Doing the Survey Two-Step: The Effects of Reticence on Estimates of Corruption in Two-Stage Survey Questions

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

August 1, 2016

Abstract: This paper develops a structural appproach for modeling how respondents answer survey questions and uses it to estimate the proportion of respondents who are reticent in answering corruption questions, as well as the extent to which reticent behaviour biases conventional estimates of corruption downwards. The context is a common two-step question, first inquiring whether a government official visited a business, and then asking about bribery if a visit was acknowledged. Reticence is a concern for both steps, since denying a visit side-steps the bribe question. This paper considers two alternative models of how reticence affects responses to two-step questions, with differing assumptions on how reticence affects the first question about visits. Maximum-likelihood estimates are obtained for seven countries using data on interactions with tax officials. Different models work best in different countries, but cross-country comparisons are still valid because both models use the same structural parameters. On average 40% of corruption questions are answered reticently, with much variation across countries. A statistic reflecting how much standard measures underestimate the proportion of all respondents who had a bribe interaction is developed. The downward bias in standard measures is highly statistically significant in all countries, varying from 12% in Nigeria to 90% in Turkey. The source of bias varies widely across countries, between denying a visit and denying a bribe after admitting a visit.

Keywords: corruption measures, structural estimation, reticence, World Bank enterprise surveys

Economics studies facts, and seeks to arrange the facts in such ways as make it possible to draw conclusions from them. As always, it is the arrangement which is the delicate operation. Facts, arranged in the right way, speak for themselves; unarranged they are as dead as mutton.

John Hicks, *The Social Framework*, 1950

1. Introduction

Cross-country estimates of the extent of corruption rely largely on the self-reports of households, business managers, and government officials. But a long-recognized problem in survey research is that individuals are reticent to tell the truth about sensitive topics, which extend from illegal activities to behaviours in which a person is simply morally invested. An old example was exaggeration in reports of possession of a library card (Locander et al. 1976), but recent examples abound. Pregnant women, especially those in higher social classes, under-report smoking, even when access to free smoking cessation services depends on such reports (Shipton et al. 2009). A significant number of girls, and even more boys, profess knowledge of the mathematical concepts of declarative fractions and subjunctive scaling, which are fictions (OECD 2015). Either conservatives exaggerate how happy they are or liberals downplay how happy they are, or both (Wojcik et al. 2015).

In attempting to counter the problem of reticence on sensitive topics, the traditional approach in survey research has been to find ways to make respondents more comfortable when answering questions, in order to elicit more candour. For example, researchers have experimented with self-administered questions, telephone interviewing, and variations in the wording of questions (Tourangeau and Yan 2007). One important approach in trying to increase respondent comfort level has been the use of unconventional questions that elicit answers that do not reveal precise facts about individuals but yield relevant sample statistics. Such techniques range from the venerable random-response question of Warner (1965) to the new item sum technique of Trappmann et. al. (2014). These different approaches have made varying contributions to alleviating the problem of reticence, but as Tourangeau and Yan (2007, p. 878) state "The need for methods of data collection that elicit accurate information is more urgent than ever."

This urgency appears no more so than in the area on which this paper focuses, measures of corruption that are comparable across countries. These measures perform a vital function. Policy makers and politicians use corruption indicators to monitor governance quality, with consequences for implementation of reforms and for the provision of aid by such entities as the Millennium Challenge Corporation, USAID, and the World Bank. Cross-country measures of corruption are often prominent in political debate, as for example in attitudes towards Greece within Europe.[1]

Innovative attempts have been made to obtain evidence on corruption that do not rely on self-reports (Reinikka and Svensson 2004; Olken 2009). However, the effort required to gather such evidence is large and there is great reliance on peculiar country-specific institutional features that create

---

[1] "Corruption still alive and well in post-bailout Greece" *The Guardian* December 3, 2014
http://www.theguardian.com/world/2014/dec/03/greece-corruption-alive-and-well

specific opportunities for measurement. Hence, self-reports from surveys will continue to provide the basis for most research on and assessment of corruption in the future. This is especially the case when the focus is on comparisons across countries.

Existing survey evidence indicates that acts of corruption are amazingly common across the globe. Rose and Peiffer (2015) estimate that over a quarter of the world's population regularly pays bribes for public services. Such acts are by definition illegal. Moreover, for obvious reasons, corruption is much more common in countries where the reach of the rule of law is tenuous. But these are exactly the countries where respondents have the most reason to fear that government officials can force survey firms to violate confidentiality agreements. They are the countries where the evidence necessary to support criminal prosecution—or simply political or administrative persecution—is the weakest. They are the places where legal defense against such prosecution or persecution is least likely to be effective. Hence, in the highest corruption environments, there is very good reason to believe that the traditional approach of survey research—increasing respondent comfort—will not eliminate respondent reticence on corruption questions.

Therefore, in aiming to develop valid, self-report-based, cross-country-comparable measures of corruption, we adopt an approach diametrically opposite to the traditional one. We accept that there will always be large numbers of respondents who have a propensity to give false answers to questions on sensitive issues, and we embrace that fact. Instead, we rely on a methodology that allows us to estimate the rate of false answers and therefore the rate of commission of the corrupt act.

The intuition behind our approach is as follows. We formulate a structural model of the behaviour of reticent respondents in answering sensitive survey questions. We then apply the model to data that simultaneously reflects the answers to two different types of questions. First, there is a conventional question (CQ) that asks explicitly whether a particular act of corruption has taken place. Naturally, if respondents are reticent, the distribution of yes/no responses on this question will depend on both the level of corruption and the level of reticence. Using this question alone one cannot disentangle the two. We therefore employ the CQ in tandem with a second type of question, a forced-response random-response question (RRQ; Warner 1965, Boruch, 1971). Importantly, we assume that responses to both the RRQ and the CQ are affected by the same degree of respondent reticence.[2] Then, the distribution of responses on the RRQ will depend on the same corruption and reticence parameters as the distribution for the CQ, but in a different way. Hence, by applying our model to the combination of the responses to the two types of questions, we obtain estimates of structural parameters that separately reflect levels of corruption and degree of respondent reticence. Our analytical approach is fully described in Sections 3 and 4, where we set up the model of respondent behaviour, apply the model to the two types of questions, and show how the key parameters of the model can be estimated.

---

[2] In Appendix B, we argue that this is a reasonable assumption and also summarize some simulation results that show that our estimates of the degree of downward bias in standard measures of corruption are lower than those that would be obtained should the RRQ work as designed.

The analysis follows and expands on that of Kraay and Murrell (2016). Their initial implementation of the above methodology focused on a CQ that asks all respondents (for example, company officials acting on behalf of firms) whether a bribe is required when government officials provide some sort of service.[3] This is a very general question that does not generate information on any specific government agency, and as such is not a good guide to specific anticorruption measures aimed at particular government agencies. To be more useful to policymakers, the CQ must be more specific, asking, for example, about levels of corruption in a named government agency. But then a problem of analysis immediately arises. Whereas most firms are likely to have had interactions with at least one government official and therefore can answer a nonspecific question, not every firm will interact with any given agency's officials. This necessitates a subtle change in the nature of the survey question asked, which in turn has significant implications for modeling how respondents answer survey questions.

Consider the subject matter of this paper. The World Bank Enterprise Surveys project (World Bank, 2015; henceforth WBES), from which our data come, asks a two-step question about each of a series of government officials. First, it asks whether the firm has been visited by a specific type of official in the past 12 months. Then, the survey goes on to inquire about whether a gift or payment was required only if the respondent acknowledged that a visit occurred. For such two-step questions, modeling respondent behaviour is more complex because respondents might be reticent in answering the question about a visit in order to forestall being asked the follow-up question about the bribe. Additionally, this reticence in acknowledging the visit might depend on whether a request for a bribe had occurred. This adds more complexity to the modeling process, because evaluating the answers to one question (the bribe) cannot be done independently of evaluating the answers to another (the visit). In this paper, we extend the Kraay-Murrell (2016) analytics to cover the case of two-step questions. We use the example of a two-step question asking about interactions with tax officials, but as we note below analogous two-step questions are very common across a whole range of well-known surveys, and so our methodology is broadly applicable in these other settings as well.

In Section 2, we provide the background to the current exercise, describing the data sets and survey questions used. We use data from seven countries included in the World Bank Enterprise Surveys project— Bangladesh, India, Nigeria, Peru, Sri Lanka, Turkey, and the Ukraine -- where we have fielded the random response questions essential to the implementation of our methodology. Thus one contribution of this paper is to compare rates of reticence across countries, then examining the degree to which the underestimation of corruption varies across nations. This has important implications for the production of valid cross-national comparisons of corruption.

Section 3 details how we model a respondent formulating a response to the two-step question. We consider two alternative models with different assumptions on how reticence affects the respondent's approach to the first part of the question on whether a visit occurred. Our first model

---

[3] Kraay and Murrell (2016) used two different data sets in their implementation. For brevity, we focus on their implementation that uses data from the World Bank's Enterprise Surveys project (World Bank, 2015), since this project also provides the data for the set of countries on which this paper focuses.

assumes respondents are reticent only on direct questions on corruption and are candid on questions about the visit of a government official. Our second model assumes that a question about a visit is innocuous to all respondents who have not subsequently experienced a bribe request, but a reticent respondent who has experienced a bribe request will be reticent about reporting visits. We show that estimates of corruption and reticence depend upon which model of respondent behaviour is used. Section 3 also describes how we model the responses to RRQs, showing that such responses are a function of the same parameters as responses to the two-part CQ.

Section 4 describes the maximum likelihood estimation procedure. Section 5 presents estimates of the two different models for each of the countries, comparing these to standard estimates of corruption that assume all answers were candid. Section 6 presents summary cross-country results based on estimates for the preferred model for each country. We estimate that the rate at which questions on bribes are answered reticently runs from a low of 27% in Bangladesh to a high of 64% in India. Not surprisingly these rates of reticence cause corruption to be significantly underestimated in all countries.

We introduce the concept of effective corruption, a derived parameter. Effective corruption is the unconditional probability of a randomly-selected firm being directly involved in a corrupt interaction with a government official from a specific agency. It is equal to the probability of a visit by an official times the probability of a bribe being solicited on a given visit. The latter probability—that is, the conditional and not the unconditional one—is the one reported in standard sources in the literature, for example by the WBES. Effective corruption adds insights by taking into account the frequency of visits by officials, which is especially important because our estimates show that respondents can be reticent in answering questions about whether visits have occurred. This implies that simple reported rates of visits by government officials cannot be taken at face value, but must be corrected for reticence in a manner similar to that used for reported rates of bribe frequency.

We obtain maximum likelihood estimates of the degree of downward bias of effective corruption in standard (non-reticence-adjusted) measures of corruption. These estimates range from a low of 12% in Nigeria to a high of 90% in Turkey. That is, only 10% of corruption by tax officials is reported in standard estimates for Turkey! Thus our methods produce startling changes in perceptions of relative corruption: Peru is perceived as more than twice as corrupt as Turkey in standard estimates, while Turkey appears more than twice as corrupt as Peru in our estimates. Estimates of the propensity of officials to make visits to firms are also affected. In India, for example, tax officials visit 50% more firms than standard estimates suggest.

Since effective corruption is the probability of a bribe being solicited by a visiting official times the probability of a visit, we can use our methods to decompose any bias in the perception of effective corruption between bias on estimates of bribes and bias on estimates of visits.In India, our estimates of corruption are 157% higher than standard estimates. This is due to the combination of a rise in the perception of visits by 49% and a rise in the perception of corruption on a visit of 72% (2.57 approximately equals 1.49 times 1.72). In other countries, Turkey, for example, there is no change in

4

the perception of visits and all the rise in effective corruption is due to an increase in the perception of levels of corruption on an official's visit.

2. The Data

We implement our methodology using data from surveys conducted by the World Bank Enterprise Survey unit (WBES, 2015) on Bangladesh, India, Nigeria, Peru, Sri Lanka, Turkey, and the Ukraine. Over the past several years, with the generous cooperation of the WBES team, we have placed forced-response random-response questions in these surveys. Each survey polled business owners or top managers in a sample of officially registered firms that is representative of the economy's formal private manufacturing and services sectors.[4] Interviews were conducted face-to-face and covered a wide range of topics, including corruption. The data from the seven surveys were collected in different waves of the WBES between 2007 and 2014. Information on the timing of interviews, the type of interview, and the number of observations included in our analysis for each country is given in Table 1. Full details of the subsample of firms used in the analysis are given in Appendix A.

We use responses to two types of question. The first is a two-step conventional question (CQ) regarding whether an interaction with a government official occurred and if it did whether a bribe interaction took place. Two-step questions with this structure are very common in survey work, appearing for example in the surveys conducted by Transparency International, the World Justice Project, the US Federal Reserve, the US Agency for International Development, and the World Health Organization, as well as the Gallup World Poll and the National Crime Victimization Survey.[5] In the World Bank's Enterprise Surveys, there are a number of such questions with identical structure, varying only in the type of official who is the subject of enquiry. The example used in this paper concerns interactions with tax officials, a question that is in the core questionnaire for the World Bank's Enterprise Surveys and has been asked in a large number of countries. This two-step CQ asks whether firms had an interaction with tax officials over the last year and if so, whether the firms were expected or requested to give gifts to the officials. Responses to this question are the basis of one of WBES's prominent corruption indicators—"Percent of firms expected to give gifts in meetings with tax

---

[4] Full details of the methodology can be found at http://www.enterprisesurveys.org/methodology. Stratified random sampling was used, with strata based on firm size, geographical location, and economic sector. Given the small sample size and the oversampling of some industries, the pattern of sampling weights is highly skewed. To prevent a small number of firms with very high weights from dominating our 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.

[5] See, for example, Transparency International's survey for the Global Corruption Barometer http://files.transparency.org/content/download/604/2549/file/2013_GlobalCorruptionBarometer_EN.pdf, the questions on gambling in the Fed's Survey of Household Economics and Decisionmaking http://www.federalreserve.gov/econresdata/2014-economic-well-being-of-us-households-in-2013-appendix-2.htm, the questions on health behaviour in the World Health Survey of the WHO http://www.who.int/healthinfo/survey/en/, the World Justice Project's questions on corruption http://worldjusticeproject.org/sites/default/files/gpp_2013_final.pdf, the questions on unwanted sexual acts in the National Crime Victimization Survey http://www.bjs.gov/index.cfm?ty=dcdetail&iid=245, and the questions on sexual behaviour in the Demographic and Health Surveys of USAID http://dhsprogram.com/What-We-Do/Survey-Types/DHS-Questionnaires.cfm.

officials".[6] Appendix A contains the precise wording of all survey questions used in this paper and further details on the samples used in the analysis, including information for each country on how many observations were dropped because questions were not answered.

For simplicity in presentation, and following previous papers, we will rather inaccurately refer to a firm as 'guilty' when it was expected or requested to give gifts in the meeting with tax officials. This is the measure of corruption. The standard approach in the literature is simply to assume that all answers to the CQs are honest and to report statistics based on such answers.[7] These we call 'standard' estimates of corruption or guilt in what follows.

The second type of question is a forced-response random-response question (RRQ). For reasons made clear below, our methodology requires responses to more than one random-response question. The ones we use are intended to be on issues that are similarly sensitive to the subject matter on the CQ. This identifying assumption is important because we will assume that the probability of reticent behaviour is the same for the CQ and the RRQs. The questions are listed in Table 2, together with summaries of the responses to each question in each of the countries.[8] Following Azfar and Murrell (2009), Clausen, Kraay, and Murrell (2011), and Kraay and Murrell (2016), survey respondents were presented with a series of ten sensitive questions. They privately toss a coin before answering each question, having previously been instructed to answer "Yes" if the coin comes up heads and otherwise answer the sensitive question. 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 2, but were not so highlighted in the questionnaire itself.

The data in Table 2 provide immediate justification for our assumption that the RRQs do little to encourage respondent candour. Absent reticent behaviour, the rate of "Yes" responses on each of the RRQs should be at least 50 percent given that half of the responses would reflect the outcome of obtaining a heads on the coin toss, which should force a "Yes" response. Yet "Yes" response rates are below 50 percent in 47 of the 49 relevant cases in Table 2 (7 RRQs for each of 7 countries). Moreover, if a positive fraction of respondents had in fact done the sensitive acts in question, we should expect even higher rates of "Yes" responses. This provides a first clear indication that a significant proportion of responses reflect reticent behaviour.

---

[6] Several measures of corruption produced and used by the World Bank are available at:
http://www.enterprisesurveys.org/data/exploretopics/corruption
[7] With one caveat: refusals to answer are sometimes treated as admissions of guilt, as for example in analysis conducted by the World Bank's Enterprise Surveys unit.
[8] Because these RRQs are not part of the core questionnaire for the World Bank's Enterprise Surveys, we placed these questions in selected Enterprise Surveys over the past several years, in collaboration with the Enterprise Survey team at the World Bank. We are particularly grateful to Giuseppe Iarossi, Jorge Rodriguez, Veselin Kuntchev and Arvind Jain for their cooperation in placing these questions.

The data underlying Table 2 also provide justification for another assumption that we use in our model, that reticent respondents do not always behave reticently, but sometimes answer questions candidly. In Sri Lanka for example, the existence of reticence itself is clear from the 8.6 percent of respondents with zero "Yes" responses on all the seven sensitive questions, since if there were no reticence less than 1 percent would answer the RRQs in this way.[9] Importantly, another 33.6 percent of respondents answer "Yes" one or two times, while if there were no reticence, only 22 percent of respondents should do so.[10] This is evidence of significant reticence but also of reticent respondents who answer some questions candidly and others reticently. These points are further amplified if we assume that some respondents have done some of the sensitive acts in question, because candour would then require more "Yes" answers.

3. Modeling Survey Responses

Our goal in this section is to provide some structure in describing the interaction between an interviewer, who would like to elicit accurate information, and the respondent, who may prefer not to disclose this information. Our particular approach in addressing the problem of respondent reticence means that providing such structure is intrinsic in our methodology. But as we proceed below, it will become obvious that *every* attempt to construct a measure of corruption from similar questions involves making assumptions, either implicitly or explicitly, on the way in which reticent behaviour influences respondents' answers. The virtue of an explicit structure is that our assumptions become clear.

We follow the Azfar-Murrell (2009) definition of reticence—a reticent respondent is one who knowingly gives 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 assume that the probability that respondents answer "Yes" to a given question depends on (i) whether they are reticent individuals, in the sense that they are sometimes unwilling to truthfully answer a sensitive question, (ii) whether those who are reticent individuals choose to behave reticently on a specific question, and (iii) whether they have in fact done the sensitive act in question, i.e. whether they are guilty, allowing for guilt rates to be different for reticent and candid respondents. These are natural assumptions. The first is implied by all of the literature on the under-reporting of sensitive acts. The second—that reticent respondents are not always reticent—is strongly suggested by patterns in the data analyzed in previous research and noted above (Azfar and Murrell 2009; Clausen et al. 2011; Kraay and Murrell 2016). The third is completely intuitive: the guilty may very well have more incentive to behave reticently.

There are five parameters in our model: the probabilities of (i) receiving a visit, (ii) being reticent, (iii) behaving reticently on a given question, (iv) being guilty if reticent, and (v) being guilty if

---

[9] Specifically, under the assumption that the respondent has done none of the sensitive acts, the probability of observing seven "No" responses is $0.5^7 < 0.01$.

[10] Specifically, under the assumption that the respondent has done none of the sensitive acts, the probability of observing one or two "Yes" responses can readily be calculated from the binomial distribution with seven trials and a success probability of 0.5.

candid. We make the identifying assumption that the answers to all questions—CQs and RRQs—can be modeled using the same parameter values. This is a strong assumption, but all approaches to using survey data in the face of reticence would need to make some such assumptions. In Appendix B, we review some literature relevant to this assumption and present results that show that our general conclusions are robust to violations of this assumption.

Reticence is an individual specific trait, with respondents being reticent or candid with probabilities r and $1 - r$. Candid respondents are always honest. However, with probability q, reticent respondents choose to behave reticently on any given sensitive question—that is, they answer in a way that obscures any possible inference of guilt. Sometimes, however, reticent respondents behave candidly, and provide honest answers with probability $1 - q < 1$. We assume that the decision to behave reticently on the two-step CQ is made once, and is independent of the similar decisions on the RRQs. Notice that this set-up makes a distinction between reticence and behaving reticently: reticence is a fixed trait for a single individual, but any reticent individual sometimes behaves reticently and sometimes candidly across different questions.

Tax officials visit or inspect businesses with probability w.[11] We differentiate between two classes of businesses in order to allow for the possibility that guilt and reticence are correlated across respondents. First, in those businesses asked for a bribe with probability $0 \le g \le 1$, respondents are reticent. Second, in those asked for a bribe with probability kg ($0 \le k \le 1$), respondents are candid. A k less than one induces the correlation between reticent behaviour and guilt. However, our estimation procedure does not preclude $k = 1$ as a special case. We can therefore test whether, in fact, the guilty are more likely to give dishonest responses.

3.1 Modeling Responses to the Conventional Question

In this subsection we develop two alternative models of how reticent behaviour affects the interview process for the two-step CQ, a boiled-down version of which is "(1) was your firm visited by tax officials and (2) if yes, did they expect an informal payment?". Importantly, if no visit is acknowledged in the first step, no bribe question is asked in the second step. This means that reticent respondents can avoid the bribe question, if they so choose, by denying that a visit occurred. Thus, in contrast with Kraay and Murrell (2016), we must not only decide how to model reticent behaviour, but also specify which parts of the two-step CQ are affected by reticent behaviour.

We consider two distinct possibilities. In the first, which we refer to as Model A, reticent respondents are always candid in responding to the first part of the CQ, i.e. the question about the visit. However, they may behave reticently when responding to the sensitive second part of the CQ, i.e. the question about the bribe. The second possibility, which we refer to as Model B, allows reticent respondents to avoid answering the sensitive second part of the CQ by responding "No" to the first part of the CQ even if a visit did occur. Although the first part of the CQ is not *per se* sensitive, a "No" answer

_____

[11] It is of course plausible that businesses of different characteristics (size, activity, etc.) have different probabilities of getting a visit or inspection by tax officials. However, we do not model tax officials' choices here.

can put the respondent in the position of avoiding an inquiry on a sensitive act.[12] Specifically, we assume that reticent respondents who received a visit but not a request for a bribe respond candidly to both parts of the CQ. In contrast, reticent respondents receiving a visit and a bribe request and choosing to behave reticently deny that the visit occurred and thereby avoid the sensitive second question.[13]

For notational convenience, we now define a random variable D that fully summarizes the possible responses to the two-part CQ. Specifically, $D = 1$ if no visit is acknowledged, $D = 2$ if a visit is acknowledged but no bribe is reported, and $D = 3$ if a visit and a bribe are both acknowledged. Note that D is defined in such a way that it is non-missing for all respondents that answered either part of the CQ. This is in contrast with the bribe question itself, which is non-missing only for those respondents who acknowledge a visit in the first part of the CQ. The use of a structural model integrating responses on both parts of the CQ therefore will allow us to use the data on all respondents in estimating rates of corruption. This is in contrast to standard estimates that focus only on the subset of the sample corresponding to respondents who acknowledge a visit.

Table 3 spells out our assumptions on the likelihood of observing these three possible responses to the CQ. The panels of the table reflect combinations of the models (A and B) and the types of respondents (candid and reticent), with only three panels needed since candid respondents behave the same in both models. Within panels, the rows reflect the three possibilities for the respondents' actual experiences on visits and bribery, i.e. (i) a visit actually did not occur, (ii) a visit did occur without a bribe request, or (iii) a visit did occur and a bribe was requested. Of course, these actual experiences are unobserved for the researcher, who only sees survey responses. The columns of the table correspond to the three possible responses to the CQ, which are observed by the researcher. The cell entries spell out the probability of the (observed) response corresponding to the column, given the model, the type of respondent, and the (unobserved) actual experience indicated in the row of the table. Of course, even though the probability of two different values of D can both be positive, the value of D is unique in any given observation in the survey.

Consider the behaviour of candid respondents, summarized in Panel 1 of Table 3. Candid respondents always answer truthfully, so there are ones on the diagonal and zeros otherwise, indicating that candid respondents always answer truthfully e.g. if there was no visit, they say there was no visit with probability 1. At the bottom of this panel we report the corresponding population probabilities of observing the three possible responses on the two-part CQ, conditional on respondents being candid, i.e. $P[D = j|C]$ for the three outcomes $j = 1,2,3$, with the conditional C indicating a candid respondent.

---

[12] On the WBES questionnaires, this tax question appears in the middle (or in Nigeria, at the end) of a series of two-part questions each of which is identical in structure to the question on taxes, but referring to other government agencies. Thus by the time the respondents reach the tax question they should know that acknowledgement of a visit will be followed by a question about a bribe request.

[13] We also considered a third model, similar to model B except that all reticence respondents who chose to behave reticently did so on the visit question whether or not a visit had resulted in a bribe request. When evaluating the performance of each model—see Section 6—this third model was the least preferred for all countries and therefore we have not reported any results for this model.

Since candid respondents respond truthfully, the probability of observing a "Yes" response to the visit question is simply $P[D = 1|C] = w$, i.e. the true probability of a visit. Similarly, the probability of reporting a visit but not a bribe is $P[D = 2|C] = w(1 - kg)$, i.e. a visit occurs with probability w and a bribe does not occur with probability $1 - kg$. Naturally, the probability of reporting a visit and a bribe is simply the product of the probability of these two events, i.e. $P[D = 3|C] = wkg$.

The second and third panels of Table 3 present the same analyses for reticent respondents, whose behaviour differs between the two models. Under the assumptions of both Model A and Model B, reticent respondents truthfully respond that there was no visit if indeed no visit occurred, as summarized by the 1 in the top-left cell of each the second and third panels. Similarly, if there was a visit and no bribe request occurred, they truthfully acknowledge the visit in the first part of the CQ, and truthfully state that there was no bribe in response to the second part of the CQ. This is represented by a 1 in the center cell of both panels. If a visit and a bribe occurred, under the assumptions of Model A reticent respondents admit the visit but behave reticently with probability q on the second part of the CQ. Specifically, with probability q the reticent respondent denies the bribe, or otherwise the bribe is acknowledged (hence with probability $1 - q$). On the other hand, under the assumptions of Model B, reticent respondents who are guilty might manifest reticent behaviour in the first part of the CQ: they deny that the visit occurred with probability q, or otherwise admit to the visit and the bribe (hence with probability $1 - q$).

The different assumptions of Model A and Model B imply different probabilities of observing the three possible responses to the CQ among reticent respondents. In Model A, reticent respondents are always candid about the visit question, and so the probability of denying a visit is just one minus the true probability of a visit, i.e. $P_A[D = 1|R] = 1 - w$, the A subscript referencing the model and the conditional R indicating a reticent respondent. In contrast, in Model B reticent respondents who (i) received a visit (with probability w), *and* (ii) received a bribe request (with probability g), *and* (iii) choose to behave reticently (with probability q) will deny that the visit occurred. These responses are in addition to those of respondents who in fact did not receive a visit (a proportion $1 - w$ of respondents), so the overall probability of a response that a visit did not occur is $P_B[D = 1|R] = 1 - w + wgq \geq P_A[D = 1|R]$.

In Model B, reticent respondents always acknowledge a visit when a bribe does not occur, so the probability of a response that a visit occurred without a bribe is simply the product of the probability of a visit and the probability of a bribe not occurring, i.e. $P_B[D = 2|R] = w(1 - g)$. However, in Model A, there are also some reticent respondents who experienced a visit and a bribe request, but decided to behave reticently on the second part of the CQ and deny that the bribe occurred. This implies a greater likelihood of observing a "No" response to the second part of the CQ in Model B, i.e. $P_A[D = 2|R] = w(1 - g + qg) \geq P_B[D = 2|R]$. Finally, in both models, reticent behaviour reduces the likelihood of respondents admitting both the visit and the bribe by the same amount, i.e. $P_A[D = 3|R] = P_B[D = 3|R] = wg(1 - q)$. In both cases, a visit *and* a bribe occur with probability wg, and reticent respondents acknowledge this with probability $1 - q$. In Model A, the remaining proportion q of

respondents admit to the visit but deny the bribe, while in Model B these respondents manifest their reticence by denying the visit, thereby avoiding the question about the bribe.

Combining these observations, we can now summarize how the presence of reticent respondents affects the interpretation of responses to both parts of the CQ. Note first that the probability that a visit is reported in the first part of the CQ is:

(1)
$$1 - P[D = 1] = \begin{cases} w, & Model\ A \\ w(1 - rgq), & Model\ B \end{cases}$$

In Model A, both candid and reticent respondents are candid in their responses to the question about visits, and so the observed rate of "Yes" responses is a valid estimate of the frequency of visits, w. However, in Model B, a fraction r of respondents are reticent, experience a bribe with probability g, and with probability q choose to behave reticently by denying that the visit occurred. As a result, the observed frequency of visits in the data is an underestimate of the true frequency of visits, and the extent of the downward bias depends on the prevalence of reticent respondents, the likelihood of reticent behaviour on the part of reticent respondents, and the frequency of bribery itself.

The interpretation of responses on the second part of the CQ is similarly clouded by reticent behaviour. Consider the rate of admission of bribery among those respondents who acknowledge a visit, which is exactly the standard estimate of corruption. Using our notation, this is:

(2)
$$\frac{P[D = 3]}{1 - P[D = 1]} = \begin{cases} (1 - r)kg + rg(1 - q), & Model\ A \\ \dfrac{(1 - r)kg + rg(1 - q)}{1 - rgq}, & Model\ B \end{cases}$$

In Model A, all respondents who were visited by an inspector admit to the visit. Among these, a proportion $1 - r$ are candid and truthfully admit that a bribe occurred with probability kg. In contrast, a proportion r of respondents are reticent. For these, a bribe occurs with probability g but only a proportion $1 - q$ of reticent respondents chooses to behave candidly and answer the question truthfully. The only difference in Model B is that reticent respondents who are guilty and choose to behave reticently deny that the visit occurred rather than admitting the visit but denying the bribe. The probability of this occurring is rgq. In Model A these respondents would have admitted the visit in the first part of the question, but would have responded "No" to the bribe question in the second stage. In Model B these respondents do not advance to the second stage and so the rate of "Yes" responses in the second stage is higher by a factor of $1/(1 - rgq) \geq 1$.

3.3 Modelling Responses to the Random-Response Questions

For both models, the interview process for RRQs is modeled as in Kraay and Murrell (2016). The key assumption here is that reticent behaviour is equally prevalent in responses to the RRQ as it is in the

CQ. Recall that on each of the questions in the RRQ, the respondent is instructed to answer the sensitive question if the coin comes up tails, and to answer "Yes" if the coin comes up heads. Our definition of a reticent respondent is one who gives knowingly false answers with a nonzero probability when honest answers to a question could generate the inference that the respondent might have committed a sensitive act. Given the assumptions above and this definition, there is therefore a probability q that a reticent respondent will respond "No" even when the coin comes up heads. Similarly, when the coin comes up tails, there is a probability q that the reticent respondent will answer "No" even if the respondent has done the sensitive act in question. The data in Table 2 provide reassurance on this assumption. If the outcome of a head on the coin toss actually encourages candour, then all entries in that table would lie above 50%. However, as discussed in the previous section, for the truly sensitive questions the responses are greater than 50% only in two out of 49 cases (seven sensitive questions in seven countries).[14]

As discussed in more detail in Kraay and Murrell (2016), these assumptions imply that the probability of a "Yes" response on any single RRQ is $0.5(1 + g)(1 - q)$. To see this, note that respondents are supposed to answer "Yes" either if they are guilty (with probability g) or if they are not guilty 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 $0.5(1 + kg)$. Candid respondents can have a lower guilt probability than reticent respondents (i.e. $kg \leq g$), but always answer honestly (i.e. q is not relevant to them).

Define the random variable X as the number of "Yes" responses on the 7 sensitive RRQs that are given in bold in Table 2. Since both the coin toss and the decision to behave reticently on a given question are independent across questions in the battery of RRQs, X is binomially distributed, with different success probabilities for reticent and candid respondents as noted above, i.e. a success probability of $0.5(1 + kg)$ for candid respondents and $0.5(1 + g)(1 - q)$ for reticent respondents.

It is also useful to briefly note why we use several RRQs and not just a single RRQ. As discussed in Section 2, in modelling the data generating process it is important to allow for the possibility that the reticent do not always behave reticently, that is to include in the model a parameter $q < 1$. But then if $k = 1$ or $k = 1 - q$, which are possibilities that we do not want to rule out *a priori*, the effects of q and r on data gathered on any single CQ or single RRQ are identical.[15] Hence, these parameters would not be separately identified in an estimation that relied on one CQ and a single RRQ. Where these two parameters do have different effects is on the variation in the number of "Yes" responses to a set of RRQs, that is X. Hence, we use a battery of seven RRQ's chosen to approximate the severity of guilt and the degree of sensitivity of the CQ.

---

[14] Since the assumption that rates of reticence on the CQ and the RRQ are the same is so critical to our procedures, and is non-standard in the context of the existing literature on RRQs, we elaborate on this point in Appendix B. Importantly, in that appendix we report the results of some simulations showing that if our assumption is not correct in the sense that reticent behaviour is less prevalent in the RRQ, then we *underestimate* the degree of downward bias in standard estimates of corruption.

[15] This point is trivial to show using equations (1) and (2) and the information appearing in the paragraphs immediately above.

The intuition for the separate identification of r and q follows from our assumptions that reticence is an individual-specific trait that is fixed across questions for a given respondent, while the event of reticent behaviour, i.e. failure to answer "Yes" when supposed to, is independent across questions. This has the implication that the presence of reticence will be reflected in a covariance across a respondent's answers to the individual questions in an RRQ battery. For example, reticent respondents are less likely to answer "Yes" to all RRQs precisely because they are reticent. This in turn will be reflected in the sample distribution of the number of "Yes" responses to a battery of individual RRQs (X). That is, for any given r, the sample distribution of X will vary with q, giving the information for identification.[16]

4. Estimation and Definitions of Composite Parameters

We estimate the parameters of our model using maximum likelihood (ML). With the notation and structure of the previous section in hand, it is straightforward to write down the likelihood function of the observed responses to the CQ and the RRQ, which is:

$$L_i(D, X; w, g, r, q, k)$$

(3)

$$= (1 - r) \left(\sum_{j=1}^{3} P[D = j|C] I_{(D=j)} \right) B \left(X; 7, \frac{1 + kg}{2} \right)$$

$$+ r \left(\sum_{j=1}^{3} P_i[D = j|R] I_{(D=j)} \right) B \left(X; 7, \frac{(1 - q)(1 + g)}{2} \right)$$

for models $i \in \{A, B\}$. $I_{(D=j)}$ is an indicator variable taking the value 1 if $D = j$ and zero otherwise. B(.) is a binomial density function for X successes in 7 trials (the number of RRQs) that have the success probability stated in the last argument.

To understand this likelihood function, note that the two lines correspond to the contribution of candid respondents (a proportion $1 - r$ of the sample) and the contribution of reticent respondents (a proportion r of the sample). For both groups, their contribution to the likelihood function is the product of the trinomial probability distribution for the three response possibilities for the CQ, and the binomial distribution summarizing the observed responses to the RRQs. Since the responses to the CQ and the RRQ are independent conditional on being candid, the product of these two distributions forms the joint distribution of responses for candid respondents. Finally note that while the response probabilities for the CQ are the same for both models for the candid respondents and are given by $P[D = j|C]$, they are different for models A and B for the reticent respondents, i.e. $P_A[D = j|R]$ and $P_B[D = j|R]$. By substituting the appropriate functions of (w, g, r, q, k) given in Table 3 for these response probabilities

---

[16] See Kraay and Murrell (2016) for more details.

and then multiplying across individuals, one obtains the likelihood function for the observed data. In the next section we report the estimates of the parameters obtained by maximizing this likelihood function.

We also define three composite parameters, which provide a more intuitive interpretation of the implications of the two models. The first we call *effective reticence*, which is the probability that a randomly-selected respondent will choose to answer a sensitive question reticently, i.e. rq. The second composite parameter is motivated by the fact that the model assumes that the reticent have a higher rate of guilt than the candid and that the rate of guilt of a randomly-selected respondent is therefore a weighted average of the rates of the two types of respondents: $g(r + (1 - r)k)$. We call this *average guilt*, which is the probability that a random respondent will be asked for a bribe given that this respondent is being visited by a tax official. Third we estimate the unconditional probability that a randomly selected respondent will be both visited by a tax official and asked for a bribe, which is average guilt times the probability of a visit, i.e. $wg(r + (1 - r)k)$. We call this *effective corruption*, reflecting the fact that in countries where corruption is widespread part of the underlying motive for visits by tax officials is to extract a bribe.

5. Results: Standard Estimates and Alternative Models of Reticence

Tables 4-10 report estimates of Model A and Model B for the seven countries in our dataset. In each Table we report estimates for the five model parameters, g, k, r, q, and w, and for the three composite parameters, *effective reticence*, *average guilt*, and *effective corruption*.[17] Our discussion focuses mainly on the three composite parameters since these are most informative about overall levels of corruption and reticence.

Effective reticence (rq) reflects the proportion of sensitive questions that are not answered candidly in the whole sample. This proportion varies from 28% for Model A in Sri Lanka to 67% for Model A in India. Variation is much greater between countries than between models within a country, the two models telling a quite consistent story about rates of effective reticence. One element of this story is that the countries seem to fall naturally into two groups. Turkey and India have rates of effective reticence approximately twice as high as the rates in all other countries. The estimates for these two countries imply that approximately 64% of the time a respondent would answer the second-part of the CQ misleadingly if the respondent had in fact been asked for a gift or informal payment. The analogous percentage for the other countries varies between around 30% and 40%. In all countries, these are high rates of misleading answers. The obvious implication is that overall rates of corruption in standard measures are downward biased, with the degree of underestimation varying markedly between countries.

---

[17] Our test statistics for the parameters are based on heteroskedasticity-consistent standard errors clustered at the strata level. Following Cameron-Miller (2010 pp. 19-20), our coefficient estimates would be consistent even in the presence of significant intra-cluster correlation of observations although they would no longer be the ML estimates. If instead we used the test statistics that were based on standard errors that were not clustered, there would be no substantive differences in our main conclusions

Examining average guilt, which is the probability that a request for a bribe is made on any visit, our estimates average approximately twice those that would be reported using standard methods. But as expected there also is large variation between countries. For Turkey, the estimate of average guilt is increased either tenfold (Model A) or fourfold (Model B) relative to standard estimates. In Peru, on the other hand, effective guilt is "only" 50 percent higher than conventional estimates. The significance of allowing for reticence is perhaps best shown by the results for Sri Lanka, because that country generally scores reasonably well in cross-country rankings of corruption, and has relatively low rates of reticent behaviour. Even then, average guilt is 69% above standard estimates in Model A and 35% above in Model B.

Effective corruption reflects the unconditional probability of being personally involved in a corrupt exchange with tax officials, that is, it is the probability of having been visited by tax officials and having been expected to give a gift on such a visit.[18] This proportion varies from 6.7% for Model A in Peru to 48% in Model B for Bangladesh. In contrast, standard estimates in these instances give 4.2% for Peru and 26% for Bangladesh. Estimates for Model A increase the perception of overall corruption by approximately 100% on average whereas the corresponding figure for Model B is 75%. Thus the risk of being personally involved in corruption is usually greatly underestimated. But it is not uniformly underestimated. In Turkey, which has low rates of corruption and high rates of reticence based on any yardstick, estimates of effective corruption are tenfold the standard estimates in Model A and fourfold in B. In contrast, Ukraine's estimate of effective corruption in Model B is marginally less than that in standard estimates.

We now turn to brief comments on the estimates of the individual parameters. For w, the probability of a visit by tax officials, estimates vary much more between countries than between models. As discussed in Section 3, Model A's estimates are by construction identical to standard estimates, whereas Model B's estimates are usually in the range of 5-10% higher than standard ones. One exception to this is India, a high corruption, high reticence country, where our estimate of visits by tax officials is more than 50% greater than conventional ones. Such an observation changes perceptions about the mechanisms of corruption in a country, suggesting that the decision of a tax official to visit a business might be an element of that mechanism.

For k, the parameter capturing (inversely) the correlation between reticence and guilt, results vary more between models than for any other of the parameters. For three countries, Model A tells a different story than Model B. For example, in Ukraine Model B finds no significant correlation (k not significantly different from 1) while Model A suggests that the reticent are four times as likely to be

---

[18] This composite parameter is different from the measure of corruption in dealing with tax officials that is usually publicized (e.g. by the WBES) in that it takes into account the fact that not every business is necessarily involved in the contexts that potentially involve corruption. Average guilt, estimates of which are provided in Tables 4-10, is the concept that is usually reported.

guilty as the candid. However, in general, estimates of k are quite low, implying a strong correlation between reticence and guilt, which is hardly surprising.[19]

6. Results: Preferred Estimates and Analysis

In this section we draw together the preferred estimates for each country. We choose either Model A or Model B as the preferred model, based on which has the largest value of the maximized log-likelihood.[20] Table 11 first indicates our preferred model and then the following five rows summarize the preferred estimates of the five parameters of the model, i.e. w, r, q, g, and k, and the corresponding estimate of effective reticence, rq. In the remaining rows we focus on accounting for the primary question of interest in this paper, which concerns the size and source of downward biases present in measures of corruption that do not take reticent behaviour into account. We focus here on estimates of effective corruption, i.e. the likelihood that a firm is visited by a tax inspector and a bribe request takes place. The standard way to estimate effective corruption without taking reticence into account would be to simply calculate the proportion of all respondents who admit to both a visit and a bribe request. The population frequency of this is $P[D = 3]$ in the notation of Section 3. Note that the standard estimate of effective corruption differs from the standard estimate of corruption conditional on a visit occurring, which is described in Equation (2) above and is commonly reported in analysis of the Enterprise Survey data.

Our goal in this paper is to contrast the standard estimate with the "true" rate of effective corruption, which is the product of the true probability of a visit, w, and the true probability that a bribe encounter occurs, $g(r + (1 - r)k)$. In addition, we would like to decompose the downward bias into the parts due to the under-reporting of visits and the under-reporting of bribe experiences. This decomposition is as follows:

(4)
$$\frac{P[D = 3]}{wg(r + (1 - r)k)} = \frac{1 - P[D = 1]}{w} \cdot \frac{\dfrac{P[D = 3]}{1 - P[D = 1]}}{g(r + (1 - r)k)}$$

The left-hand side of this expression is the ratio of measured corruption to true effective corruption, which is less than one if there is reticent behaviour. The first term on the right-hand side is the ratio of the population value of the standard estimate of the probability of a visit to the true probability, and so measures the downward bias in reported visits due to reticence. The second term on the right-hand side is the ratio of the population value of the standard estimate of bribery conditional on a visit to the corresponding true probability, and so measures the downward bias in reported bribe interactions conditional on a visit being reported.

---

[19] This general point is strongly reinforced when we examine the set of preferred estimates in the ensuing section. It is also a strong conclusion in Kraay and Murrell (2016).

[20] Formally, we perform the Vuong (1989) model selection test for non-nested models estimated by maximum likelihood. The test is very simple and involves forming the difference between the maximized value of the likelihood function between the two models for each observation, and then performing a standard t-test of the null hypothesis that the mean difference is zero. The difference in the maximized values of the log likelihoods for the two models is statistically significant at the 5% level in Ukraine.

These two sources of bias are different in the two models we have considered. In particular from Equation (1) we see that the bias in reported visits is:

(5)
$$\frac{1 - P[D = 1]}{w} = \begin{cases} 1, & Model\ A \\ 1 - rqg, & Model\ B \end{cases}$$

In Model A we assumed there is no reticence in responding to the visit question so there is no bias relative to the true incidence of visits. However in Model B reticent respondents who experienced a bribe behave reticently with probability q, and so there is a downward bias, i.e. $1 - rqg < 1$. We report standard and reticence-adjusted estimates of the probability of a visit, and the ratio of the two, in the first part of Panel B of Table 11. This ratio is equal to 1 for Sri Lanka, Turkey, and Ukraine where Model A is the preferred model. However, for the remaining four countries this ratio is less than one, and substantially so in India (67%) and Bangladesh (73%). On average, for the four countries where Model B is the preferred model, reticent behaviour implies that only 81% of visits that actually occur are acknowledged by respondents.

We next turn to the bias in reported bribes which is:

(6)
$$\frac{\frac{P[D = 3]}{1 - P[D = 1]}}{g(r + (1 - r)k)} = \begin{cases} \dfrac{r(1 - q) + (1 - r)k}{r + (1 - r)k}, & Model\ A \\[4mm] \dfrac{\dfrac{(r(1 - q) + (1 - r)k)}{1 - rqg}}{r + (1 - r)k}, & Model\ B \end{cases}$$

In both Model A and Model B, the frequency of reported bribes among those who report a visit is biased down relative to the true probability of a bribe occurring, i.e. this ratio is less than one when there is reticent behaviour. Comparing Model A and B, the downward bias is greater in Model A. As discussed in Section 3, this is because some reticent respondents simply denied the visit altogether in Model B rather than admit the visit and then deny the bribe. We summarize the estimated downward biases in the second part of Panel B in Table 11, reporting standard and reticence-adjusted estimates of rates of bribery among those who admit to a visit, as well as the ratio of the two. This ratio naturally is less than one, but varies widely across countries. At the one extreme, in Turkey we estimate that only 10% of the bribes that actually occur are reported in responses to the second part of the question. At the other extreme, in Nigeria virtually all bribe interactions are acknowledged in the second part of the question. On average across the seven countries, we estimate that 58% of bribes that actually occur are acknowledged.

Inserting Equations (5) and (6) into Equation (4), we find that the overall bias in standard estimates of effective corruption relative to actual effective corruption is the same function of the parameters in both Models A and B, i.e.

$$(7) \qquad \frac{P[D=3]}{wg(r+(1-r)k)} = \frac{r(1-q)+(1-r)k}{r+(1-r)k}$$

for both models. Standard estimates of effective corruption are biased downward as long as $rq > 0$, i.e. as long as there is reticent behaviour. We report the standard estimates, estimated true effective corruption, and the ratio of the two in the last Part of Table 11. The downward bias is substantial, with the ratio of standard estimates to actual effective corruption averaging 50% across the seven countries. Naturally, this ratio also varies considerably across countries, from a low of 10% in Turkey to a high of 88% in Nigeria. To assess the significance of these overall downward biases, we also report the z-statistic for the test of the null hypothesis that the ratio on the right-hand side of equation (7) is equal to one. This null hypothesis is overwhelming rejected in all seven countries.

Figure 1 depicts the extent to which standard estimates of corruption are biased in the seven countries, how much the degree of bias varies, and the source of the bias, in terms of reticence on visits and reticence on bribes given visits.   The solid, blue bar shows our preferred estimates of effective corruption while the white bar shows standard estimates. The varying heights of the first bar show how much corruption levels vary between countries, the contrast between Bangladesh and Peru, for example, being very striking. The ratios of the last to the first bars show how much the underestimation of corruption due to reticence varies between countries, with the contrast between Nigeria and Turkey standing out. Finally, the decomposition of bias is depicted in two steps, when moving from the white bar to the cross-hatched, red bar to the solid blue bar, with under-reporting due to reticence on the bribe question first corrected and then under-reporting due to reticence on visits corrected. While Bangladesh, India, and Nigeria evidence a significant portion of their bias from under-reporting on visits, Peru and the three countries with the preferred Model A do not.

7. Conclusions

We conclude with five points not emphasized above. First, when aiming to produce valid cross-country comparisons of corruption, we highlight how important it is to model responses in terms of structural parameters. Second, we underscore how much cross-country perceptions of corruption can change when such perceptions are cast in terms of structural parameters estimated using models that acknowledge the effect of reticence. Third, we note that similar changes may occur over time, to the extent that reticent behaviour changes over time. Fourth, we emphasize the broader applicability of these techniques to other survey settings where reticent behaviour is a concern. Finally, we comment on the plausibility of the identifying assumptions that drive our approach.

Our summary in Section 6 reflects the results from the single preferred model for each country, basing the choice on which model maximizes the likelihood. As it happens, the preferred model differs between countries. In 3 of the 7 cases the preferred model has no reticence on visits and in the remainder there is reticence on visits when a bribe occurred in the subsequent interaction.[21] This fact speaks to the virtue of our approach of estimating structural parameters that have identical meanings

---

[21] As noted above, the data for all countries reject a model that reticent respondents treat the visit as a sensitive issue in exactly the same way that they would confront a question on bribes

across countries but enter different models in different ways. Because respondent behaviour on reticence differs across countries, standard statistics, which reflect varying combinations of those structural parameters, have different meanings across countries. By focusing on the estimates of structural parameters themselves, we are able to generate valid comparisons between countries. Since the design of any anti-corruption program would need access to those structural parameters, such programs would be much better informed by the estimates we present here than by simple summary statistics from two-step questions. Indeed, as we have shown above, the intuitively appealing standard summary statistics might contain little useful information given the biases induced by reticent behaviour.

Figure 2 depicts how much perceptions of differences in effective corruption across countries can change when using comparisons of rigorously-estimated structural parameters rather than comparisons of the intuitively appealing, but informationally obscure, standard estimates. On the horizontal axis we graph the standard estimates of effective corruption and on the vertical axis we report our preferred reticence-adjusted estimates, those appearing in Table 11. The upward sloping line traces out the points where reticence-adjusted estimates equal standard estimates of guilt. The large differences between model-based estimates of effective corruption and standard estimates are readily apparent in the large vertical distance between any data point and the upward-sloping line. More notably, there are very significant reversals of orderings. In standard estimates, Nigeria appears much more corrupt than both India and Ukraine, whereas in our estimates Nigeria is less corrupt than those two countries. Whereas standard estimates place Turkey as the least corrupt of our seven countries, we place the levels of corruption of Sri Lanka and Peru below those of Turkey, with Figure 2 exhibiting a dramatic change in the relative perception of corruption in Turkey. Given that aid allocations are partially based on such rankings, it would seem prudent to use rankings based on models using structural parameters when the data are from sensitive questions asked on surveys.

Our results clearly show that the extent of reticence, and the channels through which it is manifested in survey responses, vary considerably across countries. An open question that we cannot address with our data is the extent to which reticence varies within countries over time. Just as cross-country differences in reticence cloud the interpretation of cross-country differences in survey responses to sensitive questions, it is possible that changes over time in survey responses are affected by changes in reticent behaviour. This concern may be particularly acute when comparing survey responses from before and after major political or institutional transitions. For example, it is plausible that a large shift from a repressive regime to a more open regime might simultaneously influence the incentives and opportunities for corruption, and survey respondents' willingness to acknowledge corrupt acts. Depending on whether these effects complement or offset each other, average survey responses to questions about corruption may exaggerate or understate the effect of the institutional change on actual corruption. In principle, it is possible to address this issue empirically, by applying the methodology in this paper at different points in time in the same country, ideally before and after large institutional reforms. However, at this point we can do little more than speculate about the likely direction of these changes in reticent behaviour over time, while emphasizing that our paper suggests they could be considerable.

19

It should be noted that our methodology is a general one, not restricted to study of corruption, but also applicable in other survey settings where responses to sensitive questions may be influenced by reticent behaviour. As noted earlier, two-stage questions regarding sensitive acts are common in many other settings, and our methodology provides a framework for handling reticence in either stage of the question. When considering applications of this methodology in other contexts, we can offer at least some hints of the wider external validity of this approach. For three of the countries included here, Bangladesh, India, and Sri Lanka, we have also applied our methodology in household surveys conducted by the Gallup World Poll. Despite the very different context of household (as opposed to firm) surveys, we find a rank ordering of countries by effective reticence that is remarkably consistent with this paper, with Sri Lanka the lowest, Bangladesh next, and India having much higher rates than the other two.

To be sure, our method does rely on a strong identifying assumption—that we are able to find a set of questions (used for the RRQ's) that are approximately comparable to the CQ in terms of degrees of guilt and reticence. As with most identifying assumptions, this cannot be tested directly. However, we can provide some indirect evidence on the plausibility of our identification strategy by noting that our estimates of reticence are consistent with those in the literature where scholars have been fortunate enough to be able to compare survey responses to direct evidence of the phenomenon under investigation.[22] For example, Wolter and Preisendörfer (2013) present estimates corresponding to our rq and they find a value of 0.40 for self-reports of criminal activity, whereas our country estimates for corruption range from 0.28 to 0.64. Gong (2015) reports on a number of studies of self-reported sexual activity that contain estimates analogous to our rq ranging from 0.22 to 0.48. Funk (2016) compares polls on voting behaviour to official data on Swiss referenda, finding reticence varying widely between referenda. For example, her estimates imply an rq as high as 0.45 for a measure on reducing the voting age and one of 0.26 for a measure on easing immigration restrictions. For some studies, the focus is on the proportion of wrong answers, which is comparable to our grq, which ranges from 0.07 to 0.35. The comparable estimate in the meta-analysis of six studies by Lensvelt-Mulders et al. (2005) is 0.38. Therefore, our strong identifying assumption leads to estimates that are consistent with those obtained in contexts where the availability of direct data obviates the need to make such an assumption. But these are exactly the situations where survey methods are not needed—the sought-after information is available directly. For corruption, direct evidence is near-impossible to obtain, necessitating a method such as the one we have developed above and the need for the identifying assumption that we make.

---

[22] In producing the information in the following, we made a number of assumptions on how to treat non-responses, whether voters accurately reported whether they voted, etc. While precise numbers would change if we varied these assumptions, the overall conclusion of this paragraph would be unaltered.

Referenes

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.

Boruch, R. F. (1971). "Assuring confidentiality of responses in social research: a note on strategies." *The American Sociologist* 6, 4, 308-311.

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 Søreide, Cheltenham: Edward Elgar, 428-450.

Cameron, A. Colin and Douglas L. Miller. 2010. "Robust Inference with Clustered Data" Handbook of Empirical Economics and Finance, edited by Aman Ullah and David E. A. Giles. Boca Raton: Chapman and Hall/CRC, 1-28.

Holbrook Allyson L., and Jon A. Krosnick "Measuring Voter Turnout By Using The Randomized Response Technique: Evidence Calling Into Question The Method's Validity." *Public Opinion Quarterly*, Vol. 74, No. 2, Summer 2010, pp. 328–343

Coutts, Elisabeth and Ben Jann. "Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)" *Sociological Methods & Research* 40(1) 169–193 2011.

Funk, Patricia . "How Accurate Are Surveyed Preferences for Public Policies? Evidence from a Unique Institutional Setup". *Review of Economics and Statistics* July 2016, Vol. 98, No. 3: 442–454

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

Kraay, Aart, and Peter Murrell. "Misunderestimating Corruption." *Review of Economics and Statistics*, July 2016, Vol. 98, No. 3: 455–466

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.

Lensvelt-Mulders, Gerty J.L.M.   and Hennie R. Boeije "Evaluating compliance with a computer assisted randomized response technique: a qualitative study into the origins of lying and cheating" *Computers in Human Behavior* 23 (2007) 591–608

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.

OECD. *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, PISA, OECD Publishing.

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

Reinikka R. and J. Svensson "Local Capture: Evidence from a Central Government Transfer Program in Uganda" *Quarterly Journal of Economics*, 2004.

Rose, Richard and Caryn Peiffer, *Paying Bribes for Public Services* , Palgrave Macmillan February 2015.

Shipton D., Tappin D.M., Vadiveloo T., Crossley J.A., Aitken D.A., and Chalmers J. "Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study." BMJ. 2009 Oct 29; 339:b4347.

Tourangeau, Roger, and Ting Yan. "Sensitive questions in surveys." Psychological Bulletin 133, no. 5 (2007): 859-833.

Trappmann, Mark, Ivar Krumpal, Antje Kirchner, and Ben Jann "Item Sum: A New Technique For Asking Quantitative Sensitive Questions." *Journal Of Survey Statistics And Methodology* (2014) 2, 58–77

Vuong, Q. (1989), ''Likelihood Ratio Tests for Model Selection and Nonnested Hypotheses,'' *Econometrica* 57, 307–333.

Warner, Stanley L. "Randomized response: A survey technique for eliminating evasive answer bias." Journal of the American Statistical Association 60, no. 309 (1965): 63-69.

Wojcik, Sean P., Arpine Hovasapian, Jesse Graham, Matt Motyl, and Peter H. Ditto "Conservatives report, but liberals display, greater happiness" *Science* 13 March 2015 347(6227) 1243-1247.

Wolter, Felix and Peter Preisendörfer " Asking Sensitive Questions: An Evaluation of the Randomized Response Technique Versus Direct Questioning Using Individual Validation Data" *Sociological Methods & Research*, 42(3) 321-353, 2013.

World Bank "Enterprise Surveys" (WBES). 2015. http://www.enterprisesurveys.org/

Appendix A: The Survey Questions and the Data

The Two-Step Conventional Question

1. A professional surveyor read the following to the respondent: *"Over the last year, was this establishment visited or inspected by tax officials?"* Respondents could either answer "Yes", or "No", or "Don't know" (DK) or refuse (R) to answer. Respondents answering DK or R were dropped from the sample. The incidence of DK and R responses to these questions is given in Table A.1.

2. We set $D = 1$ if the respondent did not acknowledge that the visit occurred, i.e. if the respondent answered "No".

3. If the above question was answered with a "Yes", then the interviewer read the following to the respondent: [23] *"In any of these inspections or meetings was a gift or informal payment expected or requested?"* Respondents could either answer "Yes", or "No", or DK or R.[24] Respondents who answered DK or R were dropped from the sample. The incidence of DK and R responses to these questions is given in Table A.1.

4. We set $D = 2$ if the respondent said "No" to the inquiry about the bribe request, and we set $D = 3$ if the bribe request is acknowledged.

The Random Response Questions

1. A professional surveyor 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."*

2. The ten sensitive questions used in this battery of questions are given in Table 2. Respondents who refused to respond or responded "Don't know" were dropped from the sample.

3. The variable *X* used in the analysis is equal to the number of the seven bolded questions in Table 2 for which the respondent answers "Yes".

The incidence of DK and R responses to these questions is given in Table A.1.

---

[23] The questionnaire contains one other question that is asked between these two questions if the respondent answers yes to the first question: "If visited or inspected by tax officials, over the last year, how many times was this establishment either inspected by tax officials or required to meet with them?" Information from this sub-question is not used here.

[24] The World Bank constructs the numerator of the following variable: "Percent of firms expected to give gifts in meetings with tax officials" by including both those who answer "yes" and those who refuse to answer, effectively assuming that a refusal means "yes". In contrast, we drop from the sample those who refuse to answer.

Cleaning the Data for Interviewer Effects

The RRQ battery is a key ingredient in our methodology, and therefore it is important to ensure that this unusual and cumbersome-to-administer procedure was implemented as designed. Enumerators received specific training on the RRQ methodology. As part of this training, they learned how the RRQ methodology is supposed to provide greater anonymity for respondents. However, they were not briefed on our intention to use the RRQ battery to make inferences about reticence.

Despite these precautions we do find some evidence of interviewer effects in the data that might indicate variation across interviewers in the implementation of the RRQs. In all countries we have information on the identity of the interviewer for each respondent.[25] For each interviewer, we calculated the proportion of respondents with seven "No" responses on the RRQs. For most interviewers in most countries, we found rates of seven "No" responses that were not too different from the corresponding country averages. However, we did find some interviewers with implausibly high rates of seven "No" responses. We speculate that this may reflect differences across interviewers in how the RRQ was implemented. One possibility is that the interviewer incorrectly had the respondent toss the coin only once and let a single outcome govern the responses to all questions in the RRQ battery. This could lead to an upward bias in our estimates of the prevalence of reticent behaviour. To avoid such a possibility, we drop all interviews performed by interviewers whose interviewer-specific rate of seven "No" responses on the sensitive RRQ questions was more than five standard deviations above the corresponding country average.[26] Combining all surveys except India, we drop 2 percent of interviewers who accounted for just under one tenth of all respondents who answered "No" seven times on the RRQ battery. Including India, we drop a total of 4 percent of interviewers, together accounting for over third of all of the respondents who answered "No" seven times.

This process is necessary in order to pursue the objective of focusing solely on the effects of respondent reticence. Our goal is not to evaluate the properties of survey data as a whole, but rather to investigate the effect of reticent behaviour on the CQ as a possible source of bias in estimates of corruption. The goal is furthered by focusing on a subset of the data where one can be most sure that interview procedures were followed faithfully. We also note that while dropping these interviewers naturally increases the rate of "Yes" responses on the RRQ, it only has small effects on the rate of "Yes" responses on the CQ. This treatment of interviewers did not result in any changes in the data from Turkey and Nigeria. It had the biggest effect on the rate of "Yes" responses on the second part of the CQ in the data from India where the rate increased from 9.1 percent to 11.9 percent. In Peru, Bangladesh 2011, and Bangladesh 2013, this rate increased by less than one percentage point, and in Ukraine it decreased by less than 0.1 percentage point. This suggests that our concerns about the dropped interviewers applies only to their administration of the RRQs, except in India, perhaps.

---

[25] With the exception of India where this information is available only for half of the sample and Nigeria where the interviewer code is missing for 2387 out of 5544 interviews. Therefore the procedure described below is not applicable to these observations in India and Nigeria.

[26] Specifically, if i in country c carried out n_{ic} interviews, for which a proportion p_{ic} answered "No" to all seven sensitive questions, we dropped all the interviews of this interviewer if $p_{ic} - 5\sqrt{\frac{p_{ic}(1-p_{ic})}{n_{ic}}} > \frac{\sum_i n_{ic} p_{ic}}{\sum_i n_{ic}}$

Appendix B: The Assumption that Reticence on the CQ and the RRQs is the Same

Our methodology assumes that rates of reticence on the CQ are the same as on the RRQs. This appears to be a strong assumption because the RRQ was developed with the exact purpose of reducing respondent reticence relative to that on CQs. In this Appendix we justify our assumption in two ways. First, we show that the assumption is reasonable given current evidence in the survey-research literature. Second, we show that one of our major conclusions—the underestimation of corruption—is robust when this assumption is relaxed, that is, assuming the RRQ does reduce respondent reticence.

In a meta-analysis, Lensvelt-Mulders et al. (2005) examined the few studies where RRQs and CQs were used and external validation of survey responses was possible. They found that on average RRQs had 90 percent of the reticence of conventional face-to-face interview questions (CQs). Holbrook and Krosnick (2010) and Wolter and Preisendörfer (2013) cite a large number of studies of the effects of RRQs versus CQs and both conclude that there are reasons to doubt the efficacy of RRQs. After conducting their own study showing that the use of RRQs actually increased estimates of voter turnout to impossible levels, Holbrook and Krosnick (2010, p. 336) conclude that "… among the few studies that have compared RRT and direct self-report estimates of socially admirable attributes, none yielded consistent evidence that the RRT significantly reduced reported rates…This calls into question interpretations of all past RRT studies and raises serious questions about whether the RRT has practical value for increasing survey reporting accuracy."

Coutts and Jann (2011) used exactly the technique that we used in our study—forced-response, manual-coin toss RRQ—to examine six sensitive topics. They find that admission rates for RRQs are much lower than for CQs for not buying a ticket on public transport, shoplifting, marijuana use, DUI, and infidelity, while higher only for keeping extra change when too much was given in a transaction. They attribute their results on RRQs as reflecting the fact that a forced-yes response can feel like an admission of guilt. (Indeed, a yes response should mean that the Bayesian posterior probability of guilt is higher than the prior for anybody but the respondent, such as a judgmental interviewer.) Wolter and Preisendörfer (2013) also compared a CQ to a forced-response RRQ, questioning a sample of known convicted criminals on whether they had committed an offense. Whereas 100% of the sample were guilty, 57.5% acknowledged this in a CQ and 59.6% in an RRQ, a trivial increase in candour. As the qualitative study of Lensvelt-Mulders and Boeije (2007) shows, the forced-response of yes after a coin-toss is highly unpopular among respondents, thus suggesting the reason why RRQs might not produce their desired effect.

One can also address the issue analytically. Suppose that the world is such that reticence on the RRQ is less than on the CQ, that is the RRQ has some of the effect that its proponents hoped for. In terms of the parameters of our models, there are now two values of q, one for the CQ and one for the RRQs, and $q^{CQ} > q^{RRQ}$. One can then ask what the biases in our estimation procedure would be given that our procedure embodies the assumption that reticence is the same on the two types of questions. This is easy to answer analytically in one case, when there is a one-step CQ and $k = 1$. (This is equivalent to Model A with $k = 1$, since in that model respondents are always candid about visits). Suppose that our maximum likelihood estimate of average guilt is denoted g^e and the true value of average guilt is g^a.

Then, it is straightforward to show that $\frac{g^e}{(1+g^e)}$ consistently estimates $\frac{g^a}{(1+g^a)} \cdot \frac{(1-rq^{CQ})}{(1-rq^{RRQ})}$.[27] Thus our procedures underestimate the actual rate of guilt in this special case

When we turn to model B or instances where $k < 1$, or both, the analytics is not as straightforward. Thus, we use simulations for the analysis. A single simulation is as follows. We generate a data set of 10,000 observations using one of our models, e.g. model B, and parameter values that appear in Table 11 for a particular country for which that model is preferred, e.g. India. However, when we generate the observations we make one variation on the model, we assume reticence on the RRQ is less than on the CQ. That is, q^{CQ} is set at the value of q in Table 11, but $q^{RRQ} = 0.8 * q^{CQ}$. Then when we estimate the model we incorrectly assume that the simulated world is one where $q^{CQ} = q^{RRQ}$, that is, esimation is as described in Section 4.

The results of the simulations appear in the table immediately below. Because the results are so consistent, and consistent with the analytics for the simple case above, six simulations are sufficient, each one matching our preferred model for a country. In all cases, our procedures severely underestimate the true rates of effective corruption when the world is one in which $q^{CQ} = 0.8 * q^{RRQ}$ and estimation incorrectly imposes the assumption that $q^{CQ} = q^{RRQ}$. The degree of underestimation varies between 3 standard deviations (Peru) and 43 (Ukraine). Thus, our conclusion that standard estimates of corruption are significantly underestimated is robust to the criticism that we have incorrectly assumed that the RRQ has no affect in diminishing respondent reticence.

Model used in estimation	Country whose parameter values are taken from Table 11	True value of effective corruption in model simulation (from Table 11)	Maximum-likelihood estimate of effective corruption
B	Bangladesh	0.482	0.409
			(0.009)
B	India	0.309	0.199
			(0.008)
B	Nigeria	0.261	0.234
			(0.006)
B	Peru	0.073	0.061
			(0.004)
A	Sri Lanka	0.076	0.055
			(0.004)
A	Turkey	0.164	0.030
			(0.004)
A	Ukraine	0.310	0.095
			(0.005)

Note: standard errors of estimates in parentheses.

---

[27] The version of our model with a one-step question and $k = 1$ is analyzed in Kraay and Murrell (2013). This result follows transparently from the formulae for population moments in that paper.

Figure 1: Estimates of Effective Corruption, Adjusted for Sources of Bias Due to Reticence

■ Effective Corruption, Adjusted for Under-reporting of Both Visits and Bribes

■ Effective Corruption, Adjusted for Under-reporting of Bribes but Without Adjustment for Under-reporting of Visits

□ Effective Corruption Without Adjusting for Under-reporting of Visits and Bribes = Standard Estimate of Effective Corruption
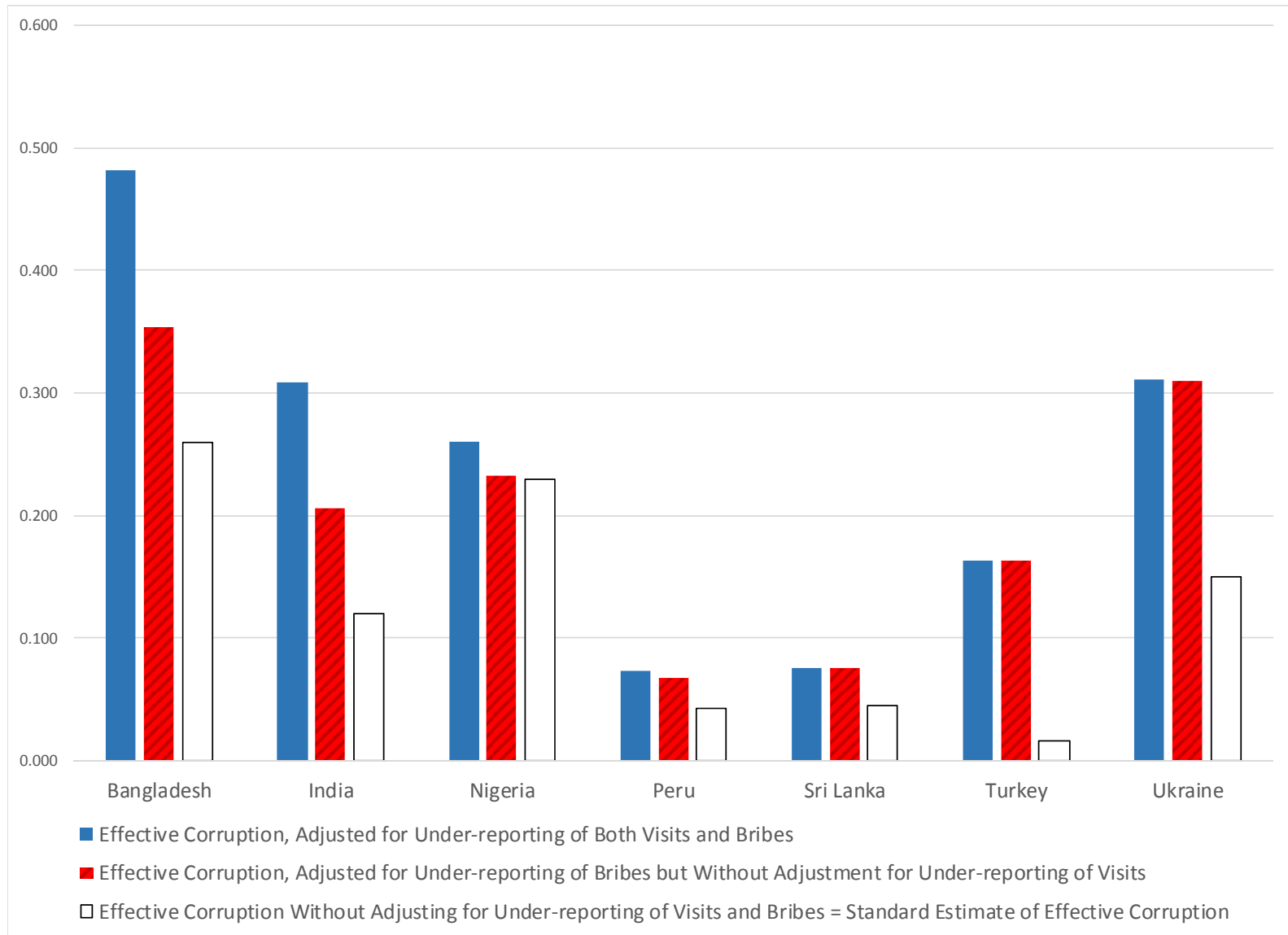
Figure 2: Standard and Reticence-Adjusted Rates of Effective Corruption
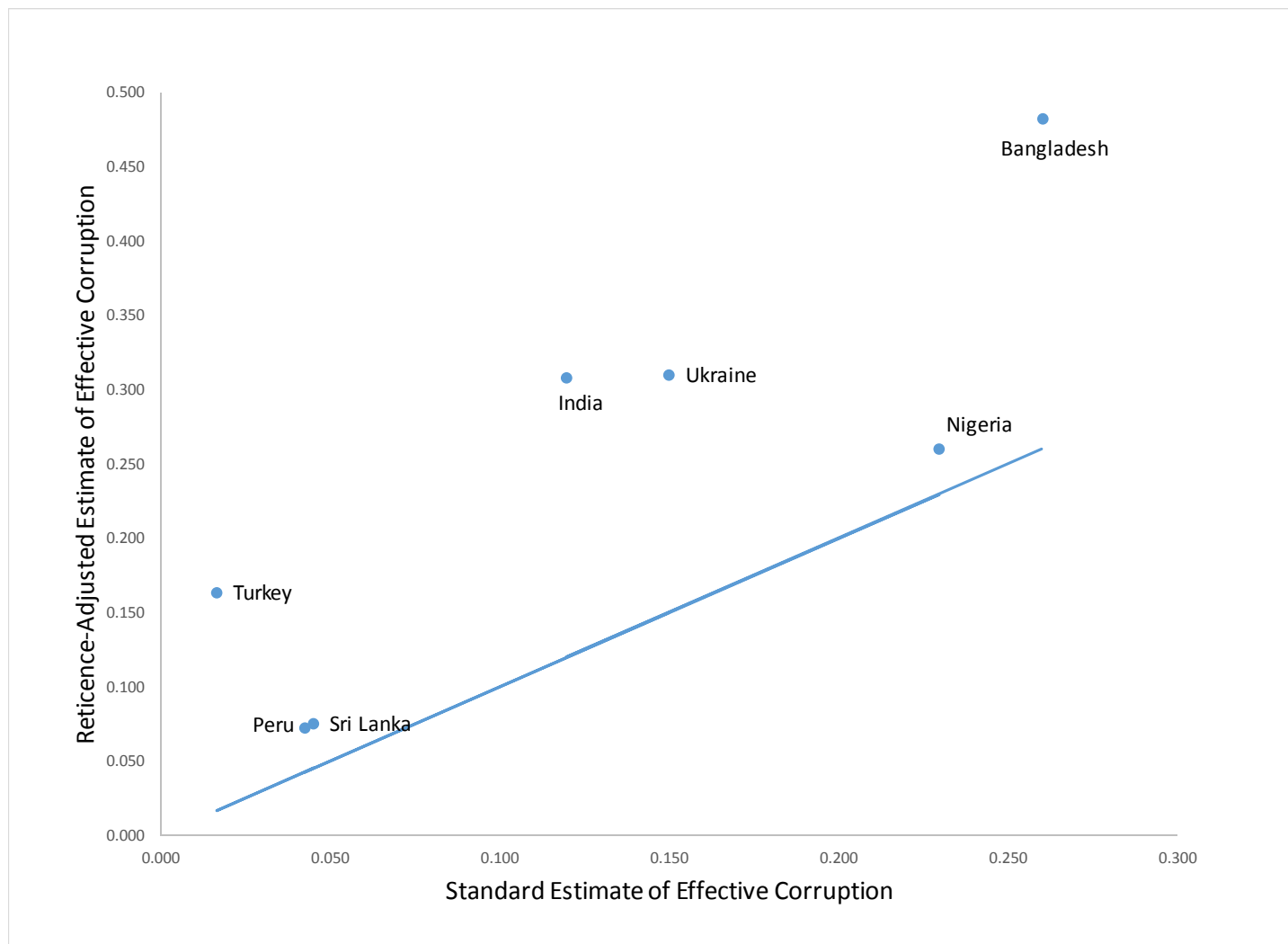
Table 1: Description of the Surveys

Country	Timing of interviews	Method of interviews	Number of observations used in the analysis
Bangladesh	from April 2013 through September 2013	face-to-face/PAPI	915
India	from June 2013 through December 2014	face-to-face/PAPI	4623
Nigeria	from January 2007 through December of 2007, and from June 2010 through December 2010	face-to-face/PAPI	5537
Peru	from April 2010 through April 2011	face-to-face/PAPI	590
Sri Lanka	from June 2011 through November 2011	face-to-face/PAPI	443
Turkey	from January 2013 through June 2014	face-to-face/mix of CAPI and PAPI	964
Ukraine	from January 2013 through November 2013	face-to-face/PAPI	467

Note: In Nigeria, firms located in different and non-overlapping regions were interviewed in 2007 and 2010 which is why the data from these different time-periods are combined.

Table 2: Summary Results from Random Response Questions

	Percentage of Respondents Answering "Yes"						
	Bangladesh	India	Nigeria	Peru	Sri Lanka	Turkey	Ukraine
Have you ever paid less in personal taxes than you should have under the law?	52	53.5	49.5	39.2	42.2	15.9	40.5
Have you ever paid less in business taxes than you should have under the law?	48.6	49.5	42.3	41.2	42.4	15	45
Have you ever made a misstatement on a job application?	34.3	13.4	41.4	36.1	39.3	14.5	37.9
Have you ever used the office telephone for personal businesses?	65.6	36.6	49.7	73.7	52.8	48.3	79
Have you ever inappropriately promoted an employee for personal reasons?	37.6	13.8	39.5	39	37.9	25.1	48.2
Have you ever deliberately not given your suppliers or clients what was due?	35.3	10.1	36	36.6	39.7	16.9	41.1
Have you ever lied in your self-interest?	58.4	28.4	50.9	53.2	49	27.2	70.5
Have you ever inappropriately hired a staff member for personal reasons?	47.2	16	39.9	39.5	39.5	25.2	48.6
Have you ever been purposely late for work?	47.5	21.7	47.2	54.7	47.6	30.4	43.9
Have you ever unfairly dismissed an employee for personal reasons?	43.2	9.7	35.1	32	33	17.1	39.2
Number of responses to questions in bold	915	4623	5537	590	443	964	467

Actual Behaviour (Unobserved)	Probability of Observed Responses on Two-Stage Corruption Question		
	No Visit ($D = 1$)	Visit, No Bribe ($D = 2$)	Visit, Bribe ($D = 3$)
Panel 1: Candid Respondents (Both Models)			
No Visit: $(1 - w)$	1	0	0
Visit, No Bribe: $w(1 - kg)$	0	1	0
Visit, Bribe: wkg	0	0	1
Probability of Observing Response:	$P[D = 1\|C] = (1 - w)$	$P[D = 2\|C] = w(1 - kg)$	$P[D = 3\|C] = wkg$
Panel 2: Reticent Respondents, Model A			
No Visit: $(1 - w)$	1	0	0
Visit, No Bribe: $w(1 - g)$	0	1	0
Visit, Bribe: wg	0	q	$1 - q$
Probability of Observing Response:	$P_A[D = 1\|R] = (1 - w)$	$P_A[D = 2\|R] = w(1 - g + qg)$	$P_A[D = 3\|R] = wg(1 - q)$
Panel 3: Reticent Respondents, Model B			
No Visit: $(1 - w)$	1	0	0
Visit, No Bribe: $w(1 - g)$	0	1	0
Visit, Bribe: wg	q	0	$1 - q$
Probability of Observing Response:	$P_B[D = 1\|R] = 1 - w + wgq$	$P_B[D = 2\|R] = w(1 - g)$	$P_B[D = 3\|R] = wg(1 - q)$

Table 3: Modelling How Reticence Affects Responses to the Conventional Question

Notes: This table summarizes the probability of observing the three different response possibilities to the two-step corruption question, for candid and reticent respondents, and for the two different models of reticent behaviour that we consider. The first column reports the possible unobserved outcomes, i.e. whether a visit occurred, and if a visit occurred whether a bribe was requested, together with the corresponding probabilities of the three events. The remaining three columns report the observed responses for candid and reticent respondents, for the two different models discussed in the text. Note that q is the probability of reticent behaviour on the two-step corruption question.

Table 4: Estimates of Reticence and Guilt from the Enterprise Survey in Bangladesh

	Standard	Model A	Model B
Prob. of interaction with tax official (w) = Pr(inspected)	0.619*** (21.49)	See note below	0.843*** (20.04)
Reticence (r)	0 .		0.952*** (21.42)
Probability reticent person answers question reticently (q)	0 .		0.476*** (21.25)
Effective reticence (rq)	0 .		0.453*** (25.41)
Guilt rate of the reticent (g)	0.420*** (17.26)		0.587*** (20.02)
Reduction in Guilt for the Candid (k)	1 .		0.468** (3.20)
Average guilt = Pr(request | inspected)	0.420*** (17.26)		0.572*** (19.65)
Effective corruption $= Pr$(inspected & request)	0.260*** (13.91)		0.482*** (11.83)
Log likelihood			-2625.048
Number of observations	915	915	915
Number of clusters	100	100	100

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

Note that z-statistics for estimates of *k* are for the null hypothesis of *k = 1*

The maximum-likelihood procedure failed to converge for model A.

Table 5: Estimates of Reticence and Guilt from the Enterprise Survey in India

	Standard	Model A	Model B
Prob. of interaction with tax official (w) = Pr(inspected)	0.429*** (23.01)	0.429*** (23.01)	0.642*** (19.50)
Reticence (r)	0 .	0.803*** (46.79)	0.815*** (51.50)
Probability reticent person answers question reticently (q)	0 .	0.835*** (81.91)	0.784*** (65.29)
Effective reticence (rq)	0 .	0.671*** (58.92)	0.639*** (54.80)
Guilt rate of the reticent (g)	0.279*** (13.71)	1.000 .	0.545*** (20.15)
Reduction in Guilt for the Candid (k)	1 .	0.182*** (38.03)	0.361*** (18.14)
Average guilt = $Pr(\text{request} \mid \text{inspected})$	0.279*** (13.71)	0.839*** (49.74)	0.481*** (18.53)
Effective corruption $= Pr(\text{inspected \& request})$	0.119*** (10.69)	0.360*** (21.23)	0.309*** (10.63)
Log likelihood		-12675.054	-12668.046
Number of observations	4623	4623	4623
Number of clusters	634	634	634

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

Note that z-statistics for estimates of *k* are for the null hypothesis of *k = 1*

Table 6: Estimates of Reticence and Guilt from the Enterprise Survey in Nigeria

	Standard	Model A	Model B
Prob. of interaction with tax official (w) = Pr(inspected)	0.811*** (154.37)	0.811*** (154.37)	0.905*** (164.04)
Reticence (r)	0 .	0.429*** (32.37)	0.402*** (31.37)
Probability reticent person answers question reticently (q)	0 .	0.805*** (125.72)	0.738*** (75.41)
Effective reticence (rq)	0 .	0.345*** (39.19)	0.297*** (41.72)
Guilt rate of the reticent (g)	0.283*** (42.15)	1.000 .	0.420*** (30.57)
Reduction in Guilt for the Candid (k)	1 .	0.205*** (134.33)	0.474*** (23.92)
Average guilt = $Pr(\text{request} \mid \text{inspected})$	0.283*** (42.15)	0.546*** (43.78)	0.288*** (40.20)
Effective corruption $= Pr(\text{inspected \& request})$	0.230*** (40.66)	0.443*** (42.12)	0.261*** (35.82)
Log likelihood		-16271.138	-16241.511
Number of observations	5537	5537	5537

Notes: the z-statistics that are reported in parentheses are based on the usual ML standard errors and not on heteroskedasticity-consistent standard errors clustered at the strata level because the sampling information included in the data is incomplete. * p<0.05 ** p<0.01 *** p<0.001

Note that z-statistics for estimates of *k* are for the null hypothesis of *k = 1*

Table 7: Estimates of Reticence and Guilt from the Enterprise Survey in Peru

	Standard	Model A	Model B
Prob. of interaction with tax official (w) = Pr(inspected)	0.551*** (20.18)	0.551*** (20.18)	0.590*** (17.81)
Reticence (r)	0 .	0.381*** (7.67)	0.367*** (7.91)
Probability reticent person answers question reticently (q)	0 .	0.754*** (13.74)	0.770*** (17.53)
Effective reticence (rq)	0 .	0.287*** (9.80)	0.282*** (9.97)
Guilt rate of the reticent (g)	0.0769*** (4.67)	0.212 (1.53)	0.231** (2.65)
Reduction in Guilt for the Candid (k)	1 .	0.308** (2.74)	0.266*** (5.58)
Average guilt = Pr(request | inspected)	0.0769*** (4.67)	0.121* (2.49)	0.124*** (3.73)
Effective corruption = Pr(inspected & request)	0.0424*** (4.40)	0.0669* (2.49)	0.0729*** (3.34)
Log likelihood		-1619.5222	-1616.9835
Number of observations	590	590	590
Number of clusters	73	73	73

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

Note that z-statistics for estimates of *k* are for the null hypothesis of *k = 1*

Table 8: Estimates of Reticence and Guilt from the Enterprise Survey in Sri Lanka

	Standard	Model A	Model B
Prob. of interaction with tax official (w) = Pr(inspected)	0.537*** (14.33)	0.537*** (14.33)	0.562*** (12.29)
Reticence (r)	0 .	0.480** (3.02)	0.519*** (3.69)
Probability reticent person answers question reticently (q)	0 .	0.574*** (3.49)	0.523*** (4.66)
Effective reticence (rq)	0 .	0.276*** (10.21)	0.271*** (10.04)
Guilt rate of the reticent (g)	0.0840*** (5.02)	0.244 (1.24)	0.168* (2.19)
Reduction in Guilt for the Candid (k)	1 .	0.193*** (3.39)	0.324** (2.60)
Average guilt = $Pr(\text{request} \mid \text{inspected})$	0.0840*** (5.02)	0.142* (2.33)	0.113*** (4.00)
Effective corruption $= Pr(\text{inspected \& request})$	0.0451*** (4.34)	0.0760* (2.12)	0.0637*** (3.31)
Log likelihood		-1175.5814	-1176.4097
Number of observations	443	443	443
Number of clusters	69	69	69

Notes: z-statistics based on heteroskedasticity-consistent standard errors clustered at the strata level are reported in parentheses * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Note that z-statistics for estimates of *k* are for the null hypothesis of *k = 1*

Table 9: Estimates of Reticence and Guilt from the Enterprise Survey in Turkey

	Standard	Model A	Model B
Prob. of interaction with tax official (w) = Pr(inspected)	0.609*** (26.27)	0.609*** (26.27)	0.662*** (21.77)
Reticence (r)	0 .	0.680*** (28.85)	0.684*** (29.04)
Probability reticent person answers question reticently (q)	0 .	0.933*** (68.31)	0.914*** (84.26)
Effective reticence (rq)	0 .	0.635*** (27.24)	0.625*** (27.46)
Guilt rate of the reticent (g)	0.0273*** (4.00)	0.387* (2.11)	0.134*** (3.86)
Reduction in Guilt for the Candid (k)	1 .	0.0480*** (29.82)	0.153*** (10.77)
Average guilt = Pr(request | inspected)	0.0273*** (4.00)	0.269* (2.17)	0.0983*** (4.06)
Effective corruption $= Pr$(inspected & request)	0.0166*** (4.00)	0.164* (2.18)	0.0650*** (3.65)
Log likelihood		-2198.3828	-2202.2151
Number of observations	964	964	964
Number of clusters	174	174	174

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

Note that z-statistics for estimates of *k* are for the null hypothesis of *k = 1*

Table 10: Estimates of Reticence and Guilt from the Enterprise Survey in Ukraine

	Standard	Model A	Model B
Prob. of interaction with tax official (w) = Pr(inspected)	0.570*** (22.88)	0.570*** (22.88)	0.605*** (18.36)
Reticence (r)	0 .	0.409*** (6.48)	0.408*** (5.32)
Probability reticent person answers question reticently (q)	0 .	0.813*** (25.08)	0.689*** (13.17)
Effective reticence (rq)	0 .	0.332*** (8.23)	0.281*** (7.79)
Guilt rate of the reticent (g)	0.263*** (9.26)	1.000*** (2726.63)	0.250*** (3.54)
Reduction in Guilt for the Candid (k)	1 .	0.230*** (28.47)	0.944 (0.26)
Average guilt = $Pr(\text{request} \mid \text{inspected})$	0.263*** (9.26)	0.544*** (8.94)	0.241*** (5.60)
Effective corruption $= Pr(\text{inspected \& request})$	0.150*** (9.02)	0.310*** (7.69)	0.146*** (4.65)
Log likelihood		-1411.875	-1420.0272
Number of observations	467	467	467
Number of clusters	94	94	94

Notes: z-statistics based on heteroskedasticity-consistent standard errors clustered at the strata level are reported in parentheses * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Note that z-statistics for estimates of *k* are for the null hypothesis of *k = 1*

Table 11: Summary of Estimates from Preferred Model for Each Country

	Bangladesh Model B	India Model B	Nigeria Model B	Peru Model B	Sri Lanka Model A	Turkey Model A	Ukraine Model A
Panel A: Estimated Parameters							
Probability of visit (w)	0.843	0.642	0.905	0.590	0.537	0.609	0.570
Reticence (r)	0.952	0.815	0.402	0.367	0.480	0.680	0.409
Probability of reticent behaviour (q)	0.476	0.784	0.738	0.770	0.574	0.933	0.813
Effective reticence (rq)	0.453	0.639	0.297	0.283	0.276	0.634	0.333
Guilt rate of the reticent (g)	0.587	0.545	0.420	0.231	0.244	0.387	1.000
Reduction in guilt for the candid (k)	0.468	0.361	0.474	0.266	0.193	0.048	0.230
Panel B: Conventional Versus Reticence-Adjusted Estimates of:							
(1) Probability of a Visit							
Conventional	0.619	0.429	0.811	0.551	0.537	0.609	0.570
Reticence-Adjusted	0.843	0.642	0.905	0.590	0.537	0.609	0.570
Ratio	0.734	0.668	0.896	0.934	1.000	1.000	1.000
(2) Probability of Bribery Conditional on Visit							
Conventional	0.420	0.279	0.283	0.077	0.084	0.027	0.263
Reticence-Adjusted	0.572	0.481	0.288	0.124	0.142	0.269	0.545
Ratio	0.734	0.581	0.983	0.623	0.593	0.100	0.483
(3) Effective Corruption							
Conventional	0.260	0.120	0.230	0.042	0.045	0.016	0.150
Reticence-Adjusted	0.482	0.309	0.261	0.073	0.076	0.164	0.311
Ratio	0.539	0.388	0.881	0.581	0.593	0.100	0.483
z-statistic for test of null hypothesis that right-hand side of equation (7) equals 1	(25.11)	(56.07)	(42.33)	(5.51)	(2.57)	(37.31)	(30.48)

Notes: z-statistics based on heteroskedasticity-consistent standard errors clustered at the strata level are reported in the last row of the table.

Table A.1: Sample attrition due to Don't Know's, Refusals, and Interviewer Effects

	Total Number of Firms Surveyed*	DK or R to the preliminary part of CQ	DK or R to the follow-up part of the CQ	DK or R to the set of RRQs	answers to both CQ and RRQ available	observations deleted due to interviewer effects	sample size used in the estimates	percentage of firms surveyed that is used in the estimates
Bangladesh	1,442	14	74	399	982	67	915	63.5%
India	9,281	64	411	1,752	7,206	2,583	4,623	49.8%
Peru	1,000	2	6	215	778	188	590	59.0%
Sri Lanka	588	7	9	127	448	5	443	75.3%
Nigeria	5,544	0	2	5	5,537	0	5,537	99.9%
Turkey	1,344	25	7	361	964	0	964	71.7%
Ukraine	1,002	12	140	458	473	6	467	46.6%

* Includes only the firms that were asked both sets of questions – CQs and RRQs – used in the analysis

Note: Some respondents answer DK/R to more than one of the questions and therefore the sum of columns 3, 4, 5, and 6 can be greater than the number in column 2.