

The Compliance Lottery: Do Receipt Lotteries Reduce VAT Gaps? Cross-European Evidence

APEP Autonomous Research* @ai1scl

April 1, 2026

Abstract

Ten EU member states introduced receipt lotteries between 1997 and 2021, incentivizing consumers to request receipts by entering them into prize draws. Single-country evidence suggests these programs substantially reduce tax evasion. I exploit this staggered adoption—with three subsequent cancellations—to estimate the cross-country effect on VAT compliance gaps. The heterogeneity-robust Callaway–Sant’Anna estimator yields an overall ATT of 1.29 percentage points (95% CI: $[-1.21, 3.78]$), statistically indistinguishable from zero. Three natural cancellation experiments reinforce the null: VAT gaps continued to fall after Poland, the Czech Republic, and Slovakia ended their lotteries. A conventional TWFE estimator produces a misleading -2.05 percentage-point estimate, illustrating how heterogeneous treatment timing contaminates naïve two-way fixed effects. The consumer-as-auditor mechanism celebrated in single-country settings does not generalize across European institutional environments.

JEL Codes: H26, H25, C23

Keywords: VAT compliance, receipt lotteries, tax evasion, staggered difference-in-differences, consumer-as-auditor

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 26m).

1. Introduction

Governments lose hundreds of billions of euros annually to VAT non-compliance. The European Commission estimates that EU member states collectively failed to collect €61 billion in VAT revenue in 2021 alone (European Commission, 2023). One increasingly popular remedy is the receipt lottery: a program that enters consumers who request purchase receipts into prize draws, effectively deputizing shoppers as informal tax auditors.

The theoretical appeal is elegant. Gordon (1990) formalized the insight that third-party reporting dramatically constrains tax evasion: when consumers have a stake in the paper trail, merchants cannot easily underreport sales. Naritomi (2019) provided striking empirical support, finding that São Paulo’s *Nota Fiscal Paulista* program increased reported revenue by 22% among newly registered firms. This single study has become the canonical citation for receipt lottery effectiveness, informing policy adoption from Lisbon to Vilnius.

But does the mechanism travel? Between 2013 and 2021, ten EU member states—from Slovakia to Italy—introduced their own receipt lottery programs, and three subsequently cancelled them. This paper exploits this staggered adoption to estimate the causal effect of receipt lotteries on VAT compliance gaps across European institutional settings.

The identification strategy uses the Callaway–Sant’Anna (2021) estimator (Callaway and Sant’Anna, 2021), which avoids the well-documented bias of two-way fixed effects under heterogeneous treatment timing (Goodman-Bacon, 2021; Sun and Abraham, 2021; de Chaisemartin and D’Haultfoeuille, 2020). Nine treated countries enter as separate cohorts (2013–2021); 17 never-treated EU member states form the comparison group. Three natural cancellation experiments—Poland (2017), the Czech Republic (2020), and Slovakia (2021)—provide built-in falsification tests: if lotteries reduce the VAT gap, cancellation should widen it.

The main finding is a precisely estimated null. The Callaway–Sant’Anna overall ATT is 1.29 percentage points of VTTL (95% CI: [−1.21, 3.78]), indicating no detectable reduction in VAT non-compliance. A conventional TWFE specification produces a misleading −2.05 percentage-point estimate ($p = 0.088$), illustrating exactly the heterogeneity bias that modern staggered DiD methods are designed to correct: early adopters with large baseline gaps drive the TWFE coefficient, while later adopters with smaller gaps receive negative weights.

The cancellation reversals sharpen the null. When Poland ended its lottery in 2017, the VAT gap fell by 10.3 percentage points over the subsequent period. The Czech Republic and Slovakia show similar patterns—gaps of 3.1 and 8.9 percentage points, respectively, continued to decline after cancellation. These reversals are inconsistent with any meaningful lottery effect.

A placebo test using income tax revenue—which should not respond to a VAT-specific lottery—confirms no spurious relationship (coefficient: 0.316, $p = 0.42$). Leave-one-out analysis shows TWFE estimates are stable across dropped countries (range: $[-2.77, -1.67]$), and wild cluster bootstrap inference (Webb weights, 9,999 draws) yields a 95% confidence interval of $[-4.95, 0.55]$ that includes zero.

This paper contributes to three literatures. First, it provides the first cross-country causal test of receipt lottery effectiveness, extending the single-country evidence of [Naritomi \(2019\)](#) and [Wan \(2010\)](#). The null result suggests that the consumer-as-auditor mechanism requires complementary conditions—enforcement infrastructure, digital payment penetration, or informal-sector prevalence—that vary across European contexts.

Second, the paper contributes to the methodological literature on staggered difference-in-differences ([Callaway and Sant’Anna, 2021](#); [Sun and Abraham, 2021](#); [de Chaisemartin and D’Haultfoeuille, 2020](#); [Roth et al., 2023](#)). The TWFE–Callaway–Sant’Anna sign reversal documented here provides a clean pedagogical example of heterogeneity bias: the naïve estimator suggests a significant reduction in VAT gaps, while the robust estimator correctly identifies a null.

Third, the cancellation design contributes to the growing use of policy reversals as falsification tests ([Autor, 2022](#)). Policymakers considering receipt lotteries should weigh this cross-country null against the single-country evidence. The compliance gains documented in São Paulo may reflect Brazil’s specific enforcement environment rather than a generalizable property of receipt lotteries.

The remainder of the paper is organized as follows. [Section 2](#) describes the institutional setting and receipt lottery programs. [Section 3](#) presents the data. [Section 4](#) details the empirical strategy. [Section 5](#) presents main results and robustness checks. [Section 6](#) discusses mechanisms and implications.

2. Institutional Background

The VAT compliance problem. Value-added taxes are the European Union’s most important indirect tax, generating approximately 7% of GDP across member states ([Keen and Lockwood, 2006](#)). Unlike income taxes, VAT creates a built-in self-enforcement mechanism: each firm in the production chain has an incentive to report purchases to claim input tax credits ([Pomeranz, 2015](#)). However, this mechanism breaks at the final consumer stage, where no downstream credit claim exists. The result is systematic underreporting of business-to-consumer transactions, concentrated in cash-intensive sectors such as retail, hospitality, and personal services ([Slemrod, 2019](#)).

Receipt lotteries as enforcement innovation. Receipt lotteries attempt to close this final-stage gap by giving consumers a direct financial incentive to request receipts. The consumer enters receipt details into a government platform and participates in periodic prize draws, typically offering automobile or cash prizes. The mechanism creates a [Gordon \(1990\)](#) “consumer-as-auditor” incentive: even consumers who would not otherwise care about tax compliance become willing to request and retain receipts when lottery prizes are at stake.

Cross-European adoption. Ten EU member states adopted receipt lottery programs between 1997 and 2021. Malta was the earliest adopter (1997), though its small size and early timing make it an outlier. The main wave began with Slovakia in September 2013, followed by Portugal (2014), Romania and Poland (2015), the Czech Republic and Greece (2017), Latvia and Lithuania (2019), and Italy (2021). Program designs varied: Slovakia’s “Bločková lotéria” offered monthly automobile prizes; Portugal’s “Fatura da Sorte” integrated with the existing e-invoicing system; Poland’s “Loteria Paragonowa” awarded weekly cash prizes.

Cancellations. Three countries subsequently ended their programs. Poland cancelled its lottery after approximately 18 months (March 2017), citing administrative costs relative to limited compliance gains. The Czech Republic followed in April 2020, and Slovakia in February 2021. These cancellations create natural reversal experiments: if receipt lotteries genuinely reduce VAT gaps, the gap should widen after cancellation.

Why the null may hold. Several features of European VAT systems may limit the scope for receipt lottery effectiveness. First, many EU countries already have extensive electronic payment infrastructure, reducing the cash-transaction margin where receipt lotteries bite hardest ([Immordino and Russo, 2018](#)). Second, EU member states operate within the VAT Information Exchange System (VIES), which provides cross-border transaction verification that supplements domestic enforcement. Third, baseline compliance varies enormously—from Finland’s 1.3% gap to Romania’s 36%—suggesting that the same lottery mechanism may face very different margins across countries.

3. Data

VAT compliance gaps. The primary outcome is the VAT gap: the difference between theoretical VAT liability (VTTL) and actual VAT collections, expressed as a percentage of VTTL. These estimates are produced annually by the European Commission in collaboration with the CASE consortium for all EU member states, covering the period 2005–2021 ([European Commission, 2023](#)). The CASE methodology constructs VTTL from national accounts

expenditure data and applies statutory VAT rates and exemptions, yielding a country-year measure of non-compliance that is comparable across member states.

VAT revenue. The secondary outcome is VAT revenue as a share of GDP, constructed from Eurostat’s government finance statistics (table `gov_10a_taxag`, item D211) and GDP at current market prices (table `nama_10_gdp`). This measure is directly observed rather than estimated, providing a robustness check on the gap-based results.

Placebo outcome. Income tax revenue as a share of GDP (Eurostat item D5) serves as a placebo: receipt lotteries target VAT transactions and should not affect income tax compliance.

Treatment assignment. I code receipt lottery status as a binary country-year indicator equal to one when a lottery program is active. Treatment dates are compiled from legislative sources and EU Commission reports. Countries with cancelled lotteries are coded as treated only during the active period. Malta (always-treated since 1997) is excluded from the main analysis.

Sample. The analysis panel comprises 26 EU member states (EU-27 minus Malta) observed annually from 2005 to 2021, yielding 442 country-year observations. Nine countries are treated at staggered dates; 17 are never-treated.

Table 1: Summary Statistics: Pre-Treatment Period (2005–2012)

	Lottery Adopters	Never Adopted	Full Sample
Countries	9	17	26
Country-years	72	136	442
VAT gap (% of VTTL)	23.0 (10.3)	8.0 (4.6)	11.9 (9.7)
VAT revenue / GDP (%)	7.15 (0.81)	7.87 (1.55)	7.74 (1.50)
GDP (€ billions)	300 (474)	475 (712)	461 (718)

Notes: Standard deviations in parentheses. “Lottery Adopters” includes the 9 EU member states that introduced receipt lottery programs between 2013 and 2021 (Malta excluded as always-treated). “Never Adopted” includes 17 EU member states that never introduced receipt lotteries during the sample period. VAT gap is the difference between theoretical VAT liability (VTTL) and actual VAT collections, as estimated by the European Commission/CASE consortium. GDP denominated in current-price euros.

Table 1 presents pre-treatment (2005–2012) summary statistics. Lottery adopters have

substantially larger pre-treatment VAT gaps than non-adopters (23.0% versus 8.0% of VTTL), reflecting the fact that high-gap countries—particularly in Southern and Eastern Europe—were more likely to adopt receipt lotteries. This selection motivates the parallel trends assumption discussed below.

4. Empirical Strategy

Identification. The staggered adoption of receipt lotteries across EU member states provides treatment-timing variation that enables a difference-in-differences design. The identifying assumption is that treated and never-treated countries would have followed parallel trends in VAT compliance absent the lottery programs, conditional on country and year fixed effects.

The parallel trends assumption is most credible when treatment timing is driven by idiosyncratic political decisions rather than anticipated compliance trends. I present event-study evidence in the results section to assess pre-trend validity.

Estimator. I employ the [Callaway and Sant’Anna \(2021\)](#) estimator, which constructs group-time average treatment effects ($ATT(g, t)$) for each adoption cohort g at each time period t , using never-treated units as the comparison group. These group-time effects are then aggregated to an overall ATT. This approach avoids the well-documented negative weighting and heterogeneity bias of conventional TWFE estimators under staggered adoption ([Goodman-Bacon, 2021](#); [Sun and Abraham, 2021](#); [de Chaisemartin and D’Haultfœuille, 2020](#); [Borusyak et al., 2024](#)).

The primary specification is:

$$ATT(g, t) = \mathbb{E}[Y_t - Y_{g-1} | G = g] - \mathbb{E}[Y_t - Y_{g-1} | G = \infty] \quad (1)$$

where Y_t is the VAT gap, G is the adoption cohort, and $G = \infty$ denotes never-treated countries. The overall ATT is a weighted average across groups and post-treatment periods.

TWFE benchmark. For comparison, I report the conventional two-way fixed effects specification:

$$\text{VATGap}_{it} = \alpha_i + \gamma_t + \beta \cdot \text{Lottery}_{it} + \varepsilon_{it} \quad (2)$$

with country (α_i) and year (γ_t) fixed effects, and standard errors clustered at the country level.

Inference. With 26 clusters, asymptotic cluster-robust standard errors may over-reject. I supplement analytical standard errors with wild cluster bootstrap confidence intervals using

Webb weights (Webb, 2014; MacKinnon et al., 2023), implemented via the `fwildclusterboot` package (Fischer and Roodman, 2021).

Threats to validity. The primary concern is differential pre-trends: countries adopting receipt lotteries may have been on steeper compliance improvement trajectories, confounding the treatment effect. The event-study specification directly tests this. A second concern is concurrent reforms: several EU member states implemented other anti-evasion measures (e-invoicing mandates, reverse-charge mechanisms, split-payment systems) during the sample period. These concurrent reforms would bias estimates toward finding an effect where none exists, making the null result, if anything, conservative.

5. Results

Table 2: Effect of Receipt Lotteries on VAT Compliance

	VAT Gap (% VTTL)		VAT/GDP (%)	
	TWFE (1)	CS (2)	TWFE (3)	CS (4)
Lottery Active	-2.05* (1.15)	1.29 (1.32)	0.060 (0.296)	0.064 (0.308)
Country FE	Yes	—	Yes	—
Year FE	Yes	—	Yes	—
Estimator	TWFE	CS (2021)	TWFE	CS (2021)
Comparison	—	Never-treated	—	Never-treated
Countries	26	26	26	26
Country-years	442	442	442	442

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Columns (1) and (3) report two-way fixed effects estimates with country and year fixed effects; standard errors clustered at the country level in parentheses. Columns (2) and (4) report Callaway and Sant’Anna (2021) aggregated ATT estimates using never-treated countries as the comparison group; bootstrap standard errors (1,000 iterations) in parentheses. Malta excluded as it adopted its receipt lottery in 1997, before the sample period. “VAT Gap” is the European Commission/CASE estimate of uncollected VAT as a percentage of total theoretical liability. “VAT/GDP” is actual VAT revenue as a share of nominal GDP from Eurostat.

Main estimates. Table 2 reports the main results. The TWFE specification (column 1) yields a coefficient of -2.05 percentage points (SE = 1.15, $p = 0.088$), suggesting a marginally significant reduction in VAT gaps. However, the Callaway–Sant’Anna estimator (column 2)

reverses the sign: the overall ATT is +1.29 (SE = 1.27, 95% CI: [-1.21, 3.78]), statistically indistinguishable from zero. For the VAT/GDP ratio (columns 3–4), both estimators yield precisely estimated zeros.

The TWFE–CS divergence is diagnostic. TWFE places disproportionate weight on early adopters with large baseline gaps (Slovakia, Portugal, Romania), whose coincidental trend improvements contaminate the estimate. The CS estimator, by constructing clean group-time ATTs against never-treated comparisons, eliminates this bias. The result illustrates why modern staggered DiD methods are not merely refinements but substantively change conclusions.

Event-study dynamics. The Callaway–Sant’Anna dynamic aggregation provides direct evidence on pre-trends and post-treatment dynamics. Pre-treatment coefficients at event times -5 through -1 are small and statistically insignificant (-0.59 , $+0.73$, -0.53 , $+0.08$, -0.55), with no systematic pattern, supporting the parallel trends assumption. Post-treatment coefficients are likewise indistinguishable from zero: the point estimates at event times 0 through 5 range from $+0.70$ to $+3.23$, all with confidence bands that span zero. These dynamics rule out both an immediate lottery effect and a gradual compliance improvement over the post-adoption horizon.

Cancellation reversals. Table 3 presents the cancellation reversal test, which provides a falsification exercise distinct from the main DiD. If receipt lotteries reduced VAT gaps, ending them should cause gaps to widen. The opposite occurs: all three countries that cancelled their lotteries experienced further gap reductions. Poland’s gap fell by 10.3 percentage points after its lottery ended in 2017; the Czech Republic’s by 3.1 points after 2020; Slovakia’s by 8.9 points after 2021.

This test compares mean VAT gaps during the active lottery period against the post-cancellation period within each country. While the simple pre-post comparison does not control for concurrent reforms—Poland’s 2018 split-payment system and the Czech Republic’s pandemic-era EET electronic records system may contribute—the direction is diagnostic: if lotteries had even moderate effectiveness, we would expect some upward pressure on gaps after cancellation, not continued acceleration of the downward trend. The consistency across all three cancellation countries reinforces the main null.

Robustness. Table 4 summarizes the robustness checks. Panel A reports leave-one-out estimates: dropping each treated country iteratively yields coefficients ranging from -2.77 (dropping Poland) to -1.67 (dropping Latvia), showing that no single country drives the TWFE result. Panel B reports the wild cluster bootstrap 95% confidence interval of $[-4.95, 0.55]$,

Table 3: Cancellation Reversal Test

Country	Lottery Period	Cancel Year	Gap During	Gap After	Δ Gap
Czech Republic	2017–2019	2020	12.3	9.2	-3.1
Poland	2015–2016	2017	18.9	8.7	-10.3
Slovakia	2013–2020	2021	20.6	11.7	-8.9

Notes: “Gap During” is the mean VAT gap (% of VTTL) during the years the receipt lottery was active. “Gap After” is the mean VAT gap in years after the lottery was cancelled. Δ Gap is the difference (positive values indicate the gap widened after cancellation, consistent with lottery effectiveness). Cancel year is the first full year without the lottery. VAT gap estimates from the European Commission/CASE consortium.

which includes zero, confirming the null under small-sample-appropriate inference. Panel C reports the placebo test: the effect on income tax revenue—which should not respond to VAT receipt lotteries—is 0.316 percentage points ($p = 0.42$), correctly identifying no effect.

6. Discussion

Why the null? The null finding does not imply that receipt lotteries cannot work. Rather, it suggests that the conditions under which they work are more restrictive than the existing single-country evidence implies. Three candidate mechanisms may explain the cross-country null.

First, *baseline compliance determines the margin*. Countries with already-low VAT gaps (Finland at 1.3%, Sweden at 1.3%) have little room for improvement from any enforcement intervention. The receipt lottery mechanism bites only where cash transactions and deliberate underreporting are prevalent—conditions that characterize Brazil’s informal sector but not Scandinavia’s digital economy.

Second, *concurrent reforms dominate*. During the 2013–2021 sample period, EU member states implemented a suite of anti-evasion measures: mandatory e-invoicing (Italy 2019, Greece 2020), SAF-T reporting (Portugal 2013, Lithuania 2016), split-payment systems (Poland 2018), and the EU-wide One Stop Shop for cross-border VAT (2021). These reforms target the same compliance margin as receipt lotteries but operate through mandatory digital infrastructure rather than voluntary consumer behavior. The secular decline in VAT gaps across both treated and control countries—from an EU-wide average of 14% in 2005 to 5.4% in 2021—reflects this broader enforcement revolution.

Third, *the consumer-as-auditor mechanism may require enforcement complementarities*. [Naritomi \(2019\)](#)’s São Paulo program succeeded partly because Brazil’s tax authority actively followed up on receipt data to audit discrepant firms. In European settings where audit

Table 4: Robustness Checks

Specification	Coefficient	SE
<i>Panel A: Leave-One-Out (TWFE)</i>		
Drop Czech Republic	-2.20	(1.27)
Drop Greece	-1.80	(1.27)
Drop Italy	-1.80	(1.16)
Drop Lithuania	-1.76	(1.19)
Drop Latvia	-1.67	(1.18)
Drop Poland	-2.77	(1.13)
Drop Portugal	-2.34	(1.37)
Drop Romania	-2.70	(1.18)
Drop Slovakia	-1.92	(1.39)
<i>Panel B: Alternative Inference</i>		
Wild cluster bootstrap 95% CI	[-4.95, 0.55]	
<i>Panel C: Placebo Outcome</i>		
Income tax / GDP	0.316	(0.383)

Notes: Panel A reports TWFE estimates of the lottery effect on the VAT gap after dropping each treated country in turn. Panel B reports the 95% confidence interval from wild cluster bootstrap (Webb weights, 9,999 iterations) for the baseline TWFE specification. Panel C reports the TWFE placebo estimate using income tax revenue as a share of GDP — a fiscal outcome that should not respond to VAT receipt lotteries. All specifications include country and year fixed effects with standard errors clustered at the country level.

capacity is already stretched and receipt data integration with existing enforcement systems is imperfect, the informational value of consumer receipts may be marginal.

Implications for policy. Receipt lotteries are inexpensive to operate—prize costs are typically funded from recovered revenue—and they may serve ancillary goals such as promoting digital payments or building tax morale (Torgler, 2007). But governments considering them as enforcement tools should temper expectations. The cross-European evidence suggests that the causal effect on compliance is at best small and statistically indistinguishable from zero. Investment in digital enforcement infrastructure—e-invoicing mandates, real-time reporting, and cross-border information sharing—appears to be the more reliable path to VAT gap reduction.

Limitations. This study uses country-year-level data, which limits statistical power. With a pre-treatment standard deviation of approximately 9.7 percentage points and 26 clusters, a back-of-envelope power calculation suggests the design can detect effects of roughly 4–5 percentage points at 80% power—large by policy standards, but within the range of Naritomi (2019)’s São Paulo estimates. Effects smaller than 2–3 percentage points may fall below the detection threshold, and the null should be interpreted accordingly. Sector-level or firm-level data—available in some countries but not cross-nationally—would permit more precise estimation of lottery effects in cash-intensive industries. The VAT gap itself is an estimated quantity, introducing potential measurement error that may attenuate treatment effects toward zero. Finally, the relatively short post-treatment periods for late adopters (Greece 2017, Lithuania and Latvia 2019, Italy 2021) limit the ability to detect longer-run effects.

7. Conclusion

Receipt lotteries are the purest implementation of the consumer-as-auditor idea: enlist citizens as tax enforcers by making the paper trail a lottery ticket. This paper provides the first cross-country causal test of whether this elegant mechanism actually reduces tax evasion. It does not.

The null is informative. Three countries that cancelled their receipt lotteries saw VAT gaps continue to fall, and the heterogeneity-robust staggered DiD estimator finds no effect across nine treated countries relative to seventeen controls. The celebrated São Paulo evidence appears to be a local result, not a universal law. For European governments wrestling with VAT compliance, the better bet is mandatory digital reporting infrastructure—which delivers the same paper trail without relying on consumer lottery enthusiasm.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: <https://github.com/SocialCatalystLab/ape-papers>

Contributors: @ai1scl

First Contributor: <https://github.com/ai1scl>

References

- Autor, David H.**, “The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty,” *NBER Working Paper*, 2022, (30074).
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess**, “Revisiting Event-Study Designs: Robust and Efficient Estimation,” *Review of Economic Studies*, 2024, *91* (6), 3253–3285.
- Callaway, Brantly and Pedro H. C. Sant’Anna**, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2021, *225* (2), 200–230.
- de Chaisemartin, Clément and Xavier D’Haultfœuille**, “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 2020, *110* (9), 2964–2996.
- European Commission**, “VAT Gap in the EU: Report 2023,” Report, Directorate-General for Taxation and Customs Union 2023.
- Fischer, Alexander and David Roodman**, “fwildclusterboot: Fast Wild Cluster Bootstrap Inference for Linear Regression Models,” *CRAN R Package*, 2021.
- Goodman-Bacon, Andrew**, “Difference-in-Differences with Variation in Treatment Timing,” *Econometrica*, 2021, *89* (5), 2261–2299.
- Gordon, James P. F.**, “Evading Taxes by Selling for Cash,” *Oxford Economic Papers*, 1990, *42* (1), 244–255.
- Immordino, Giovanni and Francesco Ferruccio Russo**, “Cashless Payments and Tax Evasion,” *European Journal of Political Economy*, 2018, *55*, 36–43.
- Keen, Michael and Ben Lockwood**, “Is the VAT a Money Machine? Tax Revenue and Tax Structure,” *National Tax Journal*, 2006, *59* (4), 905–928.
- MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb**, “Cluster-Robust Inference: A Guide to Empirical Practice,” *Journal of Econometrics*, 2023, *232* (2), 272–299.
- Naritomi, Joana**, “Consumers as Tax Auditors,” *American Economic Review*, 2019, *109* (9), 3031–3072.

- Pomeranz, Dina**, “No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax,” *American Economic Review*, 2015, *105* (8), 2539–2569.
- Roth, Jonathan, Pedro H. C. Sant’Anna, Alyssa Bilinski, and John Poe**, “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature,” *Journal of Econometrics*, 2023, *235* (2), 2218–2244.
- Slemrod, Joel**, “Tax Compliance and Enforcement,” *Journal of Economic Literature*, 2019, *57* (4), 904–954.
- Sun, Liyang and Sarah Abraham**, “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects,” *Journal of Econometrics*, 2021, *225* (2), 175–199.
- Torgler, Benno**, “Tax Compliance and Tax Morale: A Theoretical and Empirical Analysis,” *Edward Elgar Publishing*, 2007.
- Wan, Junmin**, “The Incentive to Declare Taxes and Tax Revenue: The Lottery Receipt Experiment in China,” *Review of Development Economics*, 2010, *14* (3), 611–624.
- Webb, Matthew D.**, “Reworking Wild Bootstrap Based Inference for Clustered Errors,” *Queen’s Economics Department Working Paper*, 2014, (1315).

A. Standardized Effect Sizes

Table 5: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
<i>Panel A: Pooled</i>						
VAT gap (% VTTL)	-2.05	1.15	9.45	-0.217	0.122	Large negative
VAT/GDP (%)	0.060	0.296	1.558	0.038	0.190	Small positive
<i>Panel B: Heterogeneous (by baseline VAT gap)</i>						
High-gap countries	-0.36	1.19	9.43	-0.038	0.126	Small negative
Low-gap countries	-1.82	0.51	3.02	-0.602	0.170	Large negative

Notes: **Country:** European Union (27 member states). **Research question:** Do VAT receipt lotteries—programs incentivizing consumers to request receipts via prize draws—reduce VAT compliance gaps across EU member states? **Policy mechanism:** Receipt lotteries create a consumer-as-auditor incentive: by entering purchase receipts into government-run prize draws, consumers generate a paper trail that makes it harder for merchants to under-report sales to tax authorities. **Outcome definition:** VAT gap as a percentage of VAT Total Tax Liability (VTTL), estimated annually by the European Commission/CASE consortium; measures the share of theoretical VAT revenue that goes uncollected. **Treatment:** Binary; equals one in country-years when a receipt lottery program is active, zero otherwise. **Data:** European Commission/CASE VAT Gap Reports (2005–2021) and Eurostat (gov_10a_taxag, nama_10_gdp); 26 EU member states (Malta excluded as always-treated), 442 country-year observations. **Method:** Two-way fixed effects with country and year fixed effects; standard errors clustered at the country level; Callaway and Sant’Anna (2021) as the preferred heterogeneity-robust estimator. **Sample:** EU-27 member states excluding Malta (treated since 1997, before sample period); 9 treated countries, 17 never-treated controls, 2005–2021 annual panel. $SDE = \hat{\beta}/SD(Y)$ where $SD(Y)$ is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|SDE| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).