Economics Bulletin

Volume 46, Issue 2

Could you spare me a moment? A model of primary data collection by statistics institutes

Marcelo de C. Griebeler
Luiz de Queiroz College of Agriculture, University of São Paulo

Felipe Wolk Teixeira
Brazilian Institute of Geography and Statistics

Abstract

We study the strategic interaction between a statistical institute and a survey respondent (either a firm or a household). The respondent must decide whether to acquire the information required by the institute, thereby incurring a private cost. After receiving the reported data, the institute chooses whether to audit the respondent's declaration. The key innovation of our model is to introduce two-sided asymmetric information: both the cost of acquiring the information and the cost of auditing are private and known only to the respective agent. We characterize a Perfect Bayesian equilibrium and show that there is substitutability between the magnitude of the punishment for misreporting and the probability of auditing. Our framework allows for both positive and normative analysis of the challenges faced by statistical agencies in data collection, providing insights into how enforcement mechanisms can be designed to balance deterrence and audit costs.

Author's Footnote: We thank the editor and two anonymous referees for their careful reading and valuable suggestions, which significantly improved the paper. We are also grateful to Gregory Pavlov, Filipe Stona, Jefferson Bertolai, Guilherme Stein, Jaylson da Silveira, Sabino Porto Jr., and the participants of the 47th Encontro Nacional de Economia (ANPEC) for helpful comments and discussions. The views expressed in this article are those of the authors and do not necessarily reflect the views of any agency of the Brazilian government. Felipe W. Teixeira acknowledges financial support from IBGE and the CAPES Foundation – Brazil (Grant Issue 2.187116/2018-01). This research was also supported in part by the Brazilian National Council for Scientific and Technological Development (CNPq) under Grant Issue 2/2023-4.
Editor`s Note: This paper was originally submitted under a different manuscript number on 10/24/2025 and accepted for publication on 01/08/2026. A system error required us to replace the original submission with a new manuscript with a later manuscript number.
Citation: Marcelo de C. Griebeler and Felipe Wolk Teixeira, (2026) "Could you spare me a moment? A model of primary data collection by statistics institutes", *Economics Bulletin*, Volume 46, Issue 2, pages 816-826
Contact: Marcelo de C. Griebeler - marcelo.griebeler@usp.br, Felipe Wolk Teixeira - felipe.teixeira@ibge.gov.br.
Submitted: July 14, 2026. **Published:** June 30, 2026.

1 Introduction

Statistical institutes frequently face difficulties in collecting primary data from households and firms. High non-response rates and the unreliability of self-reported information are major concerns (Massey and Tourangeau, 2013; Singleton and Straits, 2017). While individuals may misreport unintentionally due to cognitive challenges in understanding and recalling information (Tourangeau et al., 2000), some responses—especially regarding sensitive variables like income—are deliberately inaccurate (Bound et al., 2001; Moore and Welniak, 2000). For firms, the situation is more complex: acquiring the necessary information often requires effort, time, and the reallocation of personnel, introducing a cost that may not be offset by any direct benefit. For instance, consider a small manufacturing firm asked to fill out an industrial production survey. The employee responsible may not have immediate access to all relevant records and must interrupt routine activities to collect and process the requested data. In some cases, this may also require consulting other departments (e.g., accounting), further increasing internal costs.[1]

In this paper, we develop a game-theoretical model to study the strategic interaction between a statistics institute and a respondent (typically a firm). The information required by the institute is modeled as a random variable whose realization is observable only to the respondent, and only at a cost. After acquiring (or not) the information, the respondent decides what value to report. Upon receiving the report, the institute decides whether to audit. The main innovation of our model is the introduction of *two-sided asymmetric information*, whereby both the cost of acquiring the information (borne by the respondent) and the cost of auditing (borne by the institute) are private. While existing literatures on tax evasion, inspection games, and communication models often assume that only one side has private information, we argue that mutual uncertainty more accurately reflects the real-world challenges of primary data collection. This richer informational structure alters the strategic environment and leads to novel policy implications: audit decisions become belief-based on both sides, and optimal enforcement involves a trade-off between costly monitoring and penalty design.

Our framework is related to classic models of tax evasion (Allingham and Sandmo, 1972; Becker, 1968), in which agents hide income to reduce liability while facing the risk of audits and penalties. However, our setting differs in two key ways: first, respondents derive no direct benefit from misreporting other than avoiding the information cost; second, the penalty is fixed and not proportional to the deviation from the truth. These features decouple the reported value from the realization of the variable. The model is also connected to communication games, particularly Cheap Talk (Crawford and Sobel, 1982), but with the important distinction that information acquisition is costly and strategic. It also shares elements with inspection games (Avenhaus et al., 2002), yet differs by assuming fully effective audits and incomplete information on both sides.

We characterize the Bayesian equilibria of the game and study their dependence on two key parameters: the variance of the required information and the severity of the punishment for misreporting. We show that higher variance increases uncertainty and makes auditing more attractive, while stronger penalties deter misreporting and reduce the need for audits. A main insight of the model is the substitutability between punish-

[1]This situation is common, for example, in the collection of data for the *Pesquisa Industrial Mensal de Produção Física* (PIM-PF), conducted by the Brazilian Institute of Geography and Statistics (IBGE). Respondents frequently report operational burdens associated with retrieving detailed production information, especially in smaller firms where responsibilities are concentrated.

ment severity and audit probability: higher penalties reduce the equilibrium likelihood of audits, which are costly for the institute.

Our results shed light on the incentives behind misreporting and information acquisition and offer policy-relevant implications for improving data quality. In particular, we identify institutional parameters that can be adjusted to reduce the need for costly enforcement while increasing compliance. The rest of the paper proceeds as follows. Section 2 presents the model and analyzes its equilibrium. Section 3 concludes with remarks for policymakers. Proofs omitted in the main text appear in Appendix A.

2 Model

2.1 Setup

Consider a statistical institute that requires primary data in order to construct statistics that are valuable for society and for informing government policy design. This primary data, however, is not directly observable by the institute, which must conduct a survey and rely on a single economic agent—henceforth referred to as the *respondent*—to provide the information. We assume that the data being requested is a random variable, $\theta \in \{\theta_L, \theta_H\}$, with $\Pr(\theta = \theta_L) = \alpha \in (0,1)$ and $\Delta = \theta_H - \theta_L > 0$. For future reference, note that $\Delta^2 = \frac{\sigma^2}{\alpha(1-\alpha)}$, where σ^2 denotes the variance of θ. Initially, neither party knows the true realization of θ, but the respondent can acquire this information by incurring a cost, as we describe below.

The respondent makes two sequential decisions. First, she decides whether to acquire the information. To do so, she must incur an acquisition cost K. This cost is private information: the institute does not observe its value but knows that it is drawn from a strictly increasing cumulative distribution function (CDF), H, with support on $[0,\infty) = \mathbb{R}_{\geq 0}$. From this point forward, we refer to K as the respondent's *type*. Second, the respondent must report the requested data to the institute. As we will show, whenever the respondent acquires the information, it is optimal for her to report its true value (i.e., the realized value of θ). Let $\hat{\theta}$ denote the reported value; then $\hat{\theta} = \theta$ whenever the respondent reports truthfully. If, however, she chooses not to acquire the information, she must decide which value to report without knowing the true realization.

After receiving the reported primary data $\hat{\theta}$, the institute may choose to audit it. The auditing procedure, when conducted, correctly detects any misreporting. Auditing entails a fixed cost A for the institute. This cost is private information: the respondent does not observe its value but knows that it is drawn from a strictly increasing CDF, G, with support $\mathbb{R}_{\geq 0}$. This implies that A can be interpreted as the institute's *type*. If the institute audits and discovers that the respondent has misreported the data, a fixed monetary fine $F > 0$ is imposed on the respondent.[2] In such cases, the respondent must also incur the cost K to obtain the true information. We assume that the fine F is exogenously given and not directly set by the institute. Moreover, the institute does not retain any monetary benefit from levying fines.[3] All random variables in the model—θ,

[2]This penalty may also be interpreted as a reputational cost resulting from being subjected to an audit. This interpretation becomes particularly relevant in cases where statistical institutes conduct audits to obtain the true information but choose not to impose a financial penalty.

[3]Observe that, if the institute retained monetary gain from levying fines, this could create an undesirable incentive for profit-maximizing behavior. An empirically grounded justification for this assumption is that revenues from fines are typically transferred to a central government authority.

K, and A—are mutually independent.

2.2 Strategies and payoffs

Since this is a Bayesian game, strategies are functions that map types into actions. Accordingly, a strategy for the respondent maps her type K into a probability $x \in [0,1]$ of *not* acquiring the information, and a probability $y \in [0,1]$ of reporting the low value $\hat{\theta} = \theta_L$, conditional on not having acquired the information—that is, $\Pr(\hat{\theta} = \theta_L \mid x = 1) = y$. A strategy for the institute, in turn, maps its type A into a probability of auditing conditional on the reported value. Formally, this is represented by the pair $(z(\theta_L), z(\theta_H)) \in [0,1] \times [0,1]$. From this point forward, we refer to the *pure strategy* of not acquiring the information as $x = 1$, the pure strategy of reporting $\hat{\theta} = \theta_L$ conditional on $x = 1$ as $y = 1$, and the pure strategy of auditing upon observing $\hat{\theta}$ as $z(\hat{\theta}) = 1$.

The respondent's expected payoff function is given by

$$\begin{aligned} U(x, y, z(\theta_L), z(\theta_H); K) &= (1-x)(-K) \\ &+ x(-K-F)\left[yz(\theta_L)(1-\alpha) + (1-y)z(\theta_H)\alpha\right]. \end{aligned} \tag{2.1}$$

Observe that, whenever the respondent chooses to acquire the information ($x = 0$), her utility is $-K$, which is independent of the institute's auditing decision. This follows from the fact that, upon acquiring the information, it is optimal for the respondent to report truthfully. If, instead, $\hat{\theta} \neq \theta$, she may be audited—resulting in a payoff of $-K-F$—or avoid auditing, in which case her utility remains $-K$, the same as when reporting truthfully.

The institute's expected payoff depends on both the reported and the true values of θ. In particular, we assume that any deviation from the true value generates a welfare loss, captured by the quadratic term $-(\hat{\theta}-\theta)^2$. This loss function penalizes larger discrepancies more severely, placing greater weight on larger deviations between the true and reported values[4]. Formally, the institute's expected utility is given by:

$$V(x, y, z(\theta_L), z(\theta_H); A) = \begin{cases} z(\theta_L)(-A) + \beta(\theta_L)(1 - z(\theta_L))(1-\alpha)(-\Delta^2), & \text{if } \hat{\theta} = \theta_L \\ z(\theta_H)(-A) + \beta(\theta_H)(1 - z(\theta_H))\alpha(-\Delta^2), & \text{if } \hat{\theta} = \theta_H, \end{cases} \tag{2.2}$$

where $\beta(\theta_L)$ and $\beta(\theta_H)$ denote the posterior probabilities that the respondent did not acquire the information, conditional on reporting θ_L and θ_H, respectively. The belief-updating term captures the fact that auditing is only beneficial when misreporting is likely, while the welfare loss component reflects the cost of auditing relative to the social benefit of deterring false reports.

Given the respondent's strategy (x, y) and a reported value $\hat{\theta}$, the institute's belief about the probability that the respondent did *not* acquire the information is computed using Bayes' rule. Specifically:

$$\beta(\theta_L) = \frac{xy}{xy + (1-x)\alpha} \tag{2.3}$$

$$\beta(\theta_H) = \frac{x(1-y)}{x(1-y) + (1-x)(1-\alpha)}. \tag{2.4}$$

[4]This quadratic form is standard in communication games. See, for instance, the seminal paper by Crawford and Sobel (1982) and the survey by Farrel and Rabin (1996) on the Cheap Talk literature.

Note that, whenever $x = 0$ in equilibrium, we have $\beta(\theta_L) = \beta(\theta_H) = 0$: upon receiving either report $\hat{\theta} \in \{\theta_L, \theta_H\}$, the institute believes that the respondent has acquired the information and thus reported truthfully. Conversely, when $x = y = 1$ in equilibrium, we obtain $\beta(\theta_L) = 1$ and $\beta(\theta_H) = 0$. In this case, receiving the report $\hat{\theta} = \theta_L$ leads the institute to believe that the respondent has *not* acquired the information, whereas the report $\hat{\theta} = \theta_H$ is interpreted as truthful.

The following table outlines the payoffs associated with each combination of pure strategies by the respondent and the institute:

Table 1: Payoffs for the respondent and the institute

	Audit	No audit
Acquire	$-K, -A$	$-K, 0$
No acquire, reports θ_L	$(-K - F)(1 - \alpha), -A$	$0, -\beta(\hat{\theta} = \theta_L)(1 - \alpha)\Delta^2$
No acquire, reports θ_H	$(-K - F)\alpha, -A$	$0, -\beta(\hat{\theta} = \theta_H)\alpha\Delta^2$

This summary captures the core trade-offs faced by each player and clarifies how individual decisions interact to shape final outcomes. These payoff structures will underpin the equilibrium analysis in the next section.

2.3 Timing

The timing of the game unfolds as follows:

1. The respondent decides whether to acquire the information. She acquires it with probability $1 - x$ and does not acquire it with probability x.
2. If she acquires the information, she truthfully reports the realized value to the institute. If she does not acquire it (with probability x), she chooses which value to report: θ_L with probability y, and θ_H with probability $1 - y$.
3. The institute observes the reported value—without knowing whether it is truthful—updates its belief about whether the respondent acquired the information, and then decides whether to audit or not. Auditing occurs with probability $z(\hat{\theta})$, which may depend on the reported value.

2.4 Equilibrium

The solution concept used in our game is the Perfect Bayesian Equilibrium (PBE). Given the structure of the game, note that the equilibrium probability of a given action must be consistent with the aggregate behavior of types for whom that action is optimal. In other words, the equilibrium frequency of each action must coincide with the total mass of types for which that action constitutes a best response. This property allows us to formally introduce the following definition.

Definition 2.1 *A PBE of the game is characterized by a strategy profile* $\{(x, y), (z(\theta_L), z(\theta_H))\}$ *and a belief system* $\beta(\hat{\theta})$ *with* $\hat{\theta} \in \{\theta_L, \theta_H\}$*.*

Consider a PBE with $x, y \in (0,1)$. In this case, conditional on no audit being conducted, the institute's expected payoffs after observing $\hat{\theta} = \theta_L$ and $\hat{\theta} = \theta_H$ are, respectively:

$$V_{\hat{\theta}=\theta_L} = -\frac{xy}{xy + (1-x)\alpha}(1-\alpha)\Delta^2$$

and

$$V_{\hat{\theta}=\theta_H} = -\frac{x(1-y)}{x(1-y) + (1-x)(1-\alpha)}\alpha\Delta^2.$$

This implies that the institute audits if and only if $-A \geq V_{\hat{\theta}=\theta_L}$ when $\hat{\theta} = \theta_L$, and $-A \geq V_{\hat{\theta}=\theta_H}$ when $\hat{\theta} = \theta_H$. Therefore, the probability of auditing conditional on the reported value is:

$$z(\theta_L) = G\left(\frac{xy}{xy + (1-x)\alpha}(1-\alpha)\Delta^2\right), \tag{2.5}$$

$$z(\theta_H) = G\left(\frac{x(1-y)}{x(1-y) + (1-x)(1-\alpha)}\alpha\Delta^2\right). \tag{2.6}$$

Recall that $\operatorname{supp} G = \mathbb{R}_{\geq 0}$, such that equations (2.5) and (2.6) imply, for instance, that $z(\theta_L) = 0$ when $x = 0$. This is because, upon receiving the report $\hat{\theta} = \theta_L$ and knowing with certainty that the respondent has acquired the information (i.e., $\beta(\theta_L) = 0$ in this case), the optimal action for the institute is not to audit. It is also straightforward to verify that other extreme cases arise when $y = 0$ or $y = 1$, which similarly yield corner solutions for the institute's auditing strategy.

Now consider the respondent's reporting decision when she chooses not to acquire the information. Since we are assuming $y \in (0,1)$, she must be indifferent between reporting θ_L and θ_H. Given that:

$$\begin{aligned} U_{\hat{\theta}=\theta_L} &= z(\theta_L)(1-\alpha)(-F-K), \\ U_{\hat{\theta}=\theta_H} &= z(\theta_H)\alpha(-F-K), \end{aligned}$$

indifference implies the following equilibrium condition:

$$z(\theta_L)(1-\alpha) = z(\theta_H)\alpha. \tag{2.7}$$

Observe that the above condition can be rewritten as $\frac{z(\theta_L)}{z(\theta_H)} = \frac{\alpha}{1-\alpha}$. This implies that, in equilibrium, the ratio between the probability of auditing after receiving $\hat{\theta} = \theta_L$ and the probability of auditing after receiving $\hat{\theta} = \theta_H$ must equal the ratio of the prior probabilities of θ_L and θ_H, respectively. In equilibrium, the institute distributes its auditing effort in proportion to the prior likelihood of each state, ensuring that the respondent has no incentive to favor one misreport over the other. This balance is necessary to sustain mixed reporting behavior when the respondent chooses not to acquire information.

Finally, the respondent finds it optimal not to acquire the information if and only if $-K \leq z(\theta_L)(1-\alpha)(-F-K)$. This condition allows us to derive the following expression for the equilibrium probability of not acquiring the information:

$$x = 1 - H\left(\frac{Fz(\theta_L)(1-\alpha)}{1 - z(\theta_L)(1-\alpha)}\right). \tag{2.8}$$

The condition above implies that $x = 1$ only if $z(\theta_L) = 0$, and consequently $z(\theta_H) = 0$, as required by equation (2.7). In other words, it is optimal for the respondent not to acquire information only when she knows that no audit will be conducted, regardless of the report she submits.

Because, in equilibrium, the probability of each action must be consistent with the aggregate behavior of types for which that action is optimal, we can state the following proposition.

Observation 2.2 *A PBE of the game is characterized by a strategy profile and belief system that jointly satisfy equations (2.5), (2.6), (2.7), and (2.8), along with the belief updating rules given by (2.3) and (2.4).*

The reasoning developed above allows us to show that the game admits no equilibrium in which either player adopts a pure strategy. Suppose, for instance, that $x = 0$ and $y \in [0,1]$. From equations (2.5) and (2.6), it follows that $z(\theta_L) = z(\theta_H) = 0$. Note that equation (2.5) confirms that $z(\theta_L) = 0$ is consistent with $x = 0$. In this potential equilibrium, the respondent acquires the information, and the institute does not audit. The resulting payoffs are $U = -K$ for the respondent and $V = 0$ for the institute. The institute's beliefs consistent with this outcome are $\beta(\theta_L) = \beta(\theta_H) = 0$. However, by deviating and choosing $x = 1$, the respondent obtains a payoff of 0, which is strictly greater than $-K$. This profitable deviation implies that such an outcome cannot constitute an equilibrium. An alternative way of demonstrating this inconsistency is to observe that $z(\theta_L) = 0$ implies $x = 1$ in equation (2.8), contradicting the initial assumption that $x = 0$.

Proposition 2.3 *There exists no PBE in which either the respondent or the institute plays a pure strategy.*

The absence of a pure strategy equilibrium stems from the strategic uncertainty faced by both players. If the institute were to always audit, the respondent would be deterred from misreporting, making auditing unnecessary and therefore not optimal. Conversely, if the institute never audits, the respondent is tempted to avoid the information cost and report arbitrarily, undermining truthful reporting. Similarly, if the respondent always reports truthfully, the institute has no reason to audit, but then the respondent would prefer not to acquire the costly information. These conflicting incentives create a situation in which no player benefits from sticking to a deterministic action, thus requiring both to randomize in equilibrium.

A complete analysis of the model above, including comparative statics, is challenging due to the high dimensionality of the system that characterizes the equilibrium. However, we are able to reduce its dimensionality and carry out such analysis in specific cases. In particular, when $\alpha = y = \frac{1}{2}$, equation (2.7) implies that $z(\theta_L) = z(\theta_H) = z$, so the equilibrium conditions simplify to the following system:

$$z = G\left(2x\sigma^2\right), \tag{2.9}$$

$$x = 1 - H\left(\frac{Fz}{2-z}\right). \tag{2.10}$$

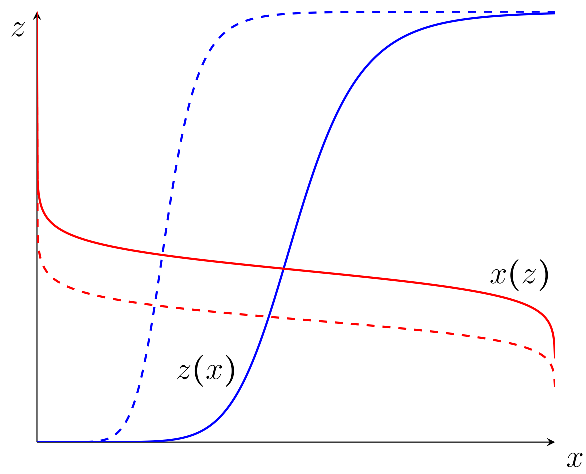
Note that we use the identity $\Delta^2 = \frac{\sigma^2}{\alpha(1-\alpha)}$ to rewrite the first equation. This reduced-form system allows us to investigate how the equilibrium probabilities of acquiring information and auditing respond to changes in the variance of the underlying information.

For this special case, we obtain the following result:

Lemma 2.4 *There exists a unique pair of probabilities $x \in (0,1)$ and $z \in (0,1)$, representing the respondent's probability of not acquiring information and the institute's probability of auditing, respectively, that jointly satisfy the system of equations (2.9)–(2.10).*

There are two ways to prove the above lemma. The first relies on the observation that equations (2.9) and (2.10) define functions $z(x)$ and $x(z)$, respectively. One can verify that $z'(x) > 0$ and $x'(z) < 0$. This implies that the probability of auditing increases with the probability that the respondent does not acquire information, while the probability of not acquiring information decreases as auditing becomes more likely. Moreover, since $z(0) = G(0) = 0$, $z(1) = G(2\sigma) > 0$, $x(0) = 1$, and $x(1) = 1 - H(F) > 0$, the two curves must intersect. Finally, given that both functions are monotonic, they intersect only once. Figure 1 illustrates this result for the case in which the distributions of both A and K are log-logistic with shape parameter $\alpha = 1$ and scale parameter $\beta = 8$. In this case, $H(w) = G(w) = \frac{1}{1+w^{-8}}$, where $w \in \{A, K\}$.

Figure 1: Equilibrium when $\alpha = y = \frac{1}{2}$



Source: Created by the authors

2.5 Comparative statics

We are interested in analyzing how changes in F and σ^2 affect the equilibrium values of x and z. We begin by examining the effect of an increase in the magnitude of the fine F. From equation (2.10), we observe that a higher F directly reduces the probability that the respondent does not acquire information. According to equation (2.9), this reduction in x leads to a lower probability of auditing z. However, a lower z, in turn, increases the incentive not to acquire information—partially offsetting the original effect. The following proposition establishes that, in equilibrium, the direct (negative) effect dominates the indirect (positive) one.

Proposition 2.5 *In equilibrium, both z and x are strictly decreasing in F.*

The result above implies that the higher the fine imposed on the respondent for misreporting—an event that occurs only when she chooses not to acquire information—the more likely the equilibrium outcome is one in which no information is acquired and no audit is conducted. One possible interpretation is that, if the institute has discretion over the choice of F, it should make use of this instrument, as it is highly effective in disciplining the respondent's behavior. We can also interpret Proposition 2.5 geometrically: an increase in F shifts the function $x(z)$, given by equation (2.10), downward. This results in a new intersection point with the function $z(x)$ at lower values of both x and z. This comparative static effect is illustrated in Figure 1, where the dashed red curve represents the shift in the equilibrium locus induced by a higher fine.

As shown in equation (2.9), an increase in σ^2 has a direct effect on z: it increases the probability that the institute chooses to audit. A higher z, in turn, reduces the respondent's incentive to avoid acquiring information, thus lowering x, as seen in equation (2.10). The following proposition establishes that, in equilibrium, the direct (positive) effect on z dominates, while the indirect (negative) effect on x also prevails.

Proposition 2.6 *In equilibrium, z is strictly increasing and x is strictly decreasing in σ^2.*

Higher variance implies greater uncertainty, which strengthens the institute's incentive to audit. This effect is strong enough to reduce the probability that the respondent chooses not to acquire information. Geometrically, an increase in σ^2 shifts the function $z(x)$ to the left, leading to a higher equilibrium value of z and a lower value of x. This suggests that environments characterized by high uncertainty make equilibria in which the institute audits and the respondent acquires the information—thus reporting truthfully—more likely. In Figure 1, this effect is represented by the dashed blue curve, which shows the impact of a higher variance on the equilibrium locus.

3 Concluding remarks

We argue that the assumption of two-sided asymmetric information between statistical institutes and respondents is not only realistic but also essential to understanding the challenges of primary data collection. Our model captures these frictions and offers policy-relevant insights: by adjusting the fine F, institutes can partially substitute costly audits while deterring misreporting. Uncertainty about the underlying information increases both the incentive to misreport and the need to audit, making this a central factor in determining equilibrium behavior.

For tractability, we restricted attention to cases with fixed values of the prior α and the respondent's reporting strategy y, and assumed that misreporting is beneficial only insofar as it allows respondents to avoid the cost of acquiring information. We also modeled penalties as fixed, reflecting the institutional reality that statistical agencies typically sanction detected misreporting or non-response in a binary manner, rather than conditioning fines on the magnitude of the deviation from the truth. Relaxing these assumptions—for instance, allowing penalties to depend on the size of the misreport or respondents to acquire partial information—constitutes a natural avenue for future research.

Additional extensions include endogenizing α and y, reversing the order of moves so that the institute announces the fine first—thereby introducing commitment considerations—and studying environments with multiple heterogeneous respondents and correlated signals. Finally, we note that a promising direction for future work is to calibrate the model using administrative or survey microdata. This would allow researchers to quantify the trade-offs between audit frequency and penalty severity under real-world legal and institutional constraints, providing richer guidance for the design of enforcement mechanisms in official statistical programs.

References

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1:323–338.

Avenhaus, R., Von Stengel, B., and Zamir, S. (2002). Inspection games. *Handbook of game theory with economic applications*, 3:1947–1987.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217.

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. In *Handbook of econometrics*, volume 5, pages 3705–3843. Elsevier.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50:1431–1451.

Farrel, J. and Rabin, M. (1996). Cheap talk. *Journal of Economic Perspectives*, 10:103–118.

Massey, D. and Tourangeau, R. (2013). Where do we go from here? nonresponse and social measurement. *The ANNALS of the American Academy of Political and Social Science*, 645(1):222–236. PMID: 25484368.

Moore, J. C. and Welniak, E. J. (2000). Income measurement error in surveys: A review. *Journal of official statistics*, 16(4):331.

Singleton, R. and Straits, B. (2017). *Approaches to Social Research*. Oxford University Press.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

A Omitted proofs

The proof of Lemma 2.4 is provided throughout the main text.

Proposition 2.3

We proceed by contradiction. Suppose there exists a PBE in which at least one player adopts a pure strategy. The proof for the case in which the respondent always acquires the information, i.e., $x = 0$, is provided in the main text.

Now consider the case in which the respondent never acquires the information, i.e., $x = 1$, and reports a fixed value (i.e., $y = 0$ or $y = 1$). Suppose that $y = 1$, so the respondent always reports $\hat{\theta} = \theta_L$. From equation (2.5), it follows that $z(\theta_L) = G((1-\alpha)\Delta^2) > 0$. Substituting this into equation (2.8) yields

$$x = 1 - H\left(\frac{FG((1-\alpha)\Delta^2)(1-\alpha)}{1 - G((1-\alpha)\Delta^2)(1-\alpha)}\right) < 1,$$

which contradicts the assumption that $x = 1$.

For the case in which $x = 1$ and $y = 0$, equations (2.5) and (2.6) imply $z(\theta_L) = 0$ and $z(\theta_H) = G(\alpha\Delta^2) > 0$. However, this contradicts equation (2.7), which requires $z(\theta_H) = G(\alpha\Delta^2) = 0$ when $z(\theta_L) = 0$.

Next, consider the case in which the institute always audits, i.e., $z(\theta_L) = z(\theta_H) = 1$. From equation (2.8), full auditing increases the expected cost of misreporting, which leads to

$$x = 1 - H\left(\frac{F(1-\alpha)}{\alpha}\right) \in (0,1).$$

Now, consider equation (2.5). In order to obtain $z(\theta_L) = 1$, the argument of the distribution function G would have to go to infinity. However, since $x, \alpha \in (0,1)$, there is no value of y that satisfies this condition. This yields a contradiction. A similar argument applies for $z(\theta_H)$ using equation (2.6).

Finally, consider the case in which the institute never audits, i.e., $z(\theta_L) = z(\theta_H) = 0$. Equation (2.8) then implies $x = 1$, meaning the respondent never acquires the information and reports arbitrarily. However, from equations (2.5) and (2.6), we see that $z(\theta_L) = 0$ requires $y = 0$, while $z(\theta_H) = 0$ requires $y = 1$. These two conditions are incompatible, leading to a contradiction. Furthermore, asymmetric cases in which $z(\theta_L) = 1$ and $z(\theta_H) = 0$, or vice versa, cannot be part of an equilibrium. When the respondent plays a mixed strategy, i.e., $y \in (0,1)$, such cases are ruled out by equation (2.7), which requires the audit probabilities to be proportional to the prior probabilities of the underlying states. When $y = 1$ or $y = 0$, they are excluded by the reasoning presented above. This reinforces the conclusion that no equilibrium in pure strategies can exist.

Since in all possible pure-strategy combinations the equilibrium conditions yield internal inconsistencies, we conclude that no PBE exists in which either player adopts a pure strategy. ■

Propositions 2.5 and 2.6

Let us start by rewriting the system (2.9)-(2.10) as

$$\begin{aligned}
\Lambda_1(z,x) &= G\left(2x\sigma^2\right) - z \\
\Lambda_2(z,x) &= 1 - H\left(\frac{Fz}{2-z}\right) - x.
\end{aligned}$$

We can then build its Jacobian matrix:

$$J_\Lambda(z,x) = \begin{bmatrix} -1 & 2\sigma^2 g\left(2x\sigma^2\right) \\ -\frac{2F}{(2-z)^2} h\left(\frac{Fz}{2-z}\right) & -1 \end{bmatrix},$$

where $g(\cdot) = G'(\cdot)$ and $h(\cdot) = H'(\cdot)$. We can also calculate the following partial derivatives:

$$\frac{\partial \Lambda_1}{\partial \sigma^2} = 2xg\left(2x\sigma^2\right), \quad \frac{\partial \Lambda_1}{\partial F} = 0, \quad \frac{\partial \Lambda_2}{\partial \sigma^2} = 0, \quad \frac{\partial \Lambda_2}{\partial F} = -\frac{z}{(2-z)^2} h\left(\frac{Fz}{2-z}\right),$$

where we omit the arguments for the sake of notation.

Now, we can use the Implicit Function Theorem to perform the comparative static exercise:

$$
\begin{aligned}
\frac{dz}{dF} &= -\frac{\begin{vmatrix} 0 & 2\sigma^2 g\left(2x\sigma^2\right) \\ -\frac{z}{(2-z)^2} h\left(\frac{Fz}{2-z}\right) & -1 \end{vmatrix}}{\left|J_\Lambda\left(Q,P\right)\right|} \\
&= -\frac{\frac{2z\sigma^2}{(2-z)^2} g\left(2x\sigma^2\right) h\left(\frac{Fz}{2-z}\right)}{\left|J_\Lambda\left(Q,P\right)\right|} < 0, \qquad \text{(A.1)} \\
\frac{dx}{dF} &= -\frac{\begin{vmatrix} -1 & 0 \\ -\frac{2F}{(2-z)^2} h\left(\frac{Fz}{2-z}\right) & -\frac{z}{(2-z)^2} h\left(\frac{Fz}{2-z}\right) \end{vmatrix}}{\left|J_\Lambda\left(Q,P\right)\right|} \\
&= -\frac{\frac{z}{(2-z)^2} h\left(\frac{Fz}{2-z}\right)}{\left|J_\Lambda\left(Q,P\right)\right|} < 0, \qquad \text{(A.2)} \\
\frac{dz}{d\sigma^2} &= -\frac{\begin{vmatrix} 2xg(2x\sigma^2) & 2\sigma^2 g\left(2x\sigma^2\right) \\ 0 & -1 \end{vmatrix}}{\left|J_\Lambda\left(Q,P\right)\right|} \\
&= \frac{2xg(2x\sigma^2)}{\left|J_\Lambda\left(Q,P\right)\right|} > 0, \qquad \text{(A.3)} \\
\frac{dx}{d\sigma^2} &= -\frac{\begin{vmatrix} -1 & 2xg(2x\sigma^2) \\ -\frac{2F}{(2-z)^2} h\left(\frac{Fz}{2-z}\right) & 0 \end{vmatrix}}{\left|J_\Lambda\left(Q,P\right)\right|} \\
&= -\frac{\frac{4xF}{(2-z)^2} g\left(2x\sigma^2\right) h\left(\frac{Fz}{2-z}\right)}{\left|J_\Lambda\left(Q,P\right)\right|} < 0, \qquad \text{(A.4)}
\end{aligned}
$$

where we use the fact that $\left|J_\Lambda\left(Q,P\right)\right| = 1 + \frac{4F\sigma^2}{(2-z)^2} g\left(2x\sigma^2\right) h\left(\frac{Fz}{2-z}\right) > 0$. One can check that (A.1), (A.2), (A.3) and (A.4) prove propositions 2.5 and 2.6. ■