NBER WORKING PAPER SERIES

DO FEMALE LEADERS REDUCE CORRUPTION? NEW EVIDENCE FROM BRAZIL

Julieta Peveri
Clemence Tricaud

Working Paper 35308
http://www.nber.org/papers/w35308

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2026

We thank Ricardo Dahis, Lucie Gadenne, Laurine Martinoty, Nico Voigtlaender, and Melanie Wasserman for their valuable comments and suggestions. We thank Geer Ang and Louis Marie for excellent research assistance. The project leading to this publication has received funding from the French National Research Agency (grant ANR-25-CE26-4450). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Do Female Leaders Reduce Corruption? New Evidence from Brazil
Julieta Peveri and Clemence Tricaud
NBER Working Paper No. 35308
June 2026
JEL No. D72, D73, H11, J16

ABSTRACT

Does female leadership reduce corruption? We study this question using close mixed-gender elections in Brazilian municipalities over two decades and multiple measures of corruption: budget-based predicted corruption scores, audit irregularities, and legal sanctions. We find no evidence that electing a woman mayor affects corruption. This null result holds across time periods, mayor characteristics, and the electoral cycle. We only detect a negative effect in the small subsample of municipalities randomly audited in early terms, which coincides with a sharp imbalance in incumbency. Because incumbency directly impacts corruption, these previously documented effects likely reflect the impact of incumbency rather than gender.

Julieta Peveri
Université Paris I Panthéon Sorbonne
Julieta.Peveri@univ-paris1.fr

Clemence Tricaud
University of California, Los Angeles
Anderson School of Management
and NBER
clemence.tricaud@anderson.ucla.edu

1 Introduction

Corruption poses a fundamental challenge to effective governance, undermining public trust, distorting resource allocation, and weakening state capacity (e.g., Fisman and Svensson 2007; Rothstein 2011; Rose-Ackerman and Palifka 2016). The growing political representation of women has raised the question of whether female leaders govern differently from men, particularly in their propensity to engage in corruption.

Providing a causal answer to this question is difficult for at least two reasons. First, the gender of elected leaders is endogenous to political, social, and institutional conditions. Places that elect women may differ systematically from those that elect men, and female politicians themselves may differ from male politicians along characteristics that also matter for corruption. Second, corruption is hard to measure. This paper addresses both challenges by studying two decades of municipal elections in Brazil, a decentralized country where mayors play a central role in public service delivery and where institutional efforts to detect and sanction corruption have generated unusually rich data.

To assess the causal impact of female leadership, we exploit thousands of close mixed-gender mayoral races and compare municipalities in which a woman barely wins against a man to those in which a man barely wins against a woman. We provide extensive evidence supporting the validity of this design: baseline municipal characteristics are smooth at the threshold, and closely elected male and female mayors are similar in terms of political experience and political orientation. Female mayors are slightly younger and significantly more educated at the threshold, but we show that these differences are unlikely to drive our results. Controlling for these characteristics leaves our estimates unchanged, and we find no independent effect of education on corruption outcomes.

To capture corruption, we assembled a broad set of measures at the municipal level: (i) predicted corruption scores based on budget accounts for the universe of municipalities, generated by extending Ash et al. (2025)'s machine learning method, (ii) corruption-related irregularities recorded by the federal randomized anti-corruption audit program, and (iii) newly collected data on legal sanctions for administrative misconduct.

Our main finding is that, over the 2000–2020 period and across six corruption measures, electing a woman mayor has no detectable effect on corruption. Having a female mayor has no significant effect on the predicted corruption score, the number of audit irregularities, the probability that the mayor is legally sanctioned, suspended from political rights, the amount of financial damage to be reimbursed, or the amount of fines imposed. Focusing on the predicted corruption score—our preferred outcome, as it provides a continuous measure of corruption available for the full sample—we can reject negative effects larger than 3.3

percent of the mean and positive effects larger than 5.5 percent.

We then examine whether this null effect masks heterogeneity along mayor characteristics or electoral context. We first show that female mayors appear no more or less corrupt than male mayors regardless of their education level, political experience, political orientation, or age. Second, we do not detect any gender differences within the electoral cycle either, by looking at the effect of female leadership on corruption separately for each year of the term. Finally, turning to different time periods and audit status, we only find a large and significant effect in one specific sample: municipalities that were randomly audited over the first two electoral terms (2000 and 2004). In this small subsample, female mayors appear less corrupt, in line with Brollo and Troiano (2016), who study this precise setting. However, this pattern disappears once later elections are included or when the analysis is extended to a broader set of municipalities.

When investigating the specificity of this sample, we uncover the role of incumbency as a potential confounder. In audited municipalities during the early electoral terms—and only in this subsample—women who narrowly win are substantially less likely to be incumbents than men who narrowly win. This matters because incumbency is likely to affect corruption directly. Incumbents may have greater opportunities for corruption due to experience and networks, and they also face weaker electoral accountability—particularly in Brazil, where a two-term limit prevents second-term mayors from running again (Ferraz and Finan 2011). We provide several pieces of evidence highlighting the role of incumbency. First, revisiting Brollo and Troiano (2016)'s findings, we show that the gender effect they uncover is no longer significant once we restrict the sample to first-term mayors or control for winner characteristics. Second, we corroborate the role of incumbency by conducting a separate regression discontinuity analysis that compares narrowly elected male incumbents with narrowly elected male non-incumbents. Incumbency increases corruption both within Brollo and Troiano's sample and in our full sample, across all our corruption measures.

Overall, leveraging a large set of municipalities, electoral terms, and corruption measures, we show that female mayors in Brazil are no more or less likely to be corrupt than male mayors. More broadly, our findings highlight the importance of systematically examining discontinuities in other covariates, to avoid conflating gender with other key drivers of political behavior.

Our paper first contributes to the literature studying the effects of women's political representation on corruption. A large empirical literature documents that greater representation of women in politics is associated with lower levels of corruption, both across and within countries (Dollar et al. 2001; Swamy et al. 2001; Jha and Sarangi 2018; Bauhr et al. 2019;

Decarolis et al. 2023).[1] However, this relationship has been shown to depend on institutional context, including the degree of accountability (Esarey and Schwindt-Bayer 2018), and to be sensitive to the inclusion of cultural factors (Debski et al. 2018). To address these endogeneity concerns, recent studies have exploited close-election regression discontinuity designs in local elections. Using this approach, Brollo and Troiano (2016), Bauhr and Charron (2021), and Baskaran et al. (2024) document that female leadership reduces corruption in Brazilian municipalities, French municipalities, and Indian states, respectively.[2] These studies face two main limitations. First, corruption is generally captured using a single and often indirect measure, such as perceived corruption based on survey data, the share of single-bid procurement contracts, or politicians' asset growth. Brollo and Troiano (2016) rely on a binary indicator for corruption based on random federal audits of Brazilian municipalities. While audits provide a direct measure of corruption, only a small share of municipalities are randomly audited. Second, previous studies rely on small analysis samples, reflecting the data limitations described above, and the fact that they typically focus on only one or two electoral terms.[3]

In contrast, we leverage several measures of corruption over a longer time horizon. On top of audit irregularities, we use predicted corruption scores based on budget accounts and newly collected data on legal sanctions and financial penalties. We can thus assess the effect of female leadership on different measures of corruption that come from distinct sources and capture both the intensive and extensive margins. Moreover, these measures are available for all Brazilian municipalities and we exploit elections over two decades, yielding thousands of observations. Using a large sample and a wide array of outcomes, we show that female leadership has no detectable effect on corruption.

We also contribute more generally to the literature using close-election designs to estimate the effects of female leadership.[4] Our findings speak to recent work emphasizing that politician-characteristic RDDs can be biased when other candidate characteristics that matter for the outcomes also change discontinuously at the threshold (Marshall 2024). We show that in the small subsample of early-audited municipalities studied in Brollo and Troiano (2016), female mayors are significantly less likely to be in their second term. Given that incumbency has a direct positive effect on corruption, the estimate likely reflects the effect

[1] See Alexander (2021) for a review.

[2] Instead of relying on close elections, Afridi et al. (2017) exploit the random assignment of village council headship in India. In contrast to Baskaran et al. (2024), they find that female leadership initially increases corruption, with the effect vanishing over time.

[3] The effective sample sizes (i.e., the number of observations within the estimation bandwidths) are 103, 283, and 59 for Brollo and Troiano (2016), Bauhr and Charron (2021), and Baskaran et al. (2024), respectively.

[4] See Hessami and da Fonseca (2020) for a review.

of incumbency instead of gender. The same mechanism may also help explain Bauhr and Charron (2021)'s findings, as the gender effect they uncover on procurement dissipates once they focus on female mayors in their second term.

Finally, we contribute to the literature on the effects of incumbency and term limits on corruption. We extend Ferraz and Finan (2011), whose findings were based on the 2000 election, to additional electoral terms, and we complement Dahis et al. (2025) by drawing on measures of corruption that go beyond audit results and are thus available for all municipalities.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting and data. Section 3 outlines our empirical strategy. Section 4.1 presents the main results. Section 4.2 examines heterogeneity effects, and Section 4.3 uncovers the role of incumbency in explaining the early-audited sample result.

2 Context and Data

2.1 Brazilian Municipalities

Brazil counts 5,570 municipalities, recognized by the Constitution as autonomous federal entities responsible for providing local public services—including sanitation, transportation, education, and health care. In 2024, municipalities accounted for 24.2% of public spending, and 67.2% of their revenues came from constitutionally mandated intergovernmental transfers (Instituto Brasileiro de Geografia e Estatística 2024).

Municipal governments are made of an executive branch led by the mayor and a legislative branch, the municipal council. While the council can amend budgets and propose legislation, the mayor retains broad discretion over spending, veto power over bills, and a dominant role in policymaking.

Voting is mandatory for adults aged 18 to 70. Mayors are elected every four years under a majority-rule system, with runoffs in municipalities above 200,000 voters if no candidate wins more than 50% of valid votes. Since 1997, mayors may serve at most two consecutive terms, implying that second-term mayors are not eligible for reelection.

Women remain substantially underrepresented. In 2000, the first election in our sample, only 7.5% of candidates and 5.6% of elected mayors were women. While these shares rose through the 2000s, they have plateaued since the 2012 election at around 12% and 10%, respectively.

2.2 Municipal Corruption

Municipal corruption in Brazil commonly involves procurement fraud (e.g., over-invoicing), diversion of funds, payment for unexecuted works or services, the use of shell companies, and falsified receipts. Based on estimates by Ferraz and Finan (2011), corruption in local governments results in losses of approximately US$550 million per year.

To address this problem, the federal government created the *Controladoria-Geral da União* (CGU) in 2003 to audit municipalities' use of federal transfers. Between 2003 and 2015, municipalities were selected at random through public lotteries; since then, selection has been based on a vulnerability score. Once selected, teams of auditors inspect financial accounts, procurement procedures, and the delivery of federally funded public services.

Audits are not the only channel through which corruption is detected. Under Brazil's Improbity Law (*Lei de Improbidade Administrativa*), public officials may be held civilly liable for acts that result in illicit enrichment, cause damage to the public treasury, or violate core principles of public administration. Investigations into acts of administrative improbity can be initiated following complaints from multiple sources, including public servants, internal oversight mechanisms within public agencies (such as disciplinary commissions, administrative proceedings, and internal control units), audit courts, and open sources such as newspapers and other media outlets.

When audits or investigations uncover irregularities, mayors and municipal officials may face administrative sanctions. These sanctions typically include the suspension of political rights—including temporary bans from holding public office and the loss of voting rights—the reimbursement of the estimated financial damage caused, and additional fines.

2.3 Data

This section describes the main datasets used in this paper. Appendix Table A1 provides the definition and source of each variable.

Electoral data. Electoral results come from *Base dos Dados* (Dahis et al. 2022), which aggregates and harmonizes data from the *Tribunal Superior Eleitoral* (TSE). We exploit five election years: 2000, 2004, 2008, 2012, and 2016, corresponding respectively to the following five mayoral terms: 2001–2004, 2005–2008, 2009–2012, 2013–2016, and 2017–2020.[5] For each candidate in each municipality, we observe gender, incumbency status, age, education level, party affiliation, and number of votes received. In addition, we use data from municipal

[5]Municipal elections are held during the last year of the term— typically in October—and the newly elected mayor takes office in January of the following year.

council elections to determine whether a given mayor had previously served as councilor.[6]

Corruption. To measure corruption, we first draw on Brazil's anti-corruption audit program conducted by the CGU. We focus on the period from 2003 to 2015, during which municipalities were selected at random. Beginning in 2006, the CGU classifies irregularities into three categories: notices, intermediate irregularities, and severe irregularities. The first category includes minor acts of mismanagement, such as improperly completed documentation (Avis et al. 2018). Intermediate irregularities typically reflect more significant administrative shortcomings, such as missing maintenance schedules for school buses or inadequate health infrastructure. In contrast, severe irregularities capture clear instances of corruption, including over-invoicing, bid-rigging, and the use of "ghost" employees for embezzlement (Bourlès et al. 2025). Accordingly, we measure corruption as the number of severe irregularities revealed by the audits.[7] This measure is only available for municipalities that were randomly audited at some point between 2006 and 2015. The data also report the fiscal year of the audited budget to which each infraction pertains, which may precede the year of the audit. This allows us to measure corruption across four municipal terms: 2001–2004, 2005–2008, 2009–2012, and 2013–2016.[8] Across these terms, the audit lottery covers 7.5% of mixed-gender elections, and 83.7% of audits reveal at least one severe irregularity, with a median of 5 severe irregularities.

Second, we constructed predicted corruption scores, following Ash et al. (2025). Using a machine-learning model trained on municipalities that were audited, Ash et al. (2025) identify patterns in budget composition that are predictive of corruption detected by auditors. The trained model is then applied to budget accounts for all municipalities to generate a predicted probability of corruption. While they restrict their predictive exercise to the period from 2001 to 2012, we generated corruption predictions up to 2020. Extending the series beyond 2012 required crosswalking variables from the old budget system to the post-2012 one. We detail our procedure and performance test in Appendix C.[9] Unlike the audit-based CGU measure, the predicted corruption score is available for the universe of Brazilian municipalities for all five electoral terms, substantially expanding the sample size. For each term, we compute the average predicted corruption score over the four years in office. It ranges from 0 to 0.55,

[6]Municipal councilors are elected at the same time as mayors, under an open-list proportional system.

[7]As robustness, we consider several alternative definitions, including the sum of intermediate and severe irregularities, or the probability that the audit reveals at least one severe irregularity (see Section 4.1).

[8]Corruption in the earliest (2001–2004) and latest (2013–2016) terms is only partially observed: the former relies exclusively on retrospective coverage from later audits, and the latter does not include irregularities from the 2016 fiscal year. All our regressions include electoral term fixed effects to account for these period-specific differences.

[9]Our extended prediction model achieves an AUC-ROC of 0.723, close to the performance reported by Ash et al. (2025) for their main specification on pre-2012 data (0.781).

with an average of 0.17.[10]

Finally, we assembled novel data on legal sanctions for administrative misconduct by web-scraping the *Cadastro Nacional de Condenações Cíveis por Atos de Improbidade Administrativa e Inelegibilidade*.[11] We constructed an indicator for whether the mayor is ever sanctioned after entering office, as well as three additional variables capturing the intensity of the sanction: (i) an indicator for whether the sanction involves the suspension of political rights; (ii) the amount of money that has to be reimbursed to compensate the public treasury for the financial loss caused; and (iii) the amount of fines imposed.[12] These measures are available for all Brazilian municipalities throughout the analysis period.[13] Between 2001 and 2020, 19.2% of mayors were legally sanctioned. Among those sanctioned, 80.4% had their political rights suspended, 53.4% were required to compensate the State for damages, and 60.2% were ordered to pay fines.

Together, we exploit six measures of corruption, drawn from distinct sources and capturing both the extensive and intensive margins. Our preferred measure is the predicted corruption score: it is available for all municipalities across all five terms, unlike the audit-based irregularity count, and it is continuous and thus more granular than the binary legal-sanction indicator. For these reasons, we use it as our main outcome in the heterogeneity analyses.

Municipalities' characteristics. The municipal socio-demographic variables we use to assess the validity of our identification strategy are drawn from the most recent census conducted prior to each election (i.e., the 2000 Census for the 2000, 2004, and 2008 elections, and the 2010 Census for the 2012, and 2016 elections). Municipal GDP comes from the Brazilian Institute of Geography and Statistics (IBGE).

[10] We also derive our main result using the original Ash et al. (2025)'s model on the 2000–2012 period (Section 4.1), as well as under alternative methodological choices for our extension exercise (Appendix C.4).

[11] See `https://www.cnj.jus.br/programas-e-acoes/combate-a-corrupcao-enccla/cadastro-nacional-de-condenados-por-crimes-de-improbidade-administrativa/`.

[12] These variables are set to zero for mayors who are not sanctioned.

[13] The registry does not provide information on the year of the underlying infraction. We therefore use the date of the sanction, and code a mayor's term as corrupt if the mayor is sanctioned at any point after entering office. Mechanically, this implies that earlier-term mayors have a higher probability of being coded as corrupt: given that the data were collected in 2026, we observe outcomes over more than 20 years for the first term, compared to 10 years for the last term. The inclusion of electoral-term fixed effects controls for this asymmetry.

3 Empirical Strategy

3.1 Sample and Specification

To assess the causal effect of having a female mayor, we implement a regression discontinuity design (RDD), comparing municipalities in which a woman narrowly won against a man with those in which a man narrowly won against a woman. Our sample consists of races in the 2000, 2004, 2008, 2012, and 2016 municipal elections in which the top two candidates were of opposite genders, representing 18.7% of all races.[14]

Our final sample includes 4,614 observations spanning 2,674 distinct municipalities, as a municipality can experience multiple mixed-gender races across the five elections we consider. As shown in Appendix Table A2, municipalities in our sample are somewhat larger than the average Brazilian municipality (37,162 vs. 34,221 inhabitants), but they are similar in terms of GDP per capita, demographic characteristics (urbanization, education, and employment), number of candidates per race, and voter turnout.

We estimate the following specification:

$$Y_{m,t} = \alpha + \tau T_{m,t} + \beta_1 X_{m,t} + \beta_2 X_{m,t} T_{m,t} + \lambda_t + \mu_{m,t} \ , \tag{1}$$

where m indexes municipalities and t a given election. The running variable $X_{m,t}$ is defined as the vote share of the female candidate minus that of the male candidate. The treatment variable $T_{m,t}$ equals 1 if the female candidate wins ($X_{m,t}>0$) and 0 otherwise. We further include electoral-term fixed effects (λ_t) to control for time-varying factors affecting corruption and its detection, such as changes in audit intensity or in the legal framework.[15] The corruption outcomes $Y_{m,t}$ are measured over the term following the election. The coefficient of interest, τ, captures the local average treatment effect (LATE) of electing a female mayor on corruption in close elections. Standard errors are clustered at the municipality level.

Following Calonico et al. (2014), we use a local linear specification, report robust confidence intervals, and rely on data-driven MSERD bandwidths. We further allow bandwidths to differ on each side of the threshold, which yields a lower asymptotic mean squared error (Arai and Ichimura 2018). Our effective sample includes around 3,000 observations within

[14] When the election took place in two rounds, we consider the top two candidates in the second round (0.67% of all mixed-gender races). Some elections were subsequently invalidated and followed by a supplementary election. In these cases, we retain the results of the supplementary election (0.04%). Finally, we remove races in which the winning candidate's candidacy was withdrawn, invalidated, or contested after election day (2.3%).

[15] Term fixed effects also control for measurement differences across electoral terms—namely the change in the set of budget variables used to predict corruption after 2012 (see Appendix C) and asymmetries between terms for audit and legal-sanction outcomes (see Section 2.3).

the estimation bandwidths.[16] Our results are robust to several specification choices, including the inclusion of controls and regional fixed effects, the use of a second-order polynomial, and alternative bandwidth selection procedures (see Section 4.1).

3.2 Validity of the design

The identification assumption is that the only discrete change at the threshold is the gender of the mayor. We provide several pieces of evidence supporting this assumption. First, Appendix Figure A1 shows no evidence of manipulation, as the running variable is smooth at the cutoff. Second, Appendix Table A3 reports balance tests for each municipal characteristic listed in Appendix Table A2. The point estimates are small and only one out of eleven is significant. Appendix Table A4 and Figure A2 confirm the absence of imbalances: we do not detect any jump at the threshold when using as outcome the treatment predicted by baseline municipal characteristics.

We next examine whether winner characteristics other than gender change discontinuously at the threshold. As shown in Table 1, closely elected female mayors are 16.9 percentage points more likely to have completed higher education than closely elected male mayors (Part I, column 3).[17] Female mayors also tend to be slightly younger (47.6 vs. 49.7 years old on average, column 1). In contrast, female and male mayors are balanced in their political experience—measured by incumbency status and prior service as a municipal councilor—and political orientation, as captured by affiliation with the four main Brazilian parties, and by a left-wing indicator.[18] Appendix Table A5 further shows balance in race and campaign contributions, although these measures are only available for a subset of electoral terms.

A potential concern for the interpretation of our main result is that education may have a direct effect on corruption, which could offset that of gender. For example, if education increases corruption while female leadership reduces it, the higher education level of female mayors could generate a null effect. Several pieces of evidence suggest this is unlikely. First, our estimates are similar when we control for winner characteristics, including education (Section 4.1), and within subsamples of mayors with similar education levels (Section 4.2).

[16] Bandwidths are selected using a data-driven approach and therefore vary across outcomes, as does the effective sample size. As described in Section 2.3, audit-based measures are only available for audited municipalities in the first four terms, yielding a much lower effective sample size of about 230 observations.

[17] Part II of Table 1 presents the same analysis for different subsamples, which we discuss in Section 4.2.

[18] We define the four main parties as those most frequently represented among the top two candidates in mayoral elections over the five elections we consider. Following Dahis et al. (2026), we classify the following parties as left-wing: Partido Democrático Trabalhista (PDT), Partido Trabalhista do Brasil (PT do B), Partido Comunista do Brasil (PC do B), Partido dos Trabalhadores (PT), Partido Socialista Brasileiro (PSB), Rede Sustentabilidade (Rede), Partido Verde (PV), Partido Socialismo e Liberdade (PSOL), Partido da Mobilização Nacional (PMN), Partido Socialista dos Trabalhadores Unificado (PSTU).

Second, a separate RD analysis comparing closely elected male mayors with and without higher education yields null effects (Appendix Table A6).

4 The Effect of Female Leadership on Corruption

4.1 Main Results

As shown in Table 2, we find no effect of the mayor's gender on any of our six corruption measures. Electing a female mayor has no detectable impact on the predicted corruption score based on budgetary accounts, the number of severe irregularities revealed by an audit, the probability that the mayor faces legal action, the probability that the mayor is suspended from political rights, the (log) amount of compensation paid for the damage caused, or the (log) amount of fines imposed. Figure 1 further confirms the absence of a discontinuity at the threshold for each outcome.

Our null result is robust to alternative definitions of our outcomes and alternative specifications. First, we also obtain a small and statistically insignificant estimate when using the original predicted corruption measure from Ash et al. (2025), which is available for the first three terms only (Appendix Table A7, column 1).[19] Second, we obtain similar results when turning to alternative audit-based outcomes, including total (instead of severe only) irregularities, an indicator for at least one severe irregularity, the log of severe irregularities, severe irregularities de-meaned relative to the lottery mean, and severe irregularities per audited program (Appendix Table A7, columns 2–6). Finally, the results are unchanged when we include region fixed effects, municipal controls, and winner controls (Appendix Table A8), as well as across alternative polynomial orders and bandwidth choices (Appendix Table A9).

Focusing on the most stringent specification and on our preferred outcome (predicted corruption), we can reject negative effects larger than 3.3 percent of the mean and positive effects larger than 5.5 percent (Appendix Table A8, Panel E, column 1).[20]

4.2 Heterogeneity analysis

We now investigate whether the null effect in Table 2 masks heterogeneous effects.

[19] Table A16 in Appendix C further shows that the point estimate based on the predicted corruption score for the full period is robust to alternative methodological choices in our extension of Ash et al. (2025).

[20] These bounds are obtained by dividing the lower and upper robust 95% confidence interval bounds by the mean outcome for male-led municipalities at the threshold. The bounds are tighter for the predicted corruption score than for the other outcomes, reflecting both the smaller effective sample for the audit measure and the higher variance of the legal-sanction outcomes.

Mayor characteristics. We split the sample by the mayor's education, term in office, political orientation, and age. As shown in Appendix Table A10, the effect remains non-significant across all subsamples. Female mayors appear no more or less corrupt than male mayors regardless of whether they hold a graduate degree, are serving their first or second term, run under a left- or right-wing party, or are above or below the median age. As discussed in Section 3.2, these splits also provide additional reassurance that our main null result is not driven by imbalances in winner characteristics at the threshold.

Electoral cycles. We next examine whether the null effect masks differences at particular points in the mayoral term, in light of prior work documenting time-varying gender differences over the electoral cycle (Accettura and Profeta 2021; Chauvin and Tricaud 2026). Appendix Table A11 estimates the effect of electing a woman mayor on the predicted corruption score measured the year before the term begins (column 1), during the first year in office (column 2), during the two middle years (column 3), and during the last year of the term, which also corresponds to the election year (column 4). Further validating the empirical strategy, we find no discontinuity in the initial value of predicted corruption, showing that female- and male-led municipalities at the threshold had the same baseline level of corruption. The estimates for each period of the term are also all statistically insignificant and close to zero, ruling out the possibility that the overall null masks offsetting effects across the electoral cycle.

Time periods and audit status. Our null result contrasts with the negative effect documented by Brollo and Troiano (2016), who study audited Brazilian municipalities over the 2000 and 2004 terms. Our sample differs from theirs along two dimensions: we cover three additional electoral terms (2008, 2012, and 2016), and we include non-audited municipalities. We therefore examine whether the gender effect varies along these two dimensions.

Table 3 presents the results. The effect is not significant in the full sample (column 1), in early elections (column 2), in post-2004 elections (column 3), and among audited municipalities in recent elections (column 5). In contrast, when we restrict the sample to audited municipalities in the 2000 and 2004 terms—the setting studied by Brollo and Troiano (2016)—we find a large negative and significant effect, corresponding to a 23.7 percent decrease in the predicted corruption score (column 4).

The fact that a negative gender effect appears *only* in this small subsample—disappearing once we extend the analysis to later elections or to non-audited municipalities—suggests that this subsample differs along some dimension that matters for the results. The balance tests in Part II of Table 1—estimated separately for each of the four subsamples from Table 3—shed light on this discrepancy. Female mayors are positively selected on education across all subsamples (column 3): estimates are consistently large and positive, although some-

times imprecise. Closely elected female mayors also tend to be younger than their male counterparts (column 1), a pattern more pronounced in earlier elections (Panels A and B). Crucially, neither imbalance is specific to audited municipalities in 2000 and 2004: education gaps appear in every panel, and age gaps appear in both early-term panels regardless of audit status. The distinctive feature of this subsample is a sharp and significant imbalance in incumbency.[21] In Panel B, closely elected female mayors are 24.4 percentage points (or 58.5 percent) less likely to be serving a second term than closely elected male mayors (column 2). In all other subsamples, the point estimate on incumbency is not statistically significant and either close to zero or positive.[22]

4.3 The Role of Incumbency

The imbalance uncovered in the previous section suggests that incumbency—rather than gender—may be driving the negative effect documented in prior work. Incumbency is likely to affect corruption directly, both through the experience and political networks that incumbents accumulate in office and through reduced electoral accountability: because of Brazil's two-term limit, second-term mayors cannot seek reelection, which has been shown to increase corruption (Ferraz and Finan 2011).[23] As a result, if female mayors are less likely to be serving a second term, the apparent negative gender effect may in fact reflect the positive effect of incumbency on corruption. We provide three pieces of evidence consistent with this interpretation.

First, we revisit Brollo and Troiano (2016)'s findings using their own corruption measure—a binary indicator for corruption based on the text analysis of audit reports, available for audited municipalities in the first two terms. Consistent with their main finding and with our subsample analysis, Appendix Table A13 reports a 23.3 percentage point decline in the probability of corruption (column 2).[24] We then show that the point estimate is more than

[21] In contrast, we find no evidence of imbalances in municipal characteristics that are specific to audited municipalities during the 2000 and 2004 terms (Appendix Table A12).

[22] The estimate is large and positive (though not significant) for audited municipalities after 2004 (Panel D)—the same subsample that exhibits a large positive (though not significant) gender effect on corruption (Table 3, column 5). The two audited subsamples thus display a striking parallel: the incumbency imbalance flips from -0.244 to 0.231 across these two subsamples, and the gender coefficient on corruption mirrors it, flipping from -0.055 to 0.051.

[23] Ferraz and Finan (2011) study only the 2000 term. We extend their analysis to later electoral terms and additional corruption measures in what follows.

[24] Our estimate differs slightly from the one of Brollo and Troiano (2016) for three main reasons. First, we do not restrict the analysis to elections with only two candidates, increasing the effective sample size from 103 to 185 observations. Second, following Calonico et al. (2014), we use a triangular kernel rather than a uniform one, assigning greater weight to observations closer to the threshold. Third, while their main analysis use the "broad" measure of corruption, we use the "narrow" one which exhibits greater variability

halved and no longer statistically significant either when we restrict the sample to first-term mayors (Appendix Table A13, column 3) or when we control for winner characteristics (column 4), although all estimates are imprecise given the small sample size.

Second, incumbency independently impacts corruption within Brollo and Troiano (2016)'s setting. Following Ferraz and Finan (2011), we conduct a separate RD analysis comparing municipalities in which a second-term mayor narrowly won the election to those in which a first-term mayor narrowly won. We restrict the sample to elections in which the top two candidates are both men, to abstract from any gender effect, and we control for other winner characteristics.[25] Within audited municipalities in the 2000 and 2004 terms, the probability of corruption is 13.8 percentage points higher when the mayor is serving a second term—a 48.8 percent increase, significant at the 10 percent level (Appendix Table A13, column 5).

Third, the incumbency effect is not specific to this subsample. Table 4 shows that incumbents are more corrupt than newly elected mayors across all of our corruption measures and electoral terms. In audited municipalities over the 2000–2012 terms, the number of severe irregularities is 61.9 percent higher under incumbents (column 2). In the full sample, the predicted corruption score is 4.2 percent higher (column 1). Second-term mayors are also 17.6 percent more likely to face legal action; although this estimate is not significant ($p = 0.106$, column 3), all other legal effects are, with second-term mayors being 23.4 percent more likely to face political suspension (column 4), reimbursing 18.8 percent higher financial damages (column 5), and incurring 25.2 percent higher fines (column 6).[26]

5 Conclusion

This paper examines the relationship between female political leadership and corruption using thousands of close mayoral elections in Brazil and multiple measures of corruption. Across five electoral terms and six corruption outcomes, we find no evidence that electing a female mayor affects corruption.

We show that the previously documented negative effect of female leadership is confined to a small subsample of randomly audited municipalities in early terms. Moreover, it coincides with a sharp discontinuity in incumbency rates between closely elected female and male

(see Appendix Table A1 for the detailed definition of each variable). The effect is similar with the former, but not significant (Appendix Table A13, column 1).

[25] Appendix Tables A14 and A15 report the balance tests on municipality and winner characteristics for the incumbency RD analysis. While we find no discontinuities in baseline municipal characteristics, we detect significant discontinuities in mayors' political affiliation. We therefore control for winner characteristics in our incumbency analysis.

[26] For the outcomes expressed in logs (damage and fines), we compute the percentage effect as: $(\exp(\tau)-1)\times 100$, where τ is the point estimate.

mayors. Because incumbency increases corruption, this effect likely reflects the impact of incumbency rather than female leadership.

Our findings shed new light on the effects of female political representation and underscore the importance of examining discontinuities in characteristics other than gender in regression discontinuity designs.

Importantly, these results should not be interpreted as suggesting that greater female political representation is undesirable or unwarranted. Equity in representation remains a fundamental democratic objective, and substantial evidence indicates that women continue to face significant barriers and discrimination in politics (see e.g. Bertrand and Duflo 2017; Le Barbanchon and Sauvagnat 2022; Fujiwara et al. 2025).

Figures and Tables

Table 1: Balance tests - Winner characteristics

Part I: Full sample

	(1) Age	(2) Incumbent	(3) Graduate	(4) Former councilor	(5) PT	(6) PMDB	(7) PSDB	(8) PFL/DEM	(9) Left-wing
Female	-2.093**	-0.047	0.169***	-0.011	-0.014	-0.031	0.027	-0.015	-0.056
	(0.952)	(0.038)	(0.039)	(0.022)	(0.025)	(0.034)	(0.029)	(0.025)	(0.040)
Observations	2,769	2,796	3,401	3,196	2,977	2,930	3,228	3,008	2,700
Robust p-value	0.020	0.195	0.000	0.670	0.437	0.266	0.363	0.445	0.152
Bandwidth left	0.138	0.152	0.219	0.192	0.157	0.215	0.197	0.188	0.143
Bandwidth right	0.153	0.140	0.178	0.161	0.167	0.095	0.163	0.129	0.135
Mean outcome	49.696	0.245	0.412	0.092	0.085	0.184	0.118	0.104	0.255

Part II: By sub-samples

	(1) Age	(2) Incumbent	(3) Graduate	(4) Former councilor	(5) PT	(6) PMDB	(7) PSDB	(8) PFL/DEM	(9) Left-wing
Panel A: 2000 and 2004 terms - All Municipalities									
Female	-3.612**	-0.072	0.128	0.005	0.010	-0.034	0.035	-0.028	-0.046
	(1.598)	(0.068)	(0.078)	(0.039)	(0.034)	(0.057)	(0.056)	(0.063)	(0.052)
Observations	865	866	922	877	1,053	1,037	946	925	871
Robust p-value	0.015	0.189	0.135	0.794	0.813	0.632	0.635	0.689	0.305
Bandwidth left	0.130	0.134	0.172	0.167	0.224	0.234	0.175	0.173	0.156
Bandwidth right	0.177	0.170	0.151	0.128	0.181	0.160	0.165	0.151	0.144
Mean outcome	51.414	0.271	0.342	0.063	0.053	0.201	0.137	0.194	0.147
Panel B: 2000 and 2004 terms - Audited Municipalities									
Female	-3.119	-0.244*	0.201	-0.011	0.017	-0.112	0.103	-0.135	-0.084
	(3.480)	(0.157)	(0.159)	(0.018)	(0.016)	(0.102)	(0.134)	(0.125)	(0.096)
Observations	188	178	227	168	114	204	227	219	214
Robust p-value	0.302	0.089	0.212	0.505	0.274	0.186	0.421	0.201	0.387
Bandwidth left	0.124	0.121	0.176	0.115	0.066	0.165	0.177	0.166	0.165
Bandwidth right	0.125	0.119	0.174	0.088	0.077	0.106	0.176	0.162	0.135
Mean outcome	52.089	0.417	0.420	0.007	-0.017	0.180	0.152	0.244	0.124
Panel C: Post 2004 terms - All municipalities									
Female	-1.330	-0.037	0.182***	-0.015	-0.021	-0.021	0.024	-0.013	-0.060
	(1.061)	(0.044)	(0.049)	(0.027)	(0.031)	(0.042)	(0.031)	(0.022)	(0.051)
Observations	2,083	2,095	2,236	2,330	1,962	1,881	2,393	2,004	1,868
Robust p-value	0.210	0.399	0.000	0.546	0.425	0.459	0.457	0.382	0.261
Bandwidth left	0.172	0.184	0.188	0.212	0.150	0.184	0.212	0.196	0.148
Bandwidth right	0.149	0.137	0.174	0.176	0.147	0.092	0.205	0.108	0.125
Mean outcome	48.839	0.239	0.451	0.105	0.096	0.177	0.107	0.062	0.306
Panel D: Post 2004 terms - Audited municipalities									
Female	-0.940	0.231	0.205	-0.061	0.003	-0.027	-0.056	-0.010	0.133
	(5.107)	(0.242)	(0.244)	(0.089)	(0.104)	(0.196)	(0.150)	(0.133)	(0.218)
Observations	106	91	97	109	88	97	93	103	96
Robust p-value	0.856	0.211	0.407	0.360	0.832	0.669	0.716	0.865	0.641
Bandwidth left	0.197	0.147	0.160	0.250	0.156	0.172	0.168	0.177	0.172
Bandwidth right	0.170	0.124	0.156	0.145	0.107	0.135	0.118	0.179	0.124
Mean outcome	50.162	0.293	0.602	0.082	0.079	0.144	0.101	0.129	0.181

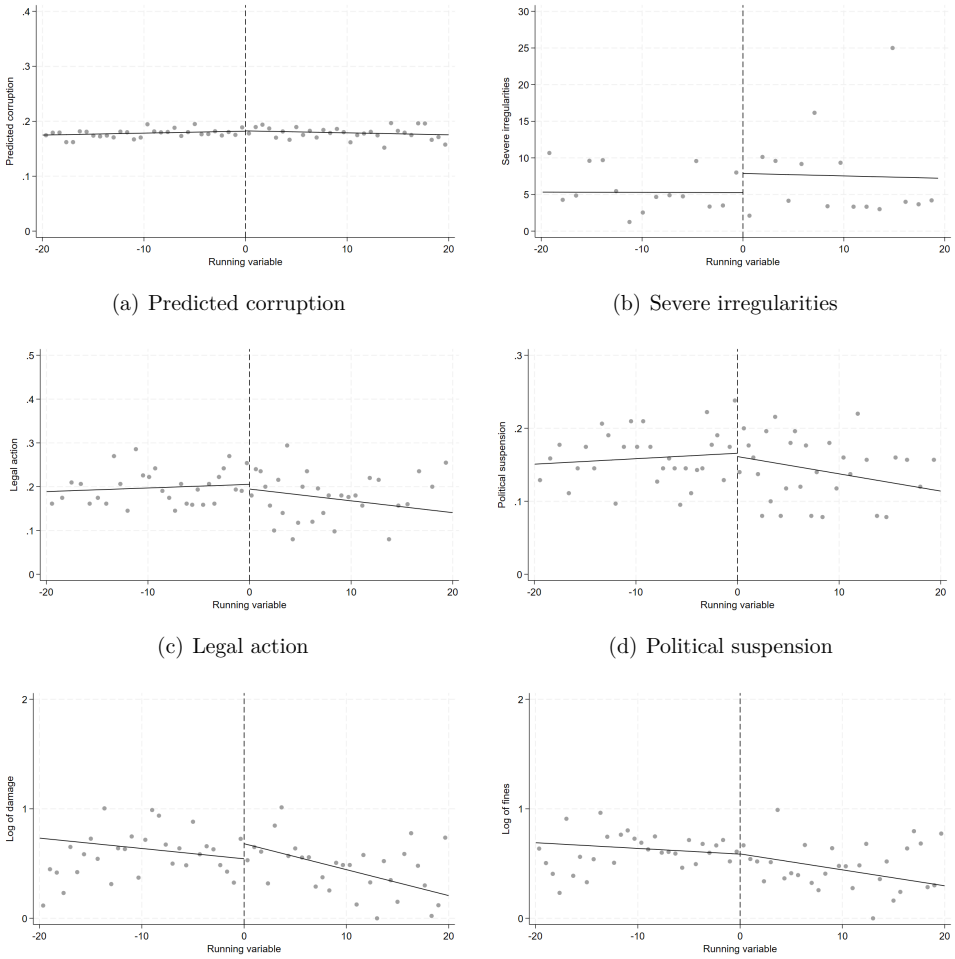
Notes: In Column 1, the outcome is the winner's age at the time of the election. In column 2 (resp. 3, 4, 6, 7, 8, 9) the outcome is an indicator equal to one if the winner is the incumbent (resp. has completed higher education, served as a city councilor in the previous term, runs under the PT, PMDB, PSDB, PFL/DEM party label, or under a party label classified as left-wing.). Part I reports results for the full sample. Part II reports four subsamples: 2000 and 2004 terms (Panel A); audited municipalities in those terms (Panel B); post-2004 terms (Panel C); and audited municipalities in those terms (Panel D). The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We use a linear fit, data-driven MSERD bandwidths, and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table 2: The effect of female leadership on corruption

	(1) Predicted corruption	(2) Severe irregularities	(3) Legal action	(4) Political suspension	(5) Log of damage	(6) Log of fines
Female	0.001	0.536	0.006	-0.004	0.175	0.046
	(0.006)	(2.703)	(0.036)	(0.030)	(0.150)	(0.145)
Observations	3,249	235	2,810	3,270	3,041	2,952
Robust p-value	0.897	0.893	0.759	0.923	0.126	0.535
Bandwidth left	0.198	0.220	0.170	0.194	0.169	0.185
Bandwidth right	0.169	0.112	0.116	0.178	0.161	0.124
Mean outcome	0.182	5.180	0.213	0.172	0.542	0.599

Notes: The outcomes are: the predicted corruption score described in Section 2.3 and Appendix C, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 1); the number of severe irregularities uncovered by CGU audits for budget years corresponding to the mayor's term, available for audited municipalities over the 2000, 2004, 2008, and 2012 terms (column 2); an indicator equal to one if the mayor appears in the National Registry of Civil Convictions for Acts of Administrative Improbity and Ineligibility after entering office, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 3); an indicator equal to one if the mayor receives a legal sanction that involves the suspension of political rights (column 4); the log(1 + x) amount of financial damage the mayor has to reimburse (column 5); the log(1 + x) amount of fines imposed (column 6). If legal action equals 0, then political suspension, damage and fines also equal 0. The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Figure 1: The effect of female leadership on corruption



(a) Predicted corruption

(b) Severe irregularities

(c) Legal action

(d) Political suspension

(e) Log of damage

(f) Log of fines

Notes: The outcomes are: the predicted corruption score described in Section 2.3 and Appendix C, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (sub-figure a); the number of severe irregularities uncovered by CGU audits for budget years corresponding to the mayor's term, available for audited municipalities over the 2000, 2004, 2008, and 2012 terms (sub-figure b); an indicator equal to one if the mayor appears in the National Registry of Civil Convictions for Acts of Administrative Improbity and Ineligibility after entering office, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (sub-figure c); an indicator equal to one if the mayor receives a legal sanction that involves the suspension of political rights (sub-figure d); the log(1 + x) amount of financial damage the mayor has to reimburse (sub-figure e); the log(1 + x) amount of fines imposed (sub-figure f). If legal action equals 0, then political suspension, damage and fines also equal 0. The running variable is defined as the difference, in percentage points, between the vote shares of the female and male candidates, with positive (negative) values indicating a female (male) victory. The dots correspond to binned averages of the outcome variable, computed using evenly spaced bins of the running variable for continuous outcomes and quantile spaced bins for binary outcomes. The figures restrict attention to observations lying within a 20 percentage-point window around the cutoff. We use a linear fit and a triangular kernel. All regressions include electoral-term fixed effects.

Table 3: The effect of female leadership on corruption - By time period

	(1)	(2)	(3)	(4)	(5)
		All municipalities		Audited municipalities	
	Full sample	2000 and 2004 terms	Post 2004 terms	2000 and 2004 terms	Post 2004 terms
Female	0.001	-0.002	0.000	-0.055***	0.051
	(0.006)	(0.011)	(0.007)	(0.020)	(0.052)
Observations	3,249	943	2,220	159	96
Robust p-value	0.897	0.752	0.866	0.001	0.369
Bandwidth left	0.198	0.213	0.203	0.111	0.156
Bandwidth right	0.169	0.118	0.152	0.084	0.147
Mean outcome	0.182	0.209	0.169	0.232	0.174

Notes: The outcome is the predicted corruption score described in Section 2.3 and in Appendix C. Columns 2 and 4 (resp. 3 and 5) focuses on the 2000 and 2004 terms (resp. 2008, 2012, and 2016 terms). Columns 4 and 5 further restrict the sample to municipalities that were randomly audited by the CGU. The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. “Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table 4: The effect of incumbency on corruption

	(1)	(2)	(3)	(4)	(5)	(6)
	Predicted corruption	Severe irregularities	Legal action	Political suspension	Log of damage	Log of fines
Incumbent	0.007*	1.999*	0.036	0.037*	0.172*	0.225**
	(0.004)	(1.458)	(0.024)	(0.022)	(0.119)	(0.105)
Observations	5,693	390	6,810	6,499	6,323	6,657
Robust p-value	0.079	0.080	0.106	0.089	0.096	0.023
Bandwidth left	0.094	0.083	0.150	0.151	0.113	0.161
Bandwidth right	0.203	0.178	0.238	0.206	0.236	0.212
Mean outcome	0.165	3.230	0.204	0.158	0.599	0.572

Notes: The sample is restricted to elections in which the top two candidates are men, and one is the incumbent. The outcomes are: the predicted corruption score described in Section 2.3 and Appendix C, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 1); the number of severe irregularities uncovered by CGU audits for budget years corresponding to the mayor’s term, available for audited municipalities over the 2000, 2004, 2008, and 2012 terms (column 2); an indicator equal to one if the mayor appears in the National Registry of Civil Convictions for Acts of Administrative Improbity and Ineligibility after entering office, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 3); an indicator equal to one if the mayor receives a legal sanction that involves the suspension of political rights (column 4); the log(1 + x) amount of financial damage the mayor has to reimburse (column 5); the log(1 + x) amount of fines imposed (column 6). If legal action equals 0, then political suspension, damage and fines also equal 0. The independent variable is an indicator equal to one if the incumbent won the election. We control for the winner’s age, college completion, party affiliation (with separate indicators for PT, PMDB, PSDB, PFL/DEM, and affiliation to a left-wing party), and for electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. “Mean outcome" reports the average outcome value for municipalities led by first-term mayors at the threshold. Standard errors are clustered at the municipality level.

References

Accettura, C. and Profeta, P. (2021). Gender differences in political budget cycles. *Working Paper*.

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

Alexander, A. C. (2021). Gender, gender equality, and corruption: A review of theory and evidence. In *The Oxford Handbook of the Quality of Government*. Oxford University Press.

Arai, Y. and Ichimura, H. (2018). Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator. *Quantitative Economics*, 9(1):441–482.

Ash, E., Galletta, S., and Giommoni, T. (2025). A machine learning approach to analyze and support anticorruption policy. *American Economic Journal: Economic Policy*, 17(2):162–193.

Avis, E., Ferraz, C., and Finan, F. (2018). Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians. *Journal of Political Economy*, 126(5):1912–1964.

Baskaran, T., Bhalotra, S., Min, B., and Uppal, Y. (2024). Women legislators and economic performance. *Journal of Economic Growth*, 29(2):151–214.

Bauhr, M. and Charron, N. (2021). Will women executives reduce corruption? Marginalization and network inclusion. *Comparative Political Studies*, 54(7):1292–1322.

Bauhr, M., Charron, N., and Wängnerud, L. (2019). Exclusion or interests? Why females in elected office reduce petty and grand corruption. *European Journal of Political Research*, 58(4):1043–1065.

Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. In *Handbook of Economic Field Experiments*, volume 1, pages 309–393. North-Holland.

Bourlès, R., Ferrali, R., Kim, G., and Peveri, J. (2025). Audits and bureaucratic corruption: Evidence from Brazilian municipalities. *Working Paper*.

Brollo, F., Nannicini, T., Perotti, R., and Tabellini, G. (2013). The political resource curse. *American Economic Review*, 103(5):1759–1796.

Brollo, F. and Troiano, U. (2016). What happens when a woman wins an election? Evidence from close races in Brazil. *Journal of Development Economics*, 122:28–45.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1):234–261.

Chauvin, J. P. and Tricaud, C. (2026). Gender and electoral incentives: Evidence from COVID-19 response. *Working Paper 32410, National Bureau of Economic Research*.

Dahis, R., Carabetta, J., Scovino, F., Israel, F., and Oliveira, D. (2022). Data basis: Universalizing access to high-quality data.

Dahis, R., de las Heras, I., and Saavedra, S. (2026). Young local politicians and deforestation in the Brazilian Amazon. *Journal of the Association of Environmental and Resource Economists (Forthcoming)*.

Dahis, R., Mattsson, M., and Sales, N. (2025). Reelection incentives and corruption: Revisiting the evidence with LLM-classified audit reports. *Working Paper*.

Debski, J., Jetter, M., Mösle, S., and Stadelmann, D. (2018). Gender and corruption: The neglected role of culture. *European Journal of Political Economy*, 55:526–537.

Decarolis, F., Fisman, R., Pinotti, P., Vannutelli, S., and Wang, Y. (2023). Gender and bureaucratic corruption: Evidence from two countries. *The Journal of Law, Economics, and Organization*, 39(2):557–585.

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

Esarey, J. and Schwindt-Bayer, L. A. (2018). Women's representation, accountability and corruption in democracies. *British Journal of Political Science*, 48(3):659–690.

Ferraz, C. and Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, 101(4):1274–1311.

Fisman, R. and Svensson, J. (2007). Are corruption and taxation really harmful to growth? Firm level evidence. *Journal of Development Economics*, 83(1):63–75.

Fujiwara, T., Hilbig, H., and Raffler, P. (2025). Biased party nominations as a source of women's electoral underperformance. *Working Paper 34396, National Bureau of Economic Research*.

Hessami, Z. and da Fonseca, M. L. (2020). Female political representation and substantive effects on policies: A literature review. *European Journal of Political Economy*, 63:101896.

Instituto Brasileiro de Geografia e Estatística (2024). Estatísticas de finanças públicas e conta intermediária de governo. `https://www.ibge.gov.br`.

Jha, C. K. and Sarangi, S. (2018). Women and corruption: What positions must they hold to make a difference? *Journal of Economic Behavior & Organization*, 151:219–233.

Le Barbanchon, T. and Sauvagnat, J. (2022). Electoral competition, voter bias, and women in politics. *Journal of the European Economic Association*, 20(1):352–394.

Marshall, J. (2024). Can close election regression discontinuity designs identify effects of winning politician characteristics? *American Journal of Political Science*, 68(2):494–510.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.

Rose-Ackerman, S. and Palifka, B. J. (2016). *Corruption and Government: Causes, Consequences, and Reform*. Cambridge University Press, New York, NY.

Rothstein, B. (2011). *The quality of government: corruption, social trust, and inequality in international perspective*. University of Chicago Press, Chicago, IL.

Swamy, A., Knack, S., Lee, Y., and Azfar, O. (2001). Gender and corruption. *Journal of Development Economics*, 64(1):25–55.

Supplemental Appendix

Contents

A Additional Tables

Table A1: Variable definitions and sources

Variable	Definition	Source
Panel A: Municipality characteristics		
Population	Population of the municipality.	IBGE.
GDP per capita	Gross domestic product of the municipality per capita in R$.	IBGE.
Urban pop.	Share of the municipality's population living in an urban area.	2000 and 2010 Census, downloaded from *Base dos Dados* (Dahis et al. 2022).
Female pop.	Share of the municipality's population that is female.	2000 and 2010 Census, downloaded from *Base dos Dados* (Dahis et al. 2022).
High-educated pop.	Share of the municipality's adult population that completed high-school..	2000 and 2010 Census, downloaded from *Base dos Dados* (Dahis et al. 2022).
Under 14 pop.	Share of the municipality's population under 14 years old.	2000 and 2010 Census, downloaded from *Base dos Dados* (Dahis et al. 2022).
Working age pop.	Share of the municipality's population between 15 and 64 years old.	2000 and 2010 Census, downloaded from *Base dos Dados* (Dahis et al. 2022).
Employed pop.	Share of the municipality's working-age population that is employed.	2000 and 2010 Census, downloaded from *Base dos Dados* (Dahis et al. 2022).
Informality	Share of the municipality's working-age population employed without a labor card.	2000 and 2010 Census, downloaded from *Base dos Dados* (Dahis et al. 2022).
Numb. of candidates	Number of mayoral candidates in the election.	TSE, downloaded from *Base dos Dados* (Dahis et al. 2022).
Turnout	Turnout in the first round of the municipal election.	TSE, downloaded from *Base dos Dados* (Dahis et al. 2022).
Panel B: Winners' characteristics		
Female	Indicator equal to one if the candidate is a female, as registered by the electoral justice, and 0 if male.	TSE, downloaded from *Base dos Dados* (Dahis et al. 2022).
Age	Age of the candidate at the time of the election, computed using the election date and the candidate's date of birth.	TSE, downloaded from *Base dos Dados* (Dahis et al. 2022).
Incumbent	Indicator equal to one if the candidate is the incumbent at the time of the election, and zero otherwise. Constructed by matching candidate identifiers and names with winners of the previous election.	TSE, downloaded from *Base dos Dados* (Dahis et al. 2022).
Graduate	Indicator equal to one if the candidate has completed tertiary-level education.	TSE, downloaded from *Base dos Dados* (Dahis et al. 2022).
Former councilor	Indicator equal to one if the candidate served as a city councilor in the previous term. Constructed by matching candidate identifiers with councilor winners of the previous election.	TSE, downloaded from *Base dos Dados* (Dahis et al. 2022).
PT, PMDB, PSDB, PFL/DEM	Indicators equal to one if the mayoral candidate ran under the respective party label in the election.	TSE, downloaded from *Base dos Dados* (Dahis et al. 2022).
Left-wing	Indicator equal to one if the candidate's party belongs to the PDT, PT do B, PC do B, PT, PSB, Rede, PV, PSOL, PMN, or PSTU.	Authors' computation based on TSE's party declarations.
Panel C: Outcomes		
Predicted corruption	Continuous corruption score predicted by a machine-learning model that link municipal budget accounts to audit result.	Authors' calculations, following Ash et al. (2025), see Appendix C.
Severe irregularities	Number of severe irregularities reported in CGU municipal audit reports for budget years audited during the mayor's term.	CGU audit data (Controladoria-Geral da União), downloaded from *Base dos Dados*.
Legal action	Indicator equal to one if the mayor appears in the National Registry of Civil Convictions for Acts of Administrative Improbity and Ineligibility after entering office.	Cadastro Nacional de Condenações Cíveis por Atos de Improbidade Administrativa e Inelegibilidade (web-scraped).
Political suspension	Indicator equal to one if the mayor receives a legal sanction that results in the suspension of her political rights. Equal zero if the mayor receives no sanction or if the sanction does not involve the suspension of her political rights.	Cadastro Nacional de Condenações Cíveis por Atos de Improbidade Administrativa e Inelegibilidade (web-scraped).
Log of damage	Amount that the politician has to reimburse to compensate for the financial damage caused, as consequence of convictions for administrative misconduct or corruption ($\log + 1$). Equal zero if the mayor is not sanctioned or if the sanction does not involve reimbursements.	Cadastro Nacional de Condenações Cíveis por Atos de Improbidade Administrativa e Inelegibilidade (web-scraped).
Log of fines	Total monetary penalties imposed on the politician as a consequence of convictions for administrative misconduct or corruption ($\log + 1$). Equal zero if the mayor is not sanctioned or if the sanction does not involve fines.	Cadastro Nacional de Condenações Cíveis por Atos de Improbidade Administrativa e Inelegibilidade (web-scraped).
Broad corruption	Indicator equal to one if the CGU audit report identifies broad corruption violations—such as illegal procurement, fraud, favoritism, over-invoicing, diversion of funds, or paid but not proven expenses—based on text analysis of audit reports.	Brollo et al. (2013), obtained from Ash et al. (2025) replication package.
Narrow corruption	Indicator equal to one if the CGU audit report identifies severe corruption violations—such as illegal procurement, fraud, favoritism, or over-invoicing—based on text analysis of audit reports.	Brollo et al. (2013), obtained from Ash et al. (2025) replication package.

Notes: Definition and sources of the variables used in the main analysis. Census variables are drawn from the most recent census before each election (2000 Census for the 2000, 2004, and 2008 elections; 2010 Census for the 2012 and 2016 elections).

Table A2: Descriptive statistics

	Mixed races (N = 4,614)		All races (N =25,128)	
	Mean	SD	Mean	SD
Population	37,162	292,420	34,221	207,462
GDP per capita	10,078	13,782	10,729	12,139
% of urban population	61.0	22.3	60.8	23.0
% of female population	49.4	1.5	49.3	1.5
% of high-educated population	18.3	9.6	17.9	9.4
% of population under 14	29.4	6.1	29.1	6.1
% of working age population	63.3	4.8	63.6	4.7
% of employed population	53.6	12.2	55.1	12.1
% of informality	18.6	6.5	18.5	6.7
Number of candidates	2.8	1.1	2.8	1.1
Turnout	86.6	5.9	87.1	5.9

Notes: The statistics are at the municipality × election level. The first two columns focuses on our analysis sample of mixed-gender races (see Section 3.1), while the last two consider all Brazilian municipal elections. We exclude elections in which the winning candidate's candidacy was withdrawn, invalidated, or judicially contested after election day. See Table A1 for the definition and source of each variable.

Table A3: Balance tests - Municipality characteristics

	(1) Population	(2) GDP pc	(3) % Urban	(4) % Female	(5) % High-educated	(6) % Under 14	(7) % Working age	(8) % Employed	(9) % Informality	(10) # candidates	(11) Turnout
Female	-0.161*	-0.060	0.000	0.000	0.008	-0.003	-0.001	0.006	0.005	-0.018	-0.522
	(0.094)	(0.067)	(0.018)	(0.001)	(0.006)	(0.005)	(0.004)	(0.011)	(0.006)	(0.086)	(0.500)
Observations	2,639	2,478	3,316	2,390	3,239	3,323	2,999	2,959	2,975	3,200	2,857
Robust p-value	0.062	0.294	0.980	0.885	0.177	0.545	0.796	0.659	0.411	0.790	0.184
Bandwidth left	0.170	0.116	0.226	0.125	0.214	0.214	0.163	0.154	0.171	0.216	0.125
Bandwidth right	0.094	0.136	0.153	0.109	0.147	0.166	0.163	0.169	0.146	0.138	0.210
Mean outcome	9.347	2.041	0.587	0.492	0.171	0.299	0.630	0.530	0.185	2.782	87.617

Notes: Each column considers a specific municipality characteristic, as defined in Panel A of Table A1. Population and GDP per capita are transformed using log(1 + x). The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We use a linear fit, data-driven MSERD bandwidths, and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table A4: General balance test

	(1) Predicted treatment
Female	0.002
	(0.002)
Observations	2,878
Robust p-value	0.123
Bandwidth left	0.171
Bandwidth right	0.127
Mean outcome	0.423

Notes: The outcome is the predicted treatment, computed as follows: we regress the treatment indicator on all 11 baseline characteristics listed in Appendix Table A2, and we use the regression coefficients to predict the treatment value. In the prediction regression, population and GDP per capita are transformed using log(1 + x). The independent variable is an indicator equal to one if the female candidate won the election. We control for electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table A5: Balance tests - Additional winner characteristics

	(1) White	(2) Contributions
Female	0.091	-0.042
	(0.089)	(0.099)
Observations	579	2,550
Robust p-value	0.312	0.544
Bandwidth left	0.127	0.194
Bandwidth right	0.118	0.127
Mean outcome	0.689	5.691

Notes: In Column 1, the outcome is an indicator equal to one if the candidate's declared race is white. This variable is only available for the 2016 election. In Column 2, the outcome is the amount of campaign contributions received by the candidate expressed in constant prices and transformed using log(1 + x). This variable is missing for the 2000 election. The independent variable is an indicator equal to one if the female candidate won the election. Column 2 includes electoral-term fixed effects. We use a linear fit, data-driven MSERD bandwidths, and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table A6: The effect of education on corruption

	(1) Predicted corruption	(2) Severe irregularities	(3) Legal action	(4) Political suspension	(5) Log of damage	(6) Log of fines
Graduate	0.004	-0.397	-0.036	-0.024	-0.149	-0.138
	(0.004)	(1.862)	(0.025)	(0.024)	(0.111)	(0.114)
Observations	4,939	368	5,777	5,536	6,135	5,605
Robust p-value	0.289	0.907	0.123	0.330	0.148	0.242
Bandwidth left	0.124	0.129	0.172	0.152	0.157	0.141
Bandwidth right	0.145	0.161	0.167	0.165	0.223	0.181
Mean outcome	0.160	4.124	0.217	0.167	0.670	0.660

Notes: The sample is restricted to elections in which the top two candidates are male, and one has completed higher education while the other has not. The outcomes are: the predicted corruption score described in Section 2.3 and Appendix C, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 1); the number of severe irregularities uncovered by CGU audits for budget years corresponding to the mayor's term, available for audited municipalities over the 2000, 2004, 2008, and 2012 terms (column 2); an indicator equal to one if the mayor appears in the National Registry of Civil Convictions for Acts of Administrative Improbity and Ineligibility after entering office, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 3); an indicator equal to one if the mayor receives a legal sanction that involves the suspension of political rights (column 4); the log(1 + x) amount of financial damage the mayor has to reimburse (column 5); the log(1 + x) amount of fines imposed (column 6). If legal action equals 0, then political suspension, damage and fines also equal 0. The independent variable is an indicator equal to one if the candidate who completed higher education won the election. All regressions include electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for non-graduate-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table A7: The effect of female leadership on corruption - Alternative corruption measures

	(1) Predicted corruption (Ash et al., 2025)	(2) Total irregularities	(3) At least one severe irregularity	(4) Log. of severe irregularities	(5) Severe irregularities de-meaned	(6) Severe irregularities per audited programs
Female	0.010	5.194	0.022	0.190	1.599	-0.218
	(0.021)	(7.766)	(0.135)	(0.347)	(2.694)	(0.389)
Observations	1,654	219	228	233	235	199
Robust p-value	0.675	0.391	0.915	0.725	0.754	0.343
Bandwidth left	0.208	0.155	0.180	0.187	0.213	0.155
Bandwidth right	0.155	0.171	0.143	0.136	0.113	0.107
Mean outcome	0.567	36.918	0.742	1.302	-0.464	0.900

Notes: The outcomes are: the predicted corruption score from Ash et al. (2025) (column 1); the total number of intermediate and severe irregularities uncovered by CGU audits (column 2); an indicator equal to one if at least one severe irregularity is uncovered (column 3); the log(1 + x) number of severe irregularities (column 4); the number of severe irregularities de-meaned relative to the lottery mean (column 5); and the number of severe irregularities divided by the total number of federal programs audited (column 6). The outcome in columns 1 is available for all municipalities over the first three terms of our sample (2000, 2004, and 2008 terms). The outcomes in columns 2 to 6 are available for audited municipalities over the 2000, 2004, 2008, and 2012 terms. The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. “Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table A8: Robustness - Fixed-effects and controls

	(1) Predicted corruption	(2) Severe irregularities	(3) Legal action	(4) Political suspension	(5) Log of damage	(6) Log of fines
Panel A: Electoral-term FE (Baseline)						
Female	0.001	0.536	0.006	-0.004	0.175	0.046
	(0.006)	(2.703)	(0.036)	(0.030)	(0.150)	(0.145)
Observations	3,249	235	2,810	3,270	3,041	2,952
Robust p-value	0.897	0.893	0.759	0.923	0.126	0.535
Bandwidth left	0.198	0.220	0.170	0.194	0.169	0.185
Bandwidth right	0.169	0.112	0.116	0.178	0.161	0.124
CI lower	-0.011	-5.663	-0.060	-0.062	-0.064	-0.195
CI upper	0.013	4.933	0.082	0.056	0.525	0.375
Mean outcome	0.182	5.180	0.213	0.172	0.542	0.599
Panel B: Electoral-term FE and region FE						
Female	0.005	-1.346	-0.000	-0.008	0.168	0.031
	(0.004)	(2.767)	(0.037)	(0.031)	(0.146)	(0.140)
Observations	2,812	232	2,602	3,078	3,196	3,160
Robust p-value	0.321	0.379	0.955	0.753	0.141	0.646
Bandwidth left	0.156	0.229	0.142	0.162	0.193	0.225
Bandwidth right	0.138	0.099	0.115	0.181	0.160	0.123
CI lower	-0.004	-7.858	-0.070	-0.070	-0.072	-0.210
CI upper	0.013	2.987	0.074	0.050	0.502	0.339
Mean outcome	0.181	5.147	0.217	0.175	0.562	0.621
Panel C: Electoral-term FE and municipality characteristics						
Female	0.001	-0.909	-0.009	-0.015	0.122	-0.007
	(0.005)	(2.717)	(0.036)	(0.031)	(0.146)	(0.139)
Observations	2,748	205	2,607	2,991	3,169	3,163
Robust p-value	0.883	0.449	0.853	0.583	0.257	0.881
Bandwidth left	0.150	0.172	0.142	0.158	0.199	0.227
Bandwidth right	0.137	0.092	0.119	0.169	0.148	0.123
CI lower	-0.009	-7.382	-0.078	-0.078	-0.121	-0.251
CI upper	0.010	3.270	0.065	0.044	0.453	0.293
Mean outcome	0.181	5.504	0.217	0.175	0.567	0.621
Panel D: Electoral-term FE, municipality and winner characteristics						
Female	0.002	-0.446	-0.005	-0.010	0.174	0.019
	(0.005)	(2.670)	(0.036)	(0.031)	(0.146)	(0.139)
Observations	2,760	198	2,630	3,015	3,133	3,138
Robust p-value	0.766	0.546	0.924	0.676	0.138	0.724
Bandwidth left	0.152	0.153	0.141	0.156	0.193	0.216
Bandwidth right	0.136	0.103	0.124	0.181	0.148	0.125
CI lower	-0.008	-6.844	-0.074	-0.073	-0.070	-0.223
CI upper	0.011	3.622	0.067	0.047	0.503	0.321
Mean outcome	0.181	5.830	0.217	0.176	0.562	0.616
Panel E: Electoral-term FE, region FE, municipality and winner characteristics						
Female	0.003	-1.145	-0.017	-0.020	0.113	-0.006
	(0.004)	(2.836)	(0.036)	(0.031)	(0.144)	(0.139)
Observations	2,752	201	2,563	2,910	3,242	3,110
Robust p-value	0.615	0.409	0.661	0.438	0.288	0.869
Bandwidth left	0.148	0.166	0.132	0.145	0.213	0.214
Bandwidth right	0.139	0.093	0.125	0.176	0.149	0.124
CI lower	-0.006	-7.902	-0.087	-0.085	-0.130	-0.250
CI upper	0.010	3.216	0.055	0.037	0.436	0.296
Mean outcome	0.181	5.581	0.220	0.178	0.580	0.615

Notes: All Panels include electoral-term fixed effects. Panel A reproduces Table 2. Panel B includes region fixed effects. Panel C controls for all municipal characteristics presented in Table A2. Panel D controls for both municipal and winner characteristics. Panel E controls for region fixed effects and both municipal and winner characteristics. The outcomes are: the predicted corruption score described in Section 2.3 and Appendix C, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 1); the number of severe irregularities uncovered by CGU audits for budget years corresponding to the mayor's term, available for audited municipalities over the 2000, 2004, 2008, and 2012 terms (column 2); an indicator equal to one if the mayor appears in the National Registry of Civil Convictions for Acts of Administrative Improbity and Ineligibility after entering office, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 3); an indicator equal to one if the mayor receives a legal sanction that involves the suspension of political rights (column 4); the log(1 + x) amount of financial damage the mayor has to reimburse (column 5); the log(1 + x) amount of fines imposed (column 6). If legal action equals 0, then political suspension, damage and fines also equal 0. The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table A9: Robustness - Bandwidths and polynomial order

	(1) Predicted corruption	(2) Severe irregularities	(3) Legal action	(4) Political suspension	(5) Log of damage	(6) Log of fines
Panel A: Distinct MSE-optimal bandwidths (Baseline)						
Female	0.001	0.536	0.006	-0.004	0.175	0.046
	(0.006)	(2.703)	(0.036)	(0.030)	(0.150)	(0.145)
Observations	3,249	235	2,810	3,270	3,041	2,952
Robust p-value	0.897	0.893	0.759	0.923	0.126	0.535
Bandwidth left	0.198	0.220	0.170	0.194	0.169	0.185
Bandwidth right	0.169	0.112	0.116	0.178	0.161	0.124
Mean outcome	0.182	5.180	0.213	0.172	0.542	0.599
Local polynomial order	1	1	1	1	1	1
Panel B: Distinct CER-optimal bandwidths						
Female	0.002	-0.990	-0.003	-0.008	0.185	0.052
	(0.007)	(2.804)	(0.040)	(0.033)	(0.168)	(0.160)
Observations	2,470	196	2,058	2,495	2,280	2,178
Robust p-value	0.752	0.575	0.977	0.818	0.208	0.648
Bandwidth left	0.132	0.165	0.113	0.129	0.113	0.123
Bandwidth right	0.112	0.084	0.077	0.119	0.107	0.082
Mean outcome	0.181	5.470	0.233	0.186	0.517	0.591
Local polynomial order	1	1	1	1	1	1
Panel C: Common MSE-optimal bandwidth						
Female	0.001	0.470	-0.006	-0.005	0.130	0.037
	(0.006)	(3.096)	(0.036)	(0.031)	(0.138)	(0.148)
Observations	3,268	190	2,836	3,219	3,428	2,690
Robust p-value	0.895	0.903	0.991	0.906	0.241	0.589
Bandwidth left	0.188	0.129	0.149	0.182	0.207	0.138
Bandwidth right	0.188	0.129	0.149	0.182	0.207	0.138
Mean outcome	0.181	6.456	0.215	0.173	0.560	0.591
Local polynomial order	1	1	1	1	1	1
Panel D: Common CER-optimal bandwidth						
Female	0.002	-0.977	-0.016	-0.011	0.185	0.031
	(0.006)	(3.401)	(0.040)	(0.034)	(0.150)	(0.167)
Observations	2,498	157	2,103	2,439	2,681	1,957
Robust p-value	0.746	0.656	0.740	0.755	0.182	0.759
Bandwidth left	0.125	0.097	0.099	0.121	0.138	0.092
Bandwidth right	0.125	0.097	0.099	0.121	0.138	0.092
Mean outcome	0.180	6.895	0.235	0.185	0.520	0.615
Local polynomial order	1	1	1	1	1	1
Panel E: 10 p.p bandwidth						
Female	0.002	-0.814	-0.016	-0.022	0.178	0.034
	(0.010)	(4.501)	(0.054)	(0.050)	(0.260)	(0.228)
Observations	2,100	160	2,110	2,110	2,110	2,110
Robust p-value	0.854	0.223	0.691	0.309	0.833	0.900
Bandwidth left	0.100	0.100	0.100	0.100	0.100	0.100
Bandwidth right	0.100	0.100	0.100	0.100	0.100	0.100
Mean outcome	0.180	6.535	0.236	0.198	0.509	0.611
Local polynomial order	1	1	1	1	1	1
Panel F: 2nd order local polynomial						
Female	0.001	-1.490	-0.002	-0.018	0.246	0.082
	(0.007)	(3.144)	(0.045)	(0.039)	(0.179)	(0.167)
Observations	3,651	273	3,273	3,417	3,709	3,703
Robust p-value	0.870	0.468	0.980	0.573	0.119	0.537
Bandwidth left	0.249	0.298	0.199	0.196	0.226	0.282
Bandwidth right	0.217	0.168	0.170	0.217	0.278	0.193
Mean outcome	0.181	5.540	0.225	0.189	0.482	0.583
Local polynomial order	2	2	2	2	2	2

Notes: Panel A uses distinct MSE-optimal bandwidths on each side of the threshold , reproducing Table 2. Panel B uses distinct CER-optimal bandwidths. Panels C and D use a common data-driven bandwidth on both sides of the cutoff—MSE-optimal in Panel C and CER-optimal in Panel D. Panel E uses the same 10 p.p bandwidth on both sides of the cutoff and for all outcomes. Panel F uses a 2nd order polynomial fit with distinct MSE-optimal bandwidths. The outcomes are: the predicted corruption score described in Section 2.3 and Appendix C, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 1); the number of severe irregularities uncovered by CGU audits for budget years corresponding to the mayor's term, available for audited municipalities over the 2000, 2004, 2008, and 2012 terms (column 2); an indicator equal to one if the mayor appears in the National Registry of Civil Convictions for Acts of Administrative Improbity and Ineligibility after entering office, available for all municipalities over the 2000, 2004, 2008, 2012, and 2016 terms (column 3); an indicator equal to one if the mayor receives a legal sanction that involves the suspension of political rights (column 4); the log(1 + x) amount of financial damage the mayor has to reimburse (column 5); the log(1 + x) amount of fines imposed (column 6). If legal action equals 0, then political suspension, damage and fines also equal 0. The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table A10: The effect of female leadership on corruption - By mayors' characteristics

	(1) Full sample	(2) Non graduate	(3) Graduate	(4) First-term	(5) Second term	(6) Right-wing	(7) Left-wing	(8) Below median age	(9) Above median age
Female	0.001	0.013	-0.003	0.000	0.002	0.000	0.007	0.000	0.003
	(0.006)	(0.009)	(0.008)	(0.007)	(0.013)	(0.006)	(0.014)	(0.009)	(0.008)
Observations	3,249	1,544	1,721	2,467	715	2,690	636	1,455	1,731
Robust p-value	0.897	0.108	0.784	0.993	0.886	0.986	0.527	0.796	0.864
Bandwidth left	0.198	0.204	0.177	0.216	0.162	0.207	0.144	0.163	0.181
Bandwidth right	0.169	0.139	0.223	0.158	0.145	0.210	0.182	0.240	0.161
Mean outcome	0.182	0.187	0.174	0.183	0.177	0.182	0.178	0.182	0.181

Notes: The outcome is the predicted corruption score described in Section 2.3 and in Appendix C. Columns 2 and 3 (resp. 4 and 5, 6 and 7, 8 and 9) split the sample according to whether the elected mayor completed higher education or not (resp. is serving a first or second term, is left- or right-wing, is below or above the median age). The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table A11: The effect of female leadership on corruption - Across the electoral cycle

	(1) Initial value	(2) First year	(3) Middle years	(4) Last year
Female	0.009	0.004	-0.000	-0.002
	(0.009)	(0.007)	(0.006)	(0.007)
Observations	2,733	2,861	3,182	3,041
Robust p-value	0.230	0.386	0.918	0.674
Bandwidth left	0.224	0.167	0.191	0.174
Bandwidth right	0.124	0.132	0.165	0.156
Mean outcome	0.177	0.182	0.185	0.178

Notes: The outcome is the predicted corruption score described in Section 2.3 and in Appendix C. Column 1 measures it in the year before the mayoral term begins (unavailable for the 2001–2004 term); column 2 in the first year in office; column 3 averaged across the two middle years of the term; and column 4 in the last year of the term. The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. Standard errors are clustered at the municipality level.

Table A12: Balance on municipality characteristics - Audited municipalities in early terms

	(1) Population	(2) GDP	(3) % Urban	(4) % Female	(5) % High-educated	(6) % Under 14	(7) % Working age	(8) % Employed	(9) % Informality	(10) # candidates	(11) Turnout
Female	0.146	-0.055	-0.016	-0.003	-0.021	-0.010	0.002	-0.014	-0.011	-0.073	-0.073
	(0.294)	(0.211)	(0.060)	(0.005)	(0.020)	(0.016)	(0.013)	(0.033)	(0.021)	(0.283)	(1.840)
Observations	168	208	223	242	209	257	221	197	211	198	242
Robust p-value	0.452	0.671	0.651	0.385	0.186	0.466	0.847	0.505	0.465	0.828	0.862
Bandwidth left	0.095	0.155	0.182	0.211	0.184	0.229	0.177	0.142	0.171	0.149	0.208
Bandwidth right	0.142	0.139	0.132	0.150	0.094	0.175	0.147	0.111	0.117	0.115	0.164
Mean outcome	9.158	1.851	0.634	0.495	0.136	0.332	0.609	0.496	0.213	2.773	86.685

Notes: The sample is restricted to audited municipalities over the 2000 and 2004 electoral terms. Each column considers a specific municipality characteristic, as defined in Panel A of Table A1. Population and GDP per capita are transformed using log(1 + x). The independent variable is an indicator equal to one if the female candidate won the election. We use a linear fit, data-driven MSERD bandwidths, and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold. See Table A1 for variable definitions and sources.

Table A13: Revisiting Brollo and Troiano (2016)'s analysis

	(1)	(2)	(3)	(4)	(5)
	Treatment = Female				Treatment = Incumbent
	Broad measure	Narrow measure	Narrow measure First-term mayors	Narrow measure + Winner characteristics	Narrow measure + Winner characteristics
Treatment	-0.191	-0.233*	-0.089	-0.111	0.138*
	(0.141)	(0.161)	(0.205)	(0.144)	(0.094)
Observations	178	185	112	194	590
Robust p-value	0.181	0.073	0.613	0.311	0.093
Bandwidth left	0.139	0.154	0.109	0.156	0.108
Bandwidth right	0.121	0.133	0.091	0.174	0.246
Mean outcome	0.889	0.479	0.288	0.478	0.283

Notes: The outcomes are: an indicator equal to one for broad corruption (column 1) and narrow corruption (columns 2–5), both based on the text analysis of CGU audit reports by Brollo et al. (2013). These variables are only available for audited municipalities over the 2000 and 2004 terms. In columns 1–4, as in the main analysis, the sample is restricted to mixed-gender races and the independent variable is an indicator equal to one if the female candidate won the election. In column 5, the sample is restricted to elections in which the top two candidates are men, and one is the incumbent, and the independent variable is an indicator equal to one if the incumbent won. Column 3 restrict the sample to first-term mayors. Columns 4 and 5 control for all winner characteristics listed in Table 1. All regressions include electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for male-led municipalities at the threshold in columns 1–4 (resp. first-term-led municipalities in column 5). Standard errors are clustered at the municipality level.

Table A14: Balance test for the incumbency RDD - Municipality characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Population	GDP	% Urban	% Female	% High-educated	% Working age	% Under 14	% Employed	% Informality	# candidates	Turnout
Incumbent	-0.056	-0.055	0.007	0.001	-0.002	-0.003	0.002	-0.005	0.002	-0.045	-0.005
	(0.066)	(0.047)	(0.014)	(0.001)	(0.005)	(0.003)	(0.003)	(0.008)	(0.004)	(0.065)	(0.388)
Observations	6,371	5,811	6,079	5,653	5,679	5,319	5,500	5,651	5,250	5,428	5,975
Robust p-value	0.477	0.224	0.514	0.272	0.647	0.332	0.641	0.425	0.648	0.575	0.831
Bandwidth left	0.142	0.104	0.159	0.117	0.132	0.111	0.142	0.112	0.099	0.113	0.138
Bandwidth right	0.202	0.197	0.168	0.172	0.162	0.155	0.141	0.176	0.162	0.159	0.174
Mean outcome	9.371	2.122	0.587	0.493	0.171	0.638	0.289	0.553	0.181	2.683	87.861

Notes: The sample is restricted to elections in which the top two candidates are men, and one is the incumbent. Each column considers a specific municipality characteristic, as defined in Panel A of Table A1. Population and real GDP are transformed using log(1 + x). The independent variable is an indicator equal to one if the incumbent won the election. All regressions include electoral-term fixed effects. We use a linear fit, data-driven MSERD bandwidths, and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for municipalities led by first-term mayors at the threshold. Standard errors are clustered at the municipality level.

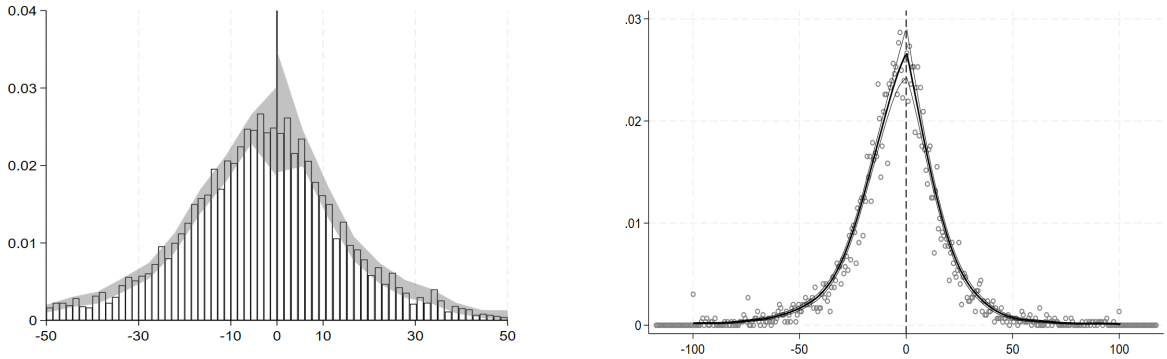
Table A15: Balance test for the incumbency RDD - Winner characteristics

	(1) Age	(2) Graduate	(3) PT	(4) PMDB	(5) PSDB	(6) PFL/DEM	(7) Left-Wing
Incumbent	1.026*	-0.019	-0.006	-0.105***	0.047**	0.026	-0.023
	(0.629)	(0.031)	(0.015)	(0.031)	(0.023)	(0.020)	(0.026)
Observations	5,978	6,411	6,719	4,223	5,664	5,910	6,294
Robust p-value	0.084	0.457	0.631	0.000	0.041	0.132	0.361
Bandwidth left	0.146	0.130	0.169	0.112	0.115	0.138	0.134
Bandwidth right	0.168	0.224	0.209	0.092	0.173	0.169	0.206
Mean outcome	49.872	0.439	0.079	0.266	0.118	0.099	0.211

Notes: The sample is restricted to elections in which the top two candidates are men, and one is the incumbent. In Column 1, the outcome is the winner's age at the time of the election. In column 2 (resp. 3, 4, 5, 6, 7) the outcome is an indicator equal to one if the winner has completed higher education (resp. runs under the PT, PMDB, PSDB, PFL/DEM party label, or under a party label classified as left-wing). The independent variable is an indicator equal to one if the incumbent won the election. All regressions include electoral-term fixed effects. We use a linear fit, data-driven MSERD bandwidths, and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. "Mean outcome" reports the average outcome value for municipalities led by first-term mayors at the threshold. Standard errors are clustered at the municipality level.

B Additional Figures

Figure A1: Density tests

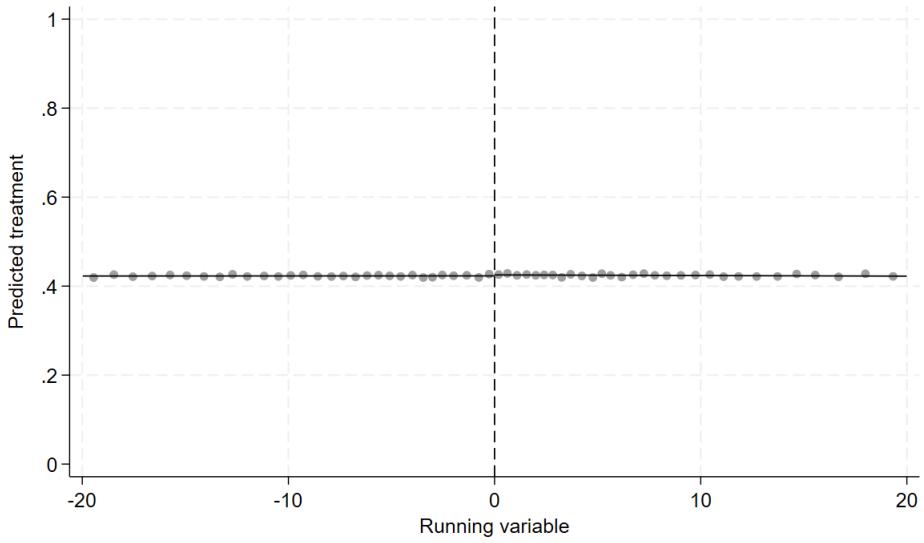


(a) Cattaneo et al. (2018)'s density test

(b) McCrary density test

Notes: Panel A tests for a discontinuity in the density of the female candidate's margin of victory at the threshold using the method developed by Cattaneo et al. (2018). The H statistic associated with the density test is 0.501, with a p-value of 0.610. The grey shaded area represents the confidence intervals. Panel B tests for a discontinuity in the density of the female candidate's victory margin at the threshold using the method developed by McCrary (2008). The black line represents the estimated density of the running variable, while the grey lines denote the confidence intervals.

Figure A2: General balance test



Notes: The outcome is the predicted treatment, computed as follows: we regress the treatment indicator on all 11 baseline characteristics listed in Appendix Table A2, and we use the regression coefficients to predict the treatment value. In the prediction regression, population and real GDP are transformed using log(1 + x). The running variable is defined as the difference, in percentage points, between the vote shares of the female and male candidates, with positive (negative) values indicating a female (male) victory. The dots correspond to binned averages of the outcome variable, computed using evenly spaced bins of the running variable. The figures restrict attention to observations lying within a 20 percentage-point window around the cutoff. We use a linear fit and a triangular kernel and control for electoral-term fixed effects.

C Predicting Municipal Corruption

To predict corruption, Ash et al. (2025) train an algorithm on municipal budget data from FINBRA (2001–2012) paired with audit-based corruption measures for the subset of municipalities audited by the CGU over the same period. The algorithm learns which budget variables, and combinations of variables, predict corruption detected in audits. Once trained, the model can be applied to all municipalities—including those never audited—to generate a predicted probability of corruption.

We extend this prediction exercise beyond 2012, which raises two challenges. First, Brazil changed its municipal budget reporting system in 2013, moving from FINBRA to SICONFI; applying the model output to later years therefore requires mapping FINBRA variables to their SICONFI equivalents. Second, the main audit data used by Ash et al. (2025) to train and test the model is not available after 2012, preventing us from evaluating model performance in the post-period. We address these challenges by (a) retraining the model on an alternative audit-based corruption measure available for later years, (b) constructing a FINBRA–SICONFI crosswalk to apply the model output to 2013–2020 budget data, and (c) testing the model performance on post-2012 audit results.

C.1 Training the Model With A New Corruption Measure

Ash et al. (2025) train their model on two alternative audit-based measure. The main measure is the narrow corruption indicator from Brollo et al. (2013), defined in Table A1, which is only available for mayors elected in 2000 and 2004. The robustness measure, drawn from Avis et al. (2018), is an indicator equal to one if the municipality-year falls in the top quartile of the ratio of audit irregularities (intermediate and severe) to service orders.

We chose to train the model using a similar top-quartile indicator. Our measure differs in two ways. First, Ash et al. exploit only the year the audit was recorded (not the year of the underlying irregularity), so each audit result is spread across the three budget years preceding the audit, pairing the same corruption value with three different budget years. Because we observe the year to which each irregularity pertains, we assign each irregularity to its actual budget year. Second, consistent with our main analysis, we restrict the numerator to severe irregularities, which more clearly capture corruption (see Section 2.3). Table A16 shows that our results are robust to using the sum of intermediate and severe irregularities instead.

Like Ash et al. (2025), we train the model on the 2001–2012 period. We then use the model output to predict corruption for all municipalities, both in the 2001–2012 period but also over 2013–2020 thanks to the budget data crosswalk we explain in the next section. Finally, because we train the model on audit data available for later years, we can test the

performance of the model on post-2012 audits, as shown in Section C.3.[1]

C.2 Extending Predictions to Post-2012 Years

Brazil transitioned from the FINBRA budget reporting system to SICONFI starting in 2013.[2] Because the model was trained on FINBRA data (2001–2012), generating predictions for 2013–2020 requires mapping the FINBRA budget items to their SICONFI equivalents. We matched budget items using the Portaria budget classification code, variable names, and variable descriptions. Out of the 797 FINBRA budget items, only 527 contribute to the prediction of the model. We were able to crosswalk 386 of them, representing 73% of the predictive items and 85% of the top 100 most predictive ones.

The crosswalked items are available for all municipalities from 2013 to 2017. Starting in 2018, naming conventions for revenue accounts changed: approximately 71 of the 153 revenue categories no longer appear under the same name. Using Portaria classification codes as a bridge, we recovered 14 matches. The remaining 57 revenue accounts are unavailable for 2018–2020. As shown in Section C.4, our results are robust to removing these years.

C.3 Model Performance

We validate the model predictive performance on post-2012 years using CGU random audits from 2013 and 2014. The validation sample comprises 119 municipalities for 2013 (lotteries 39–40) and 57 for 2014 (lottery 40 only); no random audits were conducted after lottery 40 (August 2015), so predictions for 2015–2020 cannot be validated. To assess the performance of the model, we report the AUC-ROC (area under the receiver operator characteristic curve). As explained by Ash et al. (2025), it takes values between 0.5 (random guessing) and 1.0 (perfect accuracy), and can be interpreted as the probability that a randomly sampled corrupt municipality is ranked more highly by predicted probability of corruption than a randomly sampled noncorrupt municipality.

We obtain an AUC-ROC of 0.723 which is close to the performance of Ash et al. (2025)'s main model on pre-2012 data (0.781; see Table 2 in their paper). The performance of our model on post-2012 data is also higher than the one of a model using both severe and intermediate irregularities (instead of severe only), which yields an AUC-ROC of 0.636.

[1] Because only a small number of municipalities were audited in 2013 and 2014—the last years of random audits—we do not have enough observations to split the post-2012 audited sample into a training and testing sets, and thus use these years for testing only.

[2] In 2013, some municipalities still reported under the old balance sheet format, while others had already transitioned to the new format.

C.4 Robustness

As described above, we trained the model on pre-2012 data using our own corruption indicator. We trained the model on the full set of budget items, 527 of these turning out to matter for the prediction. For the prediction stage, that we extend to all years from 2001–2020, we use only the coefficients corresponding to the 386 crosswalked items. Applying this restriction to all years (both pre- and post- 2012) ensures temporal consistency.

Table A16 shows that our null effect of female leadership on predicted corruption is robust to alternative constructions of the prediction scores. The estimate remains small and statistically insignificant when the prediction is based on all 527 predictive FINBRA features for 2001–2012 and on the 386 crosswalked features post-2012 years (column 2); when we train the prediction model on the ratio of total irregularities to service orders rather than on severe irregularities only (column 3); when we train the model exclusively on the 386 crosswalked features that we use at the prediction stage (column 4); and when we exclude the 2016 term, for which 57 budget items are missing (column 5). These exercises indicate that our findings are not driven by particular modeling choices.

Table A16: Robustness for predicted corruption

	(1)	(2)	(3)	(4)	(5)
	Baseline	FINBRA + SICONFI	Severe + Intermediate	Training on crosswalked features	Pre-2012 terms
Female	0.001	-0.001	0.005	-0.001	0.004
	(0.006)	(0.009)	(0.009)	(0.010)	(0.007)
Observations	3,249	3,509	3,133	3,401	2,523
Robust p-value	0.897	0.879	0.460	0.891	0.513
Bandwidth left	0.198	0.221	0.190	0.209	0.219
Bandwidth right	0.169	0.211	0.154	0.195	0.164
Mean outcome	0.182	0.193	0.176	0.199	0.180

Notes: The outcomes are: the baseline predicted corruption score, constructed as described in the text (column 1). Columns 2 through 4 use alternative predicted corruption scores. In Column 2, the prediction relies on all 527 predictive FINBRA features for 2001–2012 and on the 386 crosswalked FINBRA–SICONFI features for post-2012 years; in column 3, the model is trained on an indicator for the top quartile of the ratio of total irregularities to service orders rather than on severe irregularities only; and in column 4, the model is retrained exclusively on the 386 crosswalked features. In column 5, we exclude the 2016 term, for which 57 budget items are missing. The independent variable is an indicator equal to one if the female candidate won the election. All regressions include electoral-term fixed effects. We use a linear fit with data-driven MSERD bandwidths and assess statistical significance using robust p-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. “Mean outcome" reports the average outcome value for male-led municipalities at the threshold.