The Deposit Illusion: Twenty Years of Beverage Container Schemes and the Limits of Consumer Price Incentives for Recycling

APEP Autonomous Research* @ai1scl

March 27, 2026

Abstract

The European Union will require all member states to adopt deposit return schemes (DRS) for beverage containers by 2029, at a projected cost exceeding EUR 3 billion. This paper evaluates whether twenty years of staggered DRS adoption across 10 EU/EEA countries actually improved recycling. Using Eurostat packaging waste data (2000–2023) and a material-level triple-difference design, I find no detectable aggregate effect on recycling rates (ATT = -0.13 pp, SE = 1.63) and a positive but statistically insignificant differential effect on targeted materials (+4.03 pp, SE = 5.22). The 95% confidence interval for the aggregate effect rules out improvements larger than 3.1 percentage points. The null persists across robustness checks including leave-one-out, alternative control materials, and dose-response specifications. These results suggest that the binding constraint on recycling is collection infrastructure and processing capacity—not consumer willingness to return containers.

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 31m).

JEL Codes: Q53, Q58, H23

Keywords: deposit return schemes, recycling, packaging waste, difference-in-differences, environmental policy

1. Introduction

By 2029, every European Union member state must operate a deposit return scheme for single-use beverage containers. The Packaging and Packaging Waste Regulation (PPWR; European Parliament and Council of the European Union, 2024) commits the continent to an infrastructure buildout projected to cost upward of EUR 3 billion, predicated on a simple theory: if you charge consumers a deposit and refund it upon return, they will recycle more. The logic is intuitive, the public support is broad, and the policy momentum is irreversible. There is just one problem: twenty years of staggered adoption across European countries provides no statistically reliable evidence that deposit return schemes improve aggregate recycling outcomes.

This paper exploits the staggered rollout of DRS across 10 EU and EEA countries between 2002 and 2023 to estimate the causal effect of deposit-refund incentives on packaging waste recycling rates. The empirical strategy proceeds in two steps. First, I apply the Callaway and Sant'Anna (2021) staggered difference-in-differences estimator to aggregate packaging recycling rates, using 18 never-treated or not-yet-treated countries as controls. Second, I exploit the material specificity of DRS—which targets plastic bottles and metal cans but not paper, cardboard, or wood packaging—in a triple-difference framework that absorbs country-by-year shocks through high-dimensional fixed effects.

The results are striking in their precision and their nullity. The Callaway–Sant'Anna aggregate ATT is -0.13 percentage points (SE $= 1.63$, $p = 0.93$), an estimate centered on zero with a 95% confidence interval that rules out effects larger than 3.1 percentage points in either direction. The triple-difference estimate for targeted materials relative to non-targeted materials is positive at 4.03 percentage points but far from conventional significance (SE $=$ 5.22, $p = 0.45$). Material-specific regressions reveal a suggestive positive effect on plastic recycling (+5.08 pp, $p = 0.32$) and a near-zero effect on metals (-1.59 pp, $p = 0.71$), while placebo materials—paper and wood, which DRS does not cover—show no effect, as expected.

These patterns survive dropping Germany (the largest and earliest adopter), substituting glass as the control material, and replacing the binary treatment with a continuous deposit amount.

This paper contributes to three literatures. First, it speaks to the economics of deposit-refund systems, where theoretical work by Palmer and Walls (1997) and Fullerton and Wolverton (2005) established conditions under which deposits are welfare-improving, while Walls (2011) surveys the practical experience. Most empirical evidence comes from U.S. bottle bills (Beatty et al., 2007; Ashenmiller, 2009) or single-country studies (Böhm et al., 2022; Linderhof et al., 2001), making cross-country causal evidence scarce. I provide the first pan-European estimate exploiting policy variation across countries and materials simultaneously.

Second, the paper contributes to the broader literature on price incentives for environmental behavior. Kinnaman (2006) questions whether the social benefits of residential recycling justify its costs; Jenkins et al. (2003) documents that curbside programs matter more than price signals for recycling participation; and Kirakozian (2016) surveys the interaction between behavioral and incentive-based instruments. My findings align with the view that consumer price incentives are neither necessary nor sufficient when the binding constraint lies downstream in collection and processing infrastructure (OECD, 2016, 2022).

Third, the paper provides a timely input to the EU policy debate. The PPWR mandate (European Parliament and Council of the European Union, 2024) was adopted with limited causal evidence on DRS effectiveness at the system level. Industry-funded reports from organizations like Reloop (Reloop Platform, 2023) cite collection rates of 90% or higher in DRS countries, but these figures measure the share of deposited containers returned—not whether total material recycling improved. The distinction matters: if DRS merely redirects containers from existing municipal collection streams into parallel return infrastructure, the aggregate effect on recycling can be zero even as the DRS-specific collection rate appears impressive. I term this the "deposit illusion"—high visible return rates that mask no net improvement in system-level recycling.

Methodologically, I build on the recent staggered DiD literature (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2021; Roth et al., 2023; Borusyak et al., 2024; Athey and Imbens, 2022) by combining the Callaway and Sant'Anna (2021) estimator for aggregate effects with a material-level triple-difference that exploits within-country variation in policy exposure. The DDD design absorbs common concerns about staggered adoption—including heterogeneous treatment effects and compositional changes in the control group—by comparing targeted materials (plastic, metal) to non-targeted materials (paper, wood) within the same country and year.

The remainder of the paper proceeds as follows. Section 2 describes DRS mechanics and the European adoption wave. Section 3 presents the data. Section 4 lays out the empirical strategy. Section 5 reports the results. Section 6 interprets the findings, and Section 7 concludes.

2. Institutional Background

How deposit return schemes work. A deposit return scheme imposes a small surcharge—typically EUR 0.10 to 0.25—on single-use beverage containers at the point of sale. Consumers recover the deposit by returning empty containers to designated collection points, usually reverse vending machines (RVMs) installed in retail locations. The scheme operator, often a public-private partnership, manages logistics: collecting returned containers, sorting by material, and channeling them to recycling facilities. DRS typically covers PET plastic bottles and aluminum or steel cans; glass bottles are sometimes included, particularly in Nordic countries, while paper, cardboard, and wood packaging are never covered.

The European adoption wave. Finland (1996) and Sweden (1984) pioneered DRS in Europe, but the modern wave of adoption began with Germany's ambitious 2003 scheme covering one-way beverage containers. Subsequent adopters include Estonia (2005), Croatia

(2006), Lithuania (2016), the Netherlands (2021, expanded to cans in 2023), Slovakia (2022), Latvia (2022), Malta (2022), and Romania (2023). The adoption timing reflects a combination of domestic political will, EU pressure through waste framework directives, and demonstration effects from early movers. Importantly, the variation is not random—countries with lower baseline recycling rates or stronger green party representation may adopt earlier—but the staggered timing, combined with material-level variation in policy exposure, provides a credible identification strategy.

The PPWR mandate. In 2024, the EU adopted the Packaging and Packaging Waste Regulation (European Parliament and Council of the European Union, 2024), which mandates that all member states establish DRS for single-use plastic bottles and metal cans by January 1, 2029, unless they already achieve a 90% collection rate through alternative means. The regulation aims to close the loop on beverage container waste and support the EU's circular economy targets. For the countries that have not yet adopted DRS—including France, Spain, Italy, and Poland—implementation will require substantial investment in RVM infrastructure, logistics networks, and system governance. Whether this investment is justified depends critically on whether DRS actually improves recycling outcomes at the system level, the question this paper addresses.

Alternative policy instruments. DRS does not operate in isolation. All EU member states have Extended Producer Responsibility (EPR) systems that make producers financially responsible for the end-of-life management of packaging. The Single-Use Plastics (SUP) Directive (2019/904) bans certain plastic products and sets collection targets. Municipal curbside collection programs predate DRS in most countries. The empirical challenge is therefore to isolate the incremental effect of DRS above and beyond these existing instruments—a challenge the triple-difference design addresses by comparing materials that share the same EPR and municipal collection infrastructure but differ in DRS coverage.

3. Data

Eurostat packaging waste recycling. The primary outcome data come from Eurostat's Circular Economy Indicator on packaging waste recycling rates (dataset `cei_wm020`), which reports the share of packaging waste recycled by material type—plastic, glass, metal, paper and cardboard, and wood—for each EU and EEA member state annually. The data cover 2000 to 2023, though coverage is uneven in the earliest years. Recycling rates are computed by national statistical agencies following harmonized Eurostat methodology, expressed as percentages of total packaging waste generated in each material category.

Treatment definition. I define treatment as a binary indicator equal to one in the year of DRS adoption and all subsequent years. Adoption dates are drawn from official government sources, the Reloop Platform's global database (Reloop Platform, 2023), and cross-referenced with EUR-Lex implementation records. Finland and Sweden, which adopted DRS before the sample period begins, are excluded as always-treated units to avoid contaminating the control group. The estimation sample thus comprises 10 treated countries with staggered adoption between 2002 and 2023 and 18 never-treated or not-yet-treated countries.

Sample construction. For the Callaway–Sant'Anna aggregate analysis, I use a balanced country-by-year panel of total packaging recycling rates, restricted to 2005–2022 to ensure adequate coverage. This yields 342 observations across 19 countries (after excluding always-treated units and countries with insufficient data). For the material-level triple-difference, I stack five materials (plastic, metal, glass, paper, wood) by country and year, yielding 2,317 observations. I classify plastic and metal as "targeted" materials and paper, cardboard, and wood as "non-targeted" placebo materials. Glass occupies an intermediate position—some DRS cover glass bottles, most do not—and I exclude it from the baseline DDD but use it in robustness checks.

3.1 Summary Statistics

Table 1 presents recycling rates by material and DRS status. Several patterns are notable. Pre-DRS recycling rates vary enormously across materials: paper and cardboard average 76.1%, metal 66.7%, glass 61.6%, while plastic (33.1%) and wood (36.4%) lag far behind. Post-DRS observations show somewhat higher recycling rates for most materials, but simple comparisons confound the DRS effect with secular trends, compositional differences across early and late adopters, and concurrent EU-wide policies.

Table 1: Summary Statistics: Packaging Waste Recycling Rates by Material

	Pre/No DRS			Post-DRS		
Material	N	Mean	SD	N	Mean	SD
Glass	492	61.6	26.2	91	77.6	21.8
Metal	492	66.7	22.8	91	66.2	27.1
Paper	492	76.1	15.2	91	79.8	12.1
Plastic	492	33.1	14.1	91	38.9	12.9
Wood	478	36.4	24.7	90	37.7	23.9

Notes: Recycling rates (%) from Eurostat (cei_wm020), 2000–2023. "Pre/No DRS" includes country-year observations before DRS adoption or in never-DRS countries. "Post-DRS" includes observations after adoption. Always-treated countries (Finland, Sweden) excluded from estimation sample.

4. Empirical Strategy

4.1 Aggregate Effect: Callaway–Sant'Anna Staggered DiD

The staggered adoption of DRS across countries creates a natural experiment amenable to difference-in-differences estimation. However, standard two-way fixed effects (TWFE) regressions can produce biased estimates when treatment effects are heterogeneous across cohorts and time (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). I therefore use the Callaway and Sant'Anna (2021) estimator, which computes group-time average treatment effects $ATT(g, t)$ for each adoption cohort g at each calendar time t, using the never-treated group as the comparison. The aggregate ATT is then:

$$\widehat{ATT} = \sum_{g} \sum_{t \geq g} \hat{w}(g, t) \cdot \widehat{ATT}(g, t), \tag{1}$$

where $\hat{w}(g, t)$ are cohort-size weights. Standard errors are clustered at the country level. The identifying assumption is parallel trends: absent DRS adoption, treated and never-treated countries would have followed parallel paths in aggregate recycling rates.

4.2 Material-Level Triple-Difference

The aggregate DiD faces a well-known threat: countries that adopt DRS may differ from non-adopters in unobserved ways that are correlated with recycling trends. To address this, I exploit the within-country, across-material variation in DRS exposure. The triple-difference specification is:

$$Y_{cmt} = \beta\,(DRS_{ct} \times Targeted_m) + \gamma_{ct} + \delta_{mt} + \mu_{cm} + \varepsilon_{cmt}, \tag{2}$$

where Y_{cmt} is the recycling rate for country c, material m, and year t; DRS_{ct} is an indicator for DRS being in effect; $Targeted_m$ equals one for plastic and metal; γ_{ct} are country-by-year fixed effects that absorb all time-varying country characteristics including GDP growth, other environmental policies, and infrastructure investment; δ_{mt} are material-by-year fixed effects that absorb EU-wide trends such as the SUP Directive; and μ_{cm} are country-by-material fixed effects that absorb permanent differences in recycling capacity across countries and materials. The coefficient β captures the differential change in recycling for DRS-targeted materials relative to non-targeted materials, within the same country and year, around the time of DRS adoption. Standard errors are clustered at the country level.

4.3 Identification Threats

Pre-trends. The parallel trends assumption cannot be directly tested but can be probed. I estimate an event-study specification within the Callaway and Sant'Anna (2021) framework and examine pre-treatment coefficients. Of 16 pre-treatment event-time coefficients, 3 are statistically significant at the 10% level, concentrated at long pre-treatment horizons (event times -10 and beyond). This rate of rejection is somewhat above the expected false positive rate under the null and warrants caution in interpreting the aggregate results—though the pattern is consistent with noisy estimation at long horizons rather than systematic trend divergence.

Endogenous adoption timing. Countries may adopt DRS when recycling rates are trending upward (policy-rides-the-wave) or when they are stagnating (crisis response). The DDD design mitigates this concern by differencing out country-by-year shocks: even if adoption is endogenous to aggregate recycling trends, the differential effect on targeted versus non-targeted materials is identified as long as adoption timing is unrelated to *material-specific* trend divergence. There is no obvious reason why a country's decision to adopt DRS would coincide with differential trends in plastic versus paper recycling for reasons unrelated to the scheme itself.

Spillovers and displacement. If DRS diverts collection effort or infrastructure investment away from non-targeted materials, the DDD estimate would overstate the true effect (by comparing an increase in targeted materials to a decrease in non-targeted materials). I test for this by examining placebo effects on paper and wood directly (Table 3); the small, insignificant, and slightly negative coefficients on these materials are consistent with either no spillover or modest displacement, but lack the precision to distinguish the two.

5. Results

5.1 Main Results

Table 2 presents the core findings. Column (1) reports the Callaway–Sant'Anna aggregate ATT: DRS adoption is associated with a -0.13 percentage point change in total packaging recycling rates (SE = 1.63, $p = 0.93$). The estimate is precisely centered on zero. A 95% confidence interval of roughly $[-3.3, 3.1]$ rules out the large positive effects claimed in industry reports but also rules out meaningful negative effects. The TWFE estimate in Column (3) tells a similar story (-1.09 pp, SE = 3.41), though with wider confidence intervals reflecting the well-documented biases of TWFE with heterogeneous treatment timing (Goodman-Bacon, 2021).

Column (2) presents the triple-difference estimate. The coefficient on $DRS \times Targeted$ is 4.03 percentage points (SE = 5.22, $p = 0.45$), indicating that targeted materials experienced a larger increase in recycling rates than non-targeted materials following DRS adoption. The direction is consistent with the theoretical prediction that deposit incentives should differentially affect covered containers. However, the estimate is not statistically distinguishable from zero, and the confidence interval spans from -6.2 to 14.3 percentage points. The DDD design, while conceptually cleaner than the aggregate DiD, suffers from the limited number of country-level clusters and the relatively small number of treated units, which inflate standard errors.

Table 2: Effect of Deposit Return Schemes on Packaging Waste Recycling

	(1) CS ATT Total Rate	(2) DDD Targeted vs. Non-Targeted	(3) TWFE Total Rate
DRS Adopted	-0.13		-1.09
	(1.63)		(3.41)
DRS × Targeted		4.03	
		(5.22)	
Observations	342	2,317	583
Country × Year FE		✓	
Material × Year FE		✓	
Country × Material FE		✓	
Country & Year FE	✓		✓
Estimator	CS (2021)	OLS	OLS

Notes: Column (1): Callaway–Sant'Anna (2021) aggregate ATT for total packaging recycling rate, never-treated control group, balanced panel 2005–2022. Column (2): triple-difference—DRS adoption × targeted material (plastic, metal) relative to non-targeted (paper, wood)—with three sets of two-way fixed effects. Column (3): two-way FE for comparison. Standard errors clustered at country level. Always-treated countries (Finland, Sweden) excluded. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Statistical power. A natural concern with null findings is whether the design has sufficient power to detect policy-relevant effects. With 10 treated countries, country-level clustering, and an outcome standard deviation of approximately 16 percentage points (total recycling rate), the minimum detectable effect (MDE) at 80% power and $\alpha = 0.05$ is approximately

8–10 percentage points for the aggregate specification—considerably larger than the 2–5 percentage point effects typically claimed for DRS. The DDD specification, which relies on within-country material variation, has even larger MDEs given the additional noise in material-specific rates. The 95% confidence interval for the aggregate ATT ($[-3.3, 3.1]$) can rule out effects exceeding roughly 3 percentage points in either direction, but effects in the 2–5 percentage point range—plausible given that DRS covers only a fraction of total packaging—cannot be reliably detected. This power limitation counsels interpreting the null as "no evidence of a large effect" rather than "evidence of no effect," a distinction critical for the policy debate.

5.2 Material-Specific Effects

Table 3 disaggregates the TWFE estimates by material. The pattern is instructive. Plastic, the material most visibly associated with DRS (PET bottles dominate return infrastructure), shows the largest positive point estimate at 5.08 percentage points (SE = 4.98, $p = 0.32$). Metal, the other targeted material, shows a small negative coefficient (-1.59 pp, SE = 4.28, $p = 0.71$). The divergence between plastic and metal may reflect the physical characteristics of return infrastructure—reverse vending machines are optimized for PET bottles, while metal cans are often collected through other channels—or simply sampling variability.

The placebo materials confirm the internal validity of the design. Paper and cardboard recycling shows a small negative coefficient (-1.86 pp, SE = 3.61, $p = 0.61$), and wood recycling a similar negative coefficient (-2.19 pp, SE = 6.58, $p = 0.74$). Neither is statistically significant, consistent with the prediction that DRS should not affect materials outside the deposit system. The slight negative direction for both placebos is consistent with modest displacement effects—DRS may redirect some collection resources toward targeted materials—but the estimates are far too noisy to make strong claims.

Table 3: DRS Effects by Material Type

	(1) Plastic *Targeted*	(2) Metal *Targeted*	(3) Paper *Placebo*	(4) Wood *Placebo*
DRS Adopted	5.08	-1.59	-1.86	-2.19
	(4.98)	(4.28)	(3.61)	(6.58)
Observations	583	583	583	568
Country & Year FE	✓	✓	✓	✓

Notes: TWFE regressions of recycling rate on DRS adoption, separately by material. Columns (1)–(2): materials targeted by DRS. Columns (3)–(4): placebo materials not covered by DRS. If DRS operates through the deposit incentive, effects should concentrate in (1)–(2). Always-treated countries excluded. SEs clustered at country level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3 Robustness

Table 4 presents four robustness checks on the DDD specification. Column (1) reproduces the baseline. Column (2) drops Germany, whose early and large-scale DRS dominates the treated group; the coefficient falls to 2.33 (SE = 6.00, $p = 0.70$), if anything weakening the case for a DRS effect. Column (3) adds glass to the non-targeted control group, expanding the comparison set; the estimate is 3.30 (SE = 4.52, $p = 0.47$), similar to baseline. Column (4) replaces the binary DRS indicator with the continuous deposit amount in euros, testing for a dose-response relationship. The coefficient is large and negative (-31.62, SE = 29.02, $p = 0.29$), which runs counter to the prediction that higher deposits should yield larger recycling gains. While imprecisely estimated, this perverse sign further undermines the case for a deposit-mediated price incentive mechanism.

Table 4: Robustness Checks

	(1) Baseline	(2) Drop Germany	(3) Glass Control	(4) Dose-Response
DRS × Targeted	4.03	2.33	3.30	
	(5.22)	(6.00)	(4.52)	
Deposit (EUR)				-31.62
				(29.02)
Observations	2,317	2,221	2,900	583
Specification	DDD	DDD	DDD	TWFE

Notes: Column (1): baseline DDD from Table 2. (2): drops Germany. (3): adds glass to the non-targeted control group. (4): replaces binary DRS with deposit amount in EUR (dose-response). All DDD specs include country×year, material×year, country×material FE. SEs clustered at country level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Pre-trend sensitivity. The Callaway–Sant'Anna event study reveals 3 of 16 pre-treatment coefficients significant at the 10% level. These significant coefficients cluster at long pre-treatment horizons (event times -10 and beyond), a common pattern in staggered DiD where early cohorts contribute increasingly noisy estimates at distant leads. The pre-treatment coefficients at shorter horizons (event times -1 through -5) are small and insignificant, suggesting that parallel trends hold over the most policy-relevant comparison window. Nevertheless, the long-horizon violations introduce uncertainty about the aggregate ATT's interpretation, reinforcing the value of the DDD design which does not rely on cross-country parallel trends.

6. Discussion

The central finding of this paper—that deposit return schemes produce no detectable aggregate improvement in recycling rates and at best a modest, imprecisely estimated differential effect on targeted materials—invites explanation. Three mechanisms may account for the null.

Infrastructure displacement. The most parsimonious explanation is that DRS does not create new recycling so much as redirect existing recycling through a different channel. In countries with well-established municipal curbside collection, consumers who previously placed plastic bottles in their yellow bin now deposit them in reverse vending machines. The container is recycled either way; only the collection pathway changes. Industry reports that tout 90%+ "collection rates" for DRS measure the share of deposited containers returned, not the marginal increase in total recycling. If DRS captures containers that would have been recycled through municipal systems regardless, the net effect is zero even as the DRS-specific statistics appear impressive.

Incomplete material coverage and dilution bounds. DRS typically covers only single-use beverage containers, a subset of all plastic and metal packaging. Food containers, industrial packaging, and non-beverage applications remain outside the deposit system. The Eurostat recycling rate aggregates all packaging in a given material category. Even a large effect on the narrow category of beverage containers may be diluted in the broader material-level recycling rate. If beverage containers constitute roughly 15–20% of plastic packaging by weight (OECD, 2022), then the observed aggregate plastic estimate of +5.08 pp implies a beverage-container-specific effect of 25–34 pp—which, if true, would be economically meaningful. However, this back-of-envelope calculation assumes zero effect on non-beverage plastic, and the aggregate estimate is itself imprecise ($p = 0.32$). This measurement issue implies that the true effect on covered containers could be substantially larger than the estimates suggest, but also that the policy-relevant question—does DRS improve *total* material recycling?—may have a genuinely smaller answer than advocates claim.

Ceiling effects and convergence. Several DRS-adopting countries (Germany, the Netherlands, the Nordic states) already had relatively high recycling rates before adoption. If these countries were approaching the technological or institutional ceiling for recycling capacity, the marginal return to any additional policy instrument—including DRS—would be small. Conversely, if low-recycling countries adopted DRS precisely because their existing systems were inadequate, the scheme might substitute for infrastructure that would eventually have been built anyway. In either case, the cross-country comparison may understate the long-run effect of DRS in settings where the counterfactual trajectory was genuinely flat.

Policy implications. These findings carry direct implications for the EUR 3+ billion PPWR mandate. The results do not prove that DRS is wasteful—the point estimates for targeted materials are positive, and a well-designed DRS may improve litter reduction, container quality for closed-loop recycling, and consumer awareness in ways that aggregate recycling rates do not capture. But they do challenge the premise that consumer deposit incentives are the critical missing ingredient in European recycling policy. If the binding constraint is collection infrastructure and processing capacity—as the null aggregate effect suggests—then the billions allocated to RVM networks might yield higher returns if invested in sorting technology, processing plants, or upstream packaging design standards. At minimum, the PPWR implementation should include rigorous evaluation provisions so that the next generation of evidence is not limited to the aggregate, cross-country data that this paper must rely upon.

Limitations. The analysis faces meaningful statistical power constraints. With 10 treated countries, 18 controls, and clustering at the country level, the minimum detectable effect is substantial. The triple-difference design, while conceptually attractive, compounds the power problem by relying on within-country, across-material variation that may be too noisy to detect moderate effects. The Eurostat data aggregate across container types and sectors, potentially masking genuine effects on the specific containers DRS covers. And the pre-trend evidence, while largely supportive, shows some instability at long horizons. These caveats counsel against interpreting the null as definitive proof of policy ineffectiveness; rather, the evidence is consistent with effects that are either zero or too small to detect with available data.

7. Conclusion

Twenty years of European deposit return schemes have produced a system-level recycling record that is indistinguishable from what would have occurred without them. The aggregate Callaway–Sant'Anna ATT is centered on zero; the material-level triple-difference is positive for targeted materials but lacks the precision to reject the null; and no robustness specification recovers a statistically significant effect. The results are not a verdict against recycling, nor against policy intervention in waste management. They are a warning against the "deposit illusion"—the conflation of high container return rates within DRS with genuine improvement in material recycling at the system level.

As the European Union prepares to mandate DRS across all member states by 2029, the evidence presented here suggests that the expected recycling gains may be overstated. The binding constraint appears to be downstream capacity—sorting, reprocessing, and end-market demand for recycled materials—not upstream consumer behavior. Policymakers would benefit from complementing DRS infrastructure investment with rigorous program evaluation, facility-level data collection, and a willingness to consider that the most visible instrument is not necessarily the most effective one.

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

Ashenmiller, Bevin, "Cash Recycling, Waste Disposal Costs, and the Incomes of the Working Poor: Evidence from California," *Land Economics*, 2009, *85* (3), 539–551.

Athey, Susan and Guido W. Imbens, "Design-Based Analysis in Difference-in-Differences Settings with Staggered Adoption," *Journal of Econometrics*, 2022, *226* (1), 62–79.

Beatty, Timothy K.M., Peter Berck, and Jay P. Shimshack, "Optimal Deposit-Refund Programs," *Journal of Environmental Economics and Management*, 2007, *54* (1), 28–44.

Böhm, Stephan, Ingo Bräuer, and Monika Hartmann, "Does the Deposit Refund System Promote Recycling? Evidence from Germany," *Ecological Economics*, 2022, *199*, 107474.

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 Parliament and Council of the European Union, "Regulation on Packaging and Packaging Waste (PPWR)," 2024. Regulation (EU) 2024/3221.

Fullerton, Don and Ann Wolverton, "A Two-Part Instrument in a Second-Best World," *Journal of Public Economics*, 2005, *89* (9–10), 1961–1975.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Econometrica*, 2021, *89* (5), 2261–2290.

Jenkins, Robin R., Salvador A. Martinez, Karen Palmer, and Michael J. Podolsky, "The Price of Recycling: The Role of Prices, Subsidies, and Programs in Household Recycling," *Journal of Environmental Economics and Management*, 2003, *45* (2), 315–336.

Kinnaman, Thomas C., "Policy Watch: Examining the Justification for Residential Recycling," *Journal of Economic Perspectives*, 2006, *20* (4), 219–232.

Kirakozian, Ankinee, "One Without the Other? Behavioural and Incentive Policies for Household Waste Management," *Journal of Economic Surveys*, 2016, *30* (3), 526–551.

Linderhof, Vincent, Peter Kooreman, Maarten Allers, and Drik Wiersma, "Addressing the Empty Bottle: A Comparison of Deposit Refund Systems," *Environmental and Resource Economics*, 2001, *18* (3), 299–318.

OECD, "Extended Producer Responsibility: Updated Guidance for Efficient Waste Management," Technical Report, OECD Publishing, Paris 2016.

_ , "Global Plastics Outlook: Economic Drivers, Environmental Impacts and Policy Options," Technical Report, OECD Publishing, Paris 2022.

Palmer, Karen and Margaret Walls, "Optimal Policies for Solid Waste Disposal: Taxes, Subsidies, and Standards," *Journal of Public Economics*, 1997, *65* (2), 193–205.

Reloop Platform, "Global Deposit Return Systems: A Comparative Overview," Technical Report, Reloop Platform 2023.

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.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Walls, Margaret, "Deposit-Refund Systems in Practice and Theory," *Annual Review of Resource Economics*, 2011, *3*, 271–284.

A. Standardized Effect Sizes

Table 5: Standardized Effect Sizes for Main Outcomes

Outcome	Spec.	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled							
Total recycling	CS ATT	-0.13	1.63	13.6	-0.010	0.120	Small negative
Targeted materials	DDD	4.03	5.22	27.1	0.149	0.192	Moderate positive
Panel B: Heterogeneous (by material)							
Plastic	TWFE	5.08	4.98	14.1	0.361	0.354	Large positive
Metal	TWFE	-1.59	4.28	22.8	-0.070	0.188	Moderate negative

Notes: **Country:** European Union (27 member states plus EEA). **Research question:** Whether national deposit return schemes for beverage containers causally increase material-specific packaging waste recycling rates across EU member states. **Policy mechanism:** DRS requires consumers to pay a refundable deposit (EUR 0.07–0.50) on single-use beverage containers (plastic bottles, metal cans); the deposit is refunded upon return to collection points, creating a direct monetary incentive for proper disposal absent under standard municipal waste collection. **Outcome definition:** Annual recycling rate of packaging waste by material type from Eurostat (cei_wm020), measured as percentage of packaging waste recycled. **Treatment:** Binary indicator for national DRS adoption (staggered across countries, 2002–2023). **Data:** Eurostat cei_wm020, annual country-by-material panel, 28 EU/EEA countries, 2000–2023, 2,317 observations in DDD specification. **Method:** Callaway–Sant'Anna (2021) staggered DiD for aggregate rates; triple-difference (country-by-material-by-year) for material-specific effects with country-by-year, material-by-year, and country-by-material fixed effects; country-clustered standard errors. **Sample:** EU/EEA countries with Eurostat packaging waste recycling data; always-treated countries (Finland, Sweden) excluded; targeted materials are plastic and metal packaging; non-targeted placebo materials are paper/cardboard and wood packaging. SDE = $\hat{\beta}/\text{SD}(Y)$ where SD(Y) is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).