Online Appendix for "Lying About Corruption in Surveys: Evidence from a Joint Response Model"

Appendix A: Survey Methodology

The household survey consisted of face-to-face interviews of 4200 residents (18 year olds and older) of the Gran Area Metropolitana (GAM), which includes 30 cantons in the provinces of Alajuela, Cartago, Heredia, and San José. The GAM is the principal urban center in Costa Rica. It contains approximately 2.6 million residents and accounts for 60% of the country's entire population. The survey was administered by *Borge y Asociados*, the most prominent survey research firm in Central America, between October 2013 and April 2014. The survey was preceded by a pilot consisting of 48 cases, administered in October 15 and 16. The goal of the pilot was for enumerators to familiarize themselves with the questionnaire in the field, and to test their skills in administering the questionnaire, especially the crosswise questions. On average, the interviews lasted 25 minutes.

A two-stage clustered random sample based on the 2000 national census was generated (with fixed proportions for age and gender). Three hundred and fifty primary sampling units (PSUs), the smallest geographic unit in the census, were selected from the total contained within the GAM, with twelve interviews conducted in each PSU. Interviewers began from the northernmost point of the PSU and proceeded in a clockwise direction. Within each household, interviewers were selected based on quotas by gender and age, so that half of the surveys were obtained from each gender, and one third fall into each of the categories of 18-28 years old, 29-42 years old, and 43 or more years old. In cases of refusal or when no one responded, the household was replaced with the adjacent household. All survey enumerators utilized PDAs (personal digital assistants) to conduct the survey.

Survey enumerators were recruited by *Borge y Asociados* and were mostly experienced with the administration of surveys. They went through extensive training on the details and administration of the survey instrument, especially on the execution of the crosswise questions. The training for the crosswise component of the survey consisted of a thorough explanation of the logic and functioning of the technique, as well as live practice sessions in which each enumerator practiced her delivery of this section of the survey both in front of members of the

research team and administrators from *Borge y Asociados*. By contract, only enumerators that had gone through these training sessions participated in the administration of the survey. Any potential enumerator demonstrating insufficient mastery in the delivery of this component of the survey–the most challenging– in the training sessions was removed from the team of enumerators. An important feature of the delivery of this component of the survey consisted of a script describing to respondents how a hypothetical individual with a particular value on a sensitive item and a mother born in a particular month would respond to a given crosswise item. This script was given to all respondents prior to the commencement of the sensitive questions of interest.

For the purpose of survey verification, enumerators recorded the first name only and phone number of each respondent. Verification was conducted on a randomly selected subgroup of the sample (30% percent of the total) by phone, after which this information was destroyed. Team leaders also conducted verification in the field by randomly selecting households for verification the same day that the interview was conducted. If mistakes were found using either method, interviews were replaced by new ones. The contact rate for the survey was 87 percent, the response rate was 29 percent, the cooperation rate 39 percent, and the refusal rate 44 percent (Rates calculated according to the American Association of Public Opinion Research).

Focus Groups

Focus groups were conducted in San José with residents of varied backgrounds on August 6, 7, and 8, 2013, before fielding the household survey. One of the main goals of these focus groups was to evaluate each group's understanding of the logic of the crosswise questions.

Phone Survey

Prior to conducting the household survey and the focus group sessions, a nationally representative telephone survey of 1200 Costa Rican residents (older than 18) was conducted by *Borge y Asociados* between July 15 and July 20, 2013. The goal of this survey was twofold. First, we used the survey to evaluate our questions, the questions' wording, and the order of questions for the household survey. Second, we wanted to collect information about respondents' recollections of their parents' birthdays in order to be able to use that information for the crosswise questions in the household survey. We did so by asking respondents directly about the day of birth of their mother and father in the telephone survey. To check the veracity of these

self-reports, these were checked against statistical tables produced by Costa Rica's National Institute for Statistics and Censuses (INEC) on month of birth for newborns for the 2000-2011 period (the period for which the data was available). Since there should be no systematic differences in month and day of birth across sex of child, responses for mothers and fathers were pooled together. The comparison in Table A1 shows that self-reported parent's birthdays were almost identical to the actual information obtained from INEC.

Table A1: Proportion of births falling into indicated months, telephone self-reports vs. census data **(**births occurring in October, November, or December)

Actual proportion of newborn births occurring in indicated months (INEC)

2000	0.268
2001	0.262
2002	0.268
2003	0.260
2004	0.264
2005	0.265
2006	0.268
2007	0.270
2008	0.266
2009	0.261
2010	0.262
2011	0.265
avg**.**	**0.265**

Proportion of mother's and father's birthdays occurring in indicated months according to telephone survey self-reports: **0.264**

Enumerator Scripts

Main outcome variable - Crosswise Format

Interviewers explained first the logic of the crosswise format with the following script:
"Now I am going to ask you a series of questions with a special format. These are questions especially design to protect the privacy of your answers. To be able to answer them you will have to remember (and do not tell me) the birthday of some of your relatives. The technique is based precisely on the fact that neither me, nor any of the people involved on the survey know the birthday of your relatives. This is what guarantees that we cannot not know exactly what your answer was. "Let me show you an example. . . " (Ahora le voy a hacer una serie de preguntas que tienen un formato especial. Son preguntas diseñadas especialmente para proteger la privacidad de sus respuestas. Para poder contestarlas usted va a tener que recordar (y no decirme) el día del cumpleaños de algunos de sus parientes. La técnica se basa justamente en que

ni yo, ni ninguna de las personas involucradas en la encuesta, conocen el cumpleaños de sus parientes. Esto es lo garantiza que no podamos saber cual fue exactamente su respuesta. Le muestro con un ejemplo. . .). Interviewers then showed respondents the example card presented in Figure A1.

Figure A1: The Practice Crosswise survey question

How many of the following statements are true?
(*¿Cuántas de las siguientes afirmaciones son ciertas?*)

- My mother was born in OCTOBER, NOVEMBER, or DECEMBER. (*El cumpleaños de mi __madre__ es en octubre, noviembre o diciembre*)

- In order to avoid a family conflict, I would be willing to tell a lie. (*Para evitar un conflicto familiar, estaría dispuesto/a a decir una mentira*)

Mark the appropriate answer
(*Indique la respuesta correcta*):

A. __BOTH OR NEITHER__ of the statements are true. (__*LAS DOS O NINGUNA*__ *de las afirmaciones son ciertas*)
B. __ONLY ONE__ of the statements is true. (__*UNA SOLA*__ *de las afirmaciones es cierta*)

Remember: No one from the research team knows your mother's birthday. This guaranties the anonymity of your response. (*Recuerde*: *Nadie en el grupo de investigación sabe cuándo es el cumpleaños de su pariente. Esto garantiza la privacidad de su respuesta*)

Continuation of the script:

"In my case, my mother's birthday is in the month of December and I WOULD be willing to tell a lie to avoid a family conflict. So, my answer to the question: 'How many of the following statements are true?' is "A" ("Both or neither of the statements are true"). Now let's suppose that my mother's birthday was in January and I've already told you that I would be willing to tell a lie, so my answer in this case would be "B" ("Only one of the statements is true"). Finally, if my mother's birthday was in January and I would NOT be willing to tell a lie, then my answer would be "A" because neither of the statements would be true. Since nobody knows the date of my mother's birthday, it is not possible to identify my answer to the specific statement about lying. Did I explain myself clearly? Would you like me to repeat the example? (En mi caso, mi madre cumple años en el mes de diciembre y yo SI estaría dispuesto a decir una mentira para evitar un conflicto familiar. Por lo tanto, mi respuesta a la pregunta "¿Cuántas de las siguientes afirmaciones son ciertas?" es la "A" ("Las dos o ninguna de las dos afirmaciones son ciertas"). Ahora supongamos que mi madre cumpliese años en enero, y ya le dije que yo estaría dispuesto a decir una mentira, entonces mi respuesta sería la "B" ("Una sola de las afirmaciones es cierta"). Por último, si mi madre cumpliese años en enero y yo NO estuviese dispuesto a decir una

mentira, mi respuesta sería la "A" porque ninguna de las afirmaciones es cierta. Como nadie sabe cuándo es el cumpleaños de mi madre, no es posible saber realmente cual es mi respuesta a la pregunta sobre mentiras. ¿Me explico? ¿Le gustaría que le repita el ejemplo?)

Enumerators were instructed to explain the technique and repeat the example as many times as was necessary for the respondents to understand the technique. Once this was achieved, enumerators handed out the cards with the questions we care about.

Main outcome variable - Direct Questioning Format

Interviewers explained first why we were asking the same question twice with the following script: "I've just asked you a series of questions about topics that were a little sensitive by using a technique that protects the privacy of the responses. Thanks to that technique, as I was explaining before, there is no way for us to identify your precise answer to those questions. However, we know that not everyone thinks that these topics are especially sensitive. Thus, in finishing with the survey we would like to ask you directly about these same topics. Of course, if you prefer not to answer any of these questions, please just let me know. For each of these questions, please tell me if the statement is true, false, or if you would rather not answer." (Hace un rato le hice una serie de preguntas sobre temas un poco sensibles utilizando una técnica que protege la privacidad de las respuestas. Gracias a esa técnica, como le explicaba antes, no tenemos forma de saber exactamente qué es lo que Ud. nos contestó. Sin embargo, sabemos que no todo el mundo considera esos temas tan sensibles así que para finalizar la encuesta nos gustaría preguntarle nuevamente en forma directa sobre esos mismos temas. Por supuesto, si Ud. prefiere no contestar a alguna de estas preguntas, simplemente me dice. En cada caso, dígame por favor si la afirmación es verdadera, falsa o prefiere no responder.)

After the enumerators provided this explanation they asked respondents: "In order to avoid paying a traffic ticket, I would be willing to pay a bribe to a police officer," and "*I have paid, at least once, a bribe to a police officer to avoid a traffic ticket*" (Para evitar pagar una multa de tránsito, estaría dispuesto/a a pagar un soborno a un policía; He pagado, al menos una vez, un soborno a un policía para evitar pagar una multa de tránsito") Response options were: "True", "False", and "I prefer not to respond" (Prefiero no contestar).

Online Appendix B: Additional Tables and Robustness Checks

Table A2: Parameter estimates for questions about corruption (across estimation strategies), whole sample

Q1: To avoid paying a traffic ticket, I would be willing to pay a bribe to a police officer

Prevalence estimate ($\widehat{\pi}$)			Diagnostic parameters (joint response model)		
Direct only	0.18	[0.17, 0.19]	$\widehat{\lambda}^{T}_{1}$	0.61	[0.55, 0.69]
SST only	0.22	[0.18, 0.25]	$\widehat{\lambda}^{L}_{1}$	0.35	[0.28, 0.41]
Joint response	0.29	[0.26, 0.32]	$\widehat{\lambda}^{T}_{0}$	0.97	[0.96, 0.98]

Q2: I have paid, at least once, a bribe to a police officer to avoid a traffic ticket

Prevalence estimate ($\widehat{\pi}$)			Diagnostic parameters (joint response model)		
Direct only	0.09	[0.08, 0.10]	$\widehat{\lambda}^{T}_{1}$	0.54	[0.45, 0.65]
SST only	0.13	[0.10, 0.16]	$\widehat{\lambda}^{L}_{1}$	0.44	[0.32, 0.53]
Joint response	0.16	[0.13, 0.19]	$\widehat{\lambda}^{T}_{0}$	0.98	[0.97, 0.99]

Note: Ninety-five percent confidence intervals in square brackets.

Table A3: Parameter estimates for questions about corruption, by gender

Q1: To avoid paying a traffic ticket, I would be willing to pay a bribe to a police officer
Men (N=2096)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.24	[0.22, 0.26]	$\hat{\lambda}^{T}_{1}$	0.63	[0.56, 0.71]
SST only	0.28	[0.24, 0.33]	$\hat{\lambda}^{L}_{1}$	0.35	[0.27, 0.41]
Joint response	0.37	[0.33, 0.41]	$\hat{\lambda}^{T}_{0}$	0.96	[0.94, 0.98]

Women (N=2097)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.12	[0.11, 0.14]	$\hat{\lambda}^{T}_{1}$	0.58	[0.48, 0.72]
SST only	0.15	[0.10, 0.19]	$\hat{\lambda}^{L}_{1}$	0.35	[0.20, 0.46]
Joint response	0.20	[0.16, 0.25]	$\hat{\lambda}^{T}_{0}$	0.98	[0.97, 0.99]

Q2: I have paid, at least once, a bribe to a police officer to avoid a traffic ticket
Men (N=2096)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.14	[0,12, 0.15]	$\hat{\lambda}^{T}_{1}$	0.56	[0.47, 0.66]
SST only	0.20	[0.16, 0.24]	$\hat{\lambda}^{L}_{1}$	0.42	[0.32, 0.50]
Joint response	0.24	[0.20, 0.28]	$\hat{\lambda}^{T}_{0}$	0.98	[0.97, 0.99]

Women (N=2091)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.04	[0.03, 0.05]	$\hat{\lambda}^{T}_{1}$	0.48	[0.31, 0.90]
SST only	0.07	[0.03, 0.11]	$\hat{\lambda}^{L}_{1}$	0.51	[0.06, 0.67]
Joint response	0.08	[0.05, 0.12]	$\hat{\lambda}^{T}_{0}$	0.98	[0.97, 0.99]

Note: Ninety-five percent confidence intervals in square brackets.

Table A4: Parameter estimates for questions about corruption, by age

Q1: To avoid paying a traffic ticket, I would be willing to pay a bribe to a police officer
Less than 28 (N=1295)

Prevalence estimate ($\widehat{\pi}$)			Diagnostic parameters (joint response model)		
Direct only	0.26	[0.24, 0.28]	$\hat{\lambda}^{T}_{1}$	0.65	[0.58, 0.74]
SST only	0.30	[0.24, 0.36]	$\hat{\lambda}^{L}_{1}$	0.32	[0.22, 0.40]
Joint response	0.39	[0.33, 0.44]	$\hat{\lambda}^{T}_{0}$	0.97	[0.95, 0.99]

29-42 (N=1463)

Prevalence estimate ($\widehat{\pi}$)			Diagnostic parameters (joint response model)		
Direct only	0.19	[0.17, 0.21]	$\hat{\lambda}^{T}_{1}$	0.60	[0.52, 0.71]
SST only	0.22	[0.17, 0.27]	$\hat{\lambda}^{L}_{1}$	0.36	[0.24, 0.45]
Joint response	0.31	[0.26, 0.36]	$\hat{\lambda}^{T}_{0}$	0.97	[0.95, 0.99]

43 and more (N=1434)

Prevalence estimate ($\widehat{\pi}$)			Diagnostic parameters (joint response model)		
Direct only	0.10	[0.09, 0.12]	$\hat{\lambda}^{T}_{1}$	0.55	[0.42, 0.76]
SST only	0.14	[0.09, 0.19]	$\hat{\lambda}^{L}_{1}$	0.41	[0.18, 0.54]
Joint response	0.18	[0.13, 0.23]	$\hat{\lambda}^{T}_{0}$	0.97	[0.96, 0.98]

Q2: I have paid, at least once, a bribe to a police officer to avoid a traffic ticket
Less than 28 (N=1296)

Prevalence estimate ($\widehat{\pi}$)			Diagnostic parameters (joint response model)		
Direct only	0.08	[0.06, 0.09]	$\hat{\lambda}^{T}_{1}$	0.44	[0.33, 0.64]
SST only	0.14	[0.09, 0.19]	$\hat{\lambda}^{L}_{1}$	0.55	[0.35, 0.66]
Joint response	0.17	[0.12, 0.23]	$\hat{\lambda}^{T}_{0}$	0.98	[0.97, 0.99]

29-42 (N=1463)

Prevalence estimate ($\widehat{\pi}$)			Diagnostic parameters (joint response model)		
Direct only	0.11	[0.10, 0.13]	$\hat{\lambda}^{T}_{1}$	0.54	[0.43, 0.69]
SST only	0.17	[0.12, 0.23]	$\hat{\lambda}^{L}_{1}$	0.40	[0.23, 0.52]
Joint response	0.21	[0.16, 0.25]	$\hat{\lambda}^{T}_{0}$	0.99	[0.98, 1]

43 and more (N=1427)

Prevalence estimate ($\widehat{\pi}$)			Diagnostic parameters (joint response model)		
Direct only	0.07	[0.06, 0.09]	$\hat{\lambda}^{T}_{1}$	0.65	[0.45, 0.98]
SST only	0.09	[0.04, 0.14]	$\hat{\lambda}^{L}_{1}$	0.35	[0.00, 0.54]
Joint response	0.11	[0.07, 0.16]	$\hat{\lambda}^{T}_{0}$	0.97	[0., 0.99]

Note: Ninety-five percent confidence intervals in square brackets.

Table A5: Parameter estimates for questions about corruption, by education

Q1: To avoid paying a traffic ticket, I would be willing to pay a bribe to a police officer
Some university education (N=731)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.16	[0.13, 0.18]	$\hat{\lambda}^T_1$	0.67	[0.51, 0.92]
SST only	0.21	[0.14, 0.29]	$\hat{\lambda}^L_1$	0.29	[0.02, 0.46]
Joint response	0.22	[0.16, 0.29]	$\hat{\lambda}^T_0$	0.96	[0.94, 0.98]

Secondary completed and/or some technical education (N=1140)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.21	[0.18, 0.23]	$\hat{\lambda}^T_1$	0.67	[0.57, 0.81]
SST only	0.20	[0.14, 0.26]	$\hat{\lambda}^L_1$	0.29	[0.14, 0.39]
Joint response	0.30	[0.24, 0.36]	$\hat{\lambda}^T_0$	0.98	[0.97, 1.00]

Unfinished secondary education or less (N=2322)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.18	[0.16, 0.19]	$\hat{\lambda}^T_1$	0.57	[0.50, 0.65]
SST only	0.22	[0.18, 0.27]	$\hat{\lambda}^L_1$	0.40	[0.31, 0.47]
Joint response	0.30	[0.26, 0.34]	$\hat{\lambda}^T_0$	0.96	[0.95, 0.98]

Q2: I have paid, at least once, a bribe to a police officer to avoid a traffic ticket
Some university education (N=732)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.09	[0.07, 0.11]	$\hat{\lambda}^T_1$	0.75	[0.47, 1]
SST only	0.08	[0.01, 0.15]	$\hat{\lambda}^L_1$	0.25	[0.00, 0.52]
Joint response	0.12	[0.08, 0.19]	$\hat{\lambda}^T_0$	0.97	[0.96, 0.99]

Secondary completed and/or some technical education (N=1139)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.12	[0.10, 0.14]	$\hat{\lambda}^T_1$	0.62	[0.48, 0.85]
SST only	0.15	[0.10, 0.21]	$\hat{\lambda}^L_1$	0.37	[0.14, 0.51]
Joint response	0.19	[0.14, 0.25]	$\hat{\lambda}^T_0$	0.98	[0.97, 0.99]

Unfinished secondary education or less (N=2316)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.07	[0.06, 0.08]	$\hat{\lambda}^T_1$	0.43	[0.34, 0.58]
SST only	0.14	[0.10, 0.18]	$\hat{\lambda}^L_1$	0.53	[0.38, 0.62]
Joint response	0.16	[0.12, 0.20]	$\hat{\lambda}^T_0$	0.98	[0.97, 0.99]

Note: Ninety-five percent confidence intervals in square brackets.

Table A6: Parameter estimates for questions about corruption, by wealth

Q1: To avoid paying a traffic ticket, I would be willing to pay a bribe to a police officer

Low material wealth (N=1535)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.15	[0.13, 0.17]	$\hat{\lambda}^{T}_{1}$	0.51	[0.43, 0.61]
SST only	0.23	[0.17, 0.28]	$\hat{\lambda}^{L}_{1}$	0.45	[0.34, 0.53]
Joint response	0.28	[0.23, 0.32]	$\hat{\lambda}^{T}_{0}$	0.96	[0.95, 0.98]

Moderate material wealth (N=1286)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.18	[0.16, 0.20]	$\hat{\lambda}^{T}_{1}$	0.55	[0.47, 0.66]
SST only	0.23	[0.18, 0.29]	$\hat{\lambda}^{L}_{1}$	0.41	[0.30, 0.50]
Joint response	0.32	[0.27, 0.37]	$\hat{\lambda}^{T}_{0}$	0.96	[0.94, 0.98]

High material wealth (N=1372)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.22	[0.20, 0.24]	$\hat{\lambda}^{T}_{1}$	0.80	[0.67, 0.94]
SST only	0.19	[0.14, 0.24]	$\hat{\lambda}^{L}_{1}$	0.17	[0.00, 0.30]
Joint response	0.27	[0.22, 0.32]	$\hat{\lambda}^{T}_{0}$	0.98	[0.97, 1]

Q2: I have paid, at least once, a bribe to a police officer to avoid a traffic ticket

Low material wealth (N=1529)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.06	[0.04, 0.07]	$\hat{\lambda}^{T}_{1}$	0.31	[0.23, 0.44]
SST only	0.16	[0.11, 0.21]	$\hat{\lambda}^{L}_{1}$	0.66	[0.52, 0.75]
Joint response	0.18	[0.12, 0.23]	$\hat{\lambda}^{T}_{0}$	0.98	[0.97, 0.99]

Moderate material wealth (N=1285)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.08	[0.07, 0.10]	$\hat{\lambda}^{T}_{1}$	0.53	[0.39, 0.83]
SST only	0.13	[0.07, 0.18]	$\hat{\lambda}^{L}_{1}$	0.43	[0.13, 0.59]
Joint response	0.15	[0.10, 0.20]	$\hat{\lambda}^{T}_{0}$	0.98	[0.97, 0.99]

High material wealth (N=1373)

Prevalence estimate $(\widehat{\pi})$			Diagnostic parameters (joint response model)		
Direct only	0.13	[0.12, 0.15]	$\hat{\lambda}^{T}_{1}$	0.81	[0.62, 1]
SST only	0.11	[0.06, 0.17]	$\hat{\lambda}^{L}_{1}$	0.19	[0.00, 0.37]
Joint response	0.16	[0.,0.21]	$\hat{\lambda}^{T}_{0}$	0.98	[0.97, 0.99]

Note: Ninety-five percent confidence intervals in square brackets.

Additional Analysis: Education and Wealth

Given that education and wealth are characteristics that tend to be correlated with one another, we also analyzed the influence of these variables in a 2 x 2 design in which model parameters were estimated for respondents with levels of wealth above and below the median in the sample, according to whether or not they completed secondary school.[1] Table A7 below presents the full results. The findings indicate that wealth is a more powerful determinant of truthfulness under direct questioning about corruption than formal education. For respondents who completed secondary school, individuals in the higher wealth category were substantially more likely to be truthful under direct questioning than those in the lower wealth category (On Q1, $\hat{\lambda}_1^T = 0.78$ for those with above median wealth versus $\hat{\lambda}_1^T = 0.50$ for those with below median wealth. On Q2, $\hat{\lambda}_1^T = 0.88$ for the former versus $\hat{\lambda}_1^T = 0.36$ for the latter.). For those with lower formal education, the differences across wealth categories were more muted but in the same direction. Holding wealth categories constant, one finds smaller differences across levels of education. Moreover, these differences are non-negligible only for individuals in the above median wealth group.

---

[1] Nearly half of the respondents in our sampled completed secondary school or a higher level of education (45%). The procedure for measuring wealth was the same as that elaborated above, with the median of the factor scores used to assign category membership instead of the terciles.

Table A7: Parameter estimates for questions about corruption, by wealth and education

Q1: To avoid paying a traffic ticket, I would be willing to pay a bribe to a police officer

Wealth	Education						
	Complete secondary school or higher				Incomplete secondary school or less		
Above median	Prevalence estimate $\widehat{\pi}$			Diagnostic parameters (joint response model)	Prevalence estimate $\widehat{\pi}$			Diagnostic parameters (joint response model)

Direct	0.20	[0.17, 0.21]	$\hat{\lambda}^T_1$ 0.78 [0.65, 0.94]	Direct	0.22	[0.20, 0.25]	$\hat{\lambda}^T_1$ 0.66 [0.55, 0.80]
SST	0.18	[0.13, 0.24]	$\hat{\lambda}^L_1$ 0.16 [0.00, 0.30]	SST	0.23	[0.16, 0.30]	$\hat{\lambda}^L_1$ 0.32 [0.17, 0.43]
Joint	0.24	[0.20, 0.30]	$\hat{\lambda}^T_0$ 0.99 [0.97, 1.00]	Joint	0.33	[0.26, 0.40]	$\hat{\lambda}^T_0$ 0.97 [0.95, 0.99]
		N=1,250				N=831	

Wealth							
Below median	Prevalence estimate $\widehat{\pi}$			Diagnostic parameters (joint response model)	Prevalence estimate $\widehat{\pi}$			Diagnostic parameters (joint response model)
Direct	0.17	[0.14, 0.20]	$\hat{\lambda}^T_1$ 0.50 [0.39, 0.65]	Direct	0.15	[0.13, 0.17]	$\hat{\lambda}^T_1$ 0.51 [0.43, 0.62]
SST	0.25	[0.17, 0.33]	$\hat{\lambda}^L_1$ 0.48 [0.32, 0.59]	SST	0.22	[0.17, 0.27]	$\hat{\lambda}^L_1$ 0.45 [0.34, 0.53]
Joint	0.32	[0.25, 0.40]	$\hat{\lambda}^T_0$ 0.95 [0.93, 0.98]	Joint	0.28	[0.23, 0.33]	$\hat{\lambda}^T_0$ 0.96 [0.95, 0.98]
		N=621				N=1,491	

Q2: I have paid, at least once, a bribe to a police officer to avoid a traffic ticket

Wealth	Education						
	Complete secondary school or higher				Incomplete secondary school or less		
Above median	Prevalence estimate $\widehat{\pi}$			Diagnostic parameters (joint response model)	Prevalence estimate $\widehat{\pi}$			Diagnostic parameters (joint response model)
Direct	0.13	[0.11, 0.14]	$\hat{\lambda}^T_1$ 0.88 [0.64, 1.00]	Direct	0.09	[0.07, 0.11]	$\hat{\lambda}^T_1$ 0.59 [0.39, 0.95]
SST	0.10	[0.04, 0.15]	$\hat{\lambda}^L_1$ 0.12 [0.00, 0.35]	SST	0.13	[0.06, 0.19]	$\hat{\lambda}^L_1$ 0.40 [0.15, 0.58]
Joint	0.14	[0.10, 0.20]	$\hat{\lambda}^T_0$ 0.98 [0.97, 0.99]	Joint	0.16	[0.09, 0.22]	$\hat{\lambda}^T_0$ 0.98 [0.97, 0.99]
		N=1,251				N=830	

Wealth							
Below median	Prevalence estimate $\widehat{\pi}$			Diagnostic parameters (joint response model)	Prevalence estimate $\widehat{\pi}$			Diagnostic parameters (joint response model)
Direct	0.08	[0.05, 0.10]	$\hat{\lambda}^T_1$ 0.36 [0.24, 0.56]	Direct	0.06	[0.05, 0.07]	$\hat{\lambda}^T_1$ 0.36 [0.27, 0.52]
SST	0.19	[0.11, 0.27]	$\hat{\lambda}^L_1$ 0.64 [0.43, 0.75]	SST	0.15	[0.10, 0.20]	$\hat{\lambda}^L_1$ 0.59 [0.42, 0.70]
Joint	0.21	[0.13, 0.28]	$\hat{\lambda}^T_0$ 0.98 [0.96, 0.99]	Joint	0.17	[0.12, 0.22]	$\hat{\lambda}^T_0$ 0.98 [0.97, 0.99]
		N=620				N=1,486	