NBER WORKING PAPER SERIES

GENDER AND BUREAUCRATIC CORRUPTION: EVIDENCE FROM TWO COUNTRIES

Francesco Decarolis
Raymond Fisman
Paolo Pinotti
Silvia Vannutelli
Yongxiang Wang

Working Paper 28397
http://www.nber.org/papers/w28397

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2021

Decarolis gratefully acknowledges financial support from the European Research Council (ERC-2015-StG-679217). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Gender and Bureaucratic Corruption: Evidence from Two Countries

Francesco Decarolis, Raymond Fisman, Paolo Pinotti,
Silvia Vannutelli, and Yongxiang Wang

August 6, 2021

Abstract

We examine the correlation between gender and bureaucratic corruption using two distinct datasets, from Italy and from China. In each case, we find that women are far less likely to be investigated for corruption than men. In our Italian data, female procurement officials are 22 percent less likely than men to be investigated for corruption by enforcement authorities; in China, female prefectural leaders are 81 percent less likely to be arrested for corruption than men. While these represent correlations (rather than definitive causal effects), both are very robust relationships, which survive the inclusion of fine-grained individual and geographic controls, and based on Oster's (2019) test unlikely to be driven by unobservables. Using data from a survey of Italian procurement officials, we present tentative evidence on mechanism: the gender gap is partly due to women acting more "defensively" in administering their duties.

JEL Classification Numbers: J16; D73.
Key words: Gender; corruption; bureaucracy

*Decarolis: Economics Department and IGIER, Bocconi University, via Roentgen 1, Milan, Italy; email: francesco.decarolis@unibocconi.it. Fisman: Economics Department, Boston University, 270 Bay State Road, Boston, MA; email: rfisman@bu.edu. Pinotti, Bocconi University and BAFFI-CAREFIN, via Roentgen 1, Milan, Italy; email: paolo.pinotti@unibocconi.it. Vannutelli, Economics Department, Boston University, 270 Bay State Road, Boston, MA; email: svann@bu.edu. Wang, Finance Department, Marshall School of Business, University of Southern California, Los Angeles, CA; email: Yongxiang.Wang@marshall.usc.edu. We also thankfully acknowledge financial support from the European Research Council (ERC Starting Grant 679217, Decarolis).

I Introduction

There exists a very long literature on the relative probity of women versus men, and a distinct body of work on whether women govern differently from men.[1] A natural – and important – point of intersection is whether women are more corrupt than men when put in a position of public trust.

In this paper we show that, for officials in two distinct settings at two very different levels of government bureaucracy, women are far less likely to be suspected of or arrested for corruption. We employ two separate datasets obtained for unrelated research agendas to study the association between gender and corruption. Our first dataset includes the universe of Italian officials who presided over at least one procurement auction during 2000-2016 years. The second is a dataset of all Chinese bureaucrats who held the position of prefecture mayor or party secretary during 1979-2014. For our Italian data, we know whether the official has ever been flagged as suspected of corruption by any of the country's enforcement authorities. For our Chinese data, we observe whether an official has been arrested for corruption. Our data thus come from officials from distinct geographies, cultures, political systems and at very different levels in the bureaucracy.[2]

In both cases, we find far lower corruption rates among women relative to their male counterparts. In our Italian data, for men and women working within the same procurement authority, women are 22 percent less likely to be investigated for corruption by enforcement authorities. In our Chinese data, female prefecture leaders are 81 percent less likely to have been arrested for corruption than men. In both cases, we include fine-grained fixed effects to account for regional or demographic differences.

There is a range of candidate explanations for the lower observed rates of corruption investigations for women. These include gender differences in (a) *selection* into public office; (b) *opportunities* for corruption; (c) *behaviors* based on underlying gender differences in risk aversion and/or ethics; (d) *enforcement* via the judicial system.

It is beyond the scope of our paper to provide a definitive accounting of the role of each of these factors. We nonetheless take a step toward understanding *how* gender affects corruption via two further sets of analyses using extensions to our Italian data. First, we use recent survey data on Italian procurement officials to probe the existence of

[1] See Rosenbaum et al. (2014) for a survey of research on gender and honesty, and Jacobsen et al. (2018) for a summary of economics-focused work on the topic. For gender-based differences in how politicians govern, see Ferreira and Gyourko (2014) for a study of U.S. mayors, and Pande and Ford (2012) for a survey of studies based on gender quotas (including the classic work of Chattopadhyay and Duflo (2004)).

[2] Because of the richness and detail of the Italian data, as well as complementary evidence via a survey we conducted of Italian officials, we view this as the primary dataset in our analysis. Analysis from the Chinese dataset provides corroborating evidence, to show that the overall patterns we observe in Italy may generalize to other settings.

behavioral differences between male and female procurement officers. The survey presents respondents with a series of contracting scenarios, and asks for their likely course of action. For each of the four scenarios, the choices include a "defensive" choice, a "non-defensive" choice, and "I don't know what I would do." In each case the defensive option entails actions taken in order to maximize compliance and limit the risk of subsequent investigations for violation of the procedures, even if this might come at the cost of slowing down the procurement process. Women systematically select more defensive actions, and also report devoting more time to checking whether they are in compliance with relevant regulations.[3] While this does not rule out a role for other mechanisms, it does provide evidence suggesting that different (risk-avoidance) behaviors at least contribute to the gender gap in corruption. Second, we present a set of heterogeneity analyses. Generally, the gender gap in corruption is wider in more developed (Northern) regions. While on the face of it this result appears surprising, it is consistent with the findings of Decarolis et al. (2019), that the (less corrupt) North allows greater discretion in procurement, which men might exploit to a greater extent than women.

There exists a small literature that looks at gender differences in corruption, both at a macro and micro level. The earliest contributions, based on cross-country regressions relating female representation in government to corruption perception indices (Dollar et al., 2001; Swamy et al., 2001), have the natural concern of omitted country-level attributes. Researchers have also examined individual-level involvement in corruption, both in terms of likelihood of being asked for a bribe as well as willingness to pay a bribe when the opportunity arises. Women report that they are less likely to be asked for a bribe, and also are more likely to express disapproval of bribery (Swamy et al., 2001; Mocan, 2008). Given that the surveys have no payoff consequences, these findings may be explained by social desirability bias, which plausibly differs between men and women.

Survey-based evidence has also provided some insights into the reasons for gender differences in corruption. Lee and Guven (2013) emphasizes the role of "male dominant" norms, finding that in male-dominant societies men are more likely to be asked for bribes than women, suggesting that there may be gender differences in corruption *opportunities*. Naturally, this finding is subject to caveats of social desirability as well as omitted variable bias. Prior work also suggests that differential *selection* may not be the main driver of observed differences in officials' corruption, at least insofar as the supply side of officials is concerned: based on the dice-rolling honesty measure of Fischbacher and Föllmi-Heusi (2013), Gans-Morse et al. (2021) show no gender difference in honesty and selection into public service in Russia. Finally, while not focused on corruption specifically, prior

[3]Naturally, the gender difference in behaviors that we observe could result from behavioral differences between men and women for the overall population, or from a differential selection of officials by gender.

work indicates a greater leniency toward women in the judicial system more generally, potentially as a result of both "paternalism" as well as statistical discrimination (see, in particular, Starr, 2015; Bindler and Hjalmarsson, 2020 and citations therein). While these prior findings suggest that differential *enforcement* very plausibly contributes to the overall gender gap in corruption investigations, our data are not well-suited to assessing whether it is indeed a contributing factor.

We conclude our overview of the relevant literature with two papers, most closely related to our own, which both look at the corruption of male versus female political leaders, exploiting random assignment from quotas in India (Afridi et al., 2017) and quasi-experimental assignment exploiting close elections in Brazil (Brollo and Troiano, 2016). The evidence from India uses the same variation as the classic study of Chattopadhyay and Duflo (2004) based on the requirement in West Bengal that villages have a female leader. This requirement was done at random, rotating among villages with one-third treated at a time. In treated villages, survey respondents reported lower corruption. Once again, these findings rely on survey responses; furthermore, the study's design necessarily conflates turnover with gender. Brollo and Troiano (2016) combines data from Brazil's random municipal audits (Ferraz and Finan, 2008) with election results from close mixed-gender races, and reports a lower number of corruption cases for female-led municipalities. However, once one limits the sample to mixed-gender close elections, the sample size is very small (161 observations), and sensitive to the choice of specification.

We bring a number of key contributions to this literature. First, we show that for two large and diverse populations of bureaucrats, there is a far lower rate of observed corruption among women. As with Brollo and Troiano (2016), we use real measures of corruption, which are detailed in the next section, sidestepping at least some concerns of response bias. In contrast to the sizeable collection of lab experiments, we capture gender differences which combine the effects of selection and any underlying gender difference in values; this combined effect may be relevant for policy, to understand whether, conditional on reaching a particular position, women behave more or less corruptly than men. The populations we examine are also of note, as they involve very different cultural and political environments, and officials at vastly different positions in their respective hierarchies. While our data were chosen opportunistically as a result of data availability, the fact that we find a clear gender gap in corruption in both data sets suggests that we may be able to generalize from the patterns that we observe to other settings. Finally, our survey-based results on Italian procurement officials helps to get at *why* a gender gap in corruption may emerge.

It is also important to note the limitations of our analysis upfront. As already implied above, we cannot take a stand on the extent to which selection plays a role in the

gender gap in corruption investigations. That is, we do not offer any causal identification on whether a randomly selected woman is more dishonest or corrupt than a randomly selected man, or whether there is differential selection of men versus women into the positions we study. On the one hand, we capture gender differences which combine the effects of selection and any underlying gender difference in values, which is relevant to understanding whether, conditional on reaching a particular position, women behave more or less corruptly than men. However, it also means that we cannot speak to the question of whether, say, women should receive preferential treatment (for example, via gender quotas) to increase their representation in bureaucratic positions that require a high degree of probity.[4] And in this paper, we provide only limited evidence on why (conditional on selection) women are less likely to be investigated for corruption, in our Italian survey data. We identify a likely role of *behavioral* factors, in particular differences in risk aversion. But we cannot rule out a role for selection, opportunities, or enforcement – the magnitude of the effects we document suggest that further work on the mechanisms that underpin overall gender differences in corruption deserves further consideration.

Finally, we present data from two types of positions in just two countries – procurement officers in Italy and prefectural leaders in China – which naturally raises questions of generalizability. Both positions we study are relatively desirable in their respective countries. In part as a result, the prevalence of women in these positions is far lower than that of the two countries' bureaucracies more broadly. For example, only 17 percent of Italian procurement officers in our sample are women, whereas 63 percent of all public employees in Italy were female in 2013 (Baig et al., 2021); for China, just 4 percent of our prefectural leader sample are women, as compared to 43 percent of Chinese public employees in 2013. In interpreting our results, it will thus be useful to keep in mind that we study relatively desirable positions that are difficult for women to obtain. Finally, in terms of generalizability, both countries have female representation in public employment that is comparable to other nations in their respective world regions: for Europe and Central Asia, the average female share of public employment is 63 percent, and for East Asia and Pacific it is 44 percent.

Throughout the rest of the paper we will proceed with two parallel sets of analyses. We describe our Italian and Chinese datasets in Section II and in Section III we provide regression results for both settings (including analysis of our survey of Italian procurement officials). Section IV concludes.

[4]Such policies have in fact been proposed in the past. For example, in 1999 the newly installed Mexico City police chief handed over ticket-writing authority to female officers because, "I trust them" not to take bribes. See, for example, Joseph Treaster, "The World: Equal Opportunity in Mexico City; Counting on Women to Be More Honest Than Men," *The New York Times*, August 15, 1999.

II Data

II.A Data on Italian Procurement Officials

Our data draw on the same sources as Decarolis et al. (2019), a study of corruption in Italian procurement. These data include all road and building procurement auctions in Italy between 2000 and 2016, with a value of at least €40,000. The procurement authority (PA) for most auctions is a municipality, but the data also includes contracts from regional governments as well as educational institutions, hospitals, and public companies that oversee highway construction. The data we use include the identity of the contracting officer overseeing each contract (the *"Responsabile Unico del Procedimento"*, or RUP) and her social security record, from which we can identify gender, municipality of birth, and age. Crucially, we are also able to link these individuals to the *Sistema D'Indagine Interforze* (SDI) archive, which is a primary source of information that police officers and intelligence agencies use to identify potential targets for further investigation. The SDI is managed by Italy's internal intelligence and security agency, AISI, and contains reports of all individuals investigated by any of the Italian police forces: state police (*Polizia di Stato*), finance police (*Guardia di Finanza*), military police (*Carabinieri*), and environmental police (*Guardia Forestale*).

An entry in the SDI database typically occurs after a police force, based on a preliminary investigation, determines that there is sufficient evidence to open a formal investigation. This investigation might or might not lead to a court case and, if so, to a conviction. Therefore, court cases are a strict subset of the entries in the SDI database. The resulting sample of suspect offenders thus includes individuals that were convicted, acquitted, or never charged. The latter two groups plausibly comprise a large number of offenders whose guilt could not be proven in court. Indeed, corruption cases are generally complex, and convictions relatively rare, particularly in Italy.[5] Thus our Italian database is far more inclusive than standard measures of corruption based on convictions. For each RUP in our dataset, AISI searched the SDI database for any investigation in the following crime categories: corruption, malfeasance and embezzlement; abuse of power and undue influence; and violations in public auctions. The interested reader may consult the data section of Decarolis et al. (2019) for more details on the SDI database.

In Table 1, we show summary statistics for the full sample of analysis as well as a comparison of the characteristics of female versus male RUPs.[6] The first thing to note is

[5] A court case for corruption charges can only be initiated if there is direct proof of a kickback received by an official (either monetary or some other form of benefit such as hiring of a relative). Under the law, corruption represents a contract crime where two parties stipulate a contract (formal or informal) to obtain (or attempt obtaining) something unlawful.

[6] Given that a RUP might hold positions in different municipalities, an observation in our dataset is a

that men are almost twice as likely as women to have been investigated. In contrast to what we will observe in our Chinese data, here we do note substantial gender differences on a number of characteristics: women are on average 6 years younger, they are more prevalent in the North, and they are less likely to be a RUP in their region or municipality of birth, relative to males. Because of these differences, we will present some saturated specifications that include fixed effects for municipality, year of birth, and number of contracts managed by a RUP in our dataset.

II.B Data on Chinese prefectural leaders

Our sample of Chinese officials takes the data of Fisman and Wang (2017) as its starting point. This dataset includes the identities and characteristics (based on officials' resumes) of mayors and prefecture-level party secretaries who started their posts during the years 1979-2014. The identities were originally extracted from provincial yearbooks and the official website of the *People's Daily*, People.cn; additional information on politicians' qualifications and career trajectories was derived from resumes accessed via baike.Baidu.com, which is similar to Wikipedia for China (see Fisman and Wang (2017) for more details). We use the cutoff of 1979 as it is the year of transition from Mao to Deng. The sample includes a total of 3,133 officials across 289 prefectures. In addition to information on gender, the data also include information on education and, for approximately 84 percent of officials, their place of birth.

For this sample, we identify politicians that are publicly investigated for corruption. Note that investigation, arrest, and conviction are essentially synonymous in the Chinese context and we use the terms somewhat interchangeably; this is very different from the Italian setting we described above. The vast majority of the cases in our data were launched under the anti-corruption crackdown of Party Secretary Xi Jinping, which was initiated at the beginning of 2013. Almost our entire sample had already reached the position of mayor or higher by the time the anti-corruption campaign was launched. Thus, the officials we study were already quite high up in the hierarchy – with opportunities for bribe extraction. Since the campaign was entirely unanticipated, officials likely felt they could act with greater impunity during most of the period we study.

The list of officials targeted with investigation comes from the official website of the Central Commission for Discipline Inspection (`http://www.ccdi.gov.cn/scdc`), China's top anti-corruption authority. Of the 3,133 officials in our initial sample, 235 (7.5 percent) have been investigated for corruption. The vast majority of these investigations

RUP-by-procurement authority; in practice, moving across municipalities is relatively rare – 70 percent of RUPs never move, and only 10 percent move more than once.

– 209 of the 235 total – took place under Xi Jinping's anti-corruption campaign. In some specifications we will focus exclusively on these 209 post-2013 investigations as our measure of corruption. Finally, in some specifications we will limit our analysis to the 1,878 officials who started a new position as mayor or party secretary 1998 or later. This year is a natural cutoff, as it is the beginning of the 5-year Central Committee term, and because the anti-corruption crackdown targeted recently active officials. In practice, 208 of the 209 individuals targeted by Xi's campaign are included in this post-1998 subsample (results are identical if we use an earlier or later cutoff).

In Table 2, Panel A we show summary statistics for the full sample, while panel B provides summary statistics for the set of leaders that held positions starting 1998 or later (and thus were vulnerable to Xi's anti-corruption campaign). As expected, the fraction of women leaders is higher for the more recent sample (5.3 versus 3.8 percent); corruption investigations are also far higher (11.7 versus 7.5 percent) – as noted earlier, all but one of Xi's arrests in our sample are from individuals starting positions 1998 or later.

The table also provides the differences in means for male versus female leaders (the p-values in the final column are calculated using heteroskedasticity-robust standard errors). In the full sample, men are more than twice as likely as women to have been arrested (significant at the 5 percent level). However, this may understate the difference, as women are represented at a much higher rate in the post-1998 sample when corruption arrests primarily took place. When we limit the sample to this later period, the male-female gap widens to a three-fold difference (significant at the 1 percent level). It is also of note that in the post-1998 sample men and women are better balanced on other basic observables – in particular age is quite similar in the post-1998 sample; women, however, are still more educated than their male counterparts, which may suggest a higher bar for promotion for women.

III Results

III.A Evidence from investigations of Italian procurement officials

We begin by examining gender differences in the investigation of Italian procurement officials. We observe multiple procurement officials per contracting authority (generally municipalities), which means that we may have fine-grained fixed effects to compare the conduct of women and men within a relatively narrow geography (the median municipality has a population of 7,000). Our main specification is as follows:

$$InvestigatedRUP_i = \beta Female_i + RUPControls_i + PAControls_{pa(i)} + \alpha_{reg(i,t)} + \alpha_{pa(i)} + \varepsilon_i$$

for RUP i working in procurement authority $pa(i)$. The year t captures the year a RUP first appears in our dataset. RUP controls include the (log of) i's age, an indicator variable denoting whether i was born in the same municipality where she is employed, an indicator variable denoting whether i ever served as a local politician, as well as several characteristics of the auctions that i has overseen, including the (log of) number of auctions, average value of these auctions, and the average number of bidders.[7] PA controls include a set of dummies for the type of PA (Central, Region and other local authority, Hospitals and Universities, Transportation), the (log of) total number of auctions conducted by the PA during the sample period, the log of the total number of RUPs observed in the PA during the sample period, and the log of the total number of auctions managed by the RUP. Finally, we include various sets of fixed effects, depending on the specification, including 340 region-year fixed-effects $\alpha_{reg(i,t)}$ to account non-parametrically for time-varying geographic differences in corruption as well as female representation in procurement positions, and PA fixed effects $\alpha_{pa(i)}$. Note that the PA-level controls naturally drop out in specifications in which we include PA fixed effects.

In the first column of Table 3, we include only the region-year fixed effects. The point estimate on $Female$ is -0.034 (significant at the 1 percent level), indicating that female RUPs are about 40 percent less likely to be investigated for corruption relative to their male counterparts. The inclusion of PA and individual RUP controls in columns (2) and (3) reduces the $Female$ coefficient to -0.026 and -0.019 respectively, and adding PA fixed effects in column (4) further reduces it to -0.0166 (still significant at the 1 percent level). The point estimate on age is positive, which might be interpreted as an older official having more time and opportunities to have been investigated for corruption or other crimes. In column (5), our favored specification, we provide a saturated regression that includes fixed effects for birth year and also for the number of contracts overseen by i. These further controls do not affect the point estimate on $Female$, which is -0.0176. To provide a sense of magnitude, the mean of $Investigated$ is 0.08, implying a corruption gender gap of 22 percent (0.0176/0.08). In the final column, we repeat the specification from column (4), limiting the sample to municipalities, which excludes administrations that do not map to a specific local geography; excluded PAs include educational institutions, hospitals, and public companies dealing with the management

[7]The inclusion of whether a RUP was ever a local politician may suffer from a bad control problem, in the sense that selection as politician may be one channel through which corruptible men progress through the municipal bureaucracy. For results that exclude this control, see Appendix Table A.1.

of motorway sections under concession. The point estimate is very similar to our favored specification in column (5).

It is natural to consider whether women are simply assigned to manage contracts for which there is less scope for corruption. Decarolis et al. (2019) suggests that auctions that afford greater discretion to the RUP, and also foreclose competition, are most prone to abuse. In additional results reported in the appendix, we show that women are no less likely to oversee auctions that afford discretion, and those with restricted competition.[8]

It is natural to consider whether women are simply assigned to manage contracts for which there is less scope for corruption. Decarolis et al. (2019) suggests that auctions that give greater discretion to the RUP, and also foreclose competition, are most prone to abuse. In additional results reported in the appendix, we show that women are no less likely to oversee auctions that afford discretion, and those with restricted competition.[9]

In Table 4, we then consider how the gender corruption gap varies with PA- or individual-level attributes, building on our favored specification in column (5) of Table 3. Our goal in this exercise is to assess whether we can learn about the underlying reasons for the gender corruption gap by exploring where it is most prominent and/or non-existent. We selected dimensions of heterogeneity that relate to some of the underlying mechanisms we discussed in the introduction. We emphasize upfront that these are at best coarse proxies that map in a highly imperfect way to underlying mechanisms. First, we introduce several variables that could proxy for women's relative ability to tap into favor exchange networks. These include two individual-level characteristics: whether a RUP was born in the municipality in which she is employed, and whether she had been a local politician. As well, we include several measures of women's prominence in the relevant labor force, which similarly may capture gender differentials in employment and favor exchange opportunities: female labor force participation in the municipality where the PA is located, the share of female RUPs in a PA, and the share of female employees in the PA overall. We also include several measures of organized crime: whether the municipality has been subject to mafia investigations, and whether it is in the South (defined as regions to the south of, and including, Abruzzo). Finally, we include a catch-all proxy for an area's overall level of development: municipality income per capita.

The main takeaway from the heterogeneity analysis is that more developed areas have a *larger* gender corruption gap, as captured by the negative interaction of $Female$

[8]Specifically, there is no gender gap after including in the model specification time and geographic fixed effects, while, in the absence of such controls, women are *more* likely to oversee these more discretionary auctions. See Appendix Table A.2.

[9]Specifically, there is no gender gap after including in the model specification time and geographic fixed effects, while, in the absence of such controls, women are *more* likely to oversee these more discretionary auctions. See Appendix Table A.2.

and income per capita, which is the one coefficient that is stable and significant even when all covariates are included in the final column of the table. The fact that many other coefficients are much attenuated in magnitude and significance is unsurprising, given that all of the proxies we use are highly correlated with a region's level of development. On the face of it, the larger gender corruption gap in more developed municipalities is surprising. However, as we observed in the introduction, it is consistent with the greater discretion permitted in more developed areas (Decarolis et al., 2019), which men might exploit to a greater extent than women.

III.A.1 Selection on Unobservables: Discussion

There are two main unobserved factors that could account for our results, beyond a gender difference in corruptibility: differential selection into public service and differential selection into investigation and prosecution (i.e., differential enforcement).

Focusing first on selection into public administration, our regression-based strategy captures the following comparison: within a PA (given our use of PA fixed effects), what is the probability of a male versus female procurement officer being investigated for corruption, amongst the set of procurement officials employed in the PA. That is, we capture the effect of gender conditional on potential differential selection into the public administration. As we alluded to in the introduction, our data and empirical strategy are not well-suited to assessing the corruptibility of men versus women drawn from the general population. The latter is interesting from the perspective of understanding the mechanisms driving our results, and also could inform the value of policies that, say, promote the appointment of women to positions for which there may be temptations to misuse public office. Our results offer insights on the relative corruptibility of men versus women with the bureaucracy already in place, which may be relevant for a distinct set of policies, for example whom to target with more frequent audits.

We now turn to consider gender differences in selection into the set of individuals targeted with investigation, conditional on extent of dubious or suspicious behavior. We cannot rule out this potential mechanism, though we note that based on extensive conversations with enforcement authorities, this is perhaps less plausible as investigations usually originate from complaints raised by losing bidders (see Decarolis et al., 2019). On the other hand, the "defensive behavior" we document in the survey that we discuss shortly could lower the probability of procedural breaches and therefore follow-on complaints and investigations. This is again beyond the scope of our data and analysis.

We aim to assess whether selection on unobservables in general could account for our findings using the test developed in Oster (2019). The intuition for this test is that

one can learn, at least partially, about the extent of potential omitted variable bias by looking at coefficient stability across specifications that include different sets of controls. The test's key assumption is that selection on the observed controls is proportional to selection on the unobserved ones, an insight from Altonji et al. (2005). To perform the test, one needs to make an assumption on how much variation could be explained if all unobservables were accounted for (what Oster labels R_{max}), and also the degree of selection on observed versus unobserved variables (in Oster's model, δ). If $\delta = 1$ it implies that observed and unobserved covariates contribute equal explanatory power.

Oster (2019) considers two equivalent metrics for assessing extent of selection on unobservables. First, one can set bounds on R_{max} and δ which yield in turn an interval for the "true" treatment effect (in our case the effect of gender). Second, one may (equivalently) ask, given a value of R_{max}, how large the role of unobservables would need to be to drive the treatment effect to zero (see Oster, 2019 for further details of the test and its construction). We use the suggested values of $R_{max} = 1.3R$ and $\delta = 1$ in our calculations, which are reported in Table 5. We do so for two sets of comparisons. First, we consider the test if we assumed we did not have access to PA fixed effects at all (the first row), and second, we compare coefficient stability when PA fixed effects are amongst the observed controls (all estimates include region-times-year fixed effects). In both cases, we obtain a range of parameter estimates that exclude zero. Furthermore, the values of δ that would be required to generate a zero effect both exceed one, thus implying that unobservables would need to explain even more of the variation in corruption than observables, including PA fixed effects, to account for our estimated gender effect.

III.A.2 Why are women investigated less often? Evidence from RUP surveys

Finally, we provide some exploratory analyses that may shed light on *why* women are investigated less often for corruption. These findings come from a survey conducted by Decarolis and Battini (2020) of over 500 RUPs in 2020. The surveyed RUPs were mainly selected from procurement officers who participated in training courses on public procurement offered by the National School of Administration (SNA), the government body that manages the recruitment and on-the-job training of public officers.[10] Among surveyed RUPs, 47 percent are women as compared to 17 percent for the sample overall. This is driven by the SNA's objective of gender balance in SNA training programs; conditional on receiving a survey invitation, we observe no gender difference in response

[10]Specifically, the survey was sent to 1,443 RUPs, 952 of whom were the set of all RUPs who participated in recent SNA courses on public procurement. The remaining 491 public administrators were identified as RUPs through a sample analysis of public procurement calls for tenders. 417 responses were collected from members of the first category, and 121 responses from members of the second.

rate. This should be kept in mind in interpreting the results below, as there may be more competition for males to be admitted to the SNA courses and, hence, to be a respondent in the survey. Summary statistics for the survey may be found in Appendix Table A.3.

A key feature of the survey is a set of four questions that aim to assess RUP "defensiveness" based on responses to real-world scenarios that a procurement official would plausibly face; in each case, there is no clear right or wrong answer. For example, one scenario asks about a RUP's willingness to trade off delaying a project while the auction assignment is contested versus proceeding while the case is adjudicated, while another asks about whether to seek explicit approval for an agreement from the anti-corruption authority, despite already receiving a go-ahead from the Ministry of Economics and Finance. (The full set of scenarios appear in Appendix B.)

In Table 6, we compare the extent to which men versus women reacted to these scenarios by taking defensive actions. The outcome variable, $DefensiveActions$, is defined as the average probability of choosing the "defensive option" across the four scenario questions (see Appendix B). The coefficient on $Female$ is in the range of 4 to 5, representing a 15 percent increase relative to the sample mean, 25 percent of a standard deviation; the difference persists when we control for other demographics, and RUP risk references as captured by survey responses (see Appendix B). In the second half of the table, we use whether the RUP has insurance against the risk of administrative damages as the outcome, which has the further benefit of reflecting a "real" rather than hypothetical choice of the respondent.[11] We find that women are substantially *less* likely to own such insurance – whereas the mean for men is 44, for women it is 28, a difference of 16 percentage points; as shown in the table, this difference is essentially unaffected by the inclusion of controls for demographics and risk preferences. This is a natural result if one considers insurance and defensiveness to be substitutes, i.e., different approaches to risk mitigation. Overall, we interpret the survey results as indicating that the gender difference in corruption investigations may be driven in part by differences in compliance with formal protocols, and this is further reflected in a lower need to hold insurance against administrative litigation. These results together are most readily explained by differences in preferences by gender rather than differential enforcement – while the findings on insurance could result if men are targeted more with investigations, we would also then expect that men would exhibit higher defensiveness, whereas we observe the opposite.

These results, on the face of it, would appear to be in tension with our earlier findings. In particular, we observed in Appendix Table A.2 that women are just as likely

[11] Administrative damages represent the individual liability for damages of public officers who, in the performance of official duties, make decisions or take actions that damage the public administration. By law, officials must pay by themselves for this type of insurance (if they want so) as it cannot be paid for by their employer.

as men to oversee discretionary procedures and auctions with fewer than the legally mandated number of bidders. Decarolis et al. (2019) highlight that these are the two methods of procurement most likely to be associated with corruption episodes. However, as also emphasized by Decarolis et al. (2019), a well-intentioned monitor will permit less discretion if a greater probability of corruption is anticipated. To the extent that women are perceived to be less corruptible, they may be permitted greater discretion in their choice of auction, so that the overall effect on contracting procedures is ambiguous.

III.B Evidence from arrests of Chinese prefectural leaders

In this section we study whether gender is correlated with top municipal leaders' arrests for corruption, primarily under Xi Jinping's anti-corruption campaign. We employ variants on the following specification, which parallels the analyses of Italian officials:

$$Investigated_i = \beta_1 * Female_i + \beta_2 * EducationControls_i + \beta_3 * Age_i + \gamma_{p(i)} + StartCohort_i + \epsilon_i$$

for politician i, where $\gamma_{p(i)}$ is a set of fixed effects for the province p that the official first appeared as a public official in our data, $StartYear_i$ is a set of 36 fixed effects to capture the first date that a politician appears in our dataset, $EducationControls_i$ is a set of indicator variables for bachelor's, master's, and doctoral degrees (no college is the omitted category), Age is the log of i's age, and ϵ_i is the error term (we use robust standard errors throughout).

In column (1) of Table 7 we present results including only start year fixed effects, to account for the fact that women are vastly under-represented in the earlier part of the sample when few corruption arrests took place (Table 2 includes the simple difference in means for the full sample). The coefficient of -0.069 (significant at the 1 percent level) indicates that, after accounting for politician start date, women are 6.9 percentage points less likely to be arrested for corruption than males who started during the same year. The inclusion of province fixed effects in column (2) has no appreciable impact on the estimate; the inclusion of individual controls in column (3) increases the magnitude of the female effect to -0.081. (The correlation with age is near-mechanical, and results from the fact that more recent cohorts of officials were more likely to have been targeted by Xi's anti-corruption campaign.)

In the remaining columns we limit the sample to officials who started a new position as a municipal leader in 1998 or later. As observed earlier, the arrest rate for these newer politicians is much higher (11.7 percent). Given that many of the older officials in our sample were never subject to the anticorruption crackdown, we consider this to be our preferred sample. In this post-1998 subsample the coefficient on *Female* is -0.095. Given

a mean arrest rate among males in this sample of 11.7 percent, our point estimate implies that women are 81 percent less likely (−0.095/0.117) to have been arrested than men. In column (5) we add fixed effects for five year age cohorts (based on age in 2018); the point estimate on *Female* is largely unchanged. Furthermore, as with our Italian results above, we present the results of the Oster (2019) test for selection on unobservables. Table 8 clearly illustrates that, given the coefficient stability we find, unobservables would need to play an extraordinarily strong role in order to account for our results. Finally, additional evidence on the robustness of our results is provided in Table A.4. There, we repeat these analyses focusing on arrests that occurred as part of Xi's post-2013 anti-corruption crackdown. The coefficients on *Female* are essentially unchanged.

IV Conclusion

In this paper, we document a very sizeable and robust difference in rates of corruption investigations between men and women, for two very different populations of public officials. Via survey evidence on Italian procurement officials we are further able to begin to shed light on *why* such stark differences emerge – female procurement officials are more likely to follow strict legal protocols than their male counterparts. The very large effect sizes we document suggest that it is worthwhile to delve further into the underlying mechanisms that generate the gender gap in bureaucratic corruption.

As we lay out in the introduction, there are several candidate explanations – none mutually exclusive from one another – that warrant further study. The most basic explanations are based on gender differences in preferences, and we document evidence for one such differential in our paper. Beyond their differential "defensiveness" women may have a greater taste for probity, or a greater aversion to risk (e.g., Borghans et al. (2009)). Providing positive evidence for this *behavioral* mechanism does not rule out other channels through which gender differences in corruption may arise. As mentioned in the introduction, prior work shows that women may face greater leniency from enforcement authorities and the judicial system – while this evidence is based on prosecutions for other (non-corruption) criminal acts, there is reason to imagine that such gender differences could account at least in part for the gender differences in corruption investigations that we document. Differential selection may also contribute to our overall gender gap in corruption. It is easier to see how this could play a role in our Italian data, where the position of municipal procurement officer is relatively desirable for women versus men (given their outside options), and thus may attract higher-quality female candidates. We note, however, that earlier work finds no correlation between mental acuity and honesty (Hanna and Wang, 2017), so even if there are gender differences in selection on schooling

or mental ability, it does not necessarily imply selection on probity. The same argument may suggest gender differences in incentives: given their lesser outside opportunities, efficiency wage arguments may explain why women behave more honestly. One challenge to both incentive- and selection-based explanations is the consistent finding across both datasets – the officials we study in China are already high level officials, and if anything incentives for good behavior would be stronger for male city leaders, as they plausibly have stronger chances for promotion. Finally, men may simply have more opportunities for promotion, to the extent that corruption involves favor exchange that requires a network of co-conspirators. If women are less connected to such networks Fang and Huang (2017), they may have fewer opportunities for corruption.

We leave for future work the much larger enterprise of developing a broader set of explanations for the substantial gender differences in corruption that we document in this paper.

References

Afridi, Farzana, Vegard Iversen, and MR Sharan, "Women political leaders, corruption, and learning: Evidence from a large public program in India," *Economic Development and Cultural Change*, 2017, *66* (1), 1–30.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber, "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of political economy*, 2005, *113* (1), 151–184.

Baig, Faisal Ali, Xu Han, Zahid Hasnain, and Daniel Rogger, "Introducing the Worldwide Bureaucracy Indicators: A New Global Dataset on Public Sector Employment and Compensation," *Public Administration Review*, 2021, *81* (3), 564–571.

Bindler, Anna and Randi Hjalmarsson, "The persistence of the criminal justice gender gap: Evidence from 200 years of judicial decisions," *The Journal of Law and Economics*, 2020, *63* (2), 297–339.

Borghans, Lex, James J Heckman, Bart HH Golsteyn, and Huub Meijers, "Gender differences in risk aversion and ambiguity aversion," *Journal of the European Economic Association*, 2009, *7* (2-3), 649–658.

Brollo, Fernanda and Ugo Troiano, "What happens when a woman wins an election? Evidence from close races in Brazil," *Journal of Development Economics*, 2016, *122*, 28–45.

Chattopadhyay, Raghabendra and Esther Duflo, "Women as policy makers: Evidence from a randomized policy experiment in India," *Econometrica*, 2004, *72* (5), 1409–1443.

Decarolis, Francesco and Stefano Battini, "Indagine sull'Amministrazione Difensiva," *Rivista Italiana di Public Management*, 2020, *3* (2).

_ , Raymond Fisman, Paolo Pinotti, and Silvia Vannutelli, "Rules, Discretion, and Corruption in Procurement: Evidence from Italian Government Contracting," *Available at SSRN 3533368*, 2019.

Dollar, David, Raymond Fisman, and Roberta Gatti, "Are women really the "fairer" sex? Corruption and women in government," *Journal of Economic Behavior & Organization*, 2001, *46* (4), 423–429.

Fang, Lily Hua and Sterling Huang, "Gender and connections among Wall Street analysts," *The Review of Financial Studies*, 2017, *30* (9), 3305–3335.

Ferraz, Claudio and Frederico Finan, "Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes," *The Quarterly journal of economics*, 2008, *123* (2), 703–745.

Ferreira, Fernando and Joseph Gyourko, "Does gender matter for political leadership? The case of US mayors," *Journal of Public Economics*, 2014, *112*, 24–39.

Fischbacher, Urs and Franziska Föllmi-Heusi, "Lies in disguise—an experimental study on cheating," *Journal of the European Economic Association*, 2013, *11* (3), 525–547.

Fisman, Raymond and Yongxiang Wang, "The distortionary effects of incentives in government: Evidence from China's" death ceiling" program," *American Economic Journal: Applied Economics*, 2017, *9* (2), 202–18.

Gans-Morse, Jordan, Alexander Kalgin, Andrei Klimenko, Dmitriy Vorobyev, and Andrei Yakovlev, "Self-Selection into Public Service When Corruption is Widespread: The Anomalous Russian Case," *Comparative Political Studies*, 2021, *54* (6), 1086–1128.

Hanna, Rema and Shing-Yi Wang, "Dishonesty and selection into public service: Evidence from India," *American Economic Journal: Economic Policy*, 2017, *9* (3), 262–90.

Jacobsen, Catrine, Toke Reinholt Fosgaard, and David Pascual-Ezama, "Why do we lie? A practical guide to the dishonesty literature," *Journal of Economic Surveys*, 2018, *32* (2), 357–387.

Lee, Wang-Sheng and Cahit Guven, "Engaging in corruption: The influence of cultural values and contagion effects at the microlevel," *Journal of Economic Psychology*, 2013, *39*, 287–300.

Mocan, Naci, "What determines corruption? International evidence from microdata," *Economic Inquiry*, 2008, *46* (4), 493–510.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Pande, Rohini and Deanna Ford, "Gender quotas and female leadership," 2012.

Rosenbaum, Stephen Mark, Stephan Billinger, and Nils Stieglitz, "Let's be honest: A review of experimental evidence of honesty and truth-telling," *Journal of Economic Psychology*, 2014, *45*, 181–196.

Starr, Sonja B, "Estimating gender disparities in federal criminal cases," *American Law and Economics Review*, 2015, *17* (1), 127–159.

Swamy, Anand, Stephen Knack, Young Lee, and Omar Azfar, "Gender and corruption," *Journal of development economics*, 2001, *64* (1), 25–55.

Table 1: Summary statistics for the Italian data

	All	Male	Female	Difference
	(1)	(2)	(3)	(3)-(2)
Female	0.17			
	(0.38)			
Investigated	0.09	0.09	0.05	-0.04 ***
	(0.28)	(0.29)	(0.22)	[0.000]
Age RUP (in 2018)	57.83	58.97	52.45	-6.52 ***
	(9.43)	(9.16)	(8.83)	[0.000]
RUP born in same Region	0.87	0.88	0.84	-0.04 ***
	(0.33)	(0.32)	(0.37)	[0.000]
RUP born in same Municipality	0.28	0.30	0.18	-0.11 ***
	(0.45)	(0.46)	(0.39)	[0.000]
Politician	0.13	0.14	0.09	-0.05 ***
	(0.34)	(0.35)	(0.29)	[0.000]
Tot. Auctions managed by RUP	6.19	6.34	5.52	-0.82 ***
	(8.71)	(8.87)	(7.86)	[0.000]
Tot. Auctions managed by PA	153.26	151.43	161.95	10.52
	(553.26)	(544.37)	(593.66)	[0.320]
Tot. RUP in PA	18.59	18.40	19.48	1.08
	(60.44)	(59.54)	(64.57)	[0.351]
Average Value of Auctions (in 0000)	567.28	569.02	559.05	-9.96
	(3094.68)	(3283.00)	(1969.28)	[0.807]
Average Number of Bidders	25.22	25.96	21.68	-4.28 ***
	(35.58)	(36.48)	(30.71)	[0.000]
Area=North	0.51	0.47	0.65	0.17 ***
	(0.50)	(0.50)	(0.48)	[0.000]
Area=Center	0.16	0.17	0.15	-0.01 *
	(0.37)	(0.37)	(0.36)	[0.086]
Area=South	0.33	0.36	0.20	-0.16 ***
	(0.47)	(0.48)	(0.40)	[0.000]
Population (log)	13.39	13.39	13.41	0.02
	(0.84)	(0.83)	(0.88)	[0.172]
Observations	21,277	17,574	3,703	21,277

Notes: *Investigated* is an indicator variable equal to 1 if the contracting officer overseeing each contract (the *"Responsabile Unico del Procedimento"*, or RUP) has been investigated for corruption. PA stands for Procurement Authority. The sample includes the universe of RUP-PA unique pairs. See text for further details. Standard deviations in parentheses. P-values in squared brackets. Significance: $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Table 2: Summary statistics for the Chinese data

	All	Male	Female	Difference
	(1)	(2)	(3)	(3)-(2)
	Panel A: Full Sample			
Female	0.04			
	(0.19)			
Investigated for Corruption	0.08	0.08	0.03	-0.04 **
	(0.26)	(0.27)	(0.18)	0.011
Masters	0.32	0.30	0.59	0.29 ***
	(0.46)	(0.46)	(0.49)	0.000
Doctor	0.10	0.10	0.12	0.03
	(0.30)	(0.30)	(0.33)	0.360
Year of First Appointment	1997.56	1997.34	2003.13	5.80 ***
	(9.83)	(9.80)	(8.83)	0.000
Age	56.98	56.85	60.09	3.24 **
	(27.40)	(27.77)	(15.28)	0.029
Observations	3133	3013	120	3133
	Panel B: Started post $\geqslant$ 1998			
Female	0.05			
	(0.22)			
Investigated for Corruption	0.12	0.12	0.04	-0.08 ***
	(0.32)	(0.33)	(0.20)	[0.000]
Masters	0.48	0.47	0.64	0.17 ***
	(0.50)	(0.50)	(0.48)	[0.001]
Doctor	0.16	0.16	0.15	-0.01
	(0.37)	(0.37)	(0.36)	[0.756]
Year of First Appointment	2004.11	2004.00	2006.18	2.18 ***
	(5.90)	(5.90)	(5.59)	[0.000]
Age	59.59	59.69	57.79	-1.90 *
	(13.19)	(13.32)	(10.39)	[0.079]
Observations	1878	1778	100	1878

Notes: The sample in Panel A is the set of Chinese officials who held the position of prefecture mayor or party secretary during 1979-2014. In Panel B the sample is limited to individuals who started such a position 1998 or later. *Investigated* is an indicator variable denoting that the official was investigated for corruption. *Year of First Appointment* is the year the official first held a position of prefecture mayor or party secretary. See text for further details. Standard deviations in parentheses. P-values in squared brackets. Significance: * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Probability of Italian RUP investigation as a function of gender

	Dependent variable: Investigated					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0338***	-0.0261***	-0.0187***	-0.0165**	-0.0176***	-0.0155**
	[0.00560]	[0.00566]	[0.00564]	[0.00642]	[0.00653]	[0.00679]
Age (log)		0.0539***	0.0847***	0.0828***		0.0856***
		[0.0144]	[0.0144]	[0.0167]		[0.0177]
RUP born in same		-0.0187***	-0.0130**	-0.0130**	-0.0163***	-0.0128**
Municipality		[0.00503]	[0.00505]	[0.00603]	[0.00611]	[0.00611]
Tot. Auctions			0.0320***	0.0334***		0.0346***
managed by RUP (log)			[0.00216]	[0.00242]		[0.00253]
Average Value of			0.00822***	0.00201		0.00181
Auctions (log)			[0.00269]	[0.00324]		[0.00345]
Average Number of			-0.0114***	-0.00797***		-0.00867***
Bidders (log)			[0.00195]	[0.00245]		[0.00254]
RegionXYear FE	Yes	Yes	Yes	Yes	Yes	Yes
PA Controls	No	No	Yes	No	No	No
RUP Controls	No	Yes	Yes	Yes	Yes	Yes
PA FE	No	No	No	Yes	Yes	Yes
Age & N.Contracts FE	No	No	No	No	Yes	No
Muni only	No	No	No	No	No	Yes
Dep. Var. Mean	0.0867	0.0867	0.0867	0.0802	0.0802	0.0823
Observations	21277	21277	21277	17830	17822	16773
Adjusted R-sq	0.0704	0.0773	0.0963	0.371	0.381	0.374

Notes:The dependent variable, *Investigated* is an indicator equal to 1 is the public official in charge of the auction (the RUP) has been investigated. The analysis is conducted on a panel of RUP-PA observations. PA Controls include a set of dummies for the type of PA (Central, Region and other local authority, Hospitals and Universities, Transportation), the log of the total number of auctions done by the PA during the sample period, the log of the total number of RUPs observed in the PA during the sample period, the log of the total number of auctions managed by the RUP, the log of the average value of auctions managed by the RUP and the log of the average number of bidders participating in auctions managed by the RUP. Individual Controls include the log of the age, an indicator for whether the RUP was born in the same city where she operates as a RUP, an indicator for whether the RUP was ever elected as a local politician, and a set of dummies for the region of birth of the RUP. Robust standard errors clustered at the RUP level are in parentheses. Significance: $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Table 4: Probability of Italian RUP investigation as a function of gender, heterogeneity analyses

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female=1	-0.0300***	-0.0161***	-0.0356***	-0.0272***	-0.0140*	-0.0149*	-0.0144	0.00130	0.00958
	[0.00653]	[0.00618]	[0.00599]	[0.00635]	[0.00773]	[0.00816]	[0.00878]	[0.00970]	[0.0208]
Female=1 ×	0.0324**								0.0236
Born in Municipality=1	[0.0135]								[0.0193]
Female=1 ×		-0.0392*							-0.00228
Local Politician=1		[0.0210]							[0.0286]
Female=1 ×			0.0471***						-0.00504
South=1			[0.0159]						[0.0272]
Female=1 ×				0.0267					0.0168
Mafia Municipality=1				[0.0209]					[0.0238]
Female=1 ×					-0.0219***				0.0108
Female LFP (SD)					[0.00810]				[0.0160]
Female=1 ×						-0.0138			-0.0165
Share Female RUPs (SD)						[0.0112]			[0.0197]
Female=1 ×							-0.0147*		-0.00695
Share Female PA (SD)							[0.00766]		[0.00874]
Female=1 ×								-0.0600***	-0.0772***
Avg. Income (SD)								[0.0154]	[0.0263]
RegionXYear FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PA Controls	No	No	No	No	No	No	No	No	No
Indiv. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Muni only	No	No	No	No	No	No	No	No	No
Dep. Var. Mean	0.0823	0.0823	0.0828	0.0827	0.0829	0.0823	0.0884	0.0824	0.0892
Observations	16775	16775	16510	16294	16140	16775	10132	16767	9522
R-sq	0.347	0.348	0.346	0.344	0.347	0.346	0.349	0.347	0.353

Notes:The dependent variable, $Investigated$ is an indicator equal to 1 is the public official in charge of the auction (the RUP) has been investigated. $BorninMunicipality$ is a dummy equal to 1 if the RUP was born in the same municipality in which she works as RUP. $LocalPolitician$ is a dummy equal to 1 if the RUP was ever elected as a local politician. $South$ is a dummy equal to 1 if the PA is located in one of the Southern Regions of Italy (Abruzzo, Basilicata, Campania, Calabria, Molise, Puglia, Sardinia, Sicily). $MafiaMunicipality$ is a dummy equal to 1 if the PA is located in a municipality ever subject to mafia investigations. $FemaleLFP(SD)$ is the standardized value of Female Labor Force Participation in the municipality where the PA is located. $ShareFemaleRUPs(SD)$ is the standardized value of the share of female RUPs in the PA. $ShareFemalePA(SD)$ is the standardized value of the share of female employees in the PA. $Avg.Income(SD)$ is the standardized value of the average income (from tax revenues) in the municipality where the PA is located. The analysis is conducted on a panel of RUP-PA observations. PA Controls include a set of dummies for the type of PA (Central, Region and other local authority, Hospitals and Universities, Transportation), the log of the total number of auctions done by the PA during the sample period, the log of the total number of RUPs observed in the PA during the sample period, the log of the total number of auctions managed by the RUP, the log of the average value of auctions managed by the RUP and the log of the average number of bidders participating in auctions managed by the RUP. Individual Controls include the log of the age, an indicator for whether the RUP was born in the same city where she operates as a RUP, an indicator for whether the RUP was ever elected as a local politician, and a set of dummies for the region of birth of the RUP. Robust standard errors clustered at the RUP level are in parentheses. Significance: $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Table 5: Selection on Unobservables in Italian Data

	Baseline Effect [R2]	Controlled Effect [R2]	Identified Set	δ for $\beta = 0$ given Rmax
	(1)	(2)	(3)	(4)
		Panel A: Without Municipality FE		
Female	-0.033 [0.070]	-0.020 [0.092]	[-0.003,-0.020]	1.11
		Panel B: Including Municipality FE		
Female	-0.033 [0.070]	-0.018 [0.504]	[-0.006,-0.018]	1.39

We perform the test suggested by Oster (2019). In Panel A, Column 1, we report the coefficient from our baseline regression, including only region-times-year fixed effects. In Column 2, we report the coefficient from our regression, including all available controls (except for PA fixed effects). In Column 3, we calculate the identified set for the bias-adjusted treatment effect, using the suggested values of $R_{max} = 1.3R$ and $\delta = 1$. In Column 4, we calculate the value of δ s.t. $\beta = 0$, assuming again that $R_{max} = 1.3R$, where R is the observed R-squared from the regression in Column 2. Panel B repeats the same exercise but includes PA fixed effects for the regression in Column 2. R_{max} is the R-squared of a hypothetical regression including all observables and unobservables, while δ represents the degree of selection on observed versus unobserved variables. For further details, see Section III.A.1.

Table 6: Defensive Actions and Insurance in Italian Survey Data

	Defensive Action			Own Insurance		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	5.484***	4.173**	4.237**	-11.34**	-14.33***	-13.12**
	[1.711]	[1.852]	[1.898]	[4.962]	[5.307]	[6.015]
Dep. var. Mean	31.05	30.72	29.49	36.86	37.05	39.18
Dep. var. SD	18.95	18.39	17.04	48.30	48.36	48.90
Demographics	No	Yes	Yes	No	Yes	Yes
Risk	No	No	Yes	No	No	Yes
Observations	508	402	323	388	359	291
R-sq	0.0226	0.0302	0.0721	0.0531	0.0949	0.0912

Notes: In Columns 1-3, the dependent variable, $DefensiveActions$, is defined as the average probability of choosing the "defensive option" across the four scenario questions (see Appendix B). Demographics controls include tenure in the public sector, age, age squared, as well as fixed effects for the type of PA the RUP works at; Risk controls include RUP risk references as captured by survey responses (see Appendix B); The dependent variable in Columns 4-6,$Insurance$, reflects whether the RUP has insurance against the risk of administrative damages. Significance: $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Table 7: Probability of Chinese officials' investigation as a function of gender

	Dependent variable: Investigated				
	Full Sample			Started post $\geqslant$ 1998	
	(1)	(2)	(3)	(4)	(5)
Female	-0.069***	-0.067***	-0.079***	-0.092***	-0.086***
	[0.0177]	[0.0176]	[0.0178]	[0.0215]	[0.0217]
Doctor			0.057*	0.053	0.038
			[0.0252]	[0.0278]	[0.0269]
Masters			0.038*	0.041*	0.030
			[0.0154]	[0.0190]	[0.0193]
No Degree			0.007	0.010	0.026
			[0.0127]	[0.0287]	[0.0298]
Age (log)			-0.232***	-0.303*	
			[0.0688]	[0.134]	
Age Missing			-1.047***	-1.344*	
			[0.305]	[0.570]	
Province FE	No	Yes	Yes	Yes	Yes
Start year FE	Yes	Yes	Yes	Yes	Yes
Age Cohort FE	No	No	No	No	Yes
Dep. Var. Mean	0.075	0.075	0.075	0.117	0.121
Observations	3133	3133	3133	1874	1807
Adjusted R-sq	0.044	0.044	0.056	0.027	0.028

Notes: The sample in columns (1)-(3) is the set of Chinese officials who held the position of prefecture mayor or party secretary during 1979-2014; in columns (4) and (5) the sample is limited to individuals who started such a position 1998 or later. The outcome in all columns is an indicator variable denoting that the official was investigated for corruption. Please see text for further details. Robust standard errors in parentheses. Significance: $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Table 8: Selection on Unobservables in Chinese Data

	Baseline Effect [R2]	Controlled Effect [R2]	Identified Set	δ for $\beta = 0$ given Rmax
	(1)	(2)	(3)	(4)
		Panel A: Without Province FE		
Female	-0.069 [0.041]	-0.081 [0.053]	[-0.097,-0.081]	-1.16
		Panel B: Including Province FE		
Female	-0.069 [0.041]	-0.080 [0.065]	[-0.089,-0.080]	-7.95

We perform the test suggested by Oster (2019). In Panel A, Column 1, we report the coefficient from our baseline regression, including cohort fixed effects. In Column 2, we report the coefficient from our regression, including all available controls (except for Province fixed effects). In Column 3, we calculate the identified set for the bias-adjusted treatment effect, using the suggested values of $R_{max} = 1.3R$ and $\delta = 1$. In Column 4, we calculate the value of δ s.t. $\beta = 0$, assuming again that $R_{max} = 1.3R$, where R is the observed R-squared from the regression in Column 2. Panel B repeats the same exercise but including Province fixed effects for the regression in Column 2. R_{max} is the R-squared of a hypothetical regression including all observables and unobservables, while δ represents the degree of selection on observed versus unobserved variables. For further details, see Section III.A.1.

A Additional Tables and Figures

Table A.1: Probability of Italian RUP investigation as a function of gender, excluding control for whether RUP had been a local politician

	Dependent variable: Investigated					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0338***	-0.0295***	-0.0214***	-0.0188***	-0.0176***	-0.0178***
	[0.00560]	[0.00569]	[0.00565]	[0.00644]	[0.00653]	[0.00680]
Age (log)		0.0599***	0.0907***	0.0875***		0.0909***
		[0.0145]	[0.0145]	[0.0168]		[0.0177]
RUP born in same		-0.0228***	-0.0158***	-0.0151**	-0.0163***	-0.0149**
Municipality		[0.00507]	[0.00507]	[0.00605]	[0.00611]	[0.00613]
Tot. Auctions			0.0324***	0.0336***		0.0349***
managed by RUP (log)			[0.00217]	[0.00242]		[0.00253]
Average Value of			0.00818***	0.00208		0.00191
Auctions (log)			[0.00270]	[0.00325]		[0.00347]
Average Number of			-0.0119***	-0.00820***		-0.00891***
Bidders (log)			[0.00196]	[0.00245]		[0.00254]
RegionXYear FE	Yes	Yes	Yes	Yes	Yes	Yes
PA Controls	No	No	Yes	No	No	No
RUP Controls	No	Yes	Yes	Yes	Yes	Yes
PA FE	No	No	No	Yes	Yes	Yes
Age & N.Contracts FE	No	No	No	No	Yes	No
Muni only	No	No	No	No	No	Yes
Dep. Var. Mean	0.0867	0.0867	0.0867	0.0802	0.0802	0.0823
Observations	21277	21277	21277	17830	17822	16773
Adjusted R-sq	0.0704	0.0738	0.0936	0.370	0.381	0.372

Notes:The dependent variable, *Investigated* is an indicator equal to 1 is the public official in charge of the auction (the RUP) has been investigated. The analysis is conducted on a panel of RUP-PA observations. PA Controls include a set of dummies for the type of PA (Central, Region and other local authority, Hospitals and Universities, Transportation), the log of the total number of auctions done by the PA during the sample period, the log of the total number of RUPs observed in the PA during the sample period, the log of the total number of auctions managed by the RUP, the log of the average value of auctions managed by the RUP and the log of the average number of bidders participating in auctions managed by the RUP. Individual Controls include the log of the age, an indicator for whether the RUP was born in the same city where she operates as a RUP, and a set of dummies for the region of birth of the RUP. Robust standard errors clustered at the RUP level are in parentheses. Significance: $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Table A.2: Gender Differences in the type of auctions

	(1) DiscretProc	(2) $DiscretProc_{lowN}$	(3) DiscretProc	(4) $DiscretProc_{lowN}$
Female	0.0168*	0.0113**	-0.0146	-0.000689
	[0.00874]	[0.00490]	[0.00960]	[0.00548]
Population (log)	-0.191	-0.140***		
	[0.210]	[0.0471]		
Population Sq. (log)	0.00599	0.00570***		
	[0.00798]	[0.00175]		
Constant	1.854	0.906***	0.372***	0.0544***
	[1.380]	[0.317]	[0.00151]	[0.000864]
Dep. Var. Mean	0.225	0.225	0.225	0.225
Observations	109875	109875	108717	108717
R-sq	0.442	0.108	0.577	0.249
Geog. FE			PA	PA

Notes: The dependent variable is indicated on top of each column.$DiscretProc$ denotes negotiated procedures. $DiscretProc_{lowN}$ denotes negotiated procedures with fewer than the legally mandated number of bidders. The analysis is conducted at the contract level. All regressions include Year fixed effects, a linear control for reserve price (in log) Price and 5 dummies for different contract size thresholds (up to 100k, 100-150k, 150-300k, 300-500k, 500k-1mil, 1-1.5mil, over 1.5mil) as well as controls for contract characteristics: 4 dummies for category type (Civil Building, Roadworks, Specialized Works or Others), 1 dummy for whether the contract was awarded under urgency and 1 dummy for whether the object of the contract entailed maintenance. Robust standard errors clustered at the PA level are in parentheses. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. Robust standard errors clustered at the administration level are in parentheses.Significance: $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table A.3: Summary statistics for the Italian survey data

	All	Male	Female	Difference
	(1)	(2)	(3)	(3)-(2)
Female	0.47			
	(0.50)			
Defensive action (avg.)	30.56	28.51	32.84	4.33 **
	(18.76)	(20.29)	(16.64)	[0.015]
Time devoted to check compliance with regulations	68.24	62.21	74.83	12.62 ***
	(28.13)	(29.09)	(25.52)	[0.000]
Risk of Proceedings	51.78	50.58	53.18	2.61
	(28.63)	(28.48)	(28.81)	[0.377]
Risk of Proceedings, ANAC	41.91	40.60	43.43	2.83
	(29.35)	(28.22)	(30.62)	[0.359]
Risk of Proceedings, Court of Accounts	47.66	46.81	48.66	1.85
	(31.43)	(30.96)	(32.03)	[0.572]
Risk of Proceedings, Judiciary	44.28	45.12	43.23	-1.90
	(33.64)	(33.42)	(33.98)	[0.593]
Risk of Proceedings, higher admin	42.93	40.71	45.58	4.86
	(31.73)	(31.44)	(31.96)	[0.139]
Media Risk	46.62	44.94	48.55	3.61
	(28.39)	(29.29)	(27.26)	[0.200]
Complexity of Compliance with regulation	74.74	75.04	74.40	-0.64
	(20.00)	(19.77)	(20.30)	[0.743]
Knowledge of regulation	65.52	67.20	63.66	-3.54 *
	(20.80)	(19.66)	(21.90)	[0.083]
Share of own defensive actions	30.60	31.12	29.96	-1.16
	(28.33)	(27.88)	(28.94)	[0.701]
Share of defensive actions in the PA	51.08	51.80	50.18	-1.62
	(27.53)	(28.11)	(26.83)	[0.578]
Change of share in the last 5 years	29.10	32.53	24.23	-8.30 *
	(39.13)	(38.85)	(39.15)	[0.068]
Risk of civil responsibility sactions	42.96	43.09	42.81	-0.28
	(27.25)	(28.27)	(26.08)	[0.919]
Risk of admin responsibility sactions	53.30	51.83	55.04	3.21
	(28.82)	(29.78)	(27.62)	[0.265]
Has Insurance from PA for civil responsibility	17.13	20.61	13.24	-7.38 **
	(37.72)	(40.54)	(33.97)	[0.040]
Is aware of insurance for admin responsibility	0.79	0.78	0.80	0.02
	(0.41)	(0.41)	(0.40)	[0.553]
Has insurance for admin responsibility	36.75	44.44	28.42	-16.03 ***
	(48.27)	(49.82)	(45.22)	[0.001]
Observations	420	224	196	420

Notes: All variables (except Female) vary from 0 to 100. Standard deviations in parentheses. P-values in squared brackets. Significance: $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Table A.4: Probability of Chinese officials' investigation under the 2013 anti-corruption crackdown, as a function of gender

	Dependent variable: Investigated				
	Full Sample			Started post $\geqslant$ 1998	
	(1)	(2)	(3)	(4)	(5)
Female	-0.069***	-0.067***	-0.079***	-0.092***	-0.086***
	[0.0177]	[0.0176]	[0.0178]	[0.0215]	[0.0217]
Doctor			0.057*	0.053	0.038
			[0.0252]	[0.0278]	[0.0269]
Masters			0.038*	0.041*	0.030
			[0.0154]	[0.0190]	[0.0193]
No Degree			0.007	0.010	0.026
			[0.0127]	[0.0287]	[0.0298]
Age (log)			-0.232***	-0.303*	
			[0.0688]	[0.134]	
Age Missing			-1.047***	-1.344*	
			[0.305]	[0.570]	
Province FE	No	Yes	Yes	Yes	Yes
Start year FE	Yes	Yes	Yes	Yes	Yes
Age Cohort FE	No	No	No	No	Yes
Dep. Var. Mean	0.075	0.075	0.075	0.117	0.121
Observations	3133	3133	3133	1874	1807
Adjusted R-sq	0.044	0.044	0.056	0.027	0.028

Notes: The sample in columns (1)-(3) is the set of Chinese officials who held the position of prefecture mayor or party secretary during 1979-2014; in columns (4) and (5) the sample is limited to individuals who started such a position 1998 or later. The outcome in all columns is an indicator variable denoting that the official was investigated for corruption. Please see text for further details. Robust standard errors in parentheses. Significance: $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

B Additional Details on Italian Survey Data

We provide here a translation of the survey conducted by Francesco Decarolis and Stefano Battini on a sample of approximately 500 RUPs. The survey measures risk attitudes and how they translate into what is called "defensive behavior," namely the propensity of the RUP to be cautious in the management of the auction. The survey also asks whether the RUP has purchased insurance against the risk of being charged for administrative damages.

II.A Scenario Questions to assess defensive behavior

The following set of questions were used to assess defensiveness. In each question, the RUP is asked to think about how she would behave in a realistic decision-making problem, in which one of the actions is significantly more "defensive" than the other. For each question, we code the option in bold equal to 1 (i.e., more defensive), and 0 otherwise. We then construct an average index of defensiveness taking the average across responses to the four scenarios.

Scenario 1 Suppose you are appointed as Commissioner for implementation of extraordinary maintenance works. After the announcement of the winner of this tender, the award is challenged by another participant before the administrative judge. The latter decides, during the precautionary phase, not to order the suspension of the award, given the extraordinary nature of the work to be carried out. While waiting for the final judgment of the Council of State, the award provision is therefore fully effective. How do you think you would behave in this situation? **a) wait for the final judgment, even though it will delay the work** b) continue with the work, following what has been decided by the judge c) don't know.

Scenario 2 Now suppose you are the manager of a service in a local authority. You are informed about the possibility of replacing paper forms that your institution currently requires citizens to fill out, with a leaner online procedure. However, it is a new procedure, never implemented in any other similar local authority. How do you think it would behave in this situation? a) accept the use of the online procedure **b) wait for a similar local authority to implement the procedure and evaluate its impacts** c) don't know.

Scenario 3 Now suppose you are a member of a procurement awarding committee. There are many bidders who presented offers and the other members prefer to devote more time than what is requested by law in the candidate selection phase. However, this could be contested by the Court of Accounts on the basis of public damage resulting from unjustified delays in the contracting phase. How do you think you would behave in this situation? a) try to speed up the procedure **b) follow other members' approach** c) don't know.

Scenario 4 Now suppose you are the Commissioner in charge of implementation of an infrastructure project. You need to cooperate with a company of the Ministry of Economics and Finance (MEF) to evaluate the socio-economic impacts of the investment. The MEF authorizes the agreement. However, it specifies that the agreement is valid but suggests, "it could be meaningful to obtain an opinion from ANAC (Anti-corruption Authority) about correctness of

the instrument identified." How do you think you would behave in this situation? a) proceed without asking for ANAC opinion **b) ask for ANAC opinion and proceed only in the presence of a positive response** c) don't know.

II.B Risk Preferences and Perception

Q1: In general, in carrying out your work, what percentage of your time is devoted to making sure your decisions are in line with the requirements of the regulations on transparency and anti-corruption?

Q2: In your opinion, for a RUP that carries out his business in a PA similar to the one you work in, what is the risk that the surveillance and control action of one of the following actors can lead to the initiation of disciplinary proceedings – administrative or legal – for conduct related to the management of public procurement?

- Overall Risk perceived
- Risk of proceedings from ANAC (anti-corruption authority)
- Risk of proceedings from the National Court of Accounts
- Risk of proceedings from Judiciary
- Risk of proceedings from higher administration

Q3: In your opinion, does a RUP feel exposed to risk that the management of the contracts for which he is responsible may be questioned by the media (newspapers or television)?

II.C Insurance

Q1: Has the organization where you are currently employed taken steps to insure you against risks for third party civil liability?

Q2: Are you aware of the existence of private insurance policies for administrative-accounting responsibility?

Q3: Do you currently have a private insurance policy for administrative-accounting responsibility?