

A Service of



Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Balán, Pablo; Bergeron, Augustin; Tourek, Gabriel; Weigel, Jonathan

#### Working Paper Response to Adjisse, Blimpo, and Castañeda Dower (2024)

I4R Discussion Paper Series, No. 192

**Provided in Cooperation with:** The Institute for Replication (I4R)

*Suggested Citation:* Balán, Pablo; Bergeron, Augustin; Tourek, Gabriel; Weigel, Jonathan (2024) : Response to Adjisse, Blimpo, and Castañeda Dower (2024), I4R Discussion Paper Series, No. 192, Institute for Replication (I4R), s.l.

This Version is available at: https://hdl.handle.net/10419/307932

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



#### WWW.ECONSTOR.EU

# **INSTITUTE** for

No. 192 DISCUSSION PAPER SERIES

# Response to Adjisse, Blimpo, and Castañeda Dower (2024)

Pablo Balán Augustin Bergeron Gabriel Tourek Jonathan Weigel

This paper responds to:

Adjisse, Sossou, Moussa P. Blimpo, and Paul Castañeda Dower. 2024. Local Taxation and Development Finance in the DRC: A Comment on Balán et al. (2022). *IAR Discussion Paper* No. 191. Institute for Replication.

December 2024



#### **I4R DISCUSSION PAPER SERIES**

I4R DP No. 192

## Response to Adjisse, Blimpo, and Castañeda Dower (2024)

Pablo Balán<sup>1</sup>, Augustin Bergeron<sup>2</sup>, Gabriel Tourek<sup>3</sup>, Jonathan Weigel<sup>4</sup>

<sup>1</sup>Tel Aviv University/Israel <sup>2</sup>Harvard University, Cambridge/USA <sup>3</sup>University of Pittsburgh/USA <sup>4</sup>University of California, Berkeley/USA

#### DECEMBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and metascientific work in the social sciences. Provided in cooperation with EconStor, a service of the <u>ZBW – Leibniz Information Centre for Economics</u>, and <u>RWI – Leibniz Institute for Economic Research</u>, I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

		Editors	
Abel Brodeur	Anna Dreber	ool of Economics	Jörg Ankel-Peters
University of Ottawa	Stockholm Sch		RWI – Leibniz Institute for Economic Research
E-Mail: joerg.peters@rwi-es	<u>sen.de</u>	Hohenzollernstraß	e 1-3 www.i4replication.org
RWI – Leibniz Institute for E	conomic Research	45128 Essen/Gerr	

#### Response to Adjisse, Blimpo, and Castañeda Dower (2024)

Pablo Balán, Augustin Bergeron, Gabriel Tourek, and Jonathan Weigel<sup>†</sup>

November 13, 2024

#### Abstract

Adjisse, Blimpo, and Castañeda Dower (2024) reproduce the tables and figures of Balán et al. (2022) with the exception an incorrect sign on one coefficient in Table 5. Although not central to Balán et al. (2022), we are grateful to have discovered this error and have submitted a corrigendum to the AER accordingly. They then conduct three additional analyses.

- They construct standard errors with randomization inference, which does not lead to qualitatively different results.
- 2. When comparing the main three treatment arms, they control for baseline trust in the chief. This covariate is Balánced for these three treatment arms, but imBalánced in a fourth treatment arm we do not study in the paper. This variable is also uncorrelated with the outcome and thus an unlikely source of omitted variable bias. By including this control, the authors restrict the analysis to the baseline survey sample, *a sample size reduction of 91%*, which not surprisingly increases the standard errors. In this sub-sample, the magnitude of the coefficient on one treatment indicator increases slightly, suggesting if anything stronger evidence of the informational mechanism proposed in the paper. Controlling for the average level of trust in the chief in the full sample leaves our results unchanged.
- 3. They examine a prediction exercise one of four auxiliary descriptive analyses regarding the mechanism in which we compare the properties visited by different types of tax collectors across treatment arms. They add 13 *binary neighborhood-level variables* to a regression predicting tax compliance *at the household level*. This unconventional prediction approach fits noise and worsens the prediction, introducing classical measurement error into the ultimate exercise. When we redo their check with LASSO, it drops all but two of these variables and replicates our results.

Although we appreciate the interest in our paper, we question the statistical value of the latter two analyses.

<sup>\*</sup>Original paper: "Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the D.R. Congo."

<sup>&</sup>lt;sup>†</sup>Affiliations: Tel Aviv University, University of Pittsburgh, Harvard University, and University of California Berkeley.

#### **1** Introduction

We are grateful to Adjisse, Blimpo, and Castañeda Dower (2024) for their replication of Balán et al. (2022). They are able to reproduce all of the tables and figures in the paper with one exception: an error in Table 5, in which the coefficient on *Incorrect exemption* has the wrong sign (the same is true in Table A9). Although not central to our analysis, we are indebted to the replicators for discovering this error. We have filed a corrigendum with the journal accordingly, as we note in more detail below.

Concerning the rest of the report, we appreciate Adjisse, Blimpo, and Castañeda Dower (2024)'s effort to investigate the validity of the results in Balán et al. (2022). However, their robustness tests depart from conventions in ways that in our opinion limit the statistical value of their analysis. Their report also contains misleading descriptions of the analyses they undertake, despite our pointing out these issues in previous correspondence. None of their analysis, in our view, provides evidence inconsistent with the discussion of mechanisms in Balán et al. (2022).

#### 2 Review of Adjisse, Blimpo, and Castañeda Dower (2024)'s Analysis

#### 2.1 Randomization inference (Section 3.1)

As Adjisse, Blimpo, and Castañeda Dower (2024) note, using randomization inference does not lead to qualitatively different results from those in Balán et al. (2022). We have no comments on this section.

## 2.2 Controlling for baseline trust in chief and restricting the sample by 91% (Section 3.2)

Adjisse, Blimpo, and Castañeda Dower (2024) control for the covariate *Trust in chief* when reestimating the main mechanism test. When reading the abstract and introduction of the replication report, this sounds like a standard robustness check. However, what the authors do not mention in these sections of the replication report is that this variable is only available for the baseline survey sample and so controlling for it reduces the sample by 91%. Moreover, Adjisse, Blimpo, and Castañeda Dower (2024) refer to the covariate as "imBalánced." It is not imBalánced in the two main treatment comparisons of the paper, which they focus on. It is imBalánced in a separate treatment arm called Central X Local that we barely mention in the paper (and nor do Adjisse, Blimpo, and Castañeda Dower (2024)). Thus, when comparing our main treatment groups — Central v. Local, and Central v. Central + Local Information — this covariate is in fact Balánced, despite the impression a reader might get from the replication report.

We question the statistical value of controlling for a Balánced covariate that drops 91% of the sample for several reasons.

- 1. Trust in chief is uncorrelated with the outcome. The variable Trust in chief is only weakly correlated with tax compliance and revenue in our data. For example, the correlation coefficient between trust in the chief and tax compliance is very small ( $\sigma$ =-0.0095) and not statistically significant (p-value=0.5660).<sup>1</sup> The lack of a significant relationship between trust in the chief and tax compliance or revenue makes it unlikely that this variable could constitute a source of omitted variable bias. This also makes it unlikely that controlling for *Trust in chief* would affect the estimates were it not for the different and substantially smaller sample in this estimation.
- 2. Low statistical power. Reducing the sample by 91% severely limits the statistical power of the tests run by Adjisse, Blimpo, and Castañeda Dower (2024). We illustrate this point using simple power calculations with data on tax compliance in Kananga: the comparison of Central v. Local has power of 0.96 in our original specification and 0.67 when restricting to the baseline sample (Figure 1, Panel A). Adjusting the power estimation for Central v. Central + Local Information suggests power of 0.78 in the original specification and 0.42 when restricting to the baseline sample (Figure 1, Panel B). Restricting the sample by 91% results in underpowered tests.<sup>2</sup>

<sup>&</sup>lt;sup>1</sup>We find similarly small and insignificant correlations between tax compliance and trust in the chief in Central ( $\sigma$ = 0.0277, p-value=0.3545), Local ( $\sigma$ = -0.0312, p-value=0.2797), and CLI ( $\sigma$ = -0.0158, p-value=0.6574).

<sup>&</sup>lt;sup>2</sup>At times, we also conduct analysis using the endline survey sample in the paper, but only under specific conditions. First, we do this when the outcome variable is only observed in the endline survey (e.g., attitudes about the government or reported interactions with tax collectors). We do not do this when studying our main outcomes that we observe in the administrative data. Second, we are better powered to detect effects in small samples for these Likert-style survey variables than for tax compliance. To illustrate this point, we repeat the power calculations noted above using *View of government (index)* that we consider in Table 5 Panel C in Balán et al. (2022). The results presented in Figure 1 (Panel C) confirm that we are powered above 0.80 to detect effects of size 0.17 standard deviations with the endline sample of 2,411. This is in fact an *underestimate* of our statistical power because in most of the analysis using endline survey variables (e.g., Table 5), we are able to control for the baseline value of the outcome, which improves precision and raises power. We therefore view Figure 1 Panel C as a likely lower bound on power when analyzing Likert-style survey questions in our endline survey.

- 3. *A more powered test is available.* Rather than discarding 91% of the sample, we proposed to the replicators a more powered test: controlling for the average level of *Trust in chief* in the neighborhood while estimating the results in the full sample. We report the results, similar to those reported in Balán et al. (2022), in Table 1. We remain puzzled why Adjisse, Blimpo, and Castañeda Dower (2024) prefer their approach over this more powered alternative.
- 4. *Non-standard variable construction*. The variable under discussion, *Trust in chief*, is a Likertscale variable ranging from 0 (low trust) to 4 (high trust). To control for it, Adjisse, Blimpo, and Castañeda Dower (2024) introduce dummies for each value of the variable rather than simply controlling for the variable. We are unsure why they chose this approach since it diverges from the standard approach and the one we adopt in the paper.

Putting aside these remarks, even if we accept Adjisse, Blimpo, and Castañeda Dower (2024)'s approach, the results show that controlling for *Trust in chief* increases the gap between CLI and Central. This, in fact, constitutes stronger evidence of the informational mechanism. However, the report mentions, "The increase is so dramatic that we can no longer reject the null hypotheses that the effect of CLI is equal to the effect of Local." This is a misleading statement because the comparison of CLI to Local is in fact *not* the null hypothesis we test to assess mechanisms in the paper. We test the gap between CLI and Central. A reader who had not carefully engaged with Balán et al. (2022) might mistakenly read this statement as evidence against the informational mechanism when the opposite is true. We pointed out this ambiguity in prior correspondence, but the misleading characterization of our mechanism test persists in the final version.



#### Figure 1: Power Comparisons: Full Sample v. Baseline Sample



4000

Total sample size (N)

6000

8000

2000

ċ



*Notes*: Panel (a) estimates power for tax compliance in Central v. Local for different sample sizes. The first vertical line indicates the sample size of 2,439 (Adjisse, Blimpo, and Castañeda Dower, 2024); the second, 27,764 (Balán et al., 2022). Panel (b) estimates power for tax compliance in Central v. Central + Local Information for different sample sizes. The first vertical line shows a sample size of 1,797 (Adjisse, Blimpo, and Castañeda Dower, 2024); the second, 20,636 (Balán et al., 2022). Panel (c) estimates power for a Likert survey outcome for Central v. Local with a sample size of 2,411 (the baseline sample). All assumptions are based on published results, and computations use Stata's power command.

	Compliance	Revenues	Visited	Visits	Compliance	Compliance
	(1)	(2)	(3)	(4)	(5)	(6)
Central Plus Local Info	0.029**	55.773**	-0.016	-0.019	0.030**	0.025**
	(0.009)	(21.389)	(0.030)	(0.046)	(0.013)	(0.009)
Local						0.048***
2000						(0.007)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
House FE	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20636	20636	13884	13877	5283	33746
Clusters	165	165	163	163	161	267
Mean	0.051	150.66	0.387	.497	0.097	0.052
Test CLI = Local ( <i>p</i> -value)						0.011

Table 1: Central versus CLI: Controlling for Neignbornood Mean of Trust of G	ghborhood Mean of Trust of Cl	eighbor	for N	Controlling	CLI:	versus	Central	e 1:	Tab
--	-------------------------------	---------	-------	-------------	------	--------	---------	------	-----

*Notes:* Notes: This table compares the CLI arm to the Central arm, which is the excluded category. Columns 1, 5, and 6 report effects on compliance. Column 2 reports effects on revenues. Columns 3 and 4 report differences in tax visits by collectors after registration by the extensive and intensive margins, respectively. All regressions include fixed effects for house type, randomization strata, and time periods and cluster standard errors at the neighborhood level. All specifications include time fixed effects defined to maximize overlap between the treatments under comparison, as discussed in Section IV of the "Local Elites" paper. We also include as a control the neighborhood-level mean of trust of the neighborhood chief as measured in the baseline sample. Column 5 restricts to the subsample of properties that received any tax visits after registration. Column 6 includes a dummy for the Local treatment. The bottom row reports the p-value from a test for equality between the CLI and Local.

### 2.3 Predicting household-level payment propensity with neighborhood-level variables (Section 3.3)

In Balán et al. (2022), our main mechanism test compares Central to Central + Local Information (CLI). This comparison exploits experimental variation and is pre-registered in our PAP. After this test, we examine four additional observable implications consistent with the idea that chiefs' information is an important channel explaining their ability to collect more taxes than state collectors. In the second of these four additional analyses, we investigate whether the types of properties visited by collectors in CLI resemble those visited by chiefs in Local. We do this by predicting properties that chiefs would have recommended in Local and Central using a propensity score approach on a set of household characteristics.

This particular analysis is the one that Adjisse, Blimpo, and Castañeda Dower (2024) chose to study. They refer to this exercise as providing the "smoking-gun evidence" for our preferred mechanism. This diverges from the language used to refer to this analysis in the paper and could mislead readers into thinking this is our main mechanism test, which it is not. It is rather a descriptive exer-

I4R DP No. 192

cise meant to assess if chiefs were recommending visiting properties in the CLI treatment arms that are similar to those they themselves would be targeting when working as tax collectors. Even if the chiefs recommended different types of properties to those they would themselves target as collectors, this would not challenge the idea that the gap between CLI and Central provides evidence of an informational mechanism. It would just imply that chiefs are using their information differently when working as consultants and when working as collectors; it would also imply a distinct targeting pattern in CLI relative to Local because of this different information. However, the evidence suggests they did recommend similar properties as consultants compared to those they targeted as collectors.

In any case, Adjisse, Blimpo, and Castañeda Dower (2024) revisit the prediction of households' "ease of payment" and "willingness to pay" by including additional binary neighborhood-level characteristics as covariates in the prediction exercise. We again question the statistical value of this exercise for several reasons.

- 1. *Neighborhood-level binary variables are unlikely to be good predictors of household-level tax compliance.* The goal of this exercise is to assess what household-level attributes predict a chief ranking a household as high (or low) *Ease of payment.* Adding many neighborhood-level covariates is unlikely to help us achieve that goal.<sup>3</sup> What it will likely do is introduce noise into the prediction, which causes classical measurement error and thus attenuation bias. We illustrate this point in a few steps.
  - (a) First, Adjisse, Blimpo, and Castañeda Dower (2024) find that the MSE decreases slightly when they add these variables to the prediction regression (Table 8P). But they are assessing the MSE in-sample and thus potentially overfitting (i.e., fitting noise). Given that this is a prediction exercise, it would be more informative to assess prediction performance out-of-sample. To do this, we instead implement a simple test of the predictive power of the respective approaches, splitting the CLI sample in two to allow generation of the prediction in one half and correlating with actual chief consultation variables in the other. We implement this exercise across 100 iterations for the three prediction procedures con-

<sup>&</sup>lt;sup>3</sup>For example, the variables they add to the regression include dummies for chiefs being older than the median chief, for neighborhoods being above the median in distance to the city center, and 11 other indicator variables of this nature.

I4R DP No. 192

sidered by Adjisse, Blimpo, and Castañeda Dower (2024): ours, their ordered Probit, and their ordered Probit including the additional variables. This exercise reveals that adding the neighborhood-level variables to the prediction equation substantially worsens the prediction for both ease of payment and willingness to pay (Figure 2). When these noisier variables are then included in the estimations summarized in Table 8 they create classical measurement error and attenuation bias.

- (b) Second, a better way to avoid fitting noise and to select a parsimonious set of variables that are predictive of the outcome is some form of penalized regression that upweights variables closely related to the outcome and downweights those weakly related to it. To do this, we re-run the results using LASSO to determine which predictors to include and reproduce our results in Table 2.<sup>4</sup> This procedure suggests that many of the additional characteristics included by Adjisse, Blimpo, and Castañeda Dower (2024) have limited predictive power. For example, for *Ease of payment* the LASSO procedure selects all of the property owner characteristics included in our original prediction strategy but only 2 from Adjisse, Blimpo, and Castañeda Dower (2024)'s set of additional variables.<sup>5</sup> Using the LASSO-predicted measures, we replicate the results in our original Table 8, and the results are very similar (see Table 2).
- 2. Including these variables in the prediction also drops 17 neighborhoods and 167 observations (across Local and Central) from the analysis because of missing data for some of the variables. This is important to note as it could be another reason why would one expect the results reported in Table 8R2 to be different from those reported in Table 8 of Balán et al. (2022). Despite making note of this in prior communication, Adjisse, Blimpo, and Castañeda Dower (2024) chose not to mention this additional difference between their analysis and that considered in the paper.

<sup>&</sup>lt;sup>4</sup>We use the "lasso" command in Stata, for which the model selection method is cross-validation. For consistency, we categorize predicted values into a three-point-scale to match the distribution of the measures provided by chiefs in CLI neighborhoods, as we did in Balán et al. (2022) for the original prediction measure. However, if we do not impose cutoffs, the results look very similar.

<sup>&</sup>lt;sup>5</sup>These include an indicator for the chief's view of the government being above median, and an indicator for the chief having a family member who was previously chief in the neighborhood.



Figure 2: Validating Prediction Approaches in CLI Neighborhoods

*Notes*: This figure displays the estimated relationship between predicted measures of ease of payment (solid markers) and willingness to pay the property tax (hollow markers) and the actual values solicited through chief consultations in the CLI treatment arm. This relationship is estimated in three ways: using the Balán et al. approach (blue circles); Adjisse, Blimpo, and Castañeda Dower (2024)'s ordered Probit version (red triangles) as in their Table 8R3; and the Adjisse, Blimpo, and Castañeda Dower (2024)'s ordered Probit controlling for "chief characteristics" (green squares) as in their Table 8R2R3. The values displayed for each procedure are the mean coefficient values estimated for the relationship between indicators provided by chiefs and predicted indicators using each procedure across 100 iterations in each of which the sample of CLI neighborhoods is randomly split in half, the respective prediction is estimated in one half, and then the relationship between the true and predicted measure is estimated in the remaining half. Vertical gray bars display 95% confidence intervals for the 100 coefficient values estimated through all iterations.

	Visited (1)	Compliance (2)	Visited (3)	Compliance (4)	Visited (5)	Compliance (6)	Visited (7)	Compliance (8)
Panel A · Fase of payment	(-)	(-)	(-)	(-)	(2)	(*)	(.)	(*)
Ease of payment	0.045*** (0.012)	0.056*** (0.007)	0.029** (0.014)	0.044*** (0.008)				
Predicted ease of payment					0.039** (0.019)	0.050*** (0.010)	0.026 (0.018)	0.020** (0.007)
Wall quality			0.025** (0.012)	0.021** (0.007)	0.010 (0.011)	0.014** (0.007)	0.022** (0.011)	0.012** (0.005)
Roof quality			0.005 (0.006)	-0.000 (0.002)	0.006 (0.008)	0.001 (0.004)	0.018** (0.008)	-0.010 (0.006)
Erosion threat			0.017 (0.011)	-0.004 (0.004)	-0.004 (0.012)	-0.012 (0.007)	-0.001 (0.010)	-0.005 (0.005)
Observations Clusters Mean	5572 79 .376	8133 80 .072	4551 66 .352	5150 66 .065	4908 80 .448	4922 80 .112	4725 76 .41	4731 76 .059
Panel B: Willingness to pay Willingness to pay	0.034** (0.011)	0.037*** (0.007)	0.033** (0.012)	0.038*** (0.008)				
Predicted willingness to pay					0.044** (0.016)	0.027** (0.009)	0.021 (0.016)	0.037*** (0.009)
Wall quality			0.022 (0.013)	0.021** (0.009)	0.014 (0.010)	0.018** (0.006)	0.016 (0.011)	0.009* (0.005)
Roof quality			0.011 (0.008)	0.001 (0.002)	0.003 (0.006)	-0.004 (0.004)	0.019** (0.008)	-0.009 (0.006)
Erosion threat			0.016 (0.012)	-0.005 (0.005)	0.001 (0.011)	-0.008 (0.007)	0.000 (0.009)	-0.006 (0.004)
Observations	3933	5521	3929	4461	5593	5608	4903	4909
Clusters	50	50	50	50	90	90	79	79
Mean	.357	.062	.357	.066	.433	.105	.406	.058
Treatment	CLI	CLI	CLI	CLI	Local	Local	Central	Central
House FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2. The value of Chiefs Thior mation — Using LASSO in Freuleno	Table 2: The Va	alue of Chiefs' Inf	formation — Using	2 LASSO ir	1 Prediction
---	-----------------	---------------------	-------------------	------------	--------------

*Notes*: This table replicates Table 8 in Balán et al. (2022), which explores the extent to which chiefs' recommendations in Central + Local Information (CLI) predict tax visits after registration and tax payment, replacing the prediction approach used in Balán et al. (2022) with a LASSO procedure. Columns 1–4 show correlations in CLI between chiefs' recommendations and outcomes. Columns 5–8 report correlations between predicted propensity measures and outcomes in Local (Columns 5 and 6) and Central (Columns 7 and 8). Prediction measures are estimated using LASSO to select covariates. Columns 1, 3, 5, and 7 show correlations between propensity and tax visits; Columns 2, 4, 6, and 6 show correlations between propensity and compliance. All regressions include house type and randomization stratum fixed effects and cluster standard errors at the neighborhood level. Columns 3, 4, and 5–8 include controls for visible household characteristics. Number of observations and clusters differ from Table 8 in Balán et al. (2022) because of the inclusion of additional covariates identified by Adjisse, Blimpo, and Castañeda Dower (2024) and used in their Table 8R2R3, which reduces the sample by excluding neighborhoods where these variables are not available.

#### **3** Misleading Descriptions

Adjisse, Blimpo, and Castañeda Dower (2024) use language throughout their replication report that obscures (*i*) the analyses they do and (*ii*) what analyses we do in the original paper. Here we review several examples of misleading language, all of which we pointed out in previous correspondence.

- 1. When discussing controlling for baseline trust in the chief in the abstract and introduction, it is important to note that this reduces the sample size to the baseline sample, which is 9% of the total sample, so that readers can understand all of the ways the results of this test might differ from the results reported in Balán et al. (2022). Adjisse, Blimpo, and Castañeda Dower (2024) chose not to report this sample size reduction in the abstract or introduction.
- 2. Adjisse, Blimpo, and Castañeda Dower (2024) assert that the information mechanism is the "main claim" of Balán et al. (2022) (Pages 3-4). We never use this language in the paper, and it could be misleading to readers who have not read the paper. It would be more precise to refer to the information mechanism as "the mechanism for which the authors find the most evidence" rather than the "main claim" of the entire paper.
- 3. The discussion of the translation of the trust variable (Page 7) is misleading. The "mistaken translation" only occurs in the French to English translation of this variable in Balán et al. (2022)'s Appendix, where we list all variables used in the analysis. The variable was correctly posed to respondents in French or Tshiluba and correctly translated as "trust" in the main text. What Adjisse, Blimpo, and Castañeda Dower (2024) are pointing to here was a literal translation from French to English that exists *only in the online Appendix*. However, the discussion in Adjisse, Blimpo, and Castañeda Dower (2024) makes it seem like the variable was mistranslated in the field or Balán et al. (2022) in a way that would change our substantive results. This is not the case, and we pointed this out in two rounds of previous correspondence. But the same text appears in the final version.

Institute for Replication

#### 4 Comments on Discussion

- 1. On Page 16, Adjisse, Blimpo, and Castañeda Dower (2024) suggest that city chiefs might have recommended households in CLI and then, after the consultation, tried to persuade these households to pay. We think this explanation is unlikely for two reasons. First, chiefs in CLI had no obvious financial or career/promotion incentive to talk to households after the consultations (which would require costly effort). Second, the evidence is not consistent with this explanation. When asked what type of agent came to visit households for the property tax campaign, 94% of households in CLI say the state collectors, 2.6% say the city chief, and 3.4% say both. When asked the question "Aside from their visits to your compound, have you spoken with a DGRKOC<sup>6</sup> tax collector or the chef de localiteé (and/or his assistants) about the property tax since May?" 96.1% of respondents said "no" and 3.9% said "yes." According to the available data, then, it appears very rare that chiefs in CLI talked to households about property taxes, which makes this explanation unlikely.
- 2. The bribe results in Adjisse, Blimpo, and Castañeda Dower (2024)'s Table 1 include two out of four measures of bribe incidence we consider in Balán et al. (2022) Table 5: *Gap self versus admin* and *Other payments* are both missing. We provide a complete picture of the bribe results in Table 3, showing for each of our four bribe measures: (*i*) bribe incidence (extensive margin), (*ii*) bribe magnitude among payers (intensive margin), and (*iii*) total bribes (incidence x magnitude). To clarify, Adjisse, Blimpo, and Castañeda Dower (2024)'s Table 1 included only (*i*) and (*iii*) for two bribe measures out of four. We find the following results:
  - (a) <u>Incidence</u> significantly increases for three of the four measures.
  - (b) <u>Magnitude</u> significantly decreases for one of the four measures with negative and not statistically significant coefficients for two measures and a positive and not statistically significant coefficient for one measure.
  - (c) <u>Total bribes</u> significantly increases according to one measure, significantly decreases according to another measure, and the other two measures have positive but not statistically

14

<sup>&</sup>lt;sup>6</sup>The DGRKOC is the provincial tax authority.

significant coefficients.

The results are mixed, but the clearest pattern to emerge appears to be a small increase in incidence, especially given the results for the non-survey-based measures. This is the result we focus on in Balán et al. (2022).

Dependent variable	β	SE	$R^2$	Ν	$\bar{x}_{Central}$
Paid Bribe (Midline)	-0.001	0.003	0.007	18596	0.016
Bribe Amount (Midline) Payers	-367.962	316.078	0.496	266	1803.788
Bribe Amount (Midline) Total	-12.527**	5.008	0.010	18570	26.353
Gap Self v. Admin (Midline)	0.016*	0.009	0.018	14309	0.076
Gap Self v. Admin Amount (Midline) Payers	-215.759*	121.403	0.143	1644	1459.389
Gap Self v. Admin Amount (Midline) Total	20.761	19.764	0.018	13867	123.647
Paid Bribe (Endline)	0.018*	0.009	0.049	1169	0.014
Bribe Amount (Endline) Payers	-533.333	354.560	0.976	25	975
Bribe Amount (Endline) Total	19.857	12.514	0.030	1169	13.732
Other Payments (Endline)	0.031**	0.014	0.041	2407	0.094
Other Payments Amount (Endline) Payers	2476.262	1817.303	0.147	268	3514.114
Other Payments Amount (Endline) Total	420.366**	207.041	0.030	2405	330.285

Table 3: Local v. Central: Bribes — Extensive and Intensive Margins

*Notes*: Each row summarizes an OLS estimation of the main estimating equation in Balán et al., comparing Local and Central, with the dependent variable noted in the first column.  $\hat{\beta}$  is the coefficient on the treatment indicator, followed by the cluster-robust standard error,  $R^2$ , number of observations, and  $\bar{x}_{Central}$ , the Central group mean. The outcomes in rows 1 and 7 are self-reported indicators for bribe payment as measured during the midline and endline surveys, respectively. The outcome in row 4 indicates property owners who reported paying the tax but who were not recorded as having paid in the administrative data. The outcome in row 10 is self-reported payment of any informal fees at endline. The intensive margins of these measures, among those reporting (first row following) and in total (indicators multiplied by reported amount, second row following) appear in the rows below each measure. For informal payments we report the monthly values. In all panels, regressions include fixed effects for randomization strata, and cluster standard errors at the neighborhood level. Regressions estimating effects on midline outcomes include time period fixed effects described in Balán et al. and house type fixed effects. We do not include house type fixed effects for endline outcomes to maximize the analysis sample, as discussed in Balán et al. The number of observations varies across regressions due to (*i*) outcomes being drawn from different surveys, and (*ii*) non-response for specific survey questions.

3. Adjisse, Blimpo, and Castañeda Dower (2024) argue that we push an "extraction view" of chiefs. We do not agree with this characterization. We highlight the risk of local corruption in the motivation because this is the classic tradeoff focused on in the literature on delegation. But in Balán et al. (2022) we report results that are both consistent and inconsistent with this view, and we certainly are not trying to espouse the view that chiefs are mostly "extractive". As examples of evidence we provide that is inconsistent with this view, we note the lack of evidence that chiefs increased contributions for *salongo* (Table A11) — a key "social extraction" tool

I4R DP No. 192

available to chiefs (Lust and Rakner 2018) — as well as the null effects on tax morale (which one might have expected to decrease with an extractive view of chiefs, Table 5). We also provide evidence that the chiefs were not colluding with households on the property assessment margin, which is often the key source of corruption for property taxation (Khan et al. 2015). We think Adjisse, Blimpo, and Castañeda Dower (2024)'s focus on bribes here is therefore a bit misleading and leads them to mischaracterize Balán et al. (2022) as pushing a certain view of chiefs that we do not identify with.<sup>7</sup>

4. Moreover, Adjisse, Blimpo, and Castañeda Dower (2024) assert that we hypothesize that Local should cause an increase in total bribe payments — bribe incidence × bribe magnitude — whereas this is not the case. We do not claim this theoretically nor do we test for this empirically. The reason is that we believe the effect of Local on the magnitude of the equilibrium bribe is theoretically ambiguous. As discussed in Reid et al. (2023), the magnitude of the equilibrium bribe is typically modeled with a Nash bargaining setup, where the surplus from collusion with the chief could be either higher or lower than the surplus from colluding with a state collector.<sup>8</sup> Because we think this relationship is theoretically ambiguous, we do not test the hypothesis that Local would impact bribe magnitudes or total bribes paid. But this is the test that Adjisse, Blimpo, and Castañeda Dower (2024) conduct.<sup>9</sup> In sum, we believe it is misleading to attribute a view to us (the "extractive view"), assert an associated theoretical relationship that we think is ambiguous and do not test, and then argue that finding inconsistent results from midline to endline constitutes evidence against "our view."

<sup>&</sup>lt;sup>7</sup>Adjisse, Blimpo, and Castañeda Dower (2024) appear to have been influenced by the use of the term "local elites" in reaching this assessment that we seek to promote an "extractive view" of chiefs. As we noted in our previous response, it is difficult to find a term that will please everyone, and we characterize the institution of city chieftaincy in DRC in Section 3.1 in a way that we believe is inconsistent with a solely "extractive" view of chiefs.

<sup>&</sup>lt;sup>8</sup>For instance, a citizen might have weaker tax morale before a chief collector (compared to a state collector) — which would increase the magnitude of the equilibrium bribe — but the chief might also have higher psychic/social costs of bribes because he lives in the neighborhood — which would decrease the equilibrium bribe.

<sup>&</sup>lt;sup>9</sup>We also find it unconvincing to argue that a null treatment effect on bribe magnitude or total payments could be interpreted as conclusive evidence for any mechanism, including the alternative "social contract" view proposed by Adjisse, Blimpo, and Castañeda Dower (2024). Specifically, Adjisse, Blimpo, and Castañeda Dower (2024) suggest that this lack of correlation between chief collection and the amount of total bribes supports an alternative view of chiefs (in contrast with our "extractive" view), where "chiefs may use their new responsibility as an additional tool to spread risk and redistribute resources within the community." They claim that this view would predict no correlation between chief collection and total bribes, but we do not think this is a sufficiently sharp prediction to provide evidence of a given mechanism.

- 5. Also, Table 1 has a more subtle problem in that the magnitude of the equilibrium bribe is affected by both the direct effect of treatment and a selection effect because treatment causes some marginal individuals to switch from paying bribes to paying taxes instead. Currently taxpayers are coded as zeros in this estimation (as well as our replication of Table 1 in Table 3), but that is potentially misleading because it fails to capture the selection effect. This issue is described in more detail, and how to model it, in Reid et al. (2023).
- 6. The discussion of chiefs on pages 14-15 is interesting, but a bit at odds with the specific context (though surely relevant for thinking about how the results might travel to other contexts). As we note in Balán et al. (2022) on Page 9, city chiefs are more accountable to the people than to the government. For instance, the reason that some city chiefs are (very rarely) deposed is due to local community sentiment i.e., local allegations of corruption rather than top-down decisions. We learned this through focus groups and surveys with these chiefs as well as other city officials. The city government's approval of the chiefs is much more of a rubber-stamping process.
- 7. We are not sure what assumptions Adjisse, Blimpo, and Castañeda Dower (2024) make when assessing the revenue implications to reach the conclusion that they are "negligible." It is true that the fiscal deficits facing local governments in DRC are severe. That said, we find it puzzling why a 44% increase in revenue from the property tax a crucial local tax that in total accounts for about one-third of provincial tax revenue would be so readily dismissed.

#### 5 Comments on *Conclusion*

- 1. In the conclusion and at several points in their report, Adjisse, Blimpo, and Castañeda Dower (2024) suggest, sometimes indirectly, that persuasion might be an additional mechanism (which in fact was also our prior before running the experiment). If they want to push this interpretation, we would suggest they also consider, at least in passing, the evidence we provide about persuasion in the paper. Here we provide a brief summary of these results.
  - (a) First, in Balán et al. (2022), there is an overall logic of our mechanism tests. We partition

the set of all possible mechanisms into three broad families: (1) that chiefs made more visits (*more visits*), (2) that they better targeted those visits (*information/targeting*), and (3) that, conditional on visits, they better persuaded households to pay (*persuasion*). This structure is appealing because *any* mechanism has to fit into one of these families. We consider multiple empirical implications of each mechanism, and the evidence appears most consistent with the information/targeting mechanism. Below we summarize the evidence inconsistent with the persuasion mechanism.

- (b) The first piece of evidence concerning the persuasion mechanism concerns tax collection during property registration. According to a persuasion mechanism, chiefs should have been able to collect more even when their ability to target different types of households is neutralized. The structure of the tax campaign offers a way to test this: during the registration phase of the campaign, chiefs had to work in a linear house-to-house pattern, assigning tax IDs, delivering tax bills, and soliciting payments. We confirm they followed this pattern using GPS data. We can therefore examine if chiefs collected more than state collectors when we know they were unable to target their visits. We find no difference between Central and Local collection during registration (Table A26).
- (c) We also estimate heterogeneous treatment effects by baseline proxies for chiefs' power and their role in providing public goods. If chiefs collected more taxes thanks to greater powers of persuasion, then the treatment effect should be more pronounced where chiefs were more powerful, trusted, and active in service provision. Yet, we find little evidence of heterogeneity along these dimensions (Table A27).
- (d) Finally, we examine heterogeneity by cross-randomized messages on tax notices designed to test specific persuasion mechanisms by making more salient (*i*) the chief's role in local public goods provision, or (*ii*) the chief's enforcement capacity. But we again don't find evidence of meaningful heterogeneity between these messages and the Local treatment arm, as would be predicted by persuasion mechanisms.
- (e) By contrast, we find several pieces of evidence consistent with an informational/targeting mechanism:
  (a) comparing Central with Central + Local Information the main test,

built into the experimental design — (b) checking if chiefs recommended properties in CLI similar to those they themselves visited in Local, (c) examining heterogeneity in CLI based on how informed different city chiefs were, and (d) comparing Central collectors working close to their homes. All of these provide evidence consistent with an informational mechanism. When all of this evidence is considered together, we believe that most readers would likely arrive at a similar conclusion to that we reach in Balán et al. (2022).

#### 6 Discussion

A reader of Adjisse, Blimpo, and Castañeda Dower (2024) might get the impression that (*i*) we only consider the information mechanism, and (*ii*) we push this mechanism in an unqualified way. However, both of these impressions would miss the mark. As the previous section makes clear, we examine the observable implications of many potential mechanisms and conclude that the evidence most strongly supports the information story. That said, we also make clear in Balán et al. (2022) that information does not explain the entire gap between Central and Local and note: "there may thus have been other dimensions of chiefs' information useful for targeting tax visits that were not transmitted during consultations, or other mechanisms also at work" (Page 23).

Given this qualified and evidence-based position on mechanisms, we were puzzled by Adjisse, Blimpo, and Castañeda Dower (2024)'s dismissal of our analysis in this part of the paper for several reasons. First, as our response details, we do not find the specific tests run by Adjisse, Blimpo, and Castañeda Dower (2024) to be dispositive in providing evidence inconsistent with the informational mechanism. Most of their analysis reaffirms the importance of information. Second, in their previous report, Adjisse, Blimpo, and Castañeda Dower (2024) wrote that they "did not find any strong evidence supporting other explanations" (Page 10).<sup>10</sup> Finally, even if one of our supplemental tests of the informational mechanism was less robust — though we do *not* think this is the case — one would still need to judge the most likely mechanism based on the global evidence examined by Balán et al. (2022). We believe it is unlikely that readers who did this would reach a substantively different

<sup>&</sup>lt;sup>10</sup>In the Discussion Section, they speculate that persuasion could also be important without finding any direct evidence, and without discussing the empirical tests we conduct in Balán et al. (2022) to directly test the persuasion channel (Section VC).

conclusion from ours.

#### 7 Details on Corrigendum

The incorrect sign on the variable *Incorrect exemption* in Tables 5 and A9 means we must modify the phrase "chiefs were more likely to (correctly) exempt households" (p. 18) by deleting the parenthetical. It remains the case that chiefs are more likely to exempt households, and in particular, to exempt the elderly and disabled. This corrected result tells us that, if anything, chiefs were 1.2 ppts more likely to incorrectly assign exemptions, though this difference is just below the threshold for statistical significance at the 10% level (p=0.107). We have written to the AER to inform them of this mistake and the changes that must be made. We have also modified the online Appendix, which includes the literal translation of French *confiance* as "confidence" instead of the more accurate translation "trust" (as we refer to it in the text). We are very grateful to Adjisse, Blimpo, and Castañeda Dower (2024) for pointing out these inaccuracies in the paper.

#### 8 References

- Althouse, A. (2021). Post hoc power: Not empowering, just misleading. Journal of Surgical Research, 259, A3-A6.
- Gelman, A. (2019) "Don't Calculate Post-hoc Power Using Observed Estimate of Effect Size", Annals of Surgery, 269(1): pp. 9-10.
- Goodman, S. N., & Berlin, J. A. (1994) "The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results", Annals of Internal Medicine, 121(3): pp. 200-206.
- 4. Hoenig, J. M., & Heisey, D. M. (2001). "The abuse of power: The pervasive fallacy of power calculations for data analysis", The American Statistician, 55(1): pp. 19-24.
- Khan, Adnan Q., Asim I. Khwaja, Benjamin A. Olken. "Tax farming redux: Experimental evidence on performance pay for tax collectors." The Quarterly Journal of Economics 131, no. 1 (2016): 219-271.

- Lust, Ellen, and Lise Rakner. "The other side of taxation: Extraction and social institutions in the developing world." Annual Review of Political Science 21 (2018): 277-294.
- Reid O, Tshimanga Kabasubabu JF, Weigel JL. (2023) "The Supply of Bribes: Evidence from Roadway Tolls in the D.R. Congo."
- Weigel, Jonathan L. (2020) "The Participation Dividend of Taxation: How Citizens in Congo Engage More with the State When It Tries to Tax Them." Quarterly Journal of Economics 135 (4): 1849-1903.