MISUNDERESTIMATING CORRUPTION

Aart Kraay and Peter Murrell*

Abstract—Corruption estimates rely largely on self-reports of affected individuals and officials. Yet survey respondents are often reticent to tell the truth about sensitive subjects, leading to downward biases in survey-based corruption estimates. This paper develops a method to estimate the prevalence of reticent behavior and reticence-adjusted rates of corruption using survey responses to sensitive questions. A statistical model captures how respondents answer a combination of conventional and random-response questions, allowing identification of the effect of reticence. GMM and maximum likelihood estimates are obtained for ten countries. Adjusting for reticence dramatically alters the perceptions of the extent of corruption.

I. Introduction

In a classic study that compared survey responses to official records, Locander, Sudman, and Musch (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 United States, 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 more to classical music and jazz and less to 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 (2015) 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?

Received for publication August 5, 2013. Revision accepted for publication March 11, 2015. Editor: Mark W. Watson.

* Kraay: The World Bank; Murrell: University of Maryland.

We 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 its 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 of the World Bank is also gratefully acknowledged. The views expressed here are our own and do not reflect those of the Gallup Organization, the World Bank, its executive directors, or the countries they represent.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00536.

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

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) offers 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 reflecting 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 survey 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 reticence of survey respondents and simultaneously to use these estimates to determine the degree to which corruption itself has been underestimated in the past. Our results strongly confirm the previous general recognition in the literature that underestimation of corruption is a problem. Our paper's title borrows a neologism coined by former U.S. president George W. Bush, commenting on how his opponents in the 2000 presidential election had severely underestimated him.[3]

We implement our methodology using data from the 2010 World Bank Enterprise survey in Peru and in nine countries covered in the 2010 wave of the Gallup World Poll, a large cross-country public opinion survey. In the Peru survey, for example, a conventional estimate of corruption reflects the fact that 19% of firms answer that it is

[2] An exception is Olken (2009), who compared Indonesian villagers' perceptions of corruption in local road-building projects with estimates of actual "missing expenditures"—gaps between what villages reported spending on the 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.

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

The Review of Economics and Statistics, July 2016, 98(3): 455–466

doi:10.1162/REST_a_00536

common for similar firms to make informal payments to government officials. For terminological convenience, we will refer to such an answer as indicating guilt on the part of the respondent. This is because a yes response to this question is usually interpreted as an admission by the firm that it makes informal payments, even though the question itself does not specifically ask whether the respondent made such payments.

This conventional estimate assumes that all respondents are always candid when answering all questions. The major questions we address in this paper are whether this assumption is appropriate and what the quantitative implications are of a negative answer. Our estimates suggest that roughly half of respondents across the ten countries we study exhibit reticent behavior. The immediate implication of this is that estimates of guilt based on the standard interpretation of conventional questions are substantially downward biased by the presence of reticent respondents who fail to acknowledge their experiences with corruption. In Peru, we find reticence-adjusted estimates of the prevalence of corruption that are roughly three times larger than conventional estimates. In the Gallup World Poll, we find that on average, reticence-adjusted estimates are 1.7 times higher than conventional estimates. Moreover, looking across countries within the Gallup World Poll, there is a great deal of heterogeneity: reticence-adjusted estimates of the prevalence of corruption are more than two times higher than conventional rates in some countries but only 20% higher in others.

We note at the outset that these findings do not imply that survey-based estimates of corruption are without value. Indeed, the illegality of bribery implies that those involved have strong incentives to hide any evidence of such behavior, so that direct measurement of corruption is in 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, and the use of computers (Tourangeau & Yan, 2007). One important contribution was made by Warner (1965), who developed the random-response question (RRQ). In the form used in our empirical work, the respondent is asked to toss a coin privately before answering a sensitive question and then is instructed to answer yes if the coin came up heads and otherwise answer the sensitive question. We ask a battery of ten such questions, each with its own coin toss. The original motivation for RRQs was the hypothesis that a respondent will be less reticent if the interviewer and the users of the survey data do not know whether a yes response reflects the outcome of the coin toss or the response to the sensitive question. If that hypothesis is correct and respondents are candid, then it is trivial to derive unbiased estimates of the prevalence of the sensitive behavior by subtracting out the proportion of yes responses attributable to respondents obtaining a heads on the coin toss.

Unfortunately, however, evidence suggests that the RRQ methodology does not do much to reduce respondent reticence. 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, RRQs provide opportunities for other methods that do not rely at all on the candor-inducing properties that were the initial goal of the designers of the RRQ. These methods exploit the fact that the randomization probability embodied in an RRQ affects the relationship between reticence and responses. Using this insight, Clark and Desharnais (1998), Moshagen and Musch (2012), and Moshagen et al. (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 goals were 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.

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. We

[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 (nonadherents), which is thought to happen because that procedure places even innocent respondents in a position that looks as if they are admitting to the sensitive act. Moshagen, Musch, and Erdfelder (2012) estimate rates of reticence assuming that there are no nonadherents.

then develop a simple model of the interaction between an interviewer and a survey respondent in which both reticence and guilt shape responses to sensitive questions. We capture reticence by two parameters: the probability that an individual is reticent and the probability that the reticent individual behaves reticently on a particular question. We also capture guilt by two parameters, allowing guilt rates to be different for candid and reticent respondents. This reflects the intuitive idea that respondents might be reticent precisely because they are more likely to be guilty of sensitive acts and therefore have more to be reticent about.

This model leads directly to a precise specification of how different types of respondents answer sensitive questions and therefore to explicit predictions on how survey answers vary with respondent reticence and guilt. We frame these predictions in terms of observable moments in the data: average rates of yes responses to the CQ and the RRQs, the correlations of responses across the CQ and the RRQ, and the correlation of responses within the components of the RRQ battery. Equating these theoretical moments with their sample analogs in a standard method-of-moments estimator, we estimate the four parameters of the model. We then use these estimates to calculate reticence-adjusted rates of the prevalence of corruption that differ significantly from conventional estimates. As a robustness test, we also obtain maximum likelihood estimates of the model that are broadly similar to method-of-moments estimates.

Our paper proceeds in the following way. In section II, we briefly describe the Peru Enterprise Survey and the Gallup World Poll data and document key features of the data that motivate our empirical strategy. In section III, 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 reticence and guilt. Section IV describes our estimation strategy, and section V contains our results. Section VI offers concluding remarks. An online supplemental appendix provides details on the survey questions and samples of respondents used in the analyses.

II. The Context

We implement our methodology using two different data sets, one on businesses in Peru collected by the World Bank Enterprise Surveys (WBES) unit and the other consisting of household survey data from nine Asian countries included in the 2010 wave of the Gallup World Poll (GWP). In this section we first describe the two data sets separately and then document common features of the data that serve to motivate our modeling approach.

A. WBES Data on Peru

Peru is an upper-middle-income country with an economy that has been one of the fastest growing in Latin America in the past fifteen years. The survey polled business owners and top managers in a sample of 1,000 private sector firms (World Bank Enterprise Surveys, 2012).[6] Face-to-face interviews occurred from April 2010 through April 2011. Given the sensitive nature of some of the data collected, the WBES team emphasizes to respondents the efforts made to ensure confidentiality of responses.

We use a CQ that is the basis of a very common measure of corruption: the first item of data readers encounter 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." The online supplemental appendix 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 has the 44th highest reported rate of corruption. In the subsample of firms that we use, also described in the appendix, 19% of firms report that firms like their own give informal payments to government officials.[8] Absent any concerns about respondent reticence, this would be our baseline estimate of corruption in Peru. In the following discussion, we refer to such estimates as the "conventional" ones, emphasizing their common use. However, as we shall see, our estimates of the incidence of reticence imply that conventional estimates seriously underestimate the actual prevalence of corruption.

The questionnaire also presents survey participants with a series of ten sensitive random response questions, which are 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, regardless of whether they have done the sensitive act in question. If the coin comes up tails, they are instructed to answer the sensitive question. The series of ten RRQs includes three asking 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 questions used in the analysis are identified in bold in table 1 but were not so highlighted in the questionnaire itself.

B. GWP Data on Nine Asian Countries

Our GWP data set consists of household survey data from nine Asian countries in the 2010 wave of the GWP.

[6] The Peru Enterprise Survey follows a stratified random sampling approach, with strata based on firm size, geographical location, and economic sector. Full details of the methodology can be found at http://www.enterprisesurveys.org/Methodology. Sampling weights are also provided to generate results that are representative of the population of all manufacturing firms. However, given the small sample size and the oversampling of some industries, the pattern of weights is highly skewed. To prevent a small number of firms with very high weights from dominating the results, we report unweighted results throughout the paper. As a result, our results should be interpreted as representative only of the sample of firms in the data.

[7] http://www.enterprisesurveys.org/Data/ExploreEconomies/2010/peru.

[8] This "headline" prevalence of corruption figure differs slightly from the one reported in those on http://www.enterprisesurveys.org/Data/ExploreEconomies/2010/peru because of differences in the sample and sampling weights used. See the online supplemental appendix for details.

TABLE 1.—SUMMARY RESULTS FROM THE 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?	41.2
Have you ever paid less in business taxes than you should have under the law?	41.9
Have you ever made a misstatement on a job application?	36.6
Have you ever used the office telephone for personal businesses?	72.7
Have you ever inappropriately promoted an employee for personal reasons?	40.8
Have you ever deliberately not given your suppliers or clients what was due to them?	36.4
Have you ever lied in your self-interest?	53.2
Have you ever inappropriately hired a staff member for personal reasons?	40.4
Have you ever been purposely late for work?	54.8
Have you ever unfairly dismissed an employee for personal reasons?	31.9

Responses from 527 Peruvian firms, April 2010–April 2011. The seven sensitive questions used in the paper's empirics are in bold.

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, confidential, face-to-face interviews.[9] 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 nine GWP countries examined in this paper are listed in table 2. They span a wide range of levels of development: PPP GDP per capita in Malaysia, the richest, is nearly ten times that in the poorest, Cambodia. They also span a wide range of levels of corruption in the developing world from Cambodia, which is at the 84th percentile of corruption levels among all countries in the world, to Malaysia at the 32nd percentile, according to a widely used cross-country corruption rating.[10] For purposes of later comparisons, note that Peru has corruption levels in cross-country rankings similar to those of Thailand and Sri Lanka.

Among a wide variety of questions, the GWP asks a number about confidence in public institutions, including one 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. In the subsamples of respondents that we use, the percentage of households that report a personal experience with corruption range from 7% in Indonesia to 21% in Mongolia and India (see table 3). If there were no reticence, these would be our estimates of corruption. However, as we shall see, due to reticent behavior, these may seriously underestimate the actual prevalence of experiences with corruption.

With the generous collaboration of Gallup, we also placed a ten-question set of RRQs on the questionnaires used in Asian countries included in the 2010 wave of the GWP.[11] The RRQs followed the same structure as the RRQs placed in the Peru Enterprise Survey. However, the specific sensitive questions were modified to reflect the fact that the respondents were households rather than business officials. The ten specific RRQs, together with the average number of yes responses on each, are reported in table 2. The seven more sensitive questions that we use are again indicated in bold.

C. *Patterns in the Data*

The usual rationale for deploying RRQs is that they camouflage responses. Because the interviewer does not know whether a yes response is actually an admission of guilt or simply the outcome of a coin toss, RRQs are intended to encourage greater respondent candor. However, the success of RRQs in reducing reticent behavior in other settings has been limited. A glance at tables 1 and 2 suggests the same is true in our application. Absent reticent behavior, the rate of yes responses on each of the RRQs should be at least 50% given that half of the responses would reflect the outcome of obtaining a heads on the coin toss. Yet yes response rates are below 50% on all seven sensitive RRQs in Peru and in 55 of the 63 country-question pairs in GWP countries. Moreover, if the guilt rate on the questions were positive, we should expect even higher rates of yes responses. Using guilt rates for each GWP country equal to the conventional estimates from the CQ listed in table 3, only 3 of the 63 GWP country-question pairs are consistent with no reticence.[12]

Thus, we do not rely at all on the traditional claimed advantage of RRQs: increased candor. Instead, we interpret the pattern of responses to the RRQ as providing information about reticence. Following Azfar and Murrell (2009), one can obtain a simple estimate of the prevalence of reticent respondents from the proportion who answer no to all

[9] The GWP data is from a stratified random sample. Strata and PSUs are defined as geographical regions and subregions, with the precise definitions varying with the size and types of administrative divisions in each country. Within PSUs, households are selected using a random route methodology, with up to three attempts to reach selected households. Within households, an individual respondent is randomly selected using a Kisch grid methodology. In some developed countries, which do not include those studied here, the GWP uses telephone interviews. Although the GWP also reports sampling weights, for consistency with the Peru data we do not use them, and so our results should be interpreted as being representative of the sample of surveyed households only. Full details of the GWP methodology are available at http://www.gallup.com/poll/105226/world-poll-methodology.aspx.

[10] Worldwide Governance Indicators (www.govindicators.org). Data cited in the text refer to 2013.

[11] The set of RRQs was administered in Afghanistan, China, and the Philippines in addition to the nine countries listed in table 2. China is omitted because the CQ was not asked there. Afghanistan and the Philippines are omitted because the model developed in this paper does not fit the survey data from those two countries in the sense that both GMM and maximum likelihood estimators of the parameters of the model (discussed in the next sections) do not converge to interior values.

[12] The proportion of yes responses should be one-half plus one-half the guilt rate.

TABLE 2.—SUMMARY RESULTS FROM THE RANDOM RESPONSE QUESTIONS IN GALLUP WORLD POLL ASIAN COUNTRIES

	Bangladesh	Cambodia	India	Indonesia	Malaysia	Mongolia	Pakistan	Sri Lanka	Thailand
Have you ever lied to protect yourself?	62.3	69.3	66.0	67.5	68.8	62.3	67.3	68.1	69.5
Have you ever deliberately spoken ill of a member of your family or a friend?	40.8	52.9	46.8	55.3	34.6	46.6	51.8	50.2	52.4
Have you ever deliberately tried to cheat another person?	40.7	40.8	35.3	44.9	41.0	44.6	38.3	42.8	41.9
Have you ever broken a promise?	49.5	58.3	50.4	57.3	58.5	54.5	47.3	49.8	59.7
Have you ever taken something that is not yours without permission and kept it?	41.0	42.7	36.2	40.7	32.4	38.2	44.4	42.3	47.6
Have you ever bought, sold, bartered, or been given something that you knew was stolen?	39.4	37.4	32.6	42.6	27.3	38.0	35.7	38.0	42.8
Have you ever mistreated someone because they did not share your opinions or values?	57.3	42.4	44.0	44.8	47.3	63.3	41.3	46.6	52.1
Have you ever been nice to a person only because you thought it would bring you some benefit?	64.9	67.7	51.6	49.8	53.2	59.5	45.8	53.8	53.5
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?	44.4	45.2	40.8	46.1	34.0	49.1	42.7	41.8	55.8
Have you ever insulted your parents, relatives, or other elders?	41.3	38.8	36.4	44.0	24.9	49.6	41.4	39.0	38.6

Responses from surveys administered in the 2010 wave of the GWP. The seven sensitive questions used in the paper's empirics are in bold. Numbers of observations for each country are listed in table 4.

TABLE 3.—SUMMARY STATISTICS ON CQ AND RRQS FROM THE PERUVIAN ENTERPRISE SURVEY AND FOR NINE GWP ASIAN COUNTRIES

	Proportion of Yes Answers on CQ (S)	Number of Yes Answers on the Seven RRQs (X)	Proportion of Respondents Answering No Seven Times on the RRQs (X)	Correlation across Individuals of Responses on the CQ (S) and RRQs (X)
Peru	0.194	2.69	0.144	−0.041
Bangladesh	0.135	3.05	0.054	−0.001
Cambodia	0.166	3.00	0.074	0.050
India	0.207	2.72	0.130	−0.026
Indonesia	0.074	3.18	0.051	0.026
Malaysia	0.086	2.41	0.199	0.127
Mongolia	0.210	3.29	0.063	0.076
Pakistan	0.152	2.96	0.109	−0.020
Sri Lanka	0.113	3.01	0.059	0.005
Thailand	0.204	3.31	0.037	−0.013

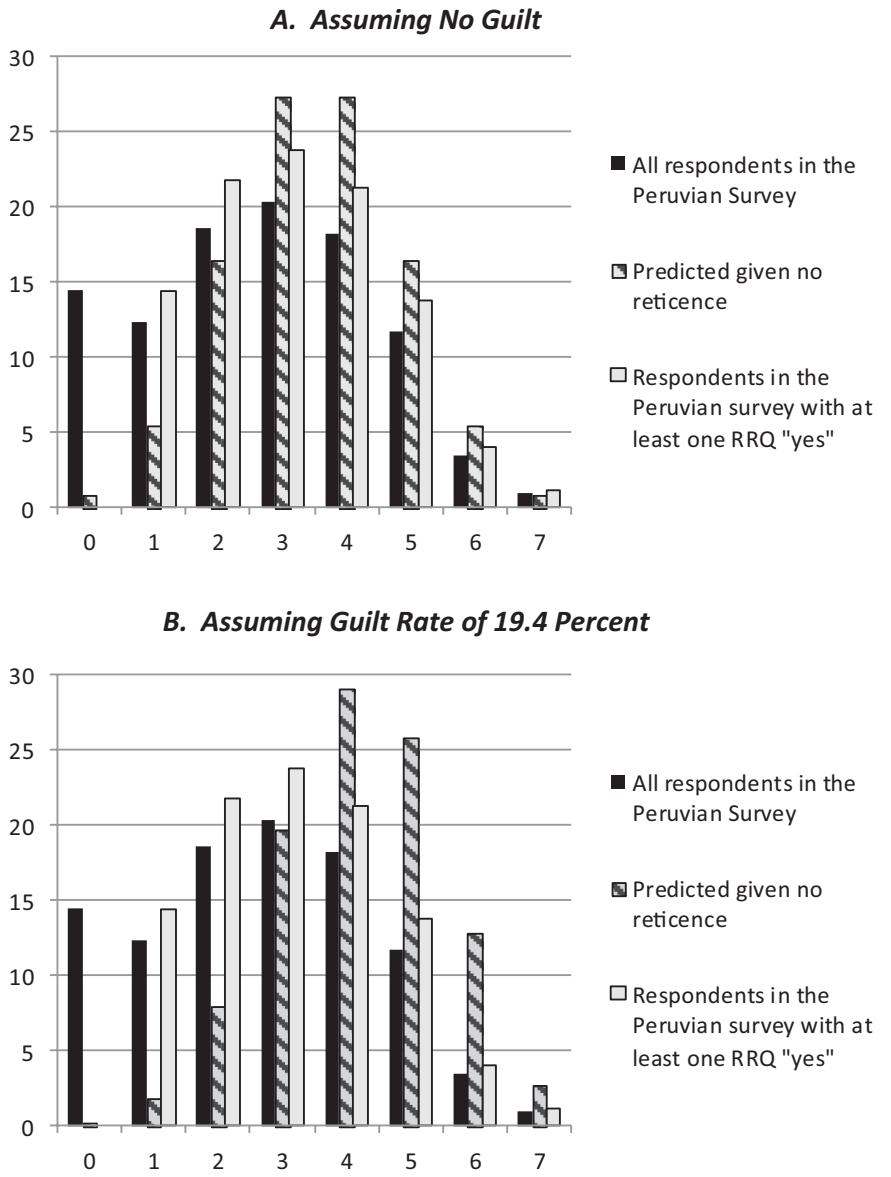
seven RRQs. Intuitively, respondents with seven no answers are highly likely to be reticent since the probability of obtaining no heads on any of seven coin tosses is very low if respondents were correctly following the protocol of the question.[13] As indicated in the summary statistics in table 3, this simple benchmark suggests that 14.4% of respondents in Peru are reticent, with corresponding rates in the GWP ranging from a low of 3.7% in Thailand to a high of 19.9% for Malaysia.

Figure 1 illustrates the Azfar-Murrell (2009) methodology and also clarifies how we improve on their methodology in this paper. In this figure and the following paragraphs, we focus on Peru, but the same distinctive features of the data that we identify are present also in the GWP countries. The top panel of figure 1 shows the distribution of yes responses on the seven sensitive RRQs in Peru. We report two such distributions: the first for all respondents and the second for only those who answered yes to at least one question. In addition, we superimpose 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. In the hypothetical, the number of yes responses should be binomially distributed with a success probability of 0.5—that is, respondents answer yes if and only if the coin comes up heads.

The actual distribution differs from this hypothetical distribution in an obvious way: there is the large mass of 14.4% of respondents with 0 yes responses identified above—those that the Azfar-Murrell (2009) methodology would specifically identify as reticent. While this approach

[13] This probability is 0.008 across seven questions if no respondent were guilty of the sensitive act. Naturally, with positive guilt rates, the probability of observing seven no responses would be even lower.

FIGURE 1.—ACTUAL AND HYPOTHETICAL DISTRIBUTIONS OF RESPONSES TO THE RRQ IN THE PERUVIAN ENTERPRISE SURVEY



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 given 0 reticence, and if the probability of guilt were 0 (top panel) or 0.194 (bottom panel), and the actual distribution of the number of yes responses among those respondents who answered yes at least once.

is intuitive, a drawback is that it does not capture reticent behavior among respondents who answer yes at least once on the RRQ. Indeed, the data do suggest that reticent behavior is common in this part of the sample as well. In the distribution of responses of those who answered yes at least once, there are still too few yes responses relative to the benchmark of no reticence. For example, of those who answer yes at least once, 30.9% answer yes only one or two times, while if there were no reticence, 22% of respondents should do so. Thus, some reticent respondents do not behave reticently on all questions, but rather answer some questions candidly and others reticently.

These points are amplified if we assume further that some respondents have in fact done some of the sensitive acts in question, requiring more yes answers if respondents were candid. This is clearly seen in the bottom panel of figure 1, which uses 19.4% as a hypothetical rate of guilt, corresponding to conventional estimates derived from the CQ. In this case, we would expect just 11% of candid respondents to answer yes only once or twice, in contrast to the 30.9% of those who had answered yes at least once. In sum, this suggests that we must allow the possibility that reticent behavior is imperfectly correlated across questions. The extent to which responses cluster at seven no's—or, more generally, the variance of responses across the RRQ—will be informative about the degree of persistence in reticent behavior across questions.

A further implication from figure 1, obtained from a comparison of the top and bottom panels, is that conclusions about the prevalence of reticent respondents depend

on assumptions about the rate of guilt, and vice versa. Therefore, the rate of reticence and the rate of guilt must be estimated jointly rather than sequentially, as in Azfar and Murrell (2009).

Finally, we note that responses to the CQ and the RRQ are correlated. At first glance, this positive correlation seems natural: reticent respondents are presumably less likely to answer yes to both the CQ and the RRQs, inducing a positive correlation in responses across questions. However, this presumption follows only when reticent and candid respondents have the same rate of guilt. To see why, suppose to the contrary that candid respondents are less likely to be guilty than reticent ones. In this case, candid respondents would have a relatively high rate of yes responses on the RRQ but a relatively low rate of yes responses on the CQ. This would tend to reduce the overall correlation of responses to the CQ and the RRQ, and might even result in the negative correlation that we see in table 3 for some countries. This points to a third way in which we elaborate on the original Azfar-Murrell (2009) methodology: while they assumed that reticent and candid respondents were equally likely to be guilty, we allow for differential rates of guilt for the two types of respondents.

In sum, a model with four parameters is needed to match the patterns we have identified in the data: the probability of reticence, the probability that a reticent respondent is reticent in a particular instance, a guilt rate for the candid, and a different guilt rate for the reticent. The following section formalizes a model incorporating these four parameters. The patterns in the data also suggest that four moments in the data will be particularly informative about these parameters: the rate of yes responses on the CQ, the rate of yes responses on the RRQs, the correlation between yes responses on the CQ and RRQs, and the variance of yes answers on the RRQs. Section IV uses these moments to estimate the four parameters.

III. Modeling the Interview Process

This section provides some structure in describing 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. In particular, the probability that respondents answer yes to a given question depends on (a) whether they are reticent in the sense that they are willing to truthfully answer a sensitive question, (b) whether they choose to behave reticently on a specific question, and (c) whether they have in fact done the sensitive act in question, that is, whether they are "guilty," with guilt potentially different for reticent and candid respondents.

We assume that the probability a respondent is reticent is $0 \leq r \leq 1$, with $1 - r$ as the complementary probability of candor. We consider reticence to be an unobserved trait that is fixed for a given respondent and influences respondent behavior across all questions. Specifically, for reticent respondents, there is a probability $0 < q \leq 1$ that a reticent respondent will behave reticently on a given question—answer no to a sensitive question when he or she is supposed to answer yes. We assume that the event of behaving reticently on a given question is independent across questions. Naturally, $q = 0$ for candid respondents. In short, reticent respondents sometimes behave reticently and candid respondents never do. The parameter q governs the persistence of reticent behavior across questions: when q is large, reticent respondents are likely to behave reticently on most questions.

For reticent respondents, the probability of guilt on a given question is $0 \leq g \leq 1$, while for candid respondents, the probability of guilt is kg, with $0 \leq k \leq 1$. The parameter k governs the correlation between reticence and guilt. As k becomes smaller, the correlation between reticence and guilt increases, and in the limit where $k = 0$, only reticent respondents are guilty. We assume that the event of being guilty on a specific sensitive question is independent across questions for both candid and reticent respondents. All assumptions apply to both the CQ and the RRQs.

These assumptions imply that, conditional on respondent type (reticent or candid), the yes/no responses to the CQ and to all of the individual questions in the RRQ are independently distributed binary random variables. However, the probability of observing a yes response is different for the CQ and the individual questions in the RRQ, and it also differs across reticent and candid respondents.

Consider first the CQ. For reticent respondents, the probability of a yes response is $p_R^{CQ} = g(1 - q)$: reticent respondents are guilty with probability g but admit their guilt only with probability $1 - q$. For candid respondents, the corresponding probability is $p_C^{CQ} = kg$: candid respondents are guilty with probability kg, and if they are, they admit to it with probability 1.

Consider next an RRQ. For a reticent respondent, the probability of a yes response is $p_R^{RRQ} = 0.5(1 + g)(1 - q)$. To see this, note that respondents are supposed to answer yes if they are guilty (with probability g) or if they are innocent and the coin comes up heads (with probability $0.5(1 - g)$). These two probabilities sum to $0.5(1 + g)$ but must be scaled down by $(1-q)$, the probability that a reticent respondent provides an honest yes response. For candid respondents, the probability of a yes response on a given RRQ is $p_C^{RRQ} = 0.5(1 + kg)$. Candid respondents can have a lower guilt probability than reticent respondents ($kg \leq g$) but always answer honestly ($q = 0$).

In the data, we cannot directly observe which respondents are reticent and which are candid. Rather, the data are a mixture of the responses of the two types. Let S be a dummy variable equal to 1 if the respondent answers yes on the CQ and let X_i be a dummy variable equal to 1 if the respondent answers yes on the ith RRQ, for the $i = 1,\ldots, 7$ questions in the RRQ battery. The expected rates of yes responses on the CQ and on an RRQ are weighted averages

of the corresponding yes rates for the two types of respondents:

$$E[S] = rp_R^{CQ} + (1-r)p_C^{CQ}, \tag{1}$$

$$E[X_i] = rp_R^{RRQ} + (1-r)p_C^{RRQ}. \tag{2}$$

Although responses are independent across questions conditional on reticence type, unconditionally the data will exhibit correlation across questions because reticence is a respondent-specific characteristic that affects responses to all questions. In particular, the covariance between the CQ and a given question in the RRQ battery is

$$\text{cov}[S, X_i] = r(1-r)\left(p_R^{CQ} - p_C^{CQ}\right)\left(p_R^{RRQ} - p_C^{RRQ}\right). \tag{3}$$

Similarly, a covariance of responses across questions in the RRQ battery is given by[14]

$$\text{cov}\left[X_i, X_j\right] = r(1-r)\left(p_R^{RRQ} - p_C^{RRQ}\right)^2. \tag{4}$$

The presence of some reticent and some candid responses is necessary to generate comovement in responses across questions in both cases: if $0 < r < 1$, then $r(1-r) > 0$. Comovement also requires reticent and candid respondents to have different rates of yes responses to the same kind of question. For example, if reticent respondents are less likely to answer yes to both types of questions, $p_R^{CQ} < p_C^{CQ}$ and $p_R^{RRQ} < p_C^{RRQ}$, there is a positive correlation in responses to the two types of questions.[15] However, this correlation need not be positive even if reticence is important. Consider, for example, the probability of observing a yes response on the CQ when $k = 1 - q$. In this case, reticent and candid respondents have the same probability of answering yes since the greater candor of the candid is precisely offset by their lower guilt. Then $p_R^{CQ} = p_C^{CQ}$, and the correlation between responses on the CQ and the RRQ is 0. Importantly, $k < 1 - q < 1$ is a necessary condition to obtain the negative correlation that we see in the data for some countries. This highlights the importance of including the possibility of correlated guilt and reticence in the model. If $k < 1 - q$, the reticent respond yes on the CQ more frequently than the candid, but if q is large, the reticent answer yes less frequently on the RRQ than the candid, leading to the negative correlation.[16]

The presence of reticent respondents with a high degree of persistence in their behavior (q large) is also crucial for capturing another key feature of the data highlighted in the previous section. There we noted that that a substantial proportion of respondents answer no to all seven RRQ questions. This implies a strong correlation in responses across individual questions in the RRQ. For the model to generate this nonzero correlation, it is necessary to have both reticent and candid respondents and to have yes response rates on the RRQ differing across the two types of respondents, $p_R^{RRQ} - p_C^{RRQ} < 0$, a sufficient condition for which is a large q.

IV. Estimation

Our goal is to estimate the four key parameters of the model: r, q, g, and k. Given that equations (1) to (4) provide moment conditions that are a function of the model's four parameters, generalized method of moments (GMM) provides a natural estimation method. Equations (1) to (4) imply a large number of moment conditions. For example, equation (2) can be applied separately to each of the seven RRQs, leading to seven moment conditions. Similarly, there are seven covariances between the CQ and each of the RRQs in equation (3) and 21 unique covariances implied by equation (4). However, because the answers to each of the seven questions in the RRQ reflect the same success probabilities, we can collapse the moment conditions into just four that are functions of only the response to the CQ, S, and the average number of yes responses on the RRQ: $X/n \equiv (1/n)\sum_{i=1}^{n} X_i$ for the $n = 7$ questions in the RRQ.

The first moment condition relates the population mean of S to its sample analog. The second equates the population mean of the number of yes responses on the n RRQs, $E[X/n] = E[X_i]$, to its sample analog. The third equates the population and sample covariances between the response to the CQ and the average number of yes responses on the RRQs: $\text{cov}[S, X/n] = \text{cov}[S, X_i]$. The fourth uses the variance of the average number of yes responses on the RRQ, which is

$$V[X/n] = \frac{E[X_i](1 - E[X_i])}{n} + \frac{n-1}{n}\text{cov}\left[X_i, X_j\right]. \tag{5}$$

Substituting equations (2) and (4) into this equation gives this final moment condition in terms of the parameters of the model.[17]

We choose to match these four moments based on our examination of the distinctive patterns in the data that we identify in section II. As a robustness check, we also estimate the model using maximum likelihood (ML). To construct the likelihood function, note that conditional on respondent type, the total number of yes responses on the RRQ is binomially distributed and, moreover, is independent of the response to the CQ, which follows a Bernoulli

[14] The derivation of these equations follows in a straightforward way from the definition of covariance applied to the model described in the previous paragraphs, plus the application of some simplifying algebra.

[15] This is the insight that drove the observations made in Azfar and Murrell (2009) and Clausen et al. (2011).

[16] The crucial role of the coin toss becomes particularly apparent in this instance: with k small and q large, the rate of yes responses on the RRQ compared to the rate on the CQ is much higher for candid than reticent respondents, exactly because few candid are guilty but all candid answer yes when the coin toss is heads.

[17] With nondegenerate data, there is always a unique real-valued solution when solving the four moment conditions for the four parameters, but the solutions can lie outside the permissible ranges of the parameters.

TABLE 4.—GMM ESTIMATES OF RETICENCE AND GUILT FROM THE PERUVIAN ENTERPRISE SURVEY AND FOR NINE GWP ASIAN COUNTRIES

	Peru	Bangladesh	Cambodia	India	Indonesia	Malaysia	Mongolia	Pakistan	Sri Lanka	Thailand
Guilt (g)	0.901**	0.266**	0.308***	0.651***	0.098	0.163**	0.401***	0.569**	0.191**	0.394**
	(2.86)	(3.28)	(4.19)	(5.90)	(0.92)	(2.63)	(6.11)	(2.64)	(3.04)	(2.78)
Reticence (r)	0.556***	0.540***	0.570***	0.648***	0.402**	0.547***	0.449***	0.431***	0.597***	0.560*
	(8.16)	(6.75)	(11.15)	(15.23)	(2.73)	(9.41)	(3.57)	(7.76)	(8.14)	(2.11)
Probability Reticent Person Answers Reticently (q)	0.763***	0.489***	0.542***	0.664***	0.409**	0.725***	0.600***	0.713***	0.425***	0.460***
	(14.64)	(7.14)	(10.77)	(16.03)	(2.99)	(16.05)	(8.04)	(10.09)	(6.99)	(4.02)
Reduction in Guilt for Candid (k)	0.187*	0.508*	0.651**	0.285***	0.870	0.841*	0.624*	0.251*	0.608*	0.489
	(2.13)	(2.28)	(3.06)	(3.86)	(0.76)	(2.17)	(2.44)	(2.09)	(2.27)	(1.67)
Effective Reticence (rq)	0.424***	0.264***	0.309***	0.430***	0.164**	0.397***	0.269***	0.307***	0.254***	0.258***
	(8.80)	(7.40)	(15.05)	(15.29)	(2.75)	(12.80)	(4.84)	(8.94)	(8.40)	(3.63)
Overall Guilt $(r+(1-r)k)g$	0.576***	0.206***	0.261***	0.487***	0.090*	0.151***	0.318***	0.326***	0.161***	0.306***
	(3.34)	(5.48)	(6.94)	(7.45)	(2.25)	(3.83)	(5.32)	(3.87)	(4.49)	(4.53)
N	527	923	907	5,447	971	891	938	838	1,003	946
Conventional estimate of corruption	0.194	0.135	0.166	0.207	0.074	0.086	0.210	0.152	0.113	0.204

The estimates for Peru are not directly comparable to those for other countries given the different types of respondents and survey questions. Z-statistics are based on heteroskedasticity-consistent standard errors, clustered at the strata level in Peru and the strata-PSU level in the GWP, are reported in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

distribution. The likelihood function for a given respondent will then be a mixture of these distributions for reticent and candid respondents:

$$L(S,X;r,q,k,g) = rB(S;1,g(1-q))B\left(X;n,\frac{(1+g)(1-q)}{2}\right)+ (1-r)B(S;1,kg)B\left(X;n,\frac{1+kg}{2}\right),$$

where X is the total number of yes responses on the RRQ for a given respondent and $B(x; n, p)$ is the binomial density function of x with n trials and a success probability p. Multiplying these likelihoods across respondents gives the overall likelihood function for the data, which can then be maximized with respect to r,q,k, and g.

V. Results

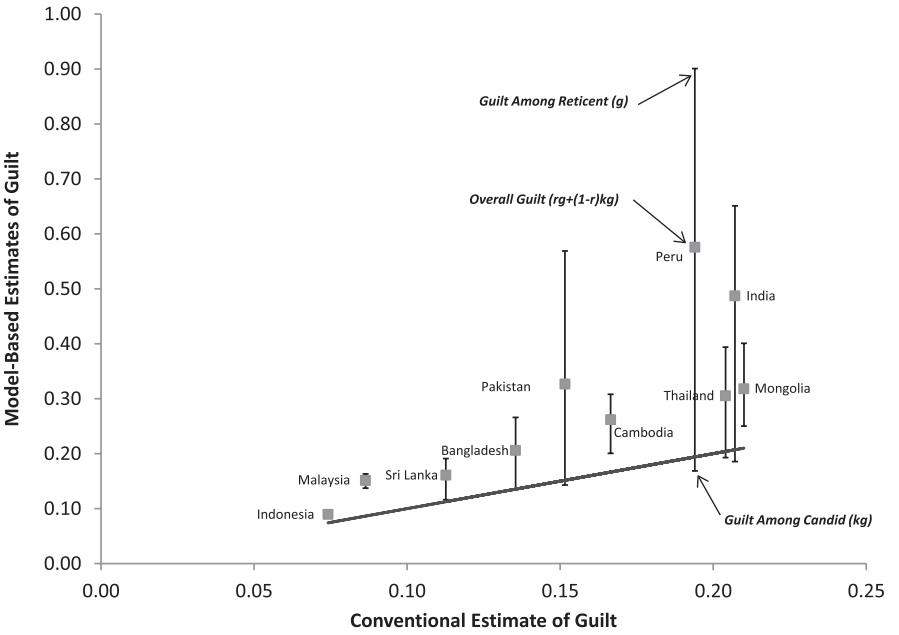
In this section, we present estimates of the parameters of our model for each of the ten countries. We are particularly interested in overall rates of reticence and guilt and how our estimates of guilt compare to conventional estimates based on the CQ alone. Although we present the results for Peru together with those of the nine countries in the GWP, it is crucial to keep in mind that the Peruvian data reflect a different environment from that captured in the GWP—corruption encountered in business operations versus corruption encountered in the daily lives of individuals.

Our core results appear in table 4. The first four rows report GMM estimates of the four key parameters in our model for all ten countries. Reticence is very common: estimated rates of reticence (r) range from 0.4 in Indonesia to 0.65 in India. Reticence rates are fairly similar across countries: for only one country (India) can we reject the null hypothesis that the proportion of reticent respondents is equal to one-half. There is more variation in the persistence of reticent behavior across questions, captured by q. Peruvian, Pakistani, and Malaysian reticent respondents answer reticently over 70% of the time, while Indonesian reticent respondents do so for only 41% of questions. Variation in the guilt rates of the reticent (g) is larger still. The highest observed guilt rate is among reticent Peruvian business officials, at 90%. In contrast, in Indonesia, Malaysia, and Sri Lanka, g is estimated at 10%, 16%, and 19%, respectively. Finally, the parameter k, which captures the differential in guilt rates between candid and reticent respondents, also varies greatly across countries. Peru (0.18) and Pakistan (0.25) exhibit the largest differentials, while k is not statistically significantly different from 1 in six of the nine GWP countries, the exceptions being Bangladesh, India, and Pakistan.

In addition to estimates of g, r, q, and k, we present estimates for two informative composite parameters. Overall rates of dishonesty in answering survey questions are captured by effective reticence (rq), which reflects the proportion of responses to survey questions that are not candid. For the businesses in the Peruvian sample, this proportion is 42%. This proportion averages 29% in the GWP countries, varying from 16% in Indonesia to 43% in India. This estimated effective reticence rate is strongly significantly different from 0 in all countries. The second composite parameter is overall guilt: the weighted average of the guilt rates of reticent and candid respondents, $(r + (1 - r)k)g$. This is the proportion of respondents in the sample who are guilty of the sensitive act in question. This proportion is highest for the businesses of Peru, at 58%, while it averages 26% in the GWP countries, ranging from 9% in Indonesia to 49% in India.

The most important message from this paper's results comes from the comparison between these estimated overall guilt rates and those that are standard in the news media and in the academic literature, the ones reflecting the mean of answers to the CQ. These conventional rates are listed in the last row of table 4. Our estimate of overall guilt is approximately three times the conventional rate in Peru. In the GWP, the mean of the ratio of our estimates to conventional estimates is 1.7. This ratio is more than 2 for India and Pakistan while in Indonesia, it is only 1.21. In Indonesia, the

FIGURE 2.—MODEL-BASED AND CONVENTIONAL ESTIMATES OF CORRUPTION



This graph plots the mean response to the CQ (horizontal axis) and the estimated rate of guilt (vertical axis) for the indicated countries and for the three indicated measures of guilt.

overall estimated guilt rate is only 21% greater than the conventional rate, a difference that is not statistically significant.

Figure 2 offers a visual summary of these results that facilitates interpretation of our estimates. On the horizontal axis, we graph the conventional estimates of the prevalence of corruption, based on simple averages of responses to the CQ for each country. On the vertical axis, we report three model-based measures of the prevalence of corruption. The upper and lower ends of the vertical bars for each country report the estimated guilt rates for the reticent and the candid respondents, g and kg, respectively. The square data point in between these two indicates the overall guilt rate for each country. The upward-sloping line traces out the points where model estimates equal conventional estimates of guilt.

The large differences between model-based estimates of overall guilt and conventional estimates are readily apparent in the large distances between the square data points and the 45 degree line. Some simple algebra shows that this distance is equal to grq. This has a natural interpretation: biases in conventional estimates of guilt reflect reticent behavior, rq, and how much this matters depends on the guilt rate of reticent respondents, g. As noted above, this overall bias reflects the differing strength across countries of the various factors highlighted in our model. While the estimated rate of reticence, r, is not that different across countries, there are substantial differences in estimates of q and g and also in the gap between the guilt rates of the reticent and the candid. These are readily apparent in the vertical ranges for each country. Thus, our results suggest that the large downward biases in conventional estimates of corruption reflect different processes in different countries.

A further interesting question is the extent to which the biases differ across countries. Would reticence-adjusted rates of corruption order countries differently from conventional estimates? A quick look at figure 2 shows that there are two cases where country ranks switch as a result of adjustments for reticence. Whereas conventional estimates place Pakistan as less corrupt than Cambodia, Thailand, and Mongolia—considerably less in the latter two cases—our estimates show corruption to be higher in Pakistan than these three countries, considerably more in the case of Cambodia. The magnitudes are large enough to lead to a significant change in the assessment of where Pakistan ranks on corruption. Peru's business officials provide the other case of reversals. In the conventional estimates, Peruvian respondents report marginally less corruption than those in India, Mongolia, and Thailand. However, there is a very large change in perceptions of corruption induced by our procedures, with Peruvian respondents now estimated to experience significantly more corruption interactions than respondents in all three of these countries. Given the differences between the Peruvian and GWP surveys, the interpretation of this second case of reversal must remain inconclusive; it could reflect characteristics of Peru, and it could reflect differences between the characteristics of surveys of businesses and of individuals.

Finally, table 5 reports the ML estimates of the parameters of the model. Comparing tables 4 and 5 reveals some systematic differences. GMM estimates of g and q tend to be lower than the ML estimates, while estimates of r and k tend to be higher for ML than for GMM. However, these differences tend to offset each other when examining the composite parameters of effective reticence and overall guilt, which are of primary interest. This is especially the case for overall guilt, where the tendency for GMM to estimate a higher rate of effective reticence is offset by the ten-

TABLE 5.—ML ESTIMATES OF RETICENCE AND GUILT FROM THE PERUVIAN ENTERPRISE SURVEY AND FOR NINE GWP ASIAN COUNTRIES

	Peru	Bangladesh	Cambodia	India	Indonesia	Malaysia	Mongolia	Pakistan	Sri Lanka	Thailand
Guilt (g)	1.000	0.345**	0.450***	1.000	0.173	0.230*	0.505***	1.000	0.245**	0.460**
	(.)	(3.03)	(3.33)	(.)	(0.77)	(2.34)	(3.36)	(.)	(2.91)	(3.07)
Reticence (r)	0.466***	0.431***	0.427***	0.475***	0.211*	0.454***	0.382***	0.326***	0.481***	0.492*
	(7.89)	(6.02)	(8.63)	(9.58)	(1.97)	(10.56)	(8.27)	(7.56)	(4.63)	(2.09)
Probability Reticent Person Answers Reticently (q)	0.804***	0.571***	0.650***	0.772***	0.632***	0.794***	0.659***	0.834***	0.495***	0.504***
	(35.14)	(7.60)	(10.69)	(27.43)	(7.45)	(18.79)	(9.93)	(32.38)	(6.15)	(4.47)
Reduction in Guilt for Candid (k)	0.140***	0.327*	0.328**	0.127***	0.376	0.407*	0.431***	0.120***	0.372*	0.373*
	(6.31)	(2.52)	(2.79)	(11.21)	(0.75)	(2.04)	(3.54)	(8.69)	(2.34)	(2.31)
Effective Reticence (rq)	0.375***	0.246***	0.278***	0.367***	0.133*	0.360***	0.252***	0.272***	0.238***	0.248***
	(9.25)	(7.13)	(12.93)	(11.98)	(2.09)	(12.87)	(6.28)	(8.24)	(7.19)	(3.39)
Overall Guilt $(r+(1-r)k)g$	0.541***	0.213***	0.277***	0.542***	0.0879	0.156***	0.328***	0.407***	0.165***	0.313***
	(9.02)	(4.91)	(5.43)	(11.91)	(1.54)	(3.37)	(4.51)	(9.96)	(4.22)	(4.30)
N	527	923	907	5,447	971	891	938	838	1,003	946
Conventional estimate of corruption	0.194	0.135	0.166	0.207	0.074	0.086	0.210	0.152	0.113	0.204

The estimates for Peru are not directly comparable to those for other countries given the different types of respondents and survey questions. Z-statistics based on heteroskedasticity-consistent standard errors, clustered at the strata level in Peru and the strata-PSU level in the GWP, are reported in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

dency of GMM to estimate a lower correlation between guilt and reticence (higher k). This means that there is little to choose between the two sets of estimates when addressing the major question of our paper, which is whether acknowledging the possibility of reticence on survey questions alters the perceptions of the extent of corruption.

VI. Conclusion

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 conventional solutions to address respondent reticence, such as random response questions, have had mixed success at best.

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 four 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, with guilt rates possibly differing between reticent and candid respondents. We show how the population frequency of these characteristics can be estimated from observable data on responses to conventional and random response questions.

We implement this methodology using the World Bank's Enterprise Survey for Peru and in a sample of nine Asian economies covered by the Gallup World Poll. In all countries, we find that reticent behavior is common: on average, roughly half of respondents in our combined sample are classified as reticent. This has important implications for the interpretation of data summarizing responses on conventional questions about corruption. Specifically, we find substantial downward biases in conventional estimates of corruption: our reticence-adjusted estimates of the prevalence of corruption in the Gallup World Poll data are on average 1.7 times higher than conventional estimates, and in Peru they are higher by a factor of 3. There are substantial differences in these biases across countries, reflecting cross-country differences in the extent to which reticent behavior is persistent across countries and the extent to which reticence tends to be particularly concentrated on the set of respondents who have in fact experienced corruption. While we do not yet have a good accounting for the reasons underlying these different mechanisms, we speculate that specific institutional and cultural features of a country will lead to different types of bias in different countries, certainly a subject that is important in future research that aims at discovering the underlying causes of cross-country variation in measured corruption.

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.[18] This

[18] 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."

research agenda is particularly important in the case of corruption, where alternatives to self-reported survey-based data are rare.

REFERENCES

Azfar, Omar, and Peter Murrell, "Identifying Reticent Respondents: Assessing the Quality of Survey Data on Corruption and Values," *Economic Development and Cultural Change* 57 (2009), 387–412.

Brenner, Philip, "Exceptional Behaviour or Exceptional Identity? Overreporting of Church Attendance in the US," *Public Opinion Quarterly* 75 (2011), 19–41.

Campbell, A., "Randomized Response Technique," *Science* 236 (1987), 1049.

Clark, S. J., and R. A. Desharnais, "Honest Answers to Embarrassing Questions: Detecting Cheating in the Randomized Response Model," *Psychological Methods* 3 (1998), 160–168.

Clausen, Bianca, Aart Kraay, and Peter Murrell, "Does Respondent Reticence Affect the Results of Corruption Surveys? Evidence from the World Bank Enterprise Survey for Nigeria" (pp. 428–450), in Susan Rose-Ackerman and Tina Søreide, eds., *International Handbook on the Economics of Corruption,* Vol. 2 (Northampton, MA: Edward Elgar Publishing, 2011).

Gong, Erick, "HIV Testing and Risky Sexual Behaviour," *Economic Journal* 125 (2015), 32–60.

Lensvelt-Mulders, G. J. L. M., J. J. Hox, P. G. M. van der Heijden, and C. J. M. Maas, "Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation," *Sociological Methods and Research* 33 (2005), 319–348.

Locander, W., S. Sudman, and N. Bradburn, "An Investigation of Interview Method, Threat and Response Distortion," *Journal of the American Statistical Association* 71 (1976), 269–275.

Moshagen, M., and J. Musch, "Surveying Multiple Sensitive Attributes Using an Extension of the Randomized-Response Technique," *International Journal of Public Opinion Research* 24 (2012), 508–523.

Moshagen, Morten, Jochen Musch, and Edgar Erdfelder, "A Stochastic Lie Detector," *Behavior Research Methods* 44 (2012), 222.

Olken, Benjamin, "Corruption Perceptions vs. Corruption Reality," *Journal of Public Economics* 93 (2009), 950–964.

Svensson, Jakob, "Who Must Pay Bribes and How Much? Evidence from a Cross Section of Firms," *Quarterly Journal of Economics* 118 (2003), 207–230.

Tourangeau, R., and T. Yan, "Sensitive Questions in Surveys." *Psychological Bulletin* 133 (2007), 859–883.

Warner, S., "Randomized-Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association* 60 (1965), 63–69.

World Bank, *World Bank Enterprise Surveys* (Washington, DC: World Bank, 2012), http://www.enterprisesurveys.org/.

——— *Worldwide Governance Indicators* (Washington, DC: World Bank, 2014), www.govindicators.org.