

Misunderestimating Corruption

Aart Kraay

Peter Murrell

The World Bank
Development Research Group
Macroeconomics and Growth Team
May 2013



Abstract

: Estimates of the extent of corruption rely largely on self-reports of individuals, business managers, and government officials. Yet it is well known that survey respondents are reticent to tell the truth about activities to which social and legal stigma are attached, implying a downward bias in survey-based estimates of corruption. This paper develops a method to estimate the prevalence of reticent behavior, in order to isolate rates of corruption that fully reflect respondent reticence in answering sensitive questions. The method is based on a statistical model of how respondents behave when answering a combination of conventional and random-response survey questions. The responses to these different types of

questions reflect three probabilities—that the respondent has done the sensitive act in question, that the respondent exhibits reticence in answering sensitive questions, and that a reticent respondent is not candid in answering any specific sensitive question. These probabilities can be estimated using a method-of-moments estimator. Evidence from the 2010 World Bank Enterprise survey in Peru suggests reticence-adjusted estimates of corruption that are roughly twice as large as indicated by responses to standard questions. Reticence-adjusted estimates of corruption are also substantially higher in a set of ten Asian countries covered in the Gallup World Poll.

This paper is a product of the Macroeconomics and Growth Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at akraay@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

MISUNDERESTIMATING CORRUPTION

Aart Kraay (The World Bank)
Peter Murrell (University of Maryland)

JEL Classification Codes: C83, O17, O43

Keywords: Corruption, reticence, random response questions

1818 H Street NW, Washington DC 20433, akraay@worldbank.org; Department of Economics, University of Maryland, College Park, MD 20742, murrell@econ.umd.edu. We would like to thank Nona Karalashvili and Tatjana Kleineberg for invaluable research assistance, and David McKenzie, Gale Muller, Luis Serven, Carlos Silva-Jauregui and Rajesh Srinivasan for helpful comments. We are grateful to the World Bank's Enterprise Survey team (and especially Jorge Meza, Federica Saliola, and David Francis) for fielding the random response questions discussed in the paper. We are indebted to Gallup for their collaboration in fielding these questions in the Gallup World Poll and making the data available to us, and especially to Nicole Naurath and Rajesh Srinivasan for their support of this project. Financial support from the Knowledge for Change Program (KCP) of the World Bank is also gratefully acknowledged. The views expressed here are the authors', and do not reflect those of the Gallup Organization, the World Bank, its Executive Directors, or the countries they represent.

1. Introduction

In a classic study that compared survey responses to official records, Locander et al. (1976) found that 19% of survey respondents in Chicago incorrectly claimed possession of a library card. Recently after radio-monitoring meters were installed in cars in the US, the radio ratings company Arbitron realized that past estimates of commuters' listening patterns had been significantly distorted by the survey responses of men who were claiming to listen to more classical and jazz, and less oldies and country music, than was actually the case.¹ Many studies have documented that survey responses indicate much higher rates of church attendance than can be verified from time use diaries, particularly in the United States (Brenner 2011). More seriously, Gong (2012) combines survey data on self-reported sexual activity with the results of tests for sexually transmitted infections, and finds that the latter provide clear evidence that survey respondents underreport their sexual activity. Imagine then how distorted responses might be if a survey asked about breaking the law in a country where privacy protections and legal rights were of concern to respondents. And how might we know the degree of distortion in the absence of pertinent official records, metering, or testing?

Despite these obvious concerns, economics research on corruption usually ignores the possibility that survey respondents are reluctant to give truthful answers to questions on sensitive topics. Svensson (2003) is a telling example of the approach within economics, both because it is a significant contribution to the literature, uncovering important relationships in the corruption behavior of developing country firms, and because of the relative emphasis it places on different methodological problems. The paper provides a careful assessment of different theories of bribe-giving and their implications for econometric specification and interpretation of results. To obtain a representative sample, data collection relied on the large stock of existing knowledge on sampling techniques. A number of convincing robustness exercises were carried out. But based on the candor of survey responses, the paper is forced to conclude that "...cases of misreporting are likely to remain in the sample. For this reason, the paper has not focused on the level of bribes per se, but rather on their correlates" (Svensson 2003 p. 225).

That sums up the current status of economics research on the reticence of survey respondents. Despite the large amount of survey data from firms that is used in empirical papers and is diffused by popular databases such as the World Bank Enterprise Surveys, the discipline does not have much to say about absolute levels of corruption when using data obtained directly from those who pay bribes or those who receive them.² Our objective in this paper is to remedy that problem by developing a methodology that allows estimation of the degree of candor and reticence of survey respondents, and simultaneously to use these estimates to determine the degree to which corruption itself has been

¹ "Never Listen to Céline? Radio Meter Begs to Differ" By Stephanie Clifford. *New York Times*, December 16th, 2009.

² An exception is Olken (2009), which compared Indonesian villagers' perceptions of corruption in local road-building projects with estimates of actual "missing expenditures", i.e. gaps between what villages reported spending on road-building projects and ex post estimates of the cost of materials based on physical audits of the roads. For obvious reasons, opportunities to directly measure corruption and contrast it with survey-based estimates are rare.

underestimated in the past. Although it is generally recognized that underestimation of corruption is a problem, our results show that it is even more of a problem than seems to have been generally recognized. Hence our paper's title, which borrows a neologism coined by former US President George W. Bush, commenting on how his opponents in the presidential election had severely underestimated him.³

We implement our methodology using data from the 2010 World Bank Enterprise survey in Peru. The baseline estimate of corruption from that survey is that 18% of firms answer that it is common for similar firms to make informal payments to government officials. For terminological convenience, we will refer to such an answer as "guilt" on the part of the respondent, because this is usually interpreted as an admission of making informal payments, even though the question itself does not specifically ask whether the respondent makes such payments. This standard interpretation of a simple yes/no question about corruption assumes that all respondents are always candid when answering all questions. In contrast, we estimate that 74% of respondents are reticent in answering sensitive questions, that reticent respondents are not candid in answering 69% of sensitive questions, and that 37% of respondents in fact find it common to make informal payments, a doubling of the estimate of corruption. For some subsets of firms the results are even more dramatic. For example, more than 90% of the respondents in small firms exhibit reticence, raising estimated guilt from 30% to more than 90%.

These results assume that reticence and guilt are uncorrelated. However, estimates of the probabilities of reticence and guilt for subsets of survey respondents—by firm sector or age, for example—show that a positive correlation is almost certainly the case. This is also intuitively plausible: the guilty have more to hide. If that is the case, reticence-adjusted estimates of corruption are even higher because the most corrupt are least likely to admit to its existence. Allowing for a positive correlation leads to corruption estimates above 50%, in contrast to the 18% indicated by applying standard analyses of the data that we use.

Of course, if rates of reticence were similar across countries, the findings from Peru would not be particularly worrisome for cross-country analysis: corruption would be underestimated in all countries, but by more or less the same amount. However, if rates of reticence vary across countries then comparisons of corruption could be quite misleading. Hence, we also apply the methodology to a set of Asian countries using data from the Gallup World Poll. Effective reticence—the proportion of reticent respondents times the probability that they are not candid on any specific sensitive question—varies dramatically between countries, with a rate of 21% in Indonesia and 48% in India, meaning that existing estimates of corruption are downward biased by less than 2 percentage points in some countries but by more than 20 percentage points in others.

We note at the outset that these findings do not imply that survey-based estimates of corruption are without value. Indeed, the illegal nature of bribery implies that those involved have strong incentives to hide any evidence of such behavior, so that direct measurement of corruption is in

³ See <http://en.wikipedia.org/wiki/Bushism>, retrieved May 11, 2013.

most cases infeasible without prohibitively costly and intrusive audits. In the absence of practical alternatives, survey data on corruption will continue to be an important source of information about corruption. This in turn underscores the need for more research with the same goal as in this paper—seeking to address potential biases in survey data on corruption.

The reticence of respondents in answering sensitive questions has been a concern of survey researchers for a long time.⁴ Much attention has been placed on techniques that aim to mitigate the problem, such as better wording of questions, the optimal structure of interviews, the use of computers, etc. (Tourangeau and Yan 2007). One fundamental contribution was made by Warner (1965), who developed the random-response question (RRQ). In the form used in our empirical work, a RRQ poses an innocuous question alongside a sensitive question and has the respondent answer only one of these. Which question is answered is dictated by the realization of an event (e.g. toss of coin) known only to the respondent, but whose probability of occurrence is known to the surveyor. The motivation for RRQs is the theory that a respondent will be less reticent if the interviewer and recipients of the survey data do not know which question was answered. If that theory is correct and respondents are candid and if the distribution of answers to the innocuous question is known a priori (e.g. probability of heads on a coin-toss), it is trivial to derive unbiased estimates of the prevalence of the sensitive behavior.

Random-response has had some successes. However, respondents often show reticence even when answering such questions. In studies where external validation of survey responses is possible, Lensvelt-Mulders et al. (2005) found that RRQs had 90% of the reticence of conventional face-to-face interview questions (CQs). RRQs performed no better than CQs on such issues as library cards, voting in elections, and arrest records. However, there is one feature of an RRQ that provides opportunities for uses different from the intentions of its designers—the randomization probability embodied in the question affects the relationship between reticence and responses. Using this insight, Clarke and Desharnais (1998), Moshagen and Musch (2012), and Moshagen, Musch, and Erdfelder (2012) suggest creating subsamples of respondents and asking them RRQs with different randomization probabilities. They then derive insights into levels of reticence and guilt.⁵

In the economics literature, Azfar and Murrell (2009) and Clausen, Kraay, and Murrell (2011) used a series of seven RRQs, in firm surveys in Romania and Nigeria respectively. They noted that a "No" answer on any single question implied a coin coming up tails. Since the occurrence of seven tails has a very low probability, these papers classified those responding with seven "No's" as reticent. In these surveys, those so classified reported significantly lower rates of commission of sensitive acts and claimed higher levels of personal ethics. These papers did not estimate population rates of reticence and guilt, since their primary goal was to show how to identify a set of respondents who were reticent with near certainty, to show that there were significant numbers of such respondents, and to examine the distinctive ways in which these respondents answered sensitive questions.

⁴ See for example, Warner (1965), Campbell (1987), Clark and Desharnais (1998), and Tourangeau and Yan (2007).

⁵ Moshagen and Musch (2012) estimate the proportion of respondents who do not follow the RRQ procedure faithfully (non-adherents), which is thought to happen because that procedure places even innocent respondents in a position that looks like they are admitting to the sensitive act. Moshagen, Musch, and Erdfelder (2012) estimate rates of reticence assuming that there are no non-adherents.

Our methodology advances on all these insights. We follow the Azfar-Murrell (2009) definition of reticence—a reticent respondent is one who gives knowingly false answers with a nonzero probability when honest answers to a specific set of survey questions could generate the inference that the respondent might have committed a sensitive act. This definition directly leads to a precise specification of how different types of respondents answer sensitive questions and therefore explicit predictions on how survey answers vary with respondent reticence and guilt. We frame these predictions in terms of three moment functions: the expected value of the response to a sensitive CQ and the expected mean and variance of the number of yes responses on a set of sensitive RRQs. The expected moments are functions of three probabilities—that a respondent is guilty on the sensitive acts, that a respondent exhibits reticence in answering sensitive questions, and that a reticent respondent is not candid in answering any specific sensitive question.

Using those three moment conditions within a standard method of moments estimator, we estimate the three probabilities. We follow Azfar and Murrell (2009) and Clausen et al. (2011) in using a set of seven RRQs, the resultant pattern of responses facilitating identification of reticence. Identification of reticence-adjusted guilt rests on the same insight as that of Clarke and Deshairnais (1998)—that identification requires RRQs with different randomization probabilities. In our implementation we use seven RRQs and one CQ, noting that a CQ is a limiting case of an RRQ.

Our paper proceeds in the following way. In Section 2, we briefly describe the Peru Enterprise Survey, discussing in some detail the CQ and RRQs that we use to develop our estimates of reticence and guilt. In Section 3, we lay out the statistical model of respondent behavior and show how observable moments in the data from the CQ and RRQs reveal information about the prevalence of respondent reticence, reticent behavior on any specific question, and guilt. Section 4 contains our main results from the Peru Enterprise Survey, while Section 5 provides cross-country evidence from the Gallup World Poll. In Section 6 we generalize our statistical model to allow for the realistic possibility that guilt and reticence are correlated across respondents and show how this leads to even higher estimates of guilt in the Peruvian data. Section 7 offers concluding remarks. Details on the survey questions and on the derivations of the relevant moments are in Appendices A and B.

2. The Context

We implement our methodology using data from the 2010 World Bank survey of Peruvian firms (World Bank Enterprise Surveys 2012). Peru is an upper middle-income country with a population of 23 million and an economy that has been one of the fastest-growing in Latin America in the last decade. The survey polled business owners and top managers in a sample of 1,000 firms representative of the economy's private sector. Interviews occurred from April 2010 through April 2011. Given the sensitive nature of some of the data collected, the World Bank Enterprise Survey team emphasizes the efforts made to ensure confidentiality of responses. Its description of the survey methodology notes *"Confidentiality of the survey respondents and the sensitive information they provide is necessary to ensure the greatest degree of survey participation, integrity and confidence in the quality of the data."*

*Surveys are usually carried out in cooperation with business organizations and government agencies promoting job creation and economic growth, but confidentiality is never compromised."*⁶

We use a CQ that is the basis of a very common measure of corruption—the first item of data that the reader encounters when perusing the World Bank's summary of results from the Peruvian survey.⁷ The question asks whether firms are expected to give gifts to public officials "to get things done". Appendix A contains the precise wording of all survey questions used in this paper. Of the 134 countries that the World Bank has surveyed on this question, Peru is ranked the 44th most corrupt. In the subsample of firms that we use—the 707 firms for which we have complete data for both the CQ and all RRQs—18% of firms expect to give informal payments to government officials.⁸ Absent any concerns about respondent reticence, this would be our baseline estimate of corruption in Peru. However, as we shall see, our estimates of the incidence of reticence imply that this is a serious underestimate of the actual prevalence of corruption.

Our RR methodology presents survey participants with a series of ten sensitive questions, listed in Table 1. Respondents privately toss a coin before answering each question and are instructed to answer "Yes" if the coin comes up heads. If the coin comes up tails, they are instructed to answer the sensitive question truthfully. The series of ten questions includes three that ask about less sensitive acts. We do not use the data from these three questions: their inclusion is to give sophisticated reticent respondents the chance to answer "Yes" occasionally without affecting the data that we use. The seven more-sensitive questions used in the analysis are identified in bold in Table 1, but were not so highlighted in the questionnaire itself.

As discussed above, the conventional rationale for deploying RRQs is that they give respondents the opportunity to "camouflage" their responses: the interviewer will not know whether a "Yes" response is actually an admission of guilt or simply the outcome of a coin toss. This in turn is expected to encourage greater candor on the part of respondents. However, as noted above, the success of RRQs in reducing reticent behavior has been shown to be limited. Instead, we initially follow the insight of Azfar and Murrell (2009) who infer the prevalence of reticent behavior from the number of respondents whose answers contain the highly improbable seven "No's" to the seven sensitive RRQs. Based on this, a baseline estimate of reticence is 23.6%.

The top panel of **Figure 1** exhibits the justification for this methodology and why this baseline estimate does not capture all those who are reticent. It shows the actual distribution of "Yes" responses in the Peru survey, and the actual distribution for those who answered yes to at least one question. In addition, it superimposes the hypothetical distribution of responses that would be observed if there were no reticent behavior, and if no respondents had actually done any of the sensitive acts described in the questions. Under this null hypothesis, the number of "Yes" responses should be binomially

⁶ Full details of the methodology can be found at <http://www.enterprisesurveys.org/Methodology>.

⁷ <http://www.enterprisesurveys.org/Data/ExploreEconomies/2010/peru>

⁸ Our estimates differ from those on <http://www.enterprisesurveys.org/Data/ExploreEconomies/2010/peru> for two reasons. First, we restrict our sample to those who have answered the RRQs. Second, we do not apply the sampling weights that are used to produce the World Bank estimates.

distributed, with a success probability of 0.5 -- i.e. respondents answer "Yes" if and only if the coin comes up heads.

The actual distribution differs from this hypothetical distribution in a very obvious way. There is a large mass of 23.6% of respondents with zero "Yes" responses. This is clear evidence of reticent behavior -- respondents are supposed to answer "Yes" every time the coin comes up heads, and so the probability of observing no "Yes" responses over seven coin tosses should only be 0.008 if respondents were correctly following the protocol of the question. Based on this insight, Azfar and Murrell (2009) identify these respondents as reticent, and the remainder as possibly candid. However, even among the possibly candid, the actual distribution of responses still yields too few "Yes" responses. For example, 37% of the possibly candid respondents answer "Yes" as few as one or two times, while under the null hypothesis that they are candid only 22% of respondents should do so. This point is amplified if we assume further that some respondents are guilty on the sensitive acts. This is because the candid who are guilty are expected to answer yes more often than the candid who are innocent. This is clearly seen in the bottom panel of **Figure 1**, which uses 18.25% as the rate of guilt, which is the sample rate of admission on the CQ. In this case, we would expect just 11% of candid respondents to answer "Yes" only one or two times, but in fact 37% of the possibly candid do so.

In sum, **Figure 1** shows that there are a significant number of reticent respondents in the sample. It also indicates that reticent respondents are not always reticent on every question, but rather are sometimes willing to give an answer that might lead to an inference of guilt. Hence, this suggests a model that has at least two parameters, a probability that a respondent is reticent and a probability that a reticent respondent answers yes in any particular instance. However, as comparison of the theoretical distributions in the two panels of **Figure 1** suggests, the distribution of yes answers is also affected by guilt rates. Thus, rates of reticent behavior cannot be estimated without taking into account the effects of a third parameter, the probability of guilt.

3. Modeling the Interview Process

In this section we develop a statistical model of the interview process for CQs and RRQs. Our objective is to provide some structure to describe the interaction between an interviewer, who would like to elicit information, and the respondent, who may prefer not to disclose this information. In our model, we focus exclusively on respondent characteristics that determine the answer to a given question.⁹ In particular, the probability that respondents answer "Yes" to a given question depends on (i) whether they have in fact done the sensitive act in question, i.e. whether they are "guilty", and (ii) whether they are willing to divulge this information, i.e. whether they are "reticent", and (iii) whether they choose to behave reticently on a specific question.

⁹ It is of course very plausible that interviewer characteristics matter as well. Indeed, survey firms typically spend considerable effort training interviewers in ways to establish rapport with respondents in order to elicit truthful responses. And it is also quite plausible that some interviewers are able to do this better than others. However, we do not seek to model interviewer characteristics here.

With respect to guilt, we assume that there is a probability $g \geq 0$ that the respondent has done the sensitive act in question. Moreover, we assume that the event of being guilty is independent across questions: the probability of having done a sensitive act referenced in one question does not affect the probability of having done another sensitive act referenced in another question. This is a strong assumption, and in Section 6 we will consider a more general framework in which there are different "types" of respondents, some of whom might have a high likelihood of guilt across questions while others have a low likelihood of guilt across questions.

With respect to reticence, we assume that reticent respondents differ from candid respondents in that they are less likely to be willing to answer "Yes" to either CQs or RRQs. Specifically, recall that all respondents are supposed to answer "Yes" to the CQ if they have done the sensitive act in question, and they are supposed to answer "Yes" to the RRQ if either they have done the sensitive act in question or if the coin toss comes up heads. Candid respondents always answer "Yes" when they are supposed to. In contrast, we assume that reticent respondents answer "Yes" when they are supposed to with probability $(1 - q) < 1$. We also assume that the event of inappropriately failing to answer "Yes" is independent across all the relevant questions. Finally, we assume that a proportion r of respondents are reticent, while a proportion $1 - r$ are candid. Thus reticence is an unobserved trait that is fixed for a given respondent, but reticent behavior, i.e. failure to answer "Yes" when supposed to, may vary across questions.

With these assumptions and notation in hand, it is straightforward to summarize the interview process using probability trees. In the top panel of **Figure 2**, we show the relevant probability tree describing how a reticent respondent answers a CQ. In this case, the tree is simple. The first branch captures whether the respondent has done the sensitive act in question or not, while the second branch captures whether the reticent respondent inappropriately fails to answer "Yes". Combining these two steps, it is clear that the respondent answers "Yes" with probability $g(1 - q)$: with probability g the respondent has done the sensitive act, and the reticent respondent admits to it with probability $1 - q$. The corresponding tree for a candid respondent is the special case where $q=0$, i.e. the candid respondent answers "Yes" if and only if he is guilty, which occurs with probability g .

In the bottom panel of **Figure 2**, we show the probability tree describing how a reticent respondent answers an RRQ. The branches of the tree corresponding to guilt and reticent behavior are identical to those in the top panel. The only difference is that each question is preceded by a coin toss. If the coin comes up heads, all respondents are supposed to answer "Yes", but only a fraction $1 - q$ of reticent respondents do so. If the coin comes up tails, all guilty respondents are supposed to answer "Yes", but only a fraction $1 - q$ of the reticent respondents do so. This means that the overall probability of observing a "Yes" response on a single RRQ is $0.5(1 - q) + 0.5(1 - q)g = 0.5(1 - q)(1 + g)$. As before, for candid respondents $q = 0$. In this case, candid respondents always answer "Yes" when the coin comes up heads (with probability 0.5), and they also answer "Yes" if the coin comes up tails and they have done the sensitive act in question (with probability $0.5g$), for a total probability of $0.5(1 + g)$.

Of course, for a given respondent, reticence and guilt are unobservable. Nevertheless, we can make inferences about the prevalence of reticence, r , reticent behavior, q , and guilt, g , from the observed responses to CQs and RRQs. Specifically, we will show how the population mean and variance of responses to the RRQ, together with the population mean response to the CQ, combine to identify the three parameters of interest. This then motivates a standard method-of-moment estimator in which sample moments are equated with population moments, and the resulting system of nonlinear equations can be solved for the parameters of interest.

A little more notation aids the exposition. Let z denote a binary random variable equal to one if a specific respondent is reticent, and zero otherwise. z will be one with probability r . Let S be a binary random variable equal to one if the respondent answers "Yes" on the CQ. From the probability tree above, S is a mixture of two Bernoulli random variables:

$$(1) \quad S = zS^R + (1 - z)S^C$$

where S^R is a Bernoulli random variable with success probability $g(1 - q)$, i.e. the reticent respondent has in fact done the sensitive act and in addition chooses to respond truthfully; and S^C is a Bernoulli random variable with success probability g , i.e. the candid respondent is guilty and accordingly answers "Yes".

Turning to the RRQ, recall that we ask a series of n questions. Let X be a random variable reflecting the number of "Yes" responses on the RRQs, which is a mixture of binomial random variables:

$$(2) \quad X = zX^R + (1 - z)X^C$$

where X^R is a binomial random variable with n trials and success probability $0.5(1 - q)(1 + g)$, capturing the number of "Yes" responses given by a reticent respondent; and X^C is a binomial random variable with n trials and success probability $0.5(1 + g)$ capturing the number of "Yes" responses given by a candid respondent. With this notation, it is straightforward to compute the population mean "Yes" responses to the CQ and the sequence of RRQs:

$$(3) \quad E[S] = g(1 - rq)$$

and

$$(4) \quad E[X/n] = 0.5(1 + g)(1 - rq)$$

The intuition for these expressions is straightforward given the preceding discussion. On the CQ, respondents answer "Yes" only if they have done the sensitive act in question, which occurs with probability g . Among these, a fraction r are reticent, and respond "Yes" with probability $1 - q$, while the remaining $1 - r$ are candid and answer "Yes" with probability 1. This gives an overall frequency of "Yes" responses of $g(r(1 - q) + (1 - r)) = g(1 - rq)$. Thus rq is an "effective" rate of reticence—the probability that an individual is reticent times the likelihood of behavior in a reticent manner on a given question. Similarly, on the sequence of RRQs, respondents are supposed to answer "Yes" with

probability $0.5(1 + g)$, reflecting both guilt and the instruction to answer "Yes" whenever the coin comes up heads. However, only candid respondents—a proportion equal to $1 - r$ —always follow these instructions, while the proportion r —the reticent respondents—do so only with probability $1 - q$, for an overall probability of $0.5(1 + g)(1 - rq)$.

A first observation that follows immediately from Equations (3) and (4) is that the frequency of guilt can be inferred from the mean "Yes" responses to the two types of questions. Specifically, taking the ratio of these two expressions and rearranging gives

$$(5) \quad \frac{g}{1 + g} = \frac{1}{2} \frac{E[S]}{E[X/n]}$$

The intuition for this expression follows from noting that we should expect that the average rate of "Yes" responses on the battery of RRQs will be higher than the rate of "Yes" responses on the CQ. This is because, for a given rate of effective reticence (rq), we will see more "Yes" responses on the RRQ coming from the cases where the respondent tosses "Heads" and follows the protocol of answering "Yes" even if not guilty.

The key point is that the difference in the rate of "Yes" responses across the two types of questions is informative about the prevalence of guilt. To see this, consider the extreme case where the average rate of "Yes" responses is the same on the CQ and the RRQ battery. The only way this can occur is if the probability of guilt is equal to one, since then there will be no inaccurate "Yes" responses in the RRQ coming from the "Heads" outcome. Any deviation from this benchmark tells us that the probability of guilt is less than one. For example, at the other extreme, suppose that we observe a much higher incidence of "Yes" responses on the RRQ battery than on the CQ. The only way this can happen is if the true rate of guilt is low.

A crucial identifying assumption here is that the rate of guilt is the same across all of the RRQs and the CQ. As with most identifying assumptions, this is untestable. We discuss the plausibility of this assumption in our empirical context in Section 4.

Given that the population frequency of guilt is identified by the ratio of the two means, it is straightforward to retrieve an estimate of the rate of effective reticence, rq , from one or the other of the two means. However, thus far we do not have sufficient information to separately identify the proportion of reticent respondents, r , and the frequency of reticent behavior, q . Intuitively, if for a given rate of guilt we observe a low rate of "Yes" responses on either the CQ or the RRQs, we cannot tell if this is because there are many reticent respondents, i.e. r is high, or if this is because reticent respondents are very likely to behave in a reticent manner on a given question, i.e. if q is high. All we know is that the effective rate of reticent behavior is high.

We next show that it is possible to separately identify r and q by drawing on additional information contained in the variance of the number of "Yes" responses in the RRQ battery. The intuition for identification comes from our assumption that reticence is an individual-specific trait which is fixed across questions for a given respondent, while the event of reticent behavior, i.e. failure to answer "Yes" when supposed to, is independent across questions. This immediately has the implication

that the presence of reticence will be reflected in some positive covariance of responses across individual questions in the RRQ battery: reticent respondents are less likely to answer "Yes" to RRQs precisely because they are reticent. This in turn will be reflected in the variance of the number of "Yes" responses over the individual RRQs.

To show this formally, the variance of the average number of "Yes" responses is:

$$(6) \quad V[X/n] = \frac{1}{n} 0.5(1+g)(1-rq)(1-0.5(1+g)(1-rq)) \\ + \frac{(n-1)}{n} (0.5(1+g)q)^2 r(1-r)$$

This equation has a straightforward intuition. The first line is recognizable as the variance of the mean of a binomial random variable with n trials and a success probability $0.5(1+g)(1-rq)$. The second line captures the covariance across responses to individual RRQs that is due to the presence of reticent respondents. To see this, consider the polar case where there are no reticent respondents, i.e. $r = 0$. In this case, all respondents are candid and answer "Yes" to each RRQ with probability $0.5(1+g)$. Since guilt and the coin toss are both independent across questions, the average rate of "Yes" responses is simply a binomial distribution with this success probability and with variance as given in the first line of Equation (6), setting $r = 0$. And in this polar case the covariance term in the second line is of course also zero. The same is true for the opposite polar case where $r = 1$. In this case all respondents are reticent, and they answer "Yes" to each RRQ with probability $0.5(1+g)(1-q)$. Again answering "Yes" is independent across questions and so Equation (6) again simplifies to the variance of a binomial distribution and the covariance term in the second line is zero.

However in the intermediate case $0 < r < 1$, the covariance term does not vanish, precisely because individual-specific reticence induces a positive covariance across RRQs, and this in turn contributes to the variance of mean "Yes" responses in the RRQ battery. Crucially, this covariance term depends on q and r separately, and this is how the variance allows us to separately identify these two parameters. Specifically, slight rearrangement of the last term in Equation (6) shows that the variance depends on g , rq , and $1/r$, and moreover is linear in $1/r$. Since we have already seen that the moment conditions in Equations (3) and (4) can be solved for g and rq , this establishes that Equations (3), (4), and (6) can together be solved for g , r , and q .

Note also the important role played by having multiple questions in the RRQ battery. If we had only one RRQ, i.e. $n = 1$, the covariance term in the second line of Equation (6) disappears, simply because there is just one RRQ and so its covariance with another is not defined. As the number of RRQs becomes large, the first term in Equation (6) disappears. This is due to the law of large numbers given our assumptions that both the coin toss and reticent behavior are independent across questions. However, the second term does not decline with n as it captures the individual-specific reticence effect which does not average out across multiple RRQs. In fact, the larger is the number of RRQ questions,

the more informative about the proportion of reticent respondents is the variance in the number of "Yes" responses in the RRQ battery.¹⁰

To summarize, in this section we have suggested a simple statistical model of how respondents behave when asked sensitive questions in both the CQ and the RRQ format. In the model, there are three key parameters of interest: the probability a respondent is guilty in the sense of having done the act referred to in the sensitive question, g , the probability that a respondent is reticent, r , and the probability that a reticent respondent chooses to behave reticently in response to a specific sensitive question. We have shown how there is a unique mapping between these three parameters and three population moments: the variance of the mean number of "Yes" responses in the RRQ battery, that mean itself, and the mean rate of "Yes" responses in the CQ. In the next section we exploit this mapping to retrieve empirical estimates of these parameters using a standard method of moments estimator that equates the population moments with their sample counterparts and then solves for the parameters of interest.

4. Results from the Peru Enterprise Survey

In this section we use the Peruvian enterprise data to estimate reticence and guilt. The data are the individual observations from the 707 respondents who answer all seven sensitive RRQs listed in Table 1 and the sensitive CQ on whether firms are expected to give gifts to public officials "to get things done". We measure the values of two variables for each respondent—a dummy indicating a "Yes" response on the CQ and the mean of the seven dummies capturing "Yes" responses on the seven RRQs. We then calculate three sample moments—the means of these two variables and the variance of the mean of the seven RRQ dummies. These are the sample analogs of the moment expressions (3), (4), and (6).¹¹ Estimation of the three parameters g , r , and q using the method of moments with three moment conditions is therefore straightforward.

The estimation uses the identifying assumption that the parameter g is the same for all seven RRQs and for the CQ. For two reasons, any application of our procedure would always require such an assumption. First, each of the seven RRQs must be distinct. Second, asking the sensitive part of one of the RRQs again as a CQ would introduce the concern that some respondents might begin to question the integrity of the survey procedure itself.

The CQ asks about instances of bribing government officials, which would normally be regarded as a worse offense than the breaking of tax laws. Judging by the distribution of responses given in Table 1, the two questions on the breaking of tax laws are the least serious in our set of seven RRQs, at least in the perception of the respondents. Thus, it is reasonable to assume that our CQ has roughly the same

¹⁰ In practical applications, this leads to a trade-off in experimental design. The above statistical considerations suggest a large number of RRQs. Practicalities in implementation—finding a large number of sensitive questions with guilt rates similar to that of the CQ and then administering all these as RRQs—suggests a small number of RRQs.

¹¹ For ease of implementation in estimation, we use the moment condition for the square of the seven dummies rather than the variance. Given that the mean of the seven dummies is one of the moment conditions also, this is formally equivalent to using the variance. The intuition of our procedure is better captured using the variance.

degree of sensitivity as the average of the RRQs. Later in this section, we present a robustness exercise that justifies this point.

Our core results appear in Table 2. In addition to estimates of g , r , and q , we present estimates for one composite parameter, effective reticence (rq), which reflects the proportion of sensitive questions that are not answered candidly for the whole sample. This proportion is approximately 50% in our estimates. Therefore the predicted guilt rate on making informal payments to public officials approximately doubles, from the 18% that comes from naively accepting the answers to the CQ to the 37% in Table 2. Remarkably, 74% of survey respondents exhibit markers of reticent behavior, with those respondents acting in this way on 69% of sensitive questions.

Although these results make a strong argument that corruption estimates must take into account reticence, results for subsets of the sample do so even more strongly. Table 3 presents estimates for five different disaggregations of the sample—sector, firm size, region, firm age, and gender of respondent.¹² Thus, for example, while baseline estimates suggest that Arequipa is 40% more corrupt than Lima, reticence-adjusted estimates suggest that the former has triple the rate of corruption of the latter. Similarly, while standard estimates suggest that very small firms have three times the rate of corruption of large firms, our results indicate the difference is six-fold. One important qualification however is that the sample sizes within some of these groups are quite small, so that the group-specific parameters are less precisely estimated. This means that in most cases we cannot reject the null hypothesis that the group-specific parameters are significantly different from the corresponding point estimates in the full sample.

One interesting way to reflect on these data is to examine whether there are any reversals in the ordering of estimates of corruption, in the sense that some sub-sample seems more corrupt than another in the baseline estimates but less corrupt in the reticence-adjusted estimates. **Figure 3** provides an easy visualization of the phenomenon. Of the 19 subsamples, 13 exhibit reversals but in most cases only in minor ways. The reason for few significant reversals is that guilt and effective reticence are so highly correlated across subsamples, as **Figure 4** clearly shows. Arequipa is an important exception to these stylized facts. According to the baseline estimates, it is ranked the thirteenth most corrupt of the nineteen subsamples but the reticence-adjusted estimates give it fourth place because it receives the highest estimated rate of effective reticence in the subsamples.

As a check on a major identifying assumption used in the empirics—that the rates of guilt on the CQ and the RRQs are roughly comparable—we obtain estimates equivalent to those in Table 2, but use a subset of four RRQs, dropping the three with the lowest rates of yes responses of the original seven that are used. (From equation (5), this will lead to a decrease in the estimated rate of guilt.) The results—in Table 4—are very similar to those of our core results, with guilt and effective reticence more than 90% of that in Table 2.

¹² Four cities are represented in the survey, but for two, Chiclayo and Trujillo, estimates of guilt are above one. We omit results from these regions.

5. Results from the Gallup World Poll

In the previous section we saw in the Peruvian data that there were both (i) considerable differences across subgroups of the population in terms of naive estimates of guilt based on a literal interpretation of conventional survey questions about corruption, and (ii) substantial differences across subgroups in the gap between reticence-adjusted and naive estimates of corruption. This suggests that naive estimates of corruption not only underestimate its true prevalence, but may even lead to an incorrect ranking of subgroups of a population according to estimates of the prevalence of corruption.

We now investigate the same issue in a cross-country setting. Our dataset consists of household survey data from a set of Asian countries included in the 2010 wave of the Gallup World Poll (GWP). The GWP is a large cross-country survey fielded annually since 2006 in over 150 countries representing 95% of the world's adult population. The GWP gathers respondents' views on a wide range of topics, using in-depth face-to-face interviews, except in some developed economies where the survey is implemented by telephone. The core GWP questionnaire is designed to be comparable across all countries. Within each country the sample is constructed to be representative of the population aged 15 and over.

The majority of questions in the GWP are non-political, focusing on respondents' well-being, happiness, life satisfaction, expectations about the future, etc. The GWP also asks a number of questions about respondents' confidence in public institutions, as well as a specific question about respondents' personal experiences with corruption. This question, which asks whether the respondent has been in a situation in the past year where a bribe was expected, is used as the CQ.

With the generous collaboration of Gallup, we also placed a 10-question RRQ on the questionnaires used in 12 Asian countries included in the 2010 wave of the GWP. The RRQ followed the same structure as the RRQ placed in the Peru Enterprise Survey. However, the specific sensitive questions were modified to reflect the fact that the respondents were randomly-selected individuals rather than firm managers. The 10 specific questions included in the coin toss battery, together with the average number of "Yes" responses, is reported in Table 5. The seven more sensitive questions that we use in the empirical analysis are indicated in bold.

The pattern of responses to the RRQ indicates that reticent behavior is likely to be present among GWP respondents. While full compliance with the coin toss question would result in "Yes" rates of at least 50%, this is rarely the case, and particularly on the sensitive questions indicated in bold. For example, in all 12 countries, the average "Yes" rate is less than 50% on three of the sensitive questions, dealing with theft of property and with insulting a family member. The same is true in 11 out of 12 countries for the questions dealing with cheating and hiding money from family members. Taken together, fully 73 out of 84 average responses to the seven sensitive questions in 12 countries result in "Yes" rates of less than 50%.

Using the methodology described in Section 3, we compute estimates of reticence and guilt for 10 of the 12 countries where we have data on both the RRQ and the CQ. We lose one country because the CQ dealing with personal experiences with corruption was not asked in China. We also do not

report results for Afghanistan where our estimate of the rate of guilt is greater than one and thus falls outside the feasible parameter space. The results appear in Table 6, which reports estimates of g , r and q , as well as effective reticence, rq . This latter measure is particularly relevant because it indicates the proportion of sensitive questions that are not answered candidly. Finally, for reference the Table also reports naive estimates of guilt.¹³

The estimated rate of effective reticence, rq , ranges from 21% in Indonesia to 48% in India. While these rates are all lower than the estimated rate of effective reticence of 51% in Peru, they still indicate that a substantial fraction of questions are not answered candidly in these countries. Naturally, this prevalence of reticent behavior implies that estimated rates of guilt are higher than those implied by average responses to the CQ. Looking across countries, the fraction of respondents reporting personal experience with corruption ranges from a low of 7% in Indonesia to a high of 23% in India. However, our estimates of guilt range from 9% in Indonesia to 44% in India.

It is also interesting to note that many of the cross-country differences in estimated guilt and reticence are large relative to the corresponding standard errors of the estimates. To illustrate this, we test the null hypothesis that each of the estimated parameters in Table 6 is equal to the corresponding point estimates that we would obtain if we were to pool data from all 10 countries in our estimation. We then indicate with * (**) (***) those cases where we can reject this null hypothesis of parameter equality at the 90 (95) (99) percent significance level. Consider for example the estimated rate of guilt, g . The pooled estimate combining data for all countries is 0.286, and we reject the null hypothesis that the country-specific estimates of g are equal to the pooled estimate in 5 out of 10 countries. For effective reticence, the pooled estimate of rq is equal to 0.381, and we find that in 8 out of 10 countries, country-specific estimates are significantly different from this pooled estimate.

Figure 5 illustrates the consequences of reticent behavior for estimates of the prevalence of corruption for the full set of 10 countries included in our analysis. The top panel compares naïve and reticence-adjusted estimates of guilt. Consider for example the comparison of Indonesia and Malaysia, which have similar naïve estimates of the prevalence of corruption of 7% and 9% respectively. However, our estimates suggest that effective reticence is much more common in Malaysia (44%) than it is in Indonesia (21%). As a result, our reticence-adjusted estimates of corruption increase much more in Malaysia (to 15%) than they do in Indonesia (to 9%). Note also that the corruption ranking of six of the ten countries differs between the naïve and the estimated rates of guilt. Overall, our estimates from the GWP show that reticent behavior is likely to lead to substantial downward biases in estimates of corruption based on CQs, and moreover that the size of these biases can differ non-trivially across countries.

¹³ In our estimates, we restrict our sample to those who have answered the RRQs, and we also do not apply the sampling weights that are used to produce the GWP estimates. As a result, the sample means of the CQ reported in the last row of Table 6 differ from those reported by Gallup.

6. Extension to Correlated Guilt and Reticence

In this section, we suggest a realistic extension of the basic statistical model we developed in Section 3 of the paper, which is motivated by the results of the empirical applications in Sections 4 and 5. This extension moves to greater realism in two dimensions. First, it relaxes our previous strong assumption that guilt is independent across questions. This captures the plausible notion that individuals who are willing to do certain bad things are also more likely to be willing to do other bad things. Second, by allowing guilt to be a persistent individual-specific trait, we are now also able to allow reticence and guilt to be correlated across individuals. This captures the plausible idea that individuals who have done bad things are also more likely to be reticent since they have something to hide. Alternatively, it allows for the case where individuals who are unprincipled in the sense of being willing to do bad things are also brazen about their misdeeds. In either case, this extension of the model allows us to consider some non-zero correlation between guilt and reticence, which was not possible in the model of Section 3.

We assume that a fraction p of respondents are "principled" in the sense that they have never done any of the sensitive acts referred to in the CQ and RRQs, while a fraction $1 - p$ of respondents are "unprincipled" in the sense that they have done all of the sensitive acts referred to in these questions. In terms of the notation of Section 3, this amounts to assuming that a fraction p of respondents has $g = 0$ while the remaining $1 - p$ has $g = 1$.

Figure 6 shows the basic probability tree that sorts respondents into types. The first branch of the tree distinguishes principled and unprincipled respondents, while the second branch distinguishes reticent from candid respondents. The key innovation here is that principled respondents are assumed reticent with probability $r - \varepsilon$, while unprincipled respondents are reticent with probability $r + \varepsilon$. Thus if $\varepsilon > 0$ ($\varepsilon < 0$) the correlation between guilt and reticence is positive (negative). We leave implicit the natural restriction that $r + \varepsilon$ and $r - \varepsilon$ are both bounded between zero and one.

Depending on the type of the individual, the remainder of the question-response process is the same as in Section 3. The probability trees describing this process are the same as those shown in **Figure 2**, but specializing to the appropriate choice of g . In particular, the decision tree for reticent and principled respondents is the same as shown in **Figure 2**, but setting $g = 0$ since we have assumed that principled respondents never do any of the sensitive acts. Similarly, the decision tree for reticent and unprincipled respondents, whom we assume have done all of the sensitive acts, is also the same as shown in **Figure 2**, but setting $g = 1$. Finally, the corresponding decision trees for candid respondents are as before in the special case of setting $g = 0$.

We now extend the notation of Section 3 to capture the responses of principled and unprincipled respondents. As before let z be a Bernoulli random variable taking the value 1 if the respondent is reticent, and zero otherwise, and now let w be a Bernoulli random variable taking the value 1 if the respondent is principled, and zero if unprincipled. We can then write the number of "Yes" responses on the CQ as a mixture of four Bernoulli random variables taking the value of 1 if the corresponding type of respondent answers "Yes":

$$(7) \quad S = zwS^{RP} + z(1-w)S^{RU} + (1-z)wS^{CP} + (1-z)(1-w)S^{CU}$$

where S^{RP} and S^{CP} have a success probability of zero (principled individuals never do any sensitive acts and so never say "Yes", whether they are reticent or candid); and S^{RU} and S^{CU} have success probabilities of $1 - q$ and 1 , respectively. (Unprincipled respondents are always guilty, but only a fraction $1 - q$ of the reticent ones admit to it.)

We can similarly write the number of "Yes" responses on the RRQ battery as a mixture of four binomial random variables corresponding to the four types of respondents:

$$(8) \quad X = zwX^{RP} + z(1-w)X^{RU} + (1-z)wX^{CP} + (1-z)(1-w)X^{CU}$$

where X^{RP} and X^{CP} have a success probability of $0.5(1 - q)$ and 0.5 respectively. (Principled individuals never do any sensitive acts, and so the candid ones answer "Yes" only when the coin comes up heads, while the reticent ones do so only when the coin comes up heads *and* they choose not to behave reticently.) X^{RU} and X^{CU} have success probabilities of $1 - q$ and 1 respectively. (Unprincipled individuals are guilty of all of the sensitive acts, and so they should say "Yes" whether the coin comes up heads or tails, but the reticent ones do so only with probability $1 - q$.)

With this notation it is straightforward to calculate the average number of "Yes" responses on the CQ and RRQ battery as before:

$$(9) \quad E[S] = (1 - p)(1 - (r + \varepsilon)q)$$

and

$$(10) \quad E[X/n] = (1 - p)(1 - (r + \varepsilon)q) + p0.5(1 - (r - \varepsilon)q)$$

These expressions are the natural generalizations of Equations (3) and (4) in Section 3. First, note that they are exactly equivalent to (3) and (4) when $\varepsilon = 0$, after making the substitution $(1 - p) = g$. Hence, in the benchmark case of zero correlation between reticence and guilt, the ratio of these two moment conditions identifies $1 - p$ and given this estimate of p , we can retrieve an estimate of rq from one of the two means. Finally, we can separately identify r and q from the variance of the RRQ, as we did in the previous model. There is however one difference here even in the case where $\varepsilon = 0$. In this generalized version of the model a fraction $1 - p$ of respondents are always guilty, and so if they were candid they would be answering "Yes" to all of the questions in the RRQ battery, thereby increasing the variance of X . Since not many respondents answer in this way, matching the variance of X entails inferring that there is a smaller proportion of candid respondents in the sample.

This is precisely what we find in Table 7, when we return to the Peruvian data. In the first column, we consider the basic case where $\varepsilon = 0$, i.e. reticence and guilt are uncorrelated. We find that the proportion of principled respondents is 0.63, implying a proportion of unprincipled respondents, $1 - p = 0.37$, which is the same as the estimated rate of guilt from our basic model in Sections 3 and 4.

Moreover, the estimated rate of effective reticence, r_q , is the same as it was in the basic model. Consistent with the intuition given above, however, we find a higher proportion of reticent respondents, with $r = 0.82$, as opposed to $r = 0.74$ in the basic model.

More generally, for , given guilt and reticence rates, the average number of "Yes" responses on the CQ will be lower if the correlation between reticence and guilt is positive, i.e. if $\varepsilon > 0$. This is because reticent behavior is now more likely among those respondents who otherwise would have answered "Yes", the unprincipled individuals (who have done the sensitive act). The average number of "Yes" responses on the RRQ battery is a weighted average of the responses of principled and unprincipled individuals. Unprincipled individuals are always guilty and are reticent with probability $+ \varepsilon$. The first term in Equation (10) is thus $1 - p$ times the average response on the RRQ in Equation (4), but with $g = 1$ and replacing r by $r + \varepsilon$. Similarly, principled individuals are never guilty, and also are reticent with probability $r - \varepsilon$, and so the second term in Equation (10) is p times Equation (4), but setting $g = 0$ and replacing r with $r - \varepsilon$.

In the remaining columns of Table 7 we impose alternative values of ε , which governs the correlation between guilt and reticence. In columns (2) and (3) we consider values of $\varepsilon = 0.2$ and $\varepsilon = 0.4$, corresponding to successively stronger correlations between reticence and guilt. The main effect of this change is to substantially decrease our estimate of the proportion of principled individuals, which falls from $p = 0.63$ in the baseline to $p = 0.34$ when $\varepsilon = 0.4$. This of course implies an increase in the estimated proportion of guilty individuals, to $1 - p = 0.66$. This figure is substantially higher than the corresponding estimate of $g = 0.37$ in the basic model in Table 2. The intuition for this increase is straightforward: if guilt and reticence are positively correlated then more of the "No" responses that we observe in the data are coming from individuals who did in fact do the sensitive act in question. Thus for a given observed rate of "No" responses, the proportion of unprincipled respondents in the data must be higher than in the case where reticence and guilt are assumed to be independent across respondents.

If on the other hand we assume that guilt and reticence are negatively correlated, we obtain the opposite finding. Under this assumption, "Yes" responses are more likely to signal actual guilt, since unprincipled respondents are assumed to be more likely to be willing to admit to having done sensitive acts. In this case, for a given observed rate of "Yes" responses, we must conclude that the proportion of unprincipled respondents is lower than under the benchmark where guilt and reticence are independent. This can be seen in column (4) of Table 7, where $\varepsilon = -0.1$. However, in this case we also find that the estimated proportion of reticent respondents becomes implausibly high. In column (4) the proportion of principled respondents who are reticent, $r - \varepsilon$, is $0.88 - (-0.1) = 0.98$, with the implication that essentially all principled respondents are reticent.¹⁴ Given that this conclusion is absurd, we think the case where reticence and guilt are positively correlated is much more likely to characterize the Peruvian observations.

¹⁴ In fact, if we go to values of $\varepsilon = -0.2$ or lower, we find that the estimated proportion of principled and reticent respondents, $r - \varepsilon$, becomes greater than one, which is in the inadmissible region of the parameter space.

The estimates of Section 4 suggest that the actual rate of corruption is more than double that derived from the usual procedure of naïvely using the responses of survey respondents. However, estimates from subsamples of the Peruvian data and the current discussion—and basic common-sense—strongly suggest that there is a positive correlation between reticence and guilt. Hence, the change from the 18% of the naïve estimates to the 37% of Section 4 is itself probably a misunderestimation of the effect of reticence on the measurement of corruption.

7. Conclusions

This paper is motivated by the uncontroversial observation that survey respondents may not always respond candidly when asked sensitive questions about their personal behavior. This is true across a broad range of topics, and we specifically focus on the implications of this observation for survey-based data on corruption. Such data, gathered systematically in many different surveys of households and firms, are intensively used in policy analysis and in public discourse about the prevalence of corruption and the success (or failure) of policies to reduce it. While there is widespread agreement that respondent reticence implies downward biases in survey-based estimates of corruption, little is known about the magnitude of these biases. Moreover, it is also well-understood in the survey-research literature that standard solutions to address respondent reticence, such as random response questions, have had, at best, mixed success.

In this paper we have proposed a novel methodology for estimating the frequency and consequences of reticent behavior. We develop a statistical model of how responses to sensitive survey questions are influenced by three characteristics of respondents that are not directly observable: whether they are reticent, whether they behave reticently in response to a particular question, and whether they are guilty in the sense of having done the sensitive act in question. We show how the population frequency of these three characteristics can be estimated using a method of moments estimator that exploits information in average responses to a CQ and RRQs as well as the variance of RRQs.

We implement this methodology in the World Bank's Enterprise Survey for Peru and in a sample of 10 Asian economies covered by the Gallup World Poll. In both applications, we find that reticent behavior is common: around two-thirds of firm managers surveyed in Peru exhibit signs of reticent behavior, while in the Gallup World Poll the fraction of individuals exhibiting reticence ranges from 40% in Indonesia to 81% in India. This in turn implies that the simple average of responses to standard questions about corruption substantially underestimates the prevalence of corruption. Our reticence-adjusted estimates of corruption in Peru are twice as high as standard estimates (37% versus 18%), and similarly we see much higher estimates of corruption in the Gallup World Poll after we adjust for reticent behavior. Moreover, corruption estimates are even higher in models that allow for a positive correlation between reticence and guilt, which is indicated strongly by our results.

An immediate implication of our findings is that self-reported survey data on the incidence of corruption substantially underestimate its actual prevalence. More practically, our findings underscore the importance of refining survey techniques to improve the measurement of corruption. This includes

finding credible and easy-to-implement markers of reticent behavior that can be routinely included in surveys that aim to gather sensitive data, as well as deploying novel survey techniques to encourage greater candor (for example using computer-assisted interviewing to reduce the sensitivity of face-to-face interactions between respondents and interviewers).¹⁵ This research agenda is particularly important in the case of corruption, where alternatives to self-reported survey-based data are rare.

¹⁵ Tourangeau and Yan (2007) provide a valuable survey of the results from many different experiments to improve the accuracy of responses on sensitive questions, concluding that "The need for methods of data collection that elicit accurate information is more urgent than ever."

References

- Azfar, Omar and Peter Murrell. 2009. "Identifying Reticent Respondents: Assessing the Quality of Survey Data on Corruption and Values" *Economic Development and Cultural Change*, January 57(2), pp. 387-412.
- Brenner, Philip. 2011. "Exceptional Behaviour or Exceptional Identity? Overreporting of Church Attendance in the US". *Public Opinion Quarterly*. 75(1):19-41.
- Campbell, A. 1987. Randomized response technique. *Science*, 236, 1049.
- Clark, S. J., & Desharnais, R. A. 1998. "Honest answers to embarrassing questions: Detecting cheating in the randomized response model". *Psychological Methods*, 3, 160–168.
- Clausen, Bianca, Aart Kraay, and Peter Murrell. 2011. "Does Respondent Reticence Affect the Results of Corruption Surveys? Evidence from the World Bank Enterprise Survey for Nigeria" *International Handbook on the Economics of Corruption, Volume 2*, edited by Susan Rose-Ackerman and Tina Sørdeide, 2011.
- Gong, Erick. 2012. "HIV Testing and Risky Sexual Behaviour". Manuscript, Middlebury College.
- Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. J. M. 2005. "Meta-analysis of randomized response research: Thirty-five years of validation." *Sociological Methods & Research*, 33, 319–348.
- Locander, W., Sudman, S., & Bradburn, N. 1976. "An investigation of interview method, threat and response distortion." *Journal of the American Statistical Association*, 71, 269–275.
- Moshagen, M., and Musch, J. 2012. "Assessing multiple sensitive attributes using an extension of the randomized-response technique." *International Journal of Public Opinion Research*. Vol. 24 No. 4.
- Moshagen, Morten, Jochen Musch, and Edgar Erdfelder. 2012. "A stochastic lie detector" *Behavior Research Methods* 44:222.
- Olken, Benjamin. 2009. "Corruption Perceptions vs. Corruption Reality." *Journal of Public Economics* 93: 7, August pp. 950-964.
- Svensson, Jakob. 2003. "Who Must Pay Bribes and How Much? Evidence from a Cross Section of Firms" *Quarterly Journal of Economics* Volume 118, Issue 1 pp. 207-230
- Tourangeau, R., & Yan, T. 2007. "Sensitive questions in surveys." *Psychological Bulletin*, 133, 859–883.
- Warner, S. 1965. "Randomized-response: A survey technique for eliminating evasive answer bias." *Journal of the American Statistical Association*, 60, 63–69.
- World Bank Enterprise Surveys. 2012. World Bank. Washington D.C.
<http://www.enterprisesurveys.org/>

Appendix A: The Survey Questions

The Random Response Questions in Peru

The questionnaire was administered in a face-to-face interview with a professional surveyor. The interviewer was asked to read the following to the respondent:

"We have designed an alternative experiment which provides the opportunity to answer questions based on the outcome of a coin toss. Before you answer each question, please toss this coin and do not show me the result. If the coin comes up heads, please answer "yes" to the question regardless of the question asked. If the coin comes up tails, please answer in accordance with your experience. Since I do not know the result of the coin toss, I cannot know whether your response is based on your experience or by chance."

The ten sensitive questions used in this battery of questions are given in Table 1. Respondents who refused to respond were dropped from the sample. This left 785 respondents who answered all seven sensitive questions.

The Conventional Question in Peru

The variable we use is constructed in the same way that the World Bank constructs the following variable: "Percent of establishments that consider that firms with characteristics similar to theirs are making informal payments or giving gifts to public officials to "get things done" with regard to customs, taxes, licenses, regulations, services, etc." See page 22 of <http://www.enterprisesurveys.org/Data/ExploreEconomies/2010/~media/FPDKM/EnterpriseSurveys/Documents/Misc/Indicator-Descriptions.pdf>.

The interviewer reads the following to the respondent:

"It is said that establishments are sometimes required to make gifts or informal payments to public officials to "get things done" with regard to customs, taxes, licenses, regulations, services etc. On average, what percentage of total annual sales, or estimated total annual value, do establishments like this one pay in informal payments or gifts to public officials for this purpose?"

The respondent is then given the option of responding either as a percentage of annual sales or a total monetary amount in domestic currency units, but not both. The constructed dummy variable then equals one if either the percentage or the monetary amount is greater than zero, or if the respondent refuses to answer. If the response was either 0% or a zero monetary amount then the dummy variable equals zero. Don't knows are treated as missing. This resulted in 881 observations. 707 of these also answered the RRQ battery.

The Random Response Questions in the Gallup World Poll

The questionnaire was administered in a face-to-face interview with a professional surveyor. The interviewer was asked to read the following to the respondent:

I am going to read out a set of questions that describes acts or behaviours that people have expressed. Unlike other questions where you would just respond with a "yes" or "no," this set has a slight variation to it. Before you answer each question, you will toss this coin, and based on which side comes up, I will give you an instruction to provide the appropriate response. Are you ready? I will now read the first question. Please toss the coin, and if the coin comes up heads, just say YES regardless of whether you have done this or not. If the coin comes up tails, please just answer the question. Please do not let me see the coin. This is very important.

The 10 subsequent questions posed using this random response methodology are listed in Table 5. Response rates were exceedingly high, with less than 100% response rates on the RRQ battery only in China, Mongolia, and the Philippines.

The Conventional Question in the Gallup World Poll

The conventional question seeks to obtain information on the respondent's personal experience with corruption, as follows:

Sometimes people have to give a bribe or a present in order to solve their problems. In the last 12 months, were you, personally, faced with this kind of situation, or not (regardless of whether you gave a bribe/present or not)?

Responses included Yes/No/Don't Know/Refused. We coded all "Yes/Refused " responses as 1 and "No " as zeros, and treated Don't Know's as missing. This is consistent with the coding of the CQ for Peru. Response rates were very generally very high, less than 2% in the median country.

Appendix B: Derivations

In this appendix we show how to derive the expressions for the means of the CQ and RRQ, and the variance of the RRQ. These derivations all rely on application of the law of iterated expectations to the random variables defined in Equations (1),(2),(7) and (8), combined with the expressions for the mean and variance of binomial random variables, i.e. if $X \sim B(n, b)$, $E[X] = nb$ and $V[X] = nb(1 - b)$.

Mean of CQ in Section 3 (Equation (3))

$$\begin{aligned}
 E[S] &= E[E[S|z]] = E[zE[S^R] + (1 - z)E[S^C]] \\
 (11) \quad &= E[zg(1 - q) + (1 - z)g] \\
 &= g(r(1 - q) + (1 - r)) = g(1 - rq)
 \end{aligned}$$

The first line is the application of the law of iterated expectations. The second line uses the fact that $S^R \sim B(1, g(1 - q))$ and $S^C \sim B(1, g)$. The last line uses the fact that $z \sim B(1, r)$.

Mean of RRQ in Section 3 (Equation (4))

$$\begin{aligned}
 E[X/n] &= E[E[X/n|z]] = E[zE[X^R/n] + (1 - z)E[X^C/n]] \\
 (12) \quad &= E[z0.5(1 - q)(1 + g) + (1 - z)0.5(1 + g)] \\
 &= 0.5(1 + g)(r(1 - q) + (1 - r)) = 0.5(1 + g)(1 - rq)
 \end{aligned}$$

The first line is the application of the law of iterated expectations. The second line uses the fact that $X^R \sim B(n, 0.5(1 - q)(1 + g))$ and $X^C \sim B(n, 0.5(1 + g))$. The last line uses the fact that $z \sim B(1, r)$.

Variance of RRQ in Section 3 (Equation (6))

$$\begin{aligned}
 V[X/n] &= E[V[X/n|z]] + V[E[X/n|z]] \\
 &= \frac{1}{n^2} E[z^2 V[X^R] + (1 - z)^2 V[X^C]] \\
 &\quad + V[zE[X^R/n] + (1 - z)E[X^C/n]] \\
 &= \frac{1}{n^2} E[z^2 n 0.5(1 - q)(1 + g)(1 - 0.5(1 - q)(1 + g)) \\
 (13) \quad &\quad + (1 - z)^2 n 0.5(1 + g)(1 - 0.5(1 + g))] \\
 &\quad + V[z(0.5(1 - q)(1 + g) - 0.5(1 + g))] \\
 &= \frac{1}{n} \left(r 0.5(1 - q)(1 + g)(1 - 0.5(1 - q)(1 + g)) \right. \\
 &\quad \left. + (1 - r) 0.5(1 + g)(1 - 0.5(1 + g)) \right) \\
 &\quad + r(1 - r)(0.5(1 - q)(1 + g) - 0.5(1 + g))^2
 \end{aligned}$$

The first equality is the law of iterated expectations applied to variances. The second equality uses the definition of X . The third equality follows from using the fact that $X^R \sim B(n, 0.5(1 - q)(1 + g))$ and $X^C \sim B(n, 0.5(1 + g))$. The fourth equality follows from the fact that $z \sim B(1, r)$. Finally, Equation (6) in the main text follows from rearranging the resulting terms in the last line.

The derivation of the corresponding moments in the model of Section 6 follows the same structure. In particular, Equation(9) follows by applying the law of iterated expectations to Equation (7), and relying on the fact that $S^{RP} \sim B(1,0)$, $S^{CP} \sim B(1,0)$, $S^{RU} \sim B(1,1 - q)$ and $S^{CU} \sim B(1,1)$. To derive the mean and variance of the RRQ, it is useful to re-write Equation (8) as:

$$(14) \quad X = wX^P + (1 - w)X^U$$

where $X^P \equiv zX^{RP} + (1 - z)wX^{CP}$ and $X^U \equiv zX^{RU} + (1 - z)X^{CU}$.

Next observe that X^P corresponds to the number of "Yes" responses for a principled individual, i.e. one who is reticent with probability $r - \varepsilon$ and is guilty with probability 0. This means that the mean and variance of X^P are given by Equations (4) and (6) in the main text, but replacing r with $r - \varepsilon$, and setting $g = 0$. Similarly, X^U corresponds to the number of "Yes" responses for an unprincipled individual, and so the mean and variance of X^U are given by Equations (4) and (6) in the main text, but replacing r with $r + \varepsilon$, and setting $g = 1$.

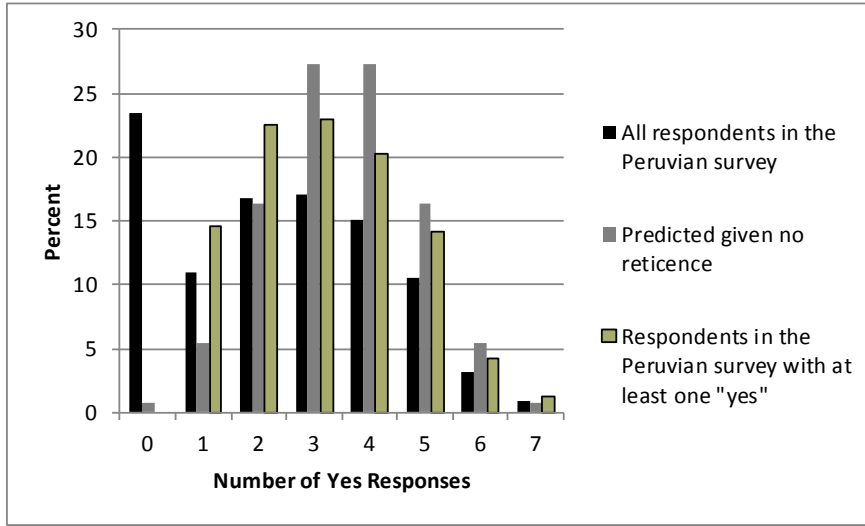
The expression for the mean of the RRQ in (10) then simply follows from applying the law of iterated expectations to Equation (14), while the variance of the RRQ (which is not reproduced in the text) follows from applying the variance form of the law of iterated expectations in Equation (13) to Equation (14), i.e.

$$(15) \quad \begin{aligned} V\left[\frac{X}{n}\right] &= E[V[X/n|w]] + V[E[X/n|w]] \\ &= \frac{1}{n^2} E[w^2 V[X^P] + (1 - w)^2 V[X^U]] \\ &\quad + V\left[zE\left[\frac{X^P}{n}\right] + (1 - z)E\left[\frac{X^U}{n}\right]\right] \\ &= \frac{1}{n^2} (pV[X^P] + (1 - p)V[X^U]) \\ &\quad + p(1 - p)(E[X^P/n] - E[X^U/n])^2 \end{aligned}$$

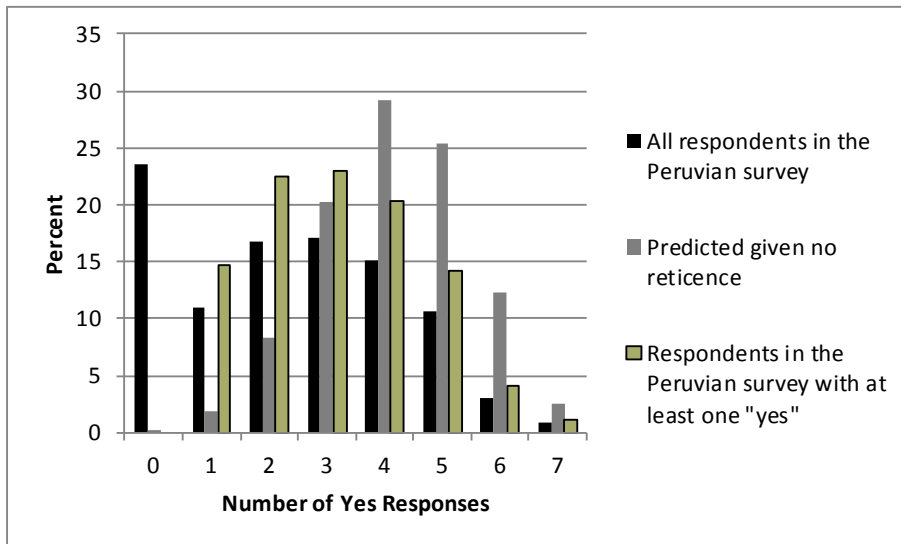
Inserting the means and variances of X^P and X^U described above gives the variance of the average number of "Yes" responses on the RRQ in the model of Section 6.

Figure 1: Actual and Hypothetical Distributions of Responses to the RRQ in the Peruvian Enterprise Survey

Panel A: Assuming No Guilt



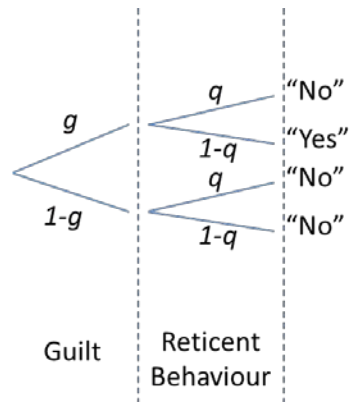
Panel B: Assuming Guilt Rate of 18.25%



Notes: This figure shows the actual distribution of the total number of "Yes" responses across seven sensitive RRQs, the hypothetical distribution of responses that would be observed if respondents were candid and if the probability of guilt were zero (top panel) or 0.1825 (bottom panel), and the actual distribution of the number of "Yes" responses among those respondents who answered "Yes" at least once.

Figure 2: Probability Trees for Reticent Respondents

Panel A: Conventional Question (CQ)



Panel B: Random Response Question (RRQ)

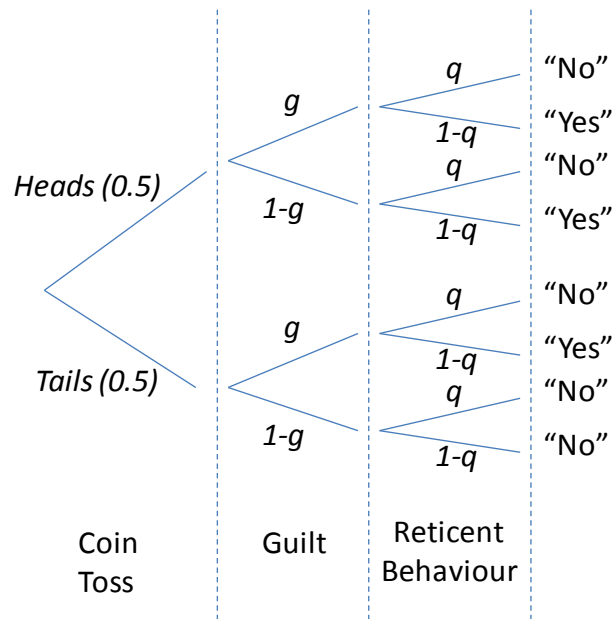
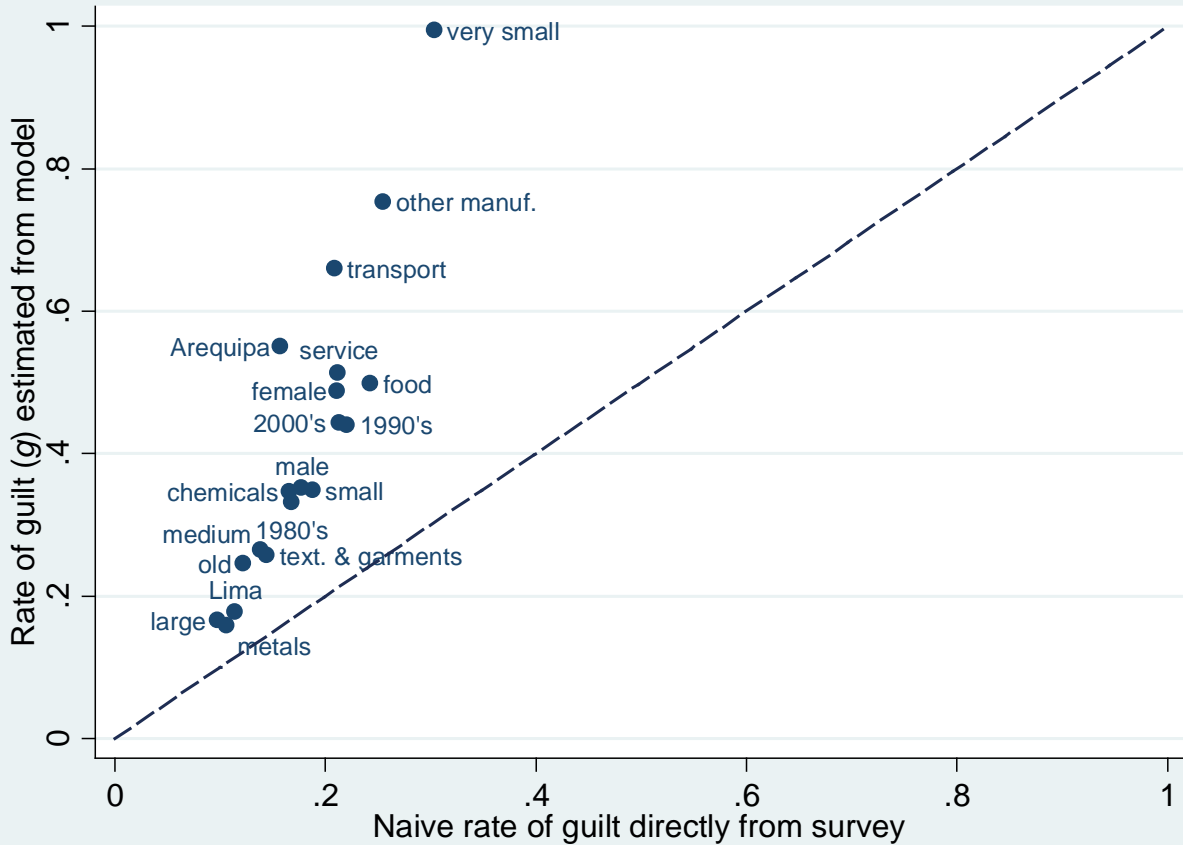
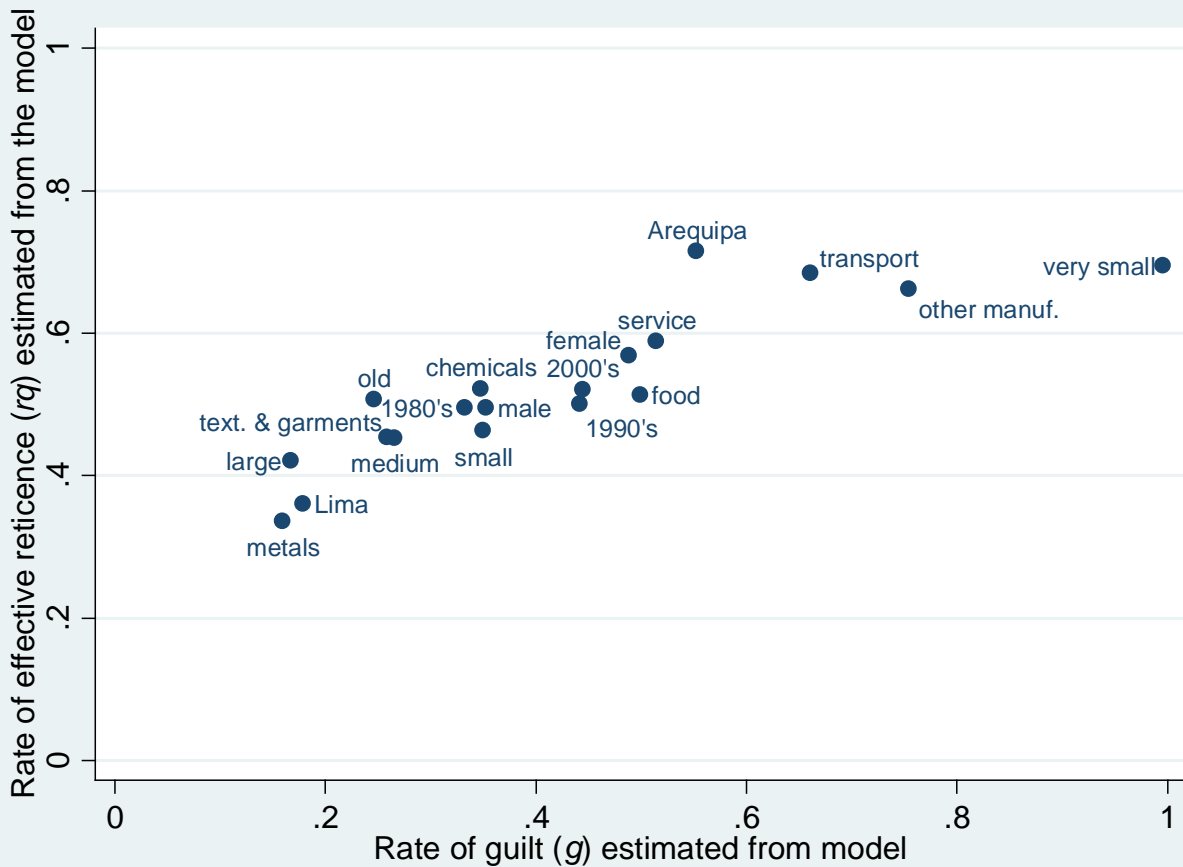


Figure 3: Reticence-Adjusted and Unadjusted (Naïve) Estimates of Guilt for Subsamples of the Peruvian Data



Note: This graph plots the mean response to the CQ (horizontal axis) and the estimated rate of guilt (vertical axis), for the indicated subsets of the Peru Enterprise Survey data. The corresponding data are also reported in Table 3.

Figure 4: Effective Reticence and Reticence-Adjusted Estimates of Guilt for Subsamples of the Peruvian Data



Note: This graph plots the estimated rate of guilt (horizontal axis) against the estimated effective reticence rate (vertical axis), for the indicated subsets of the Peru Enterprise Survey data. The corresponding data are also reported in Table 3.

Figure 5: Reticence-Adjusted and Unadjusted (Naïve) Estimates of Guilt and Effective Reticence for the Gallup World Poll Countries

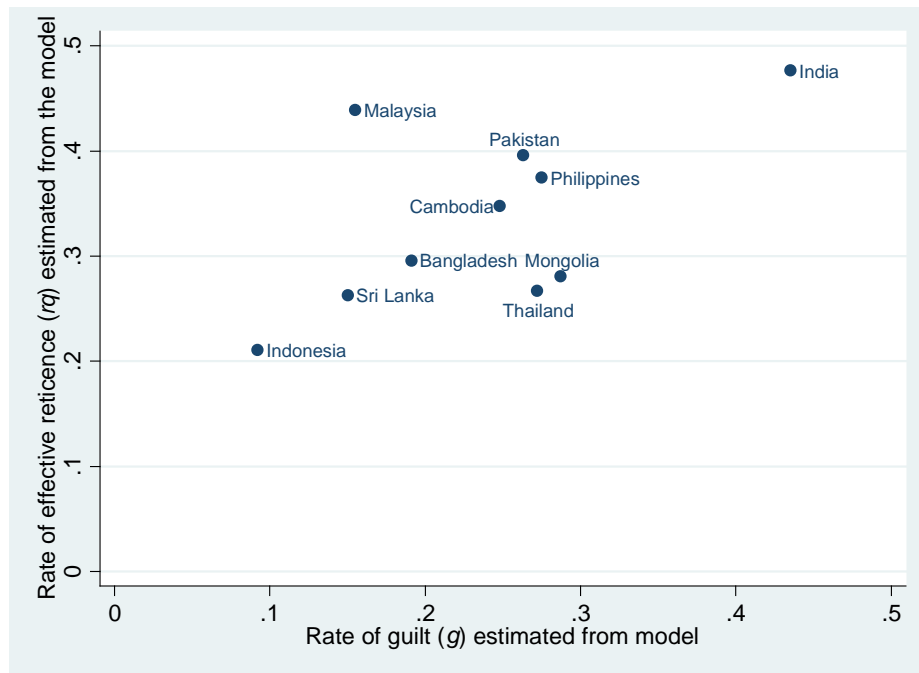
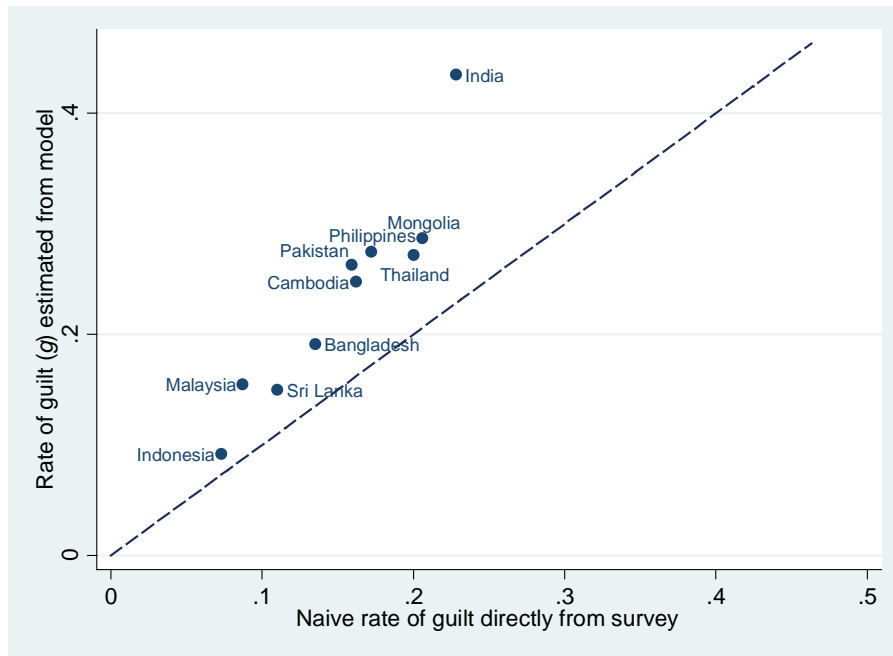


Figure 6: Probability Tree for Reticent and Principled Respondents

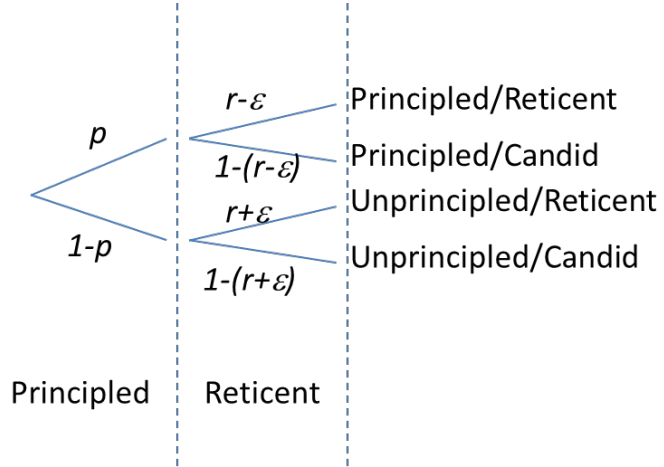


Table 1: Summary Results from Random Response Questions in Peru

	Percentage of Respondents Answering "Yes"
Have you ever paid less in personal taxes than you should have under the law?	37.8
Have you ever paid less in business taxes than you should have under the law?	36.8
Have you ever made a misstatement on a job application?	31.7
Have you ever used the office telephone for personal businesses?	74.8
Have you ever inappropriately promoted an employee for personal reasons?	36.1
Have you ever deliberately not given your suppliers or clients what was due to them?	31.5
Have you ever lied in your self-interest?	50.3
Have you ever inappropriately hired a staff member for personal reasons?	33.8
Have you ever been purposely late for work?	48.4
Have you ever unfairly dismissed an employee for personal reasons?	28.4

Note: Responses from 707 Peruvian firms, April 2010-April 2011.

The seven sensitive questions used in the paper's empirics are in bold.

Table 2: Estimates of Reticence and Guilt from the Peruvian Enterprise Survey

<i>Reticence (r)</i>	0.737*** (14.12)
<i>Probability reticent person answers question reticently (q)</i>	0.689*** (40.62)
<i>Guilt (g)</i>	0.371*** (5.11)
<i>Effective reticence (rq)</i>	0.508*** (13.05)
<i>Number of Observations</i>	707

Notes: z-statistics based on heteroskedasticity-consistent standard errors clustered at the strata level (for a total of 99 clusters) are reported in parentheses

* p<0.05 ** p<0.01 *** p<0.001

Table 3: Estimates of Reticence and Guilt for Subsets of Responding Firms in the Peruvian Enterprise Survey

	<i>"Naïve"</i> estimate of guilt rate	<i>Reticence</i> <i>r</i>	<i>Prob. question</i> answered reticently	<i>Guilt</i> <i>g</i>	<i>Effective</i> reticence <i>rq</i>	Number of observations
Sectors						
food	0.242	0.814***	0.632***	0.499**	0.514***	128
text. & garments	0.144	0.668***	0.681***	0.264*	0.455***	139
chemicals	0.165	0.759***	0.689***	0.347***	0.523***	121
metals	0.106	0.433***	0.778***	0.159*	0.337***	104
service	0.211	0.819***	0.719***	0.514*	0.589***	123
transport	0.208	0.884***	0.775***	0.66	0.685**	24
other manuf.	0.255	0.881***	0.752***	0.754	0.663***	55
Size of firm						
very small	0.303	0.919***	0.757***	0.995**	0.696***	175
small	0.188	0.673***	0.689***	0.349***	0.464***	176
medium	0.144	0.633***	0.717***	0.265***	0.454***	180
large	0.097	0.646***	0.653***	0.167***	0.422***	176
Regions						
Lima	0.114	0.529***	0.683***	0.178***	0.361***	510
Arequipa	0.157	0.831***	0.862***	0.552*	0.716***	83
Age of firm						
old	0.121	0.694***	0.731***	0.246***	0.507***	173
1980's	0.168	0.741***	0.669***	0.332***	0.496***	173
1990's	0.220	0.744***	0.674***	0.441***	0.501***	200
2000's	0.213	0.766***	0.681***	0.444**	0.522***	160
Gender of respondent						
male	0.177	0.721***	0.688***	0.352***	0.496***	593
female	0.211	0.813***	0.700***	0.488**	0.569***	114

Notes: z-statistics based on heteroskedasticity-consistent standard errors clustered at the strata level are reported in parentheses

* p<0.05 ** p<0.01 *** p<0.001

Estimated guilt rates for Chiclayo and Trujillo were above 1.0. Therefore results for these regions are omitted.

Table 4: Using only Four RRQs to Obtain Estimates of Reticence and Guilt from the Peruvian Enterprise Survey

<i>Reticence (r)</i>	0.677*** (9.71)
<i>Probability reticent person answers question reticently (q)</i>	0.680*** (28.99)
<i>Guilt (g)</i>	0.338*** (5.12)
<i>Effective reticence (rq)</i>	0.460*** (10.78)
Number of Observations	707

Notes: z-statistics based on heteroskedasticity-consistent standard errors clustered at the strata level (for a total of 99 clusters) are reported in parentheses

* p<0.05 ** p<0.01 *** p<0.001

Table 5: Summary Results from the Random Response Questions in Gallup World Poll Asian Countries

<i>Proportion of Respondents Answering "Yes" to:</i>	<u>Afghanistan</u>	<u>Bangladesh</u>	<u>Cambodia</u>	<u>China</u>	<u>India</u>	<u>Indonesia</u>	<u>Malaysia</u>	<u>Mongolia</u>	<u>Pakistan</u>	<u>Philippines</u>	<u>Sri Lanka</u>	<u>Thailand</u>
1. Have you ever lied to protect yourself?	63.9	61.2	68.3	59.1	64.3	66.7	68.5	61.2	64.1	63.8	67.7	68.8
2. Have you ever deliberately spoken ill of a member of your family or a friend?	56.1	40.1	51.7	45.5	45.2	51.5	32.3	45.9	47.6	56.8	49.7	51.4
3. Have you ever deliberately tried to cheat another person?	54.9	39.2	38.0	42.9	34.1	43.0	38.6	44.4	34.4	32.4	42.0	41.4
4. Have you ever broken a promise?	31.4	47.9	57.6	51.2	48.7	55.9	56.9	54.1	44.6	61.2	49.4	59.4
5. Have you ever taken something that is not yours without permission and kept it?	32.1	39.8	40.6	37.7	34.6	38.9	30.8	37.9	39.7	31.9	41.6	46.3
6. Have you ever bought, sold, bartered or been given something that you knew was stolen?	31.0	37.9	35.0	37.8	31.4	39.4	25.8	37.8	32.7	31.1	37.3	42.2
7. Have you ever mistreated someone because they did not share your opinions or values?	47.8	54.7	40.1	37.2	42.6	42.2	44.2	62.1	37.2	49.6	45.6	50.9
8. Have you ever been nice to a person only because you thought it would bring you some benefit?	78.6	62.8	66.5	46.0	49.6	48.4	52.2	59.0	41.3	52.0	53.2	52.5
9. If you received some extra money that your family did not know about, would you ever hide it from them and spend it on your own enjoyment?	38.5	42.6	42.5	40.3	39.0	44.0	32.2	48.6	38.4	37.8	41.4	54.5
10. Have you ever insulted your parents, relatives or other elders ?	28.3	40.0	36.8	34.6	35.3	41.0	24.0	49.1	36.8	37.6	39.1	39.2

Note: The seven sensitive questions used in the empirical part of this paper are indicated in bold.

Table 6: Estimates of Reticence and Guilt For Asian Gallup World Poll Countries

	Pakistan	Indonesia	Bangladesh	India	Philippines	Sri Lanka	Thailand	Cambodia	Malaysia	Mongolia
<i>Guilt (g)</i>	0.263 (5.80)	0.0919*** (2.86)	0.191*** (7.24)	0.435*** (8.51)	0.275 (4.51)	0.150*** (5.40)	0.272 (3.80)	0.248 (9.15)	0.155*** (4.18)	0.287 (3.88)
<i>Reticence (r)</i>	0.607 (9.03)	0.413** (3.26)	0.617 (10.93)	0.814*** (29.48)	0.528 (5.08)	0.624 (9.50)	0.666 (3.54)	0.631 (17.77)	0.588* (11.77)	0.507* (5.09)
<i>Probability reticent person answers question reticently (q)</i>	0.653*** (20.34)	0.510 (5.80)	0.479* (9.58)	0.586 (20.89)	0.709*** (19.42)	0.422*** (10.61)	0.400*** (10.18)	0.552 (29.08)	0.746*** (26.32)	0.555 (12.48)
<i>Effective reticence (rq)</i>	0.396 (9.18)	0.211** (2.93)	0.296** (7.23)	0.477*** (14.88)	0.375 (6.04)	0.263*** (8.46)	0.267* (3.84)	0.348* (17.93)	0.439* (13.67)	0.281* (5.13)
Number of Observations	937	1048	987	5760	495	1024	967	983	956	969
Number of Clusters	72	14	49	99	38	31	21	105	121	8
<i>Naive guilt rate estimated directly from survey responses</i>	0.159	0.073	0.135	0.228	0.172	0.110	0.200	0.162	0.087	0.206

Notes: z-statistics based on heteroskedasticity-consistent standard errors clustered at the primary sampling unit level are reported in parentheses.

The corresponding number of clusters is reported in the penultimate row of the table.

All coefficients are significantly different from zero at 99% levels.

* p<0.05 ** p<0.01 *** p<0.001 denote the p-values for the null hypothesis that the country-specific estimate of the indicated parameter is equal to the corresponding point estimate obtained pooling data for all countries. These pooled estimates are $\hat{g} = 0.286$, $\hat{r} = 0.676$, $\hat{q} = 0.563$, and $\hat{r}\hat{q} = 0.381$.

Table 7: The Effects of Varying the Correlation Between Guilt and Reticence

<i>Value of ε</i>	0	0.2	0.4	-0.1
<i>Probability of being principled (p)</i>	0.629*** (8.67)	0.481*** (5.81)	0.341*** (5.26)	0.678*** (10.89)
<i>Reticence (r)</i>	0.824*** (22.96)	0.690*** (45.81)	0.518*** (67.94)	0.879*** (18.00)
<i>Probability reticent person answers question reticently (q)</i>	0.616*** (22.36)	0.729*** (43.84)	0.787*** (54.87)	0.555*** (19.49)
<i>Effective reticence (rq)</i>	0.508*** (13.05)	0.508*** (13.05)	0.508*** (13.05)	0.508*** (13.05)

Note: 707 observations in all four estimates

z-statistics based on heteroskedasticity-consistent standard errors clustered at the strata level (for a total of 99 clusters) are reported in parentheses

* p<0.05 ** p<0.01 *** p<0.001