Using Digital Records and Machine Learning to Detect Tax Avoidance in Customs Declarations

A Complete Working Paper Coming Soon (Waiting for Third-Party Approval)

1 Empirical Strategy

1.1 Data

We analyze the universe of Haitian customs declarations for standard imports (procedure code IM4) from 2017 to 2024. The sample includes 2.16 million sea cargo shipments across 4,333 HS6 products, 92 months, 15 customs offices, 185 origin countries, and the top 5,000 importers by transaction volume. These data represent approximately 76% of official merchandise imports reported by the World Bank.

We construct unit values in USD per kilogram by converting declared customs values from Haitian gourdes using the median monthly exchange rate, then dividing by net weight. We trim the top and bottom 1% of the unit value distribution within each HS4-year cell to remove outliers. The tax rate is computed as the ratio of total taxes paid to declared customs value, capped at 150%.

1.2 Baseline Specification

Our baseline specification estimates the elasticity of declared import values with respect to tax rates:

$$\log(\mathrm{UV}_{ijkot}) = \beta\tau_{ijkot} + \alpha_j + \gamma_t + \delta_k + \theta_o + \mu_i + \varepsilon_{ijkot} \tag{1}$$

The unit value UV_{ijkot} corresponds to importer i importing HS6 product j in month t through office k from origin country o. We include fixed effects for products (α_j), months (γ_t), offices (δ_k), origin countries (θ_o), and importers (μ_i).

This specification identifies β from variation in tax rates within the same importer importing the same product from the same country through the same office in the same month. Tax rate variation arises from differences in product characteristics within HS6 codes, changes in tariff schedules over time, and variation in customs procedures applied to nominally similar goods. Standard errors are clustered by importer. Observations are weighted by shipment value (USD).

1.3 Office Heterogeneity

To test for office heterogeneity, we interact tax rates with office indicators:

$$\log(\mathrm{UV}_{ijkot}) = \sum_k \beta_k(\tau_{ijkot} \times \mathbb{1}_k) + \alpha_j + \gamma_t + \theta_o + \mu_i + \varepsilon_{ijkot} \tag{2}$$

The indicator $\mathbb{1}_k$ equals one for office k. Office fixed effects δ_k are absorbed by the interactions and therefore omitted. The office-specific elasticity β_k captures how responsive declared values are to taxes at office k. Identification comes from comparing the same product cleared by the same importer at different offices. We test the null hypothesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_K$ using a Wald test.

1.4 Addressing Alternative Explanations

Office heterogeneity could reflect two factors other than genuine enforcement differences. First, offices may specialize in different products with different tariff structures. Saint-Marc handles 72% vehicles (high-tax goods, median 83% rate). Port-au-Prince handles diverse general merchandise (median 29% rate).

We address this by restricting the sample to products traded at multiple offices (common support) and re-estimating the heterogeneity specification. This ensures we compare identical HS6 products across offices.

Second, different types of firms may select into different offices. Sophisticated evaders might preferentially use permissive offices while compliant firms use the main port. We address this in two ways.

First, we split the sample by importer type:

Single-office:

$$\log(\mathrm{UV}_{ijkot}) = \beta^S \tau_{ijkot} + \alpha_j + \gamma_t + \delta_k + \theta_o + \mu_i + \varepsilon_{ijkot} \tag{3}$$

Multi-office:

$$\log(\mathrm{UV}_{ijkot}) = \beta^M \tau_{ijkot} + \alpha_j + \gamma_t + \delta_k + \theta_o + \mu_i + \varepsilon_{ijkot} \tag{4}$$

Single-office importers use only one customs office throughout 2017 to 2024. Multi-office importers use two or more. If office heterogeneity is stronger for multi-office firms, this suggests they strategically exploit enforcement differences.

Second, we estimate office effects within multi-office firms:

$$\log(\mathrm{UV}_{ijkot}) = \sum_k \beta_k^W (\tau_{ijkot} \times \nVdash_k) + \alpha_{jt} + \theta_o + \mu_i + \varepsilon_{ijkot} \tag{5}$$

The term α_{jt} denotes product-by-month fixed effects, capturing time-varying product characteristics. The sample is restricted to multi-office importers. Importer fixed effects μ_i absorb all time-invariant firm characteristics: sophistication, product mix, tax attitudes, management quality. Any remaining office variation must be causal. The same firm behaves differently depending on which office clears the shipment.

2 Results

Table 1 presents our main findings. Column (1) shows the baseline elasticity. A 10 percentage point increase in the tax rate leads to a 10.1% decrease in declared unit values ($\beta = -1.009$, SE = 0.217). This unit elasticity indicates strategic under-invoicing to reduce tax liability. The specification identifies from within importer-product-month-origin-office variation, with over 4,000 product fixed effects, 92 month fixed effects, and 5,000 importer fixed effects. The large sample and tight standard errors provide precise estimates despite this demanding specification.

Column (2) reveals heterogeneity across customs offices. A Wald test strongly rejects equality of office-specific elasticities ($F = 154.4$, $p < 0.001$). The estimated elasticities range from -0.886 at Port-au-Prince (the main port handling 92.5% of import value) to -2.462 at Miragoane. Smaller secondary ports show elasticities 2.5 to 2.8 times larger than the main port. Why do offices differ so dramatically?

Is it product composition? Office heterogeneity could simply reflect different product mixes. Column (3) tests this by restricting to the 2,128 HS6 codes (59.6% of unique products, 98.7% of observations) that appear at multiple offices. This ensures we compare identical products across offices. The baseline elasticity is unchanged (-1.010, SE = 0.218). Office heterogeneity remains strongly significant ($F = 151.2$, $p < 0.001$).

Is it firm selection? Perhaps bad firms systematically select into bad offices. Table 2 tests this by classifying importers as single-office (use only one office throughout 2017 to 2024) or multi-office (use two or more offices). Single-office importers comprise 3,837 firms controlling 46% of import value. Multi-office importers comprise just 1,164 firms (23% of firms) but control 54% of import value.

Panel B compares baseline elasticities. Single-office importers show -0.713 (SE = 0.136). Multi-office importers show -1.115 (SE = 0.387), which is 56% more tax-responsive. This suggests selection matters. Firms with office access are more sophisticated evaders.

But the office heterogeneity pattern is more striking. For single-office firms, office interactions are barely significant ($F = 5.7$, $p = 0.0001$). These captive firms show minimal office-specific variation. They are stuck where they are. For multi-office firms, office interactions are highly significant ($F = 603.9$, $p < 0.001$), a 105-fold increase. Firms with office choice exhibit dramatically stronger office-specific responses. This suggests offices genuinely differ, and sophisticated importers exploit those differences.

Within-firm identification. The cleanest evidence comes from within-firm comparisons. Table 3 restricts to multi-office importers and includes importer fixed effects. Importer fixed effects absorb all firm characteristics: sophistication, product mix, tax attitudes, management quality. Any remaining office variation must be causal. The same firm behaves differently depending on which office clears the shipment.

The Wald test strongly rejects equality of office-specific elasticities ($F = 627.0$, $p < 0.001$). The same firm importing the same product declares systematically different values at different offices. This rules out firm selection. Offices genuinely differ in enforcement or corruption. Column (2) provides supporting evidence of strategic routing. Multi-office importers send higher-value shipments to secondary offices ($13,281 median) than to their main office ($4,184 median), though this difference is only marginally significant ($t = 1.89$, $p = 0.059$).

These results establish three findings. First, the aggregate import elasticity of -1.0 masks substantial spatial heterogeneity, with office-specific elasticities ranging from -0.9 to -2.5. Second, this heterogeneity reflects genuine enforcement differences, not product composition or firm types. Third, sophisticated multi-office importers (23% of firms controlling 54% of trade) strategically exploit this heterogeneity, showing 105-fold stronger office-specific responses than captive firms. Offices differ in enforcement, creating opportunities for spatial tax arbitrage within a single country's customs system.

Table 1: Import Elasticities and Office Heterogeneity

	Dependent variable: log(unit value)		
	Baseline	Office het.	Common support
	(1)	(2)	(3)
Panel A: Baseline elasticity			
Tax rate	-1.009^{***}	-0.886^{***}	-1.010^{***}
	(0.217)	(0.174)	(0.218)
Panel B: Office-specific elasticities			
Tax rate × Port-au-Prince (10DP)		-0.886^{***}	
		(0.174)	
Tax rate × Cap-Haïtien (20CH)		-1.726^{***}	
		(0.328)	
Tax rate × Saint-Marc (40SM)		-2.410^{***}	
		(0.150)	
Tax rate × Miragoane (65DM)		-2.462^{***}	
		(0.126)	
Tax rate × Gonaïves (30GO)		-1.425^{***}	
		(0.132)	
Wald test: office het. (F-stat)		154.4	151.2
p-value		< 0.001	< 0.001
HS6 × Month FE	Yes	Yes	Yes
Office FE	Yes	No	Yes
Origin country FE	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes
Observations	2,155,037	2,155,037	2,127,672
R^2	0.765	0.766	0.762
Within R^2	0.015	0.017	0.015

Notes: Dependent variable is log unit value in USD per kg. Tax rate is ratio of taxes paid to declared customs value. Standard errors clustered by importer in parentheses. Observations weighted by shipment value. Column (3) restricts to HS6 codes traded at multiple offices (59.6% of products, 98.7% of observations). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 2: Strategic Routing: Single vs. Multi-Office Importers

	Baseline elasticity		Office heterogeneity	
	Single	Multi	Single	Multi
	(1)	(2)	(3)	(4)
Panel A: Sample composition				
Number of firms	3,837	1,164	3,837	1,164
Share of import value	46.1%	53.9%	46.1%	53.9%
Panel B: Elasticity estimates				
Tax rate	-0.713^{***}	-1.115^{**}		
	(0.136)	(0.387)		
Wald test: office het. (F-stat)			5.7	603.9
p-value			< 0.001	< 0.001
Ratio: multi/single				105×
HS6 × Month FE	Yes	Yes	Yes	Yes
Office FE	Yes	Yes	No	No
Origin country FE	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes
Observations	1,122,684	1,032,353	1,122,684	1,032,353
R^2	0.738	0.786	0.740	0.789
Within R^2	0.008	0.017	0.011	0.021

Notes: Single-office importers use only one customs office throughout 2017 to 2024. Multi-office importers use two or more. Columns (3) and (4) include office-specific tax rate interactions (coefficients not shown). Standard errors clustered by importer in parentheses. Observations weighted by shipment value. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3: Within-Firm Identification of Office Effects

	Within-firm elasticities (1)	Value sorting (2)
Panel A: Within-firm office effects		
Wald test (F-stat)	627.0	
p-value	< 0.001	
Panel B: Value sorting pattern		
Median value: secondary offices		$13,281
Median value: main office		$4,184
Difference		$9,097
t-statistic		1.89
p-value		0.059
HS6 × Month FE	Yes	
Origin country FE	Yes	
Importer FE	Yes	
Sample	Multi-office firms only	Multi-office firms only
Observations	1,032,353	2,720
R^2	0.791	
Within R^2	0.022	

Notes: Column (1) restricts to multi-office importers and includes importer fixed effects, identifying office effects from within-firm variation. Column (2) tests whether multi-office importers send higher-value shipments to secondary vs. main offices. Unit of observation is importer-office pair. Standard errors clustered by importer. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.