**Referee Reports – *Journal of Institutional Economics***

**1. Data**

**R1(3):** In looking at your summary statistics, I can’t figure out why your share of female employers, managers/directors/executives, and women in leadership would have a min and max of 0 and 1. A better explanation of this would be warranted in the text, otherwise this stands out like a weird measurement error. Also, there are some municipalities have 97% of their economic activity taking place in the informal sector?! I can actually buy that, but it’s wild.

**R3(1):** Why do sample sizes differ? In Table 1, the number of municipalities for the share of females among employers is 878 but the number of municipalities for the share of females among managers, directors, or executives is 934 (and it is 930 in Table 2). The descriptive statistics and analysis have to be based on the same sample.

**R3(2):** It would be good to know from the first sentence in Section 3.1 that the cross-sectional data come from the 2010 Census because otherwise the discussion on the years over which the corruption measure is calculated becomes difficult to follow. Additionally, the robustness done in Appendix A, which only considers corruption data before 2010 should be the main analysis (and not the robustness check). How can corruption cases that occur after 2010 influence female leadership representation in 2010? Otherwise, I would try to make a stronger argument for considering corruption cases after 2010 in the corruption measure (for instance, that it is quite persistent over time by showing statistics over time for municipalities that were audited several times and arguing that because lotteries only select some municipalities each year you have to use the full sample of lotteries, even those after 2010 – if that’s the case).

**R3(5):** Is it possible to see descriptive statistics on municipalities that were excluded from the analysis as compared to those that were included (in particular in terms of female leadership representation)? Are the large municipalities included in the analysis (those that were excluded from the lottery)? How do these municipalities differ, as compared to those included in the analysis? Why are there corruption scores missing for some municipalities? How do these municipalities differ, as compared to those included in the analysis? These descriptive statistics could give a sense of whether results on all municipalities are likely to be weaker or stronger. Additionally, it would be interesting to know why Avis et al. (2018) only use 967 municipalities, while 1949 municipalities experienced at least one audit (excluding the last two years of lotteries would reduce the number of municipalities by more than half?).

**R3(d):** Could you provide a discussion on how likely the lottery is itself likely to be corrupted? Descriptive statistics on the municipalities that were audited (and those that were audited several times), as compared to the municipalities that were never audited could help the discussion.

**R5(3):** The paper could better explain the rationale behind the choice of control variables. Please cite some relevant literature to support the selection. For instance, why is the informal sector included as a control? A possible logic is that informal sectors tend to be more prone to corruption, and if you don’t control for this, you might miss capturing the true effect of corruption on the outcomes.

**R5(5):** How did you conclude that the sectors are corrupt? Explain the methodology and include some summary statistics comparing corruption levels in these selected sectors (extractive, manufacturing, construction, transportation/communication).

**R5(6)**. In the sentence: “It is quite telling that while the female labor force participation in these sectors is almost one-tenth of the overall economy (3.8% vs. 33%), the share of leadership positions held by women is ‘only’ three-fifths of that for the full sample (19.7% in corrupt sectors versus 32.8% overall; see Table 1)” — this comparison is interesting but could use more explanation. What exactly is "telling"? Is it that leadership representation is relatively higher than participation, or that both are disproportionately low?

**R6(4):** What is the role of audits? Are all of them related to tax collection or the use of public funds? It is unclear.

**R6(5):** Is there any information about the size and ownership type of the firms used to comprise the sample? It also affects female participation.

**R6(6):** I am reluctant to buy the argument that the sectoral share of females in leadership is driven by corruption. The authors should present and, if applicable, utilize the share of females employed in the sector and participating in leadership roles.

**R6(8):** Table 1, which describes the data and ratios, should be more clearly defined. What is the basis for getting different percentages for leadership positions? What are the bases in each row? Starting number of observations and the aggregation method. Additionally, when one considers the standard deviations of each mean, it suggests that there is no difference between the “corrupted” and non-corrupted sectors. Also, which ratio is eventually used for the regression analysis? The share of labor force participation is way too low if they use the somewhat recalculated ratio; it should be clearly defined and explained with a different number of observations.

**2. Econometrics**

**2.1 Clarifications / Interpretation**

**R3(4):** Industry share controls: it would be good to see coefficients on the industry share controls, in particular to see whether they correspond to the “corrupted” sectors. Relatedly, it would be good to see results on the “non-corrupted” sample; if the “selection” and “quantitative” mechanisms are taking place as suggested, we should see a “selection” effect in the “non-corrupted” sample. [I am slightly confused with the use of the term “selection effect” here; I would rather call the “selection effect” the fact that women select out from corrupted sectors. Maybe the other effect can then be called a “glass-ceiling effect”, given that corruption prevents women from reaching the higher ranks of the hierarchy?] Some z-tests on the proportion of females among employers and in management in “corrupted” versus “non-corrupted” samples might help the discussion.

**R3(6):** Is it the case that one lottery draws several municipalities to be audited, or only one? This would clarifies what the lottery fixed effects capture exactly. Related to this question, it is not clear to me why there is the lottery subscript to the corruption variable.

**R3(7):** Why not using municipality fixed effects? Instead or on top of State fixed effects. Along this line, why not clustering standard errors by municipality?

**R3(8)**: Do you include a dummy variable that captures the fact that some municipalities were audited several times?

**R5(1):** The first outcome variable is the ratio of women in leadership roles out of all leadership positions, while the second is the ratio of women in the labor force out of all working-age women. These two outcomes are not directly comparable because the denominators are inconsistent. The first one shows women’s leadership compared to both men and women in leadership, but the second one shows labor force participation relative only to women. So, when interpreting results, saying that corruption reduces both women’s leadership and labor force participation can be misleading, since one is relative to men and the other is not. To make them comparable, you could redefine the second outcome as women’s labor force participation rate divided by the male labor force participation rate ((women in labor force/working age women)/(men in labor force/working age men)). That way, if you argue that corruption reduces both participation and leadership, it would imply that men are taking more of the jobs and that in corrupt sectors, men may find it easier to navigate and climb to leadership roles. To make this claim strong, you’ll need to control for working-age population in the analysis.

**R5(2):** The third outcome variable actually strengthens the interpretation. Even if women's overall labor force participation doesn’t change—or even increases—corruption can still be a barrier specifically to attaining leadership. This helps justify that in corrupt environments, even when women manage to get jobs, it’s much harder for them to reach top positions.

**R5(4):** Sectoral shares are not included in the baseline regressions, but a sub-sample is used for specific "corrupt" sectors later. Please justify why a separate sub-sample makes sense here. It’s likely because the regressions are at the municipality level rather than at the individual level, and sectoral composition differs across municipalities.

**R6(3):**The model, methodology, and data structure are not adequately described. Initially, I was surprised that the primary model does not include the time dimension, because the authors later describe the differences that other authors used with similar data across different periods. Given the FE, I presumed there was some cross-sectional time variation that the authors could use. I realized later that the primary data sample is the cross-sectional data coming from the 2010 census. It must be described clearly; maybe first, the authors should explain the data, the sampling frame they used, any aggregation, etc. Then their methodology would not be misleading.

**R6(9):** The regression tables should include a list of all controls, with details provided in the Appendix.

**R6(11):** Overall, authors should discuss the identification strategy and underlying methodology in more detail.

**2.2 IV**

**R1(1):** [Regarding IV], since those variables are measured years before, is there any concern that they might not reflect the reality at the time your corruption variables were observed? Judiciary district and management capacity are probably relatively stable over time, but the existence of local councils and whether they are active could be more subject to change.

**R1(2):** [Regarding IV], might management capacity be endogenous to corruption? This is along the lines of the grease the wheels hypothesis – less functional areas might see more corruption just in the everyday course of business. There is literature that suggests an inverse relationship between corruption and state capacity, so maybe just citing that will overcome this particular qualm.

**R3(3):** On the instrumental variable strategy: can you provide concrete examples of “unobservable causes of corruption that could result in omitted variables biasing the estimates”? Given the outcome variable is measured for year 2010, would it be possible to build instruments from later data (1998 is more than 10 years before and corruption has likely changed over a 10 year period)? The measure of management capacity is at State level and not at municipal level, correct? How are the instruments used? Do they form a unique indicator? It would be good to include them in the descriptive statistics and to see first stage results. Why would the IV strategy only work for the full sample and the leadership category?

**R5(7):** There should be more explanation for why each of the IVs satisfies the exogeneity condition. Right now, it’s mentioned that they are measured before corruption, which helps with reverse causality. Please explain the theoretical link between these IVs and the outcomes, and cite supporting literature to justify why they affect corruption but not female leadership directly.

**R6(7):** Some measures used as instruments (e.g., an indicator of a juridical district) may be associated with municipality size, as well as with the number of economically active individuals.

**2.3 Further tests:**

**R3(e):** It would be good to have the non-logged (per capita and absolute level) figures for the corruption measure in the descriptive statistics.

**R3(i):** It would be good to see robustness results on the “corrupted” sample as well.

**R5(2)**: To address other possibilities, spatial regression analysis might be valuable. Women pushed out of employer and other leadership roles because of corruption might search nearby for less corrupt environments to apply their efforts. This might shed light on the full effect of corruption – does it squash these leadership (and entrepreneurial) opportunities for women or reallocate them elsewhere?

**R6(10):** To control for the size and possible endogeneity concerns, the authors should employ some variant of differences, using those in a “corrupt” environment at the treatment group and those in a non-corrupt environment as the control group. Perform matching based on the municipality, population size, age, and education structure, and compute the ATET (average treatment effect on the treated) while evaluating the matching quality. Given an inadequately described methodology and data construction, one may be curious whether the observed effect is not driven by selection and sampling issues. Therefore, the authors should be careful when describing the data, sample construction, and variable definitions.

**3. Theory, Literature, and General Discussion/Clarifications:**

**R3(a):** Last paragraph of page 3 and first paragraph of page 4 are too detailed for the introduction. I would shorten them. However, I would extend the sentence on the corruption measure (using a first sentence along the lines of the first sentence of Section 3.2).

**R4(1):** I think the authors undersell the potential importance of the employers variable. In the women-as-employee situation, it may be that men navigate the red tape to start the business that employs leaders. In this case, it seems women’s exposure to corruption affecting their placement in leadership roles as an employee is somewhat limited to whoever they might interact with in that firm to acquire a leadership role. However, employers potentially face the municipality level bureaucratic process of starting a business – which plausibly opens them up to greater corruption exposure. It would be interesting to explore that, separate from the rest. That is, inspect the corruption effect on women employers as a percent of total employers. Currently, it appears the authors are looking at women employers as a percent of all leadership roles or aggregate leadership roles held by women. (If the authors are currently regressing on women employers as a percent of all employer, then that should probably be stated clearly).

**R6(1):** The authors should carefully discuss whether the low presence of females in the leadership of firms (or regions) associated with higher corruption results from their own choice or a firm-level choice. It has been noted that corruption is often associated with men's activities; therefore, owners may prefer a corrupt CEO in a corrupt environment.

**R6(2):** The authors cite the literature studying the effect of corruption on female-led firms, criticizing that the focus of this literature is on the firm rather than on the individuals, and highlight additional firm-level effects, such as women being more likely to hire other women. Nevertheless, while these studies employed sizeable firm-level data with existing cross-sectional and time variation, the causal link between female participation and corruption is hard to identify. Existing corruption literature tends to agree that it is driven by firm owners who prefer “somebody” able to act in corrupt circles. In terms of political representation, the situation could be different.

**Minor Comments (clarification, minor edits, typos, citations, etc.):**

**R1(a):** Bottom of page 4, into page 5 – Those first two sentences are confusing. They seem to be saying the same thing, but suggesting there is a difference. I think you’re trying to get at the selection vs. quantitative effects, but this is worded confusingly.

**R1(b):** Middle of page 5 – “Corruption has no effect on female leadership presence, but female leadership representation is much lower across the board in these corruption prone sectors.” This is another case of weird wording. It sounds like the two parts of the sentence are contradicting each other.

**R2(1):** Cite the papers listed below

**R3(b):** In the first paragraph of page 10 “we scale this number by population” and in footnote 8 “we additionally weight our data by population”: what is the difference between these two? Is the population weighting done for the main analysis or for the robustness analysis? It is not clear to me what is the difference between results in Table 2, Panel A and results in Table A1, Panel A. I would use the population weighting in the main analysis and maybe, you can present a separate analysis for large and small municipalities in the robustness checks?

**R3(c):** It would be good to have more details on the corruption measure from Ferraz and Finan (2011) and how it differs with the main corruption measure used in the analysis.

**R3(f):** It would be good to add the significance levels in tables’ notes.

**R3(g):** Footnote 10 should appear earlier on in the paper.

**R3(h):** Please explain what you mean by “bad controls à la Angrist and Pischke (2009)”.

**R3(j):** Do you have data on the public sector? To look at the female share in leadership positions in municipalities.

**R5(8):** This sentence is confusing: “i.e., employees that are in a manager, general director, or executive role and employers - that are occupied by women.” Reword for clarity.

**Cite these (reviewer 2):**

Nishioka, Sharma, and Le's (2023) "Political Regimes and Firms' Decisions to Pay Bribes" would provide an excellent comparative perspective on how political systems influence corruption at the firm level, complementing this paper's municipal-level analysis.

Belloc, Burdin, and Landini's (2023) "Corporate Hierarchies and Workplace Voice" examines institutional structures within firms that affect power dynamics—directly relevant to understanding how corruption might reinforce or alter these hierarchies along gender lines.

Mahn, Poblete, Wang, and Heaton's (2024) "The Role of Culture as an Informal Institution in Cross-Border Venture Capital Investments" offers insights on how informal institutions like culture shape economic decisions, which could enrich the discussion of how corruption functions as an informal institution affecting gender outcomes.

Arif and Dutta's (2024) "Legitimacy of Government and Governance" explores how government legitimacy affects governance quality, providing a valuable lens for understanding the institutional context of corruption in Brazilian municipalities.

Aimar's (2023) work on "Destructive Entrepreneurship" could inform the analysis of how corruption diverts entrepreneurial activity, potentially explaining some of the observed effects on female employers in corruption-prone sectors.