$, but this is not possible since in such a case, by assumption, type (t_i, v_1^*) has an incentive to report type (t_2, v_2^*). Therefore, the constraint *IC* $1v_1^*$-$2\hat{v}_2$ binds. Furthermore, $q_{2\hat{v}_2} < 1$. If not, the seller can reduce $q_{2\hat{v}_2}$ by a small amount which almost does not decrease the revenue of selling to type $(t_2, \hat{v}_2)$ but increases the profits coming from information type t_1 since the seller can reduce $q_{1\hat{v}_1}$ and type $(t_1, \hat{v}_1)$'s virtual value is negative.

By continuity of the RHS of the constraints *IC* $1v_1^*$-$2v'$, they have to bind in a neighborhood to the right of $\hat{v}_2$; if not the firm should be able to pick $q_{2v} = 1$ in this neighborhood but this contradicts the continuity of the RHS. As these constraints bind, they hold with equality and the derivative of the RHS with respect to v' has to be zero, that is,

$$v_1^* g'(q_{2v'}) \frac{\partial q_{2w}}{\partial w}|_{w=v'} - v' \frac{\partial q_{2w}}{\partial w}|_{w=v'} = 0.$$

Therefore, in this interval either $\frac{\partial q_{2w}}{dw}|_{w=v'} = 0$ or $g'(q_{2v'}) = \frac{v'}{v_1^*}$. As $\lim_{x \to 1} g'(x) = \infty$, $q_{2v'}$ is bounded away from 1 for all v' in this neighborhood to the right of $\hat{v}_2$.

Suppose that $\mathring{v}$ is the first value at which the constraint *IC* $1v_1^*$-$2v'$ does not bind, so that the seller will pick $q_{2\mathring{v}} = 1$. Since $\lim_{v \to \mathring{v}^-} q_{2v} < 1$ and the RHS of the constraints *IC* $1v_1^*$-$2v'$ is continuous, it is not possible that $q_{2\mathring{v}} = 1$ is feasible; if not, for v close and to the left of $\mathring{v}$ the seller should be able to pick q_{2v} close to 1. Therefore, for all $v' > \hat{v}_2$ the constraint *IC* $1v_1^*$-$2v'$ binds.

Since all the constraints *IC* $1v_1^* - 2v'$ for $v' \geq \hat{v}_2$ bind, either $\frac{\partial q_{2w}}{dw}|_{w=v'} = 0$ or $g'(q_{2v'}) = \frac{v'}{v_1^*}$ for all $v' \geq \hat{v}_2$. I show that q_{2v} satisfies the first condition for v' close to $\hat{v}_2$ and that, in some cases, it satisfies the second condition for large v' values.

Since $cos^{-1}(q_{2v}) + \beta \in [\beta, \pi/2]$ and it is decreasing in q_{2v}, it can be shown that the second derivative of $g(q_{2v})$ is positive. Therefore, $g'(q_{2v})$ is increasing. This means that the solution of the equation $g'(q_{2v'}) = \frac{v'}{v_1^*}$ is increasing. Since, q_{2v} has to be increasing, and the virtual value for $v > \hat{v}_2$ is positive, there exists $\tilde{v}_2 > \hat{v}_2$ such that $q_{2v} = q_{2\hat{v}_2}$ for $v < \tilde{v}_2$ and it is equal to the solution of the equation $g'(q_{2v'}) = \frac{v'}{v_1^*}$ for all other values of v. $\tilde{v}_2$ and $\hat{v}_2$ are equal only if $g'(q_{2\hat{v}_2}) = \frac{g(q_{2\hat{v}_2})}{q_{2\hat{v}_2}}$. After some algebra this condition can be written as $tan(cos^{-1}(q_{2\hat{v}_2}) + \beta) - \sqrt{\frac{1-q_{2\hat{v}_2}}{q_{2\hat{v}_2}}} = 0$ and the LHS is strictly decreasing with $\lim_{q_{2\hat{v}_2} \to 1} LHS = tan(\beta)$. Therefore, there is not value of $q_{2\hat{v}_2}$ that satisfies this condition, and $\tilde{v}_2 > \hat{v}_2$. Finally, since $\tilde{v}_2$ does not necessarily belongs to the support of the valuations for type t_2, it might be that q_{2v} is just a constant function.

Proof Theorem 3

I only need to check that the solution actually satisfies the constraints that I have not imposed in the relaxed problem. It is straightforward to see that this is the case.

Proof Proposition 3

That none of the IC constraints bind is equivalent to the optimal mechanism being the one that gives recommendations $\psi_i = \mathbb{E}[\theta_i] + Cov(\theta_i)Var(X)^{-1}(X - \mathbb{E}[X])$ and charges prices $\gamma_i^T \gamma_i$. For such mechanism the IC constraint for type t_i reporting type t_j with $i < j$ is satisfied iff

$$\frac{(\gamma_i^T \gamma_j)^2}{\gamma_j^T \gamma_j} - \gamma_j^T \gamma_j < 0 \Leftrightarrow (\gamma_i^T \gamma_j)^2 > (\gamma_j^T \gamma_j)^2 \Leftrightarrow cos^2(\beta_{ij}) \left\|\gamma_i\right\|^2 \left\|\gamma_j\right\|^2 < \left\|\gamma_j\right\|^4 .$$

Now, in such a mechanism the upward constraint for type t_k reporting type t_l with $k > l$ is satisfied iff

$$\frac{(\gamma_l^T \gamma_k)^2}{\gamma_l^T \gamma_l} - \gamma_l^T \gamma_l < 0 \Leftrightarrow (\gamma_k^T \gamma_l)^2 < (\gamma_l^T \gamma_l)^2 \Leftrightarrow \left\|\gamma_l\right\|^2 \left\|\gamma_k\right\|^2 cos^2(\beta_{lk}) < \left\|\gamma_l\right\|^4 ,$$

where β_{lk} is the angle between vectors γ_k and γ_l. The last inequality always holds since $cos(\beta_{lk}) < 1$ and by assumption $\left\|\gamma_l\right\|^2 \geq \left\|\gamma_k\right\|^2$.

Since all IC constraints are satisfied by the proposed mechanism, it is feasible and none of the IC constraints bind.

Proof Theorem 4

1. Suppose all types receive a positive surplus and let $P = \min_i \left\{ \frac{(L_i^T Cov(\theta_i, X))^2}{L_i^T Var(X) L_i + \ell_i \sigma^2} - p_i \right\} > 0$. Then the data broker can charge new prices $\tilde{p}_i = p_i + P$, so that all IC constraints are unaffected and the IR constraints are satisfied. This modification of the mechanism clearly increases the seller's profits.
2. Suppose by contradiction that in the optimal mechanism (ψ, p), the statistic $\psi(t_i)$ is not sufficient to learn the conditional expectation of θ_i given the data X, that is,

there is not $k \neq 0$ such that $L_i = kVar(X)^{-1}Cov(\theta_i, X)$ or $\ell_i \neq 0$.

Let $R_i = Var(\theta_i) - Var(\theta_i \mid \psi(t_i)) - p(t_i)$ be the rent that type t_i receives in this mechanism, and let $i^* \in \arg\max_i \gamma_i^T \gamma_i - R_i$ be the type that in net terms is willing to pay more for all the data.

Consider the alternative mechanism (ψ', p') with $\psi'(t_{i^*}) = \mathbb{E}[\theta_{i^*} \mid X]$, $p'(t_{i^*}) = \gamma_{i^*}^T \gamma_{i^*} - R_{i^*}$, and for any $t_i \neq t_{i^*}$, $\psi'(t_i) = \psi(t_i)$ and $p'(t_i) = p(t_i)$. First, type t_{i^*} does not have incentive to report any other type since in the modified mechanism she obtains the same rent as in the original mechanism. Furthermore, any type $t_i \neq t_{i^*}$ does not want to report type t_{i^*} since by Claim 2 and Lemma 1

$$Var(\theta_i) - Var(\theta_i \mid \psi'(t_{i^*})) - p'(t_{i^*}) < \gamma_i^T \gamma_i - (\gamma_i^T \gamma_i - R_i) = R_i.$$

Finally, $p'(t_{i^*}) = \gamma_{i^*}^T \gamma_{i^*} - R_{i^*} > Var(\theta_i) - Var(\theta_i \mid \psi(t_i)) - R_i = p(t_i)$. Therefore the mechanism (ψ', p') is feasible and gives higher profits to the seller than mechanism (ψ, p), a contradiction.

Proof Theorem 5

Fix $\delta > 0$ with $\delta < \min\{\gamma_i^T \gamma_i : i \in \{1, \ldots, n\}\}$. For the subset of types $I \subseteq \{1, \ldots, n\}$, consider the modified problem

$$\begin{aligned}
V(I,\delta) = \max_{\{p_i, L_i, \ell_i\}_{i \in I}} \quad & \textstyle\sum_{i \in I} \alpha_i p_i \\
\text{s.t.} \quad & \frac{(L_i^T Cov(\theta_i, X))^2}{L_i^T Var(X) L_i + \ell_i^2 \sigma^2} - p_i \geq \frac{(L_j^T Cov(\theta_i, X))^2}{L_j^T Var(X) L_j + \ell_j^2 \sigma^2} - p_j \quad && \forall i, j \in I \\
& \frac{(L_i^T Cov(\theta_i, X))^2}{L_i^T Var(X) L_i + \ell_i^2 \sigma^2} - p_i \geq 0 && \forall i \in I \\
& p_i \geq \delta && \forall i \in I
\end{aligned}$$

and let $V(\delta) = \max\limits_{I \subseteq \{1, \ldots, n\}} V(I, \delta)$ and $I^*(\delta) = \arg \max\limits_{I \subseteq \{1, \ldots, n\}} V(I, \delta)$. This problem must have a solution in which some L_i are non-zero and some $p_i > \delta$ since, by Lemma 1, the vector of coefficients $\tilde{L}_i = Var(X)^{-1} Cov(\theta_i, X)$, the noise coefficient $\tilde{\ell}_i = 0$ and prices $\tilde{p}_i = \min\{\gamma_i^T \gamma_i : i \in \{1, \ldots, n\}\}$ satisfy all the constraints. In other words, one of the points that satisfy the Khun-Tucker conditions is the optimal solution.

In the optimal mechanism $\min\{p_i : p_i > 0\} > 0$. Then as $\delta \to 0$, the sequences $V(\delta)$

and $I^*(\delta)$ are finally constant, and equal to the seller's optimal profits and equal to the set of types to which the seller sells in the optimal mechanism.

Let I^* be the set of types to which the data broker sells in the optimal mechanism and δ small enough such that the constraints $p_i \geq \delta$ do not bind. When considering the modified problem with I^* and such small δ, the solution is exactly equal to the optimal mechanism for the data broker. For this problem let λ_{ij} be the Lagrange multiplier corresponding to the IC constraint from type t_i to type t_j and μ_i be the Lagrange multiplier corresponding to the IR constraint for type t_i for $i, j \in I^*$. Defining $a_{ji} \equiv \frac{L_i^T Cov(\theta_j, X)}{L_i^T Var(X) L_i + \ell_i^2 \sigma^2}$, the First Order Condition with respect to L_i can be written as

$$\left(\sum_{j\neq i} \lambda_{ij} + \mu_i\right)\left(a_{ii} Cov(\theta_i, X) - a_{ii}^2 Var(X) L_i\right) = \sum_{j\neq i} \lambda_{ji}\left(a_{ji} Cov(\theta_j, X) - a_{ji}^2 Var(X) L_i\right)$$

$$\Leftrightarrow \quad L_i = \frac{Var(X)^{-1}}{\frac{a_{ii}}{\sum_{j\neq i}(\lambda_{ij}+\mu_i)} - \frac{\sum_{j\neq i}\lambda_{ji} a_{ji}^2}{\sum_{j\neq i}(\lambda_{ij}+\mu_i) a_{ii}}}\left(Cov(\theta_i, X) - \frac{\sum_{j\neq i}\lambda_{ji} a_{ji}}{\sum_{j\neq i}(\lambda_{ij}+\mu_i) a_{ii}} Cov(\theta_j, X)\right).$$

Since the value of the objective function is scale invariant, L_i can be normalized such that $\frac{a_{ii}}{\sum_{j\neq i}(\lambda_{ij}+\mu_i)} - \frac{\sum_{j\neq i}\lambda_{ji} a_{ji}^2}{\sum_{j\neq i}(\lambda_{ji}+\mu_i) a_{ii}} = 1$. Further, define $c_{ji} \equiv \frac{a_{ji}}{a_{ii}}$, and since the problem is bang-bang with respect to ℓ_i, it has to be that $\ell_i = 0$ for all $i \in I^*$ and $\ell_i = \infty$ for all $i \notin I^*$. Therefore, I obtain the desired expression.

I still need to show that the mapping from the vector $(c_{ij})_{i,j} \in \mathbb{R}^{2n}$ into itself has a least one fixed point. The mapping is clearly continuous. I argue that each c_{ij} in the codomain can be bounded, and that in such a case the image is bounded as well. Then the domain and the codomain of the mapping can be restricted to be the same compact convex subset of $\mathbb{R}^{2n}$ and by Brouwer Fixed-Point theorem the mapping has at least one fixed point.

Since $p_i \geq \delta > 0$ and $(L_i^T Cov(\theta_i, X))^2 \geq p_i$, $(L_i^T Cov(\theta_i, X))^2 > 0$ for each $i \in I^*$. In particular this implies that $L_i \neq 0$, and since $Var(X)$ is positive definite, $L_i^T Var(X) L_i > 0$. Let $\zeta = \min_{i\in I^*}\left\{\frac{(L_i^T Cov(\theta_i, X))^2}{L_i^T Var(X) L_i}\right\} > 0$. By Lemma 1

$$\frac{(L_i^T Cov(\theta_j, X))^2}{L_i^T Var(X) L_i} \leq Cov(\theta_j, X)^T Var(X)^{-1} Cov(\theta_j, X).$$

Let $M = \max_{i\in I^*}\left\{Cov(\theta_i, X)^T Var(X)^{-1} Cov(\theta_i, X)\right\}$. Then it has to be that for each i and

j, $c_{ij}^2 \leq \frac{M}{\zeta}$. Then the mapping can be restricted to go from the domain $\left[-\sqrt{\frac{M}{\alpha}}, \sqrt{\frac{M}{\alpha}}\right]^{2n}$ to itself. This restricted map satisfies all the hypothesis of Brouwer Fixed Point theorem, so it has at least one fixed point. Since the problem is guaranteed to have a solution, its solution needs to satisfy the Kuhn-Tucker conditions. Therefore, one of the fixed points has to be the solution to the original problem.