Fraud Detection Under Limited State Capacity: Experimental Evidence form Senegal

Léo Czajka,[1] Bassirou Sarr,[2] Mattea Stein[3]

[1]Paris School of Economics
[2]Ministère des Finances et du Budget du Sénégal
[3]University of Naples Federico II and CSEF

July 2, 2025

Motivation

In 2019, Tax-to-GPD ratio = 12.8% in LICs, vs 32.% in HICs (OECD)

- ↑ tax revenue = ↑ public goods, ↓ aid/debt
- harmonized policy frameworks, but need for better enforcement

Tax enforcement is subject to severe **constraints** in LICs:

- harder to *find*, *reach out to*, and *convince* taxpayers (widespread poverty, imperfect infrastructure, limited welfare state)

Tax administrations (TAs) in LICs cannot go after *all* tax evaders and need **information**:

- to detect where evasion is most severe and strategically allocate scarce resources

This paper

Question: Can TAs efficiently collect information to identify *large* tax evaders?

Result: Yes! By better enforcing third-party reporting by the largest firms.

$\Rightarrow$ cheap, efficient, and easily scalable intervention

Institutional Context

Senegal - Tax/GDP $\approx$ 17% - standard policy framework - narrow, very concentrated tax base.

Tax-registered firms must submit a **list** of their **service suppliers**; & for each supplier:

- total **amount** paid to them
- their **identification number** (**ID**):
  - tax-ID if they are tax-registered (*formal*)
  - personal-ID otherwise (*informal*)
- + name, nature of the service, address and phone number

Used by the TA to:

1) check **deductible amounts** when assessing the Corporate Income Tax of the **client**.
2) **cross-check** amounts received by tax-registered **suppliers** vs. their self-declared revenue.

$\Rightarrow$ always performed on a **case by case basis**

Supplier Lists at Baseline

We collected 6,706 lists submitted by the 4,000 largest firms (clients) for 2018-2020.

- largest firms: vast supply network, communication compliance

Misreporting is **very prevalent**:

- 89% of the clients (3,487) misreported at least once over 2018-2020
  - *Misreporting* = omitting a supplier's ID, or submitting a formally incorrect ID
- on average, 84% of their supplier were reported without any ID.

...Without ID, suppliers lists are useless

Intervention

Sample: 3,487 taxpayers **previously misreporting** suppliers lists - summary stats

- stratified randomization (tax center, N. of suppliers, % of misreported suppliers)

Treatment: personalized emails + verification calls:

- TA has **detected misreporting** (individualized years), demands rectifications and invites recipient to do better in the future.
- Reminder of fiscal and reporting obligations related to suppliers list, and **warning** about possible sanctions provided by the law.
- Usual communication channels.

No sanctions applied.

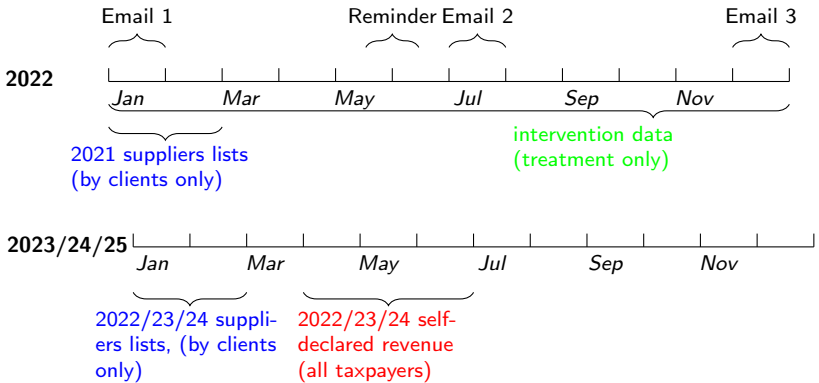
Data

Long-term collaboration with the TA, multi-year data collection effort.

We use four datasets, $>$ 2.000.000 administrative records.

- **suppliers lists**: Baseline (2018-2020) and endline (2021-2022)
  - for our sample of targeted clients
- **self-declared revenues**: we consolidate the universe of tax declarations with information on total gross revenue since (2018-2022).
  - Income Tax (Corporate, Small Business, Personal), VAT, Business Property Tax
  - for all taxpayers
- **intervention-data**: to assess the take-up
  - treatment group only
- **directories**: tax directory, business directory

Timeline: Linking Intervention and Data



Estimation

At the **client level** i :

$$y_i = \alpha + \beta Treat_i + \boldsymbol{\gamma X_i} + \delta_s + \epsilon_i \quad (1)$$

- $\boldsymbol{\gamma X_i}$: vector of controls
- δ_s: strata fixed-effects (tax centre, N. of suppliers, % misreported suppliers)

Main outcomes y_i:

- share of suppliers with a (in)valid ID.
- ... identified as informal/formal, formal and under reporting
- number of suppliers, average payment etc.

1 - Impact on Suppliers Identification - Main Results

		β	Std. err.	Lee Bounds lower	Lee Bounds upper	μ_c	N
	Probability to declare supplier list	0.017	0.013			0.772	3487
(A) Share of suppliers ...	... with a valid ID	0.247***	0.014	0.243	0.263	0.469	2633
	... with invalid ID	-0.058***	0.013	-0.079	-0.056	0.285	2633
(B) Structure of service supply	Number of suppliers	4.590	7.203	-19.075	4.853	49.903	2633
	Sum of payments (ihs)	0.036	0.056	-0.095	0.111	17.934	2633
	Average payment (ihs)	-0.021	0.043	-0.098	0.033	15.472	2633
	Median payment (ihs)	-0.079*	0.043	-0.133	-0.018	14.765	2633
	Standard Deviation (ihs)	0.105	0.171	0.014	0.425	13.928	2633

Large positive impact on the prevalence of valid IDs (+52%)

+ reduction of invalid IDs (+ pos. impact on *phone numbers*)
+ overall service supply seems unaffected

In aggregate, this allows to "better track" USD$1 billion $\approx$ 4% GDP.

2 - Impact on the Detection of Tax Evading Suppliers

Does the additional identification information point toward **tax-evading suppliers**?

We identify tax evading suppliers as:

- all **informal** suppliers (the *extensive* margin)
  - a formally correct ID that does not match the tax directory
- all **under-reporting formal** suppliers (the *intensive* margin)
  - the ID matches the tax-directory, BUT
  - the sum of payment received $>$ total revenue declared - or over-claiming clients?
    (cross-checking supplier lists with revenue database)

2 - Impact on the Detection of Tax Evading Suppliers

	β	Std. err.	Lee Bounds		μ_c	N
			lower	upper		
Share of suppliers identified as						
... informal	0.175***	0.012	0.163	0.182	0.229	2633
... formal	0.069***	0.012	0.057	0.077	0.232	2633
... ... and compliant	0.043***	0.010	0.029	0.049	0.175	2633
... ... and under-reporting	0.025***	0.005	0.008	0.027	0.056	2633

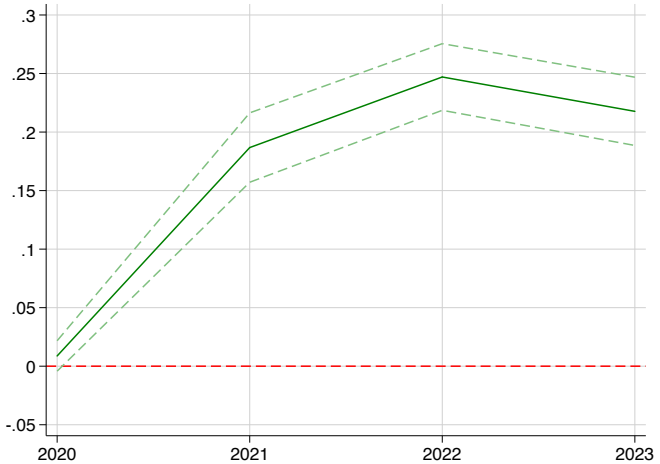
Increase in the detection of tax evading suppliers along both margins Graphical Illustration

- +76% for informal suppliers
- +44% for under-reporting formal suppliers

In aggregate, reveal \$145.5m in unreported revenue ($\approx$ 0.5% GDP).

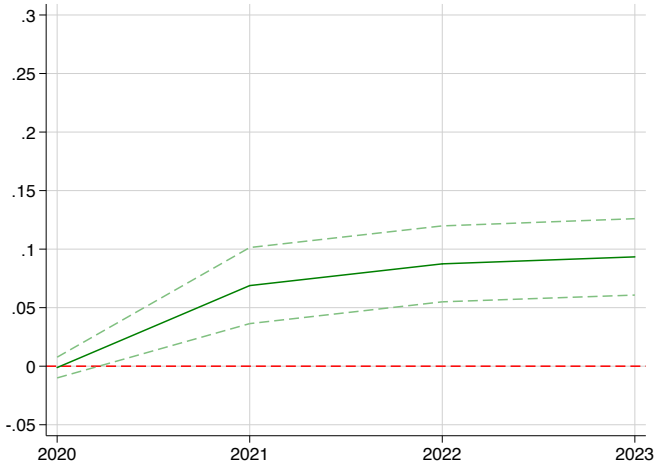
3 - All Impacts Persist After One Year

Impact on the share of suppliers with a **valid ID**

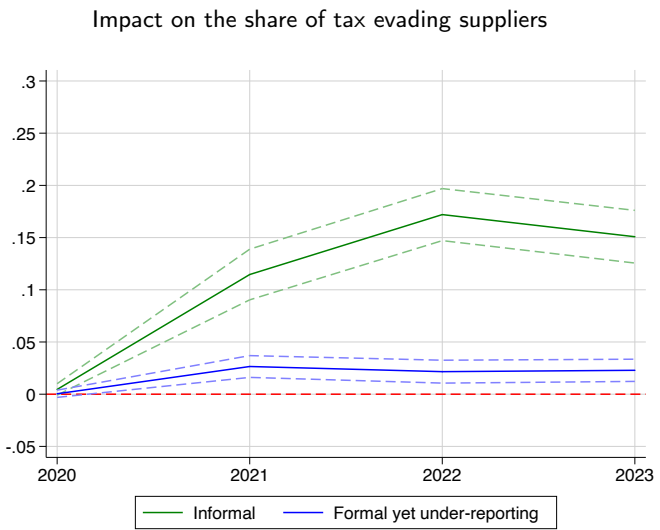


3 - All Impacts Persist After One Year

Impact on the share of suppliers with a **phone number**

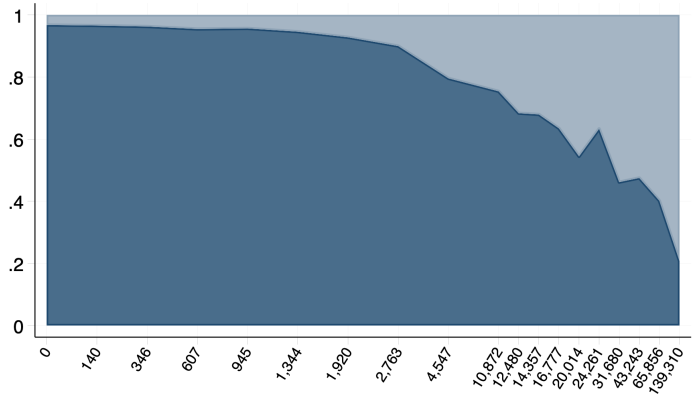


3 - All Impacts Persist After One Year



4 - Are There *Large Enough* Tax Evaders?

Ex-post distributions of tax-evading suppliers: unreported revenues are very concentrated.



4 - Simulating Additional Audit Returns - Selection Procedures

Would using the newly available information allow to better target audits?

Compare **expected total tax revenue** collected when

S_0 auditing the n largest newly detected under-reporting formal suppliers – *ranked by under-reported income*.

... versus when:

S_1 auditing of randomly selected taxpayers

S_2 auditing of the largest taxpayers (ranked by declared income)

S_3 auditing of the largest taxpayers, by centre – *ranked by declared income*

4 - Simulating Additional Audit Returns - Definitions

For each procedure S_k (with k $\in$ [0,3]), we define

$$T_k^{nr} = \sum_{i=1}^{n} (r \cdot u_i^k), \text{ s.t}$$

- $(u_i^k)_{0 \leq i \leq n}$ are unreported incomes of the n taxpayers selected by S_k
- r is the audit recovery rate

For some n, r and selection S_k, we want to estimate

$$\Delta_k^{nr} = (T_0^{nr} - T_k^{nr}) / T_k^{nr}$$

Conservative assumption: r is constant across selection procedure – so it simplifies into:

$$\Delta_k^{n} = (U_0^{n} - U_k^{n}) / U_k^{n}$$

where $U_k^n = \sum_{i=1}^{n} u_i^k$

4 - Simulating Additional Audit Returns - Hypothesis

For each i, k, we define $u_i^k = c_i^k + a_i^k$ where

- c_i^k = unreported income revealed by cross checking (measured thanks to intervention)
- a_i^k = unreported income revealed by audits themselves (to be estimated)

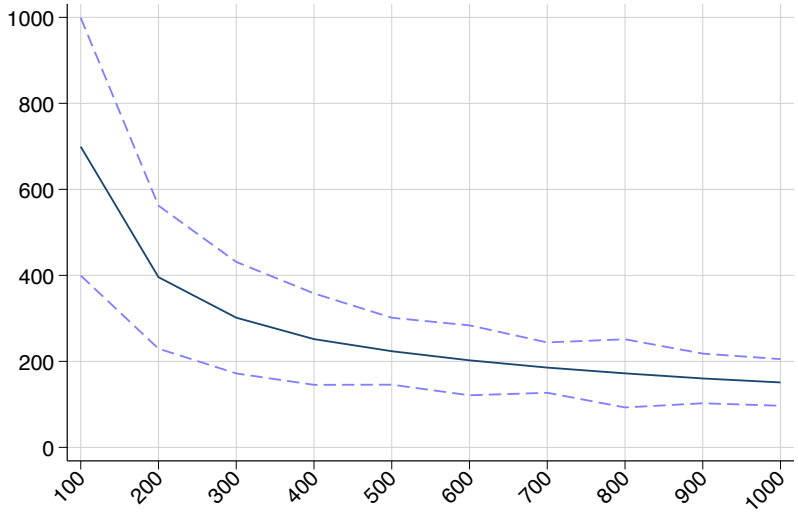
We use average (unconditional) under-reporting rates by decile of declared revenue to proxy a_i^k for all taxpayers i data

Do we under-estimate a_i^k?

- Not when compared to the available evidence from a LICs (Best et al., 2021)
- No if cross-checking is indeed a more efficient technique to discover true u_i^k

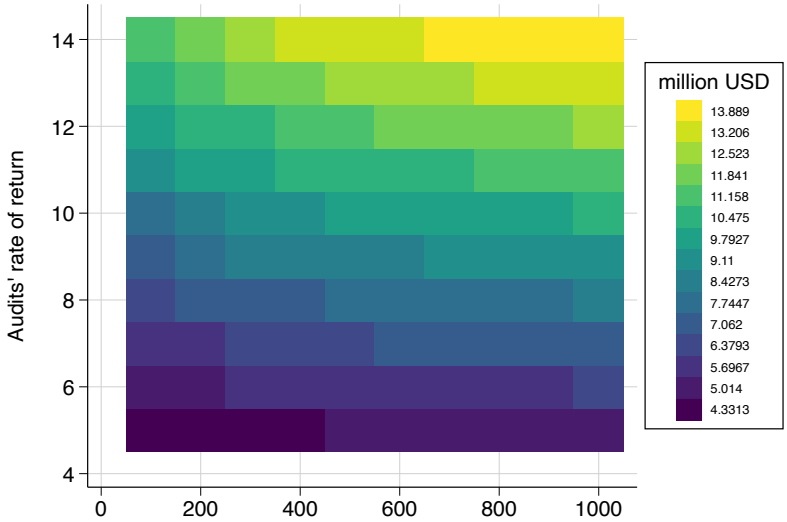
4 - Simulating Additional Audit Returns - Δ_n^3 (%)

Benchmark compared to auditing the largest firms by tax center.



Simulated Audit Returns - Δ_1^{nr} in USD

Very large immediate gains, even under conservative assumptions on audit returns.



Contributions to the Literature

(1) **Third-party reporting** has been crucial for the development of modern tax systems (Besley and Persson (2014), Jensen (2022)) yet imperfect implementation LICs (Almunia et al. (2022), Mascagni et al. (2023))

- Large scale personalized communication can improve third-party reporting

(2) Most revenue are *collected from* or *remitted by* **firms** (Milanez (2017), Slemrod and Velayudhan (2018), Best et al. (2021), Bachas et al. (2023))

- Leveraging trading networks of a just few large firms allows to access rich information.
- Conditional on being "caught" larger firms tend to be more compliant.

(3) On the **distribution of tax evasion** in LICs (Best et al., 2021)

- in the short run, TAs should strengthen enforcement among tax-registered firms rather than broaden the taxbase.

Summary

Question: Can TAs efficiently collect information to identify large non-compliant actors?

Method:

- Digitise **lists of suppliers** submitted by the largest Senegalese firms.
- Document very **prevalent misreporting**.
- **Randomise** low-cost intervention to increase the perception of sanctions' likelihood
- Estimate gains in **audit returns** generated by better targeting.

Findings:

- Large impact on the prevalence of identification information - stronger for the largest firms.
- New information reveals significant amount of unreported amounts.
- All impacts persist over time
- Unreported revenue are concentrated among a few tax-registered firms.
- New information would allow to better target audits.

Next Step and Future Research

Next steps:

- add 2024 (declared in 2025)
- explore heterogeneity using Wager & Athey (2018)
- (try to) disentangle the respective role of **spillovers** (through accountants' networks), **time**, and **online reporting** for observed trends in the control group

Future Research:

- Ongoing full-scale RCT targetting 1,500 randomly selected under-reporting formal firms
- How efficient (short run) and deterrent (long run) are audits based on this information?

Summary Statistics and Balance Tests - Full Sample - back

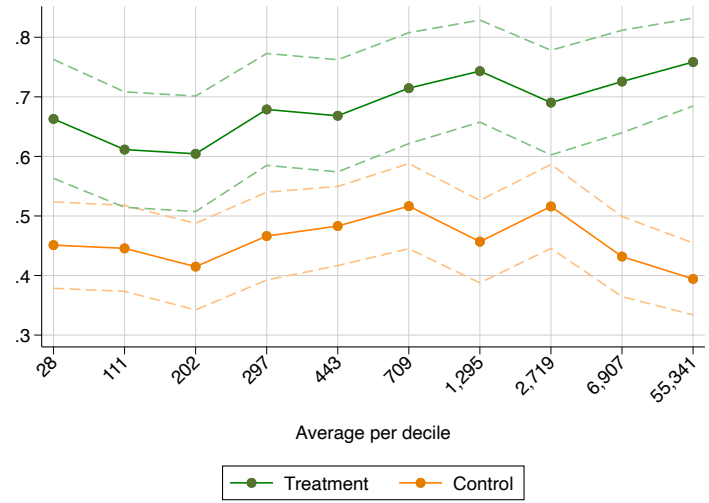
Source	Covariates/Outcomes	Summary Statistics			Balance tests	
		Mean	Std. dev.	Median	β	Std. err.
(A) Suppliers lists)	Number of suppliers	39.663	152.777	10	-6.773	4.635
	Share suppliers without any ID	0.867	0.215	1	-0.008	0.006
	... with a positive WTS	0.115	0.239	0	-0.002	0.007
	... in the Hospitality sector	0.001	0.013	0	-0.000	0.000
	... Senegalese	0.989	0.047	1	0.001	0.001
	... with transaction bellow taxable threshold	0.017	0.067	0	-0.000	0.002
	N. of y. for which we observe a suppliers list	1.705	0.775	2	-0.001	0.019
	Declares suppliers list online	0.326	0.468	0	0.004	0.011
(B) Other Declarations)	Av. Sum of WTS remitted - 2018-2020 (USD'000)	4.142	22.982	0.620	-0.894	0.784
	Av. turnover - 2018-2020 (USD'000)	5480.939	30678.406	407.394	-1213.954	981.129
	Av. number of employees - 2018-2020	56.564	319.884	6.699	-22.606**	11.135
(C) Tax Registry)	Large Taxpayers Centre (binary)	0.229	0.420	0	-1.810	1.277
	Upper-Middle Taxpayers Centre (binary)	0.126	0.332	0	-6.950	5.998
	Lower-Middle Taxpayers Centre (binary)	0.286	0.452	0	-1.209	3.065
	Regulated Sector Centre (binary)	0.358	0.479	0	-1.054	1.777
	Legal person (binary)	0.708	0.454	1	-0.006	0.011
	Number of years since first digital admin. record	11.345	3.275	13	-0.016	0.109
	Firm created before 1990 (binary)	0.299	0.457	0	0.005	0.015
	... btw 1990 and 2000 (binary)	0.076	0.266	0	0.000	0.009
	... btw 2000 and 2010 (binary)	0.274	0.446	0	-0.005	0.014
	... after 2010 (binary)	0.349	0.476	0	0.000	0.016

Take-up: Did Treated Firms Receive the Message?

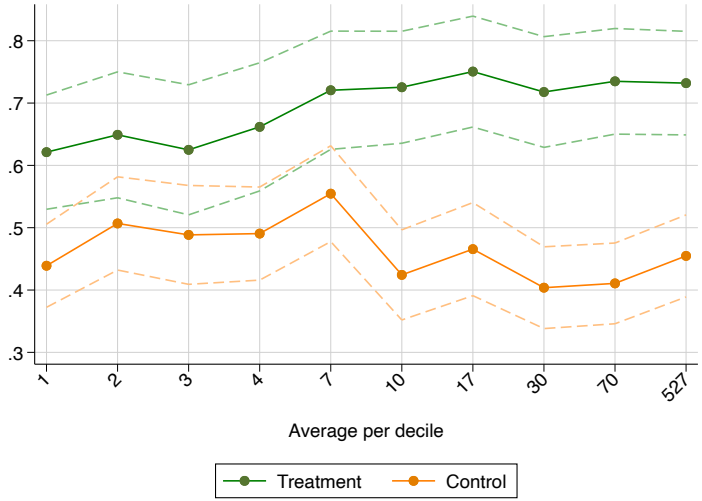
	N	%
Rectification submitted for 2018-2020 suppliers list	1061	60.8
No rectification submitted but message reception confirmed	557	31.9
... ask for more time	286	16.4
... cannot get the required information	124	7.1
... ask for assistance	62	3.5
... ceased activity	18	1.0
... other explanations	67	3.8
No rectification submitted, no confirmation	125	7.1
Total	1743	100

Treated sample only - using intervention data

1 - Additional Results: Heterogeneity by Revenue (USD '000)



1 - Additional Results: Heterogeneity by N. of Employees



1 - Impact on Suppliers Identification - Additional Results

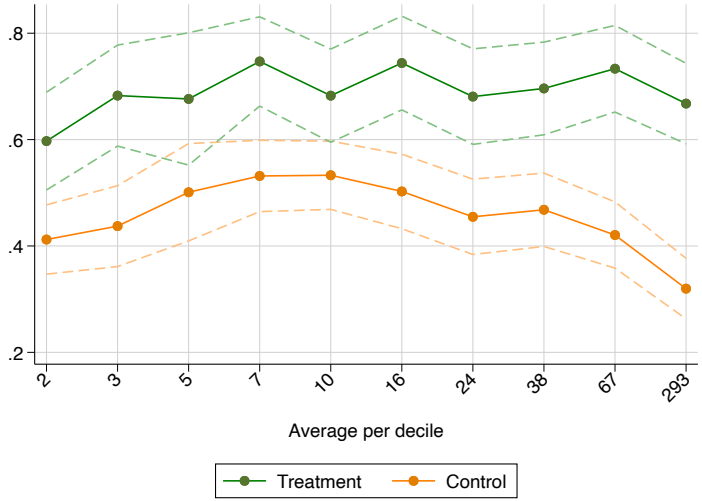
Additional results suggest that:

- stronger impact on the largest firms - link
- several firms cared to comply - sharing phone numbers, not mandatory - link
- suppliers could be reluctant to share their information with clients - link

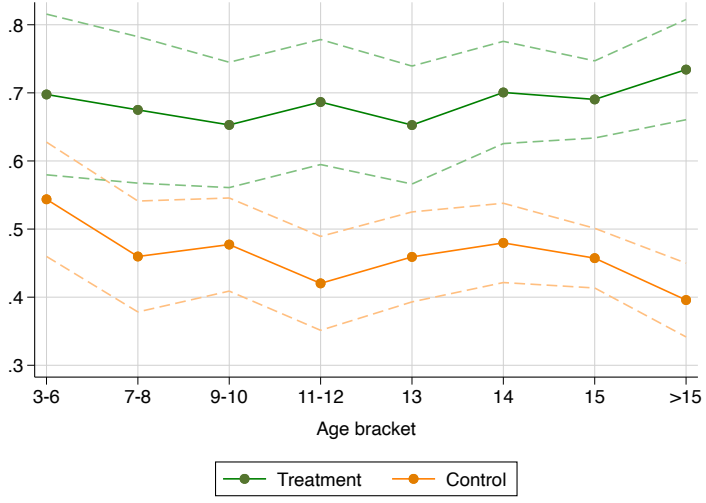
No detectable impact on suppliers' declarative behavior - link

- some of the benefits of third-party reporting are not automatic.

1 - Additional Results: Heterogeneity by N. of Service Suppliers



1 - Additional Results: Heterogeneity by Age - back



1 - Impact on Suppliers Identification - 2022 - back

		β	Std. err.	Lee Bounds lower	Lee Bounds upper	μ_c	N
	Probability to declare supplier list	0.017	0.013			0.772	3487
(A) Share of suppliers ...	... with a valid ID	0.247***	0.014	0.243	0.263	0.469	2633
	... with intentionally wrong ID	-0.003	0.005	-0.019	-0.003	0.026	2633
	... with a telephon number	0.092***	0.016	0.084	0.104	0.410	2633
(B) Structure of service supply	Number of suppliers	4.590	7.203	-19.075	4.853	49.903	2633
	Sum of payments (ihs)	0.036	0.056	-0.095	0.111	17.934	2633
	Average payment (ihs)	-0.021	0.043	-0.098	0.033	15.472	2633
	Median payment (ihs)	-0.079*	0.043	-0.133	-0.018	14.765	2633
(C) Probability to	... misreport all suppliers	-0.195***	0.014	-0.214	-0.194	0.289	2633
	... misreport no suppliers	0.101***	0.015	0.089	0.109	0.182	2633
	... misreport some suppliers	0.093***	0.018	0.085	0.105	0.527	2633

1 - Additional Results: on Suppliers Behavior

Econ. literature: income subject to third-party reporting is better declared.

Do suppliers adjust their declarative behavior after 2023?

Secondary model - at the supplier level j (restricted to formal suppliers identifiable ex-ante) :

$$y_j = \alpha + \beta IndirectTreat_j + \boldsymbol{\gamma X_j} + \delta_s + \epsilon_j \quad (2)$$

- y_j = the probability to declare; revenue declared (ihs)
- $IndirectTreat_j$ is a dummy equal to 1 if j is (mostly) a supplier of a treated client
- $\boldsymbol{\gamma X_j}$ and δ_s similar to main model

1 - Additional Results: on Suppliers Behavior - back

	2022		2023	
	Submits a declaration	Revenue declared	Submits a declaration	Revenue declared
	(binary)	(ihs)	(binary)	(ihs)
	(1)	(2)	(3)	(4)
Restricted Sample	0.006	0.056	0.004	0.086
	(0.013)	(0.256)	(0.011)	(0.195)
	[0.594]	[9.849]	[0.226]	[3.559]
	5645	5645	5645	5645
Full Sample	0.005	0.014	0.007	0.094
	(0.011)	(0.235)	(0.010)	(0.194)
	[0.631]	[10.822]	[0.272]	[4.478]
	6853	6853	6853	6853

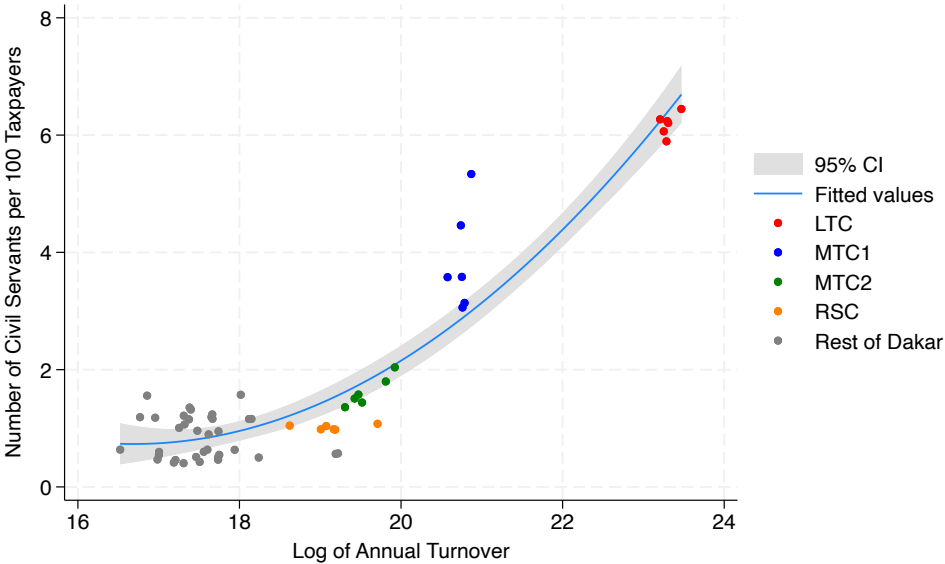
Under-reporting Suppliers or Over-claiming Clients? back

Over-claiming by clients is **less likely**

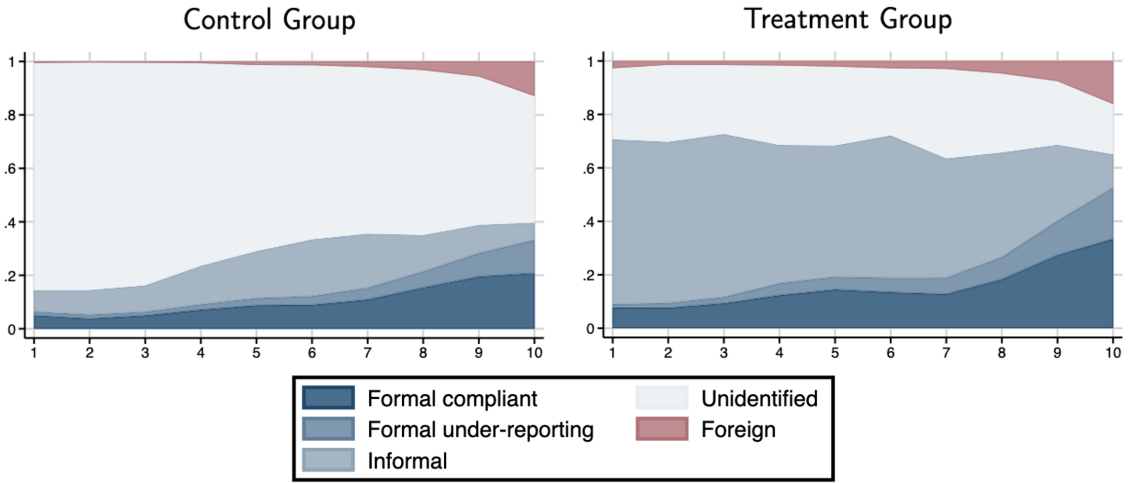
- **past deterrence** has concentrated on largest firms (i.e. the clients) while under-reporting formal suppliers most often are smaller firms.
- much easier to misreport some other, less scrutinized margins:
  - **under-report sales** to parties that do not reported to the administration
  - **over-report other deductibles** for which less details is required (purchase of informal goods, employees bellow taxable threshold, imports)
- according to **tax inspectors** themselves: amounts reported in supplier list are reliable

Importantly: over-claimed amounts would also qualify as unreported revenue, but on the client's side.

Distribution of Civil Servants



2 - Impact on the Detection of Tax Evading Suppliers - back

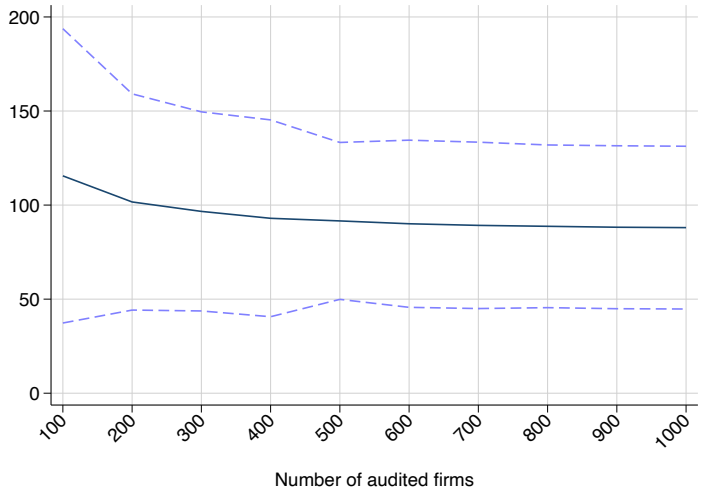


3 - Under-Reporting Rates - back

Decile	Average Declared Revenue (USD)	Number of Under-reporting Taxpayers	Average under-reporting rate as a % of declared revenue	
			conditional on under-reporting	*unconditional*
1	330	1,395	99.37	49.27
2	1,970	138	88.63	4.34
3	3,916	171	82.20	5
4	7,465	169	81.27	4.88
5	13,666	163	81.91	4.73
6	25,245	195	77.84	5.38
7	49,986	189	72.22	4.84
8	112,452	145	78.51	4.05
9	301,713	125	82.66	3.67
10	26,687,004	102	74.54	2.70
All	320,375	2,792	81.91	8.89

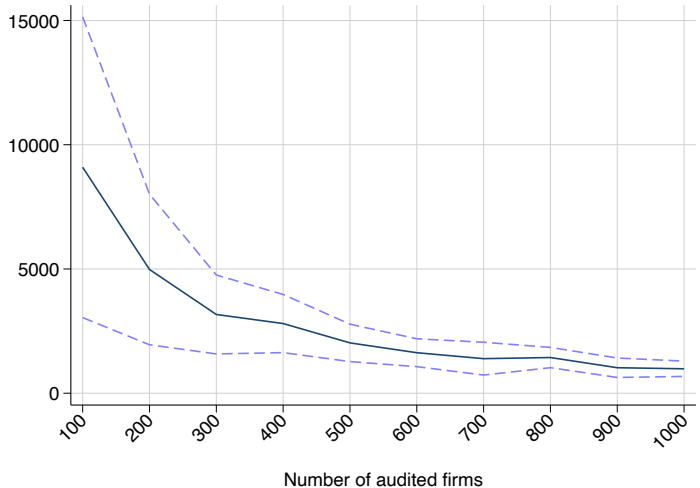
3 - Simulating Additional Audit Returns - Δ_n^1 (%)

... compared to auditing largest under-reporting formal suppliers of the control group.



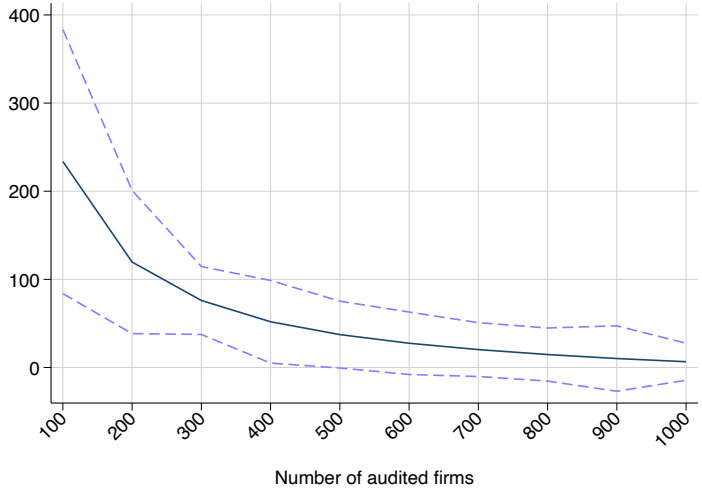
3 - Simulating Additional Audit Returns - Δ_n^2 (%)

... compared to auditing randomly selected taxpayers.

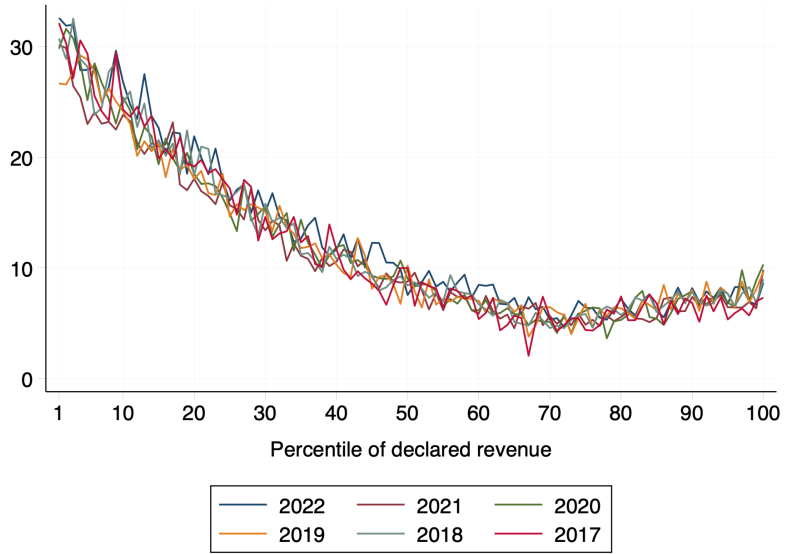


3 - Estimating Additional Audit Returns - Δ_n^3 (%)

... compared to auditing the largest firms.



3 - Simulating Additional Audit Returns - Average Tax Rate



Biblio

Almunia, M., Hjort, J., Knebelmann, J., and Tian, L. (2022). Strategic or confused firms? evidence from "missing" transactions in uganda. *Review of Economics and Statistics*, pages 1–35.

Bachas, P., Brockmeyer, A., Dom, R., and Semelet, C. (2023). Effective tax rates and firm size.

Besley, T. and Persson, T. (2014). Why do developing countries tax so little? *Journal of economic perspectives*, 28(4):99–120.

Best, M., Shah, J., and Waseem, M. (2021). Detection without deterrence: Long-run effects of tax audit on firm behavior.

Jensen, A. (2022). Employment structure and the rise of the modern tax system. *American Economic Review*, 112(1):213–34.

Mascagni, G., Dom, R., Santoro, F., and Mukama, D. (2023). The vat in practice: Equity, enforcement, and complexity. *International Tax and Public Finance*, 30(2):525–563.

Milanez, A. (2017). Legal tax liability, legal remittance responsibility and tax incidence: Three dimensions of business taxation.

Slemrod, J. and Velayudhan, T. (2018). Do firms remit at least 85% of tax everywhere? new evidence from india. *Journal of Tax Administration*, 4(1):24–37.