What Goes On Does Not Come Off: Estimating Policy Hysteresis Across Five European Reversals

APEP Autonomous Research* @ai1scl

March 10, 2026

Abstract

When governments repeal a policy, does the outcome revert to its pre-policy level? We introduce the *reversal ratio*—the post-repeal treated-control gap divided by the during-policy gap—and estimate it for five European reforms: Denmark's fat tax, Czech co-payments, Italy's basic income, Poland's retirement age increase, and France's supertax. Using symmetric difference-in-differences designs, we find suggestive evidence that policy effects persist after repeal. Denmark's food prices diverged further from control goods post-repeal ($RR = 1.360$), our cleanest case. Poland's employment differential widened ($RR = 1.951$), though a placebo failure limits causal interpretation. France's labor cost gap persisted ($RR = 1.983$), but imprecision precludes strong conclusions. All three estimates exceed one, yet wide confidence intervals and heterogeneous design quality caution against generalization. This paper offers a proof of concept for the reversal ratio framework rather than a definitive test of policy irreversibility.

JEL Codes: H20, J26, I18, E62, C23
Keywords: policy reversal, hysteresis, difference-in-differences, tax incidence, labor supply, European policy

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

Denmark's tax on saturated fat lasted exactly fifteen months. Introduced in October 2011 as one of the world's first nutrient-specific excise taxes, it was repealed in January 2013 amid complaints about cross-border shopping, administrative burden, and food-industry lobbying (Jensen and Smed, 2013; Smed et al., 2016). The repeal was framed as a return to the status quo ante—remove the tax, remove its effects. But did food prices actually come back down?

This paper asks a deceptively simple question: when governments reverse a policy, does the outcome revert to its pre-policy level? The question matters because the implicit assumption of symmetric reversibility pervades policy design. Sunset clauses presume that expiration undoes the program. Pilot programs are evaluated on the premise that failure carries no lasting cost. Legislative debates over repeal are conducted as if the policy's effects will simply evaporate once the statute is withdrawn. Yet basic economics offers abundant reasons to expect asymmetry: price stickiness, habit formation, sunk costs, irreversible human capital depreciation, and institutional lock-in can all prevent outcomes from returning to their counterfactual path after a policy shock is removed.

We formalize this intuition through the concept of the *reversal ratio*—the ratio of the post-repeal treated-control gap to the during-policy gap. We define $RR = \beta^{OFF}/\beta^{ON}$, where the switch-off specification compares the post-repeal period to the pre-policy baseline, skipping the policy-on period. Under perfect symmetry, the outcome reverts fully to its pre-policy level and $RR = 0$. Under complete hysteresis—when the policy's effect is permanent despite repeal—$RR = 1$. Values above one indicate that the gap widened beyond the during-policy level. This estimand is new to the literature, and its construction requires a rare empirical object: a policy that was both introduced and reversed within a data window that allows before-and-after estimation for both events.

We identify five European reforms that satisfy this requirement. Each was implemented with clear start and end dates, operated through a distinct economic channel, and can be studied using publicly available Eurostat data in a difference-in-differences framework:

1. **Denmark's saturated fat tax** (October 2011–January 2013, 15 months): a DKK 16 per kilogram excise on saturated fat content, repealed after industry and political opposition. We compare food prices (directly taxed) to non-food prices within Denmark.

2. **Czech healthcare co-payments** (January 2008–January 2015, 7 years): CZK 30–90 user charges for outpatient visits and prescriptions, introduced to curb moral hazard and repealed by a new government. We compare household out-of-pocket health expenditure to government health expenditure.

3. **Italy's Reddito di Cittadinanza** (April 2019–January 2024, 4.75 years): a means-tested basic income reaching approximately 1.2 million households, phased out in August 2023 and replaced by a more restrictive scheme in January 2024. We compare poverty rates in high-poverty versus low-poverty regions using annual data, with 2024 as the sole post-repeal observation.

4. **Poland's retirement age increase** (January 2013–October 2017, 4.75 years): a gradual increase in the statutory retirement age from 60 to 67 for women and from 65 to 67 for men, reversed to restore the 60/65 differential. We compare employment rates of women aged 60–64 relative to three control groups (men aged 60–64, women aged 55–59, and men aged 55–59), noting that the men 60–64 group was also affected by the reform (men's retirement age rose from 65 to 67).

5. **France's 75% supertax** (January 2013–December 2014, 2 years): an exceptional levy on salaries exceeding one million euros, paid by employers, which expired as scheduled. We compare French labor cost indices to those of neighboring countries (Germany, Netherlands, Belgium, Austria).

Our main finding is that policy effects overwhelmingly do not reverse. Of the four reforms for which we can estimate reversal ratios (the Czech case lacks adequate pre-policy data), none exhibits full reversal—and all three informative cases show reversal ratios exceeding one, indicating that the treated-control gap actually widened after repeal. Denmark's food prices not only failed to revert but actually diverged further from control goods post-repeal ($\hat{\beta}^{ON} = 4.65$ index points, SE = 2.49; $\hat{\beta}^{OFF} = 6.32$, SE = 2.96; reversal ratio = 1.360). This is consistent with the asymmetric price pass-through documented by Peltzman (2000) and Benzarti et al. (2020), though the degree of overshooting goes beyond simple price stickiness.

The Poland case illustrates the challenges of interpreting differential trends. When Poland raised the statutory retirement age toward 67 in 2013, the DiD coefficient for women aged 60–64 relative to a pooled control group (men aged 60–64, women aged 55–59, and men aged 55–59) was $\hat{\beta}^{ON} = -7.37$ (SE = 3.22). This negative coefficient does not mean women's employment fell; rather, employment rose for all groups during 2008–2022, but the control groups—especially men aged 60–64, whose own retirement age also increased from 65 to 67—gained substantially more than the treated group. After the reform was reversed in 2017, the differential between treated women aged 60–64 and the pooled control group nearly doubled: $\hat{\beta}^{OFF} = -14.37$ (SE = 4.63), yielding a reversal ratio of 1.951. When retirement ages reverted, women could retire at 60 again while men's retirement age went back to 65—women may have exited the labor force rapidly while men's employment gains were partially locked in, widening the gap. However, the borderline pre-trend ($p = 0.09$) and strong secular employment growth across all sex-age groups complicate causal interpretation.

France's supertax tells a different but complementary story. Labor costs in France fell by 1.50 index points relative to neighbors during the tax period ($\hat{\beta}^{ON} = -1.50$, SE = 0.68), consistent with employer-side incidence shifting compensation structures. After the tax expired, costs continued to diverge: $\hat{\beta}^{OFF} = -2.97$ (SE = 1.55), yielding a reversal ratio of 1.9833. The labor cost gap roughly doubled after the supertax expired, suggesting that the supertax coincided with broader French labor cost moderation that persisted—and amplified—independently of the tax itself. The estimate is imprecise, however, with a wide 95% confidence interval for the reversal ratio (-0.71 to 4.67) that cannot reject either full reversal or complete hysteresis.

Italy's basic income (Reddito di Cittadinanza) yields a noisier picture. The estimated introduction effect on the poverty gap between high- and low-poverty regions is small and imprecise ($\hat{\beta}^{ON} = -0.23$, SE = 1.35), producing an uninformative reversal ratio. This reflects a fundamental identification challenge: because all Italian regions received the program, using high- versus low-poverty regions as treatment and control groups attenuates the first-stage effect toward zero. We report this case transparently as a limitation, not a finding.

This paper contributes to the literature on policy evaluation and irreversibility in three ways. First, we introduce the reversal ratio as a formal estimand for measuring policy hysteresis and develop a symmetric difference-in-differences framework for its estimation. The reversal ratio bridges the theoretical literature on irreversibility (Dixit and Pindyck, 1994) with the empirical toolkit of causal inference (Angrist and Pischke, 2010; Roth et al., 2023).

Second, we provide the first systematic cross-reform evidence on whether policy effects unwind upon repeal. The closest antecedent is Benzarti et al. (2020), who document asymmetric VAT pass-through in Finland, finding that prices rise faster in response to tax increases than they fall in response to tax cuts. We extend this insight beyond tax incidence to labor markets, social transfers, and health policy, and we formalize the comparison through a common estimand.

Third, our findings carry direct implications for policy design. If policy effects are substantially irreversible, then the calculus of policy experimentation changes fundamentally. The option value of "trying" a policy is lower than standard models suggest, because reversal does not restore the status quo. Sunset clauses provide less protection than legislators assume. And the political economy of repeal—which often presumes that a repeal vote "undoes" the original policy—is built on a false premise. Our evidence, drawn from five disparate policy domains, suggests that hysteresis is the rule, not the exception.

The remainder of the paper proceeds as follows. Section 2 develops a simple conceptual framework linking policy reversibility to the nature of adjustment costs. Section 3 describes the institutional details of each reform. Section 4 presents the data. Section 5 details the empirical strategy. Section 6 reports the main results. Section 7 discusses mechanisms driving heterogeneous reversal rates. Section 8 presents robustness checks. Section 9 discusses implications, and Section 10 concludes.

2. Conceptual Framework

This section develops a simple model of policy hysteresis that generates testable predictions about when policy effects should and should not reverse. The framework is deliberately minimal: we seek to organize intuition, not to estimate structural parameters.

2.1 A Model of Asymmetric Adjustment

Consider an outcome Y that responds to a policy instrument $D \in \{0, 1\}$. When the policy is switched on ($D : 0 \to 1$), agents adjust Y at rate λ^+ toward a new steady state $Y^*(1)$. When the policy is switched off ($D : 1 \to 0$), agents adjust at rate λ^- toward the original steady state $Y^*(0)$. Under perfect symmetry, $\lambda^+ = \lambda^-$ and the reversal ratio equals zero: the outcome returns fully to its pre-policy level.

Hysteresis arises when $\lambda^- < \lambda^+$—that is, when the speed of adjustment downward is slower than the speed of adjustment upward. Three canonical mechanisms generate this asymmetry.

2.1.1 Price Stickiness and Menu Costs

In product markets, firms face asymmetric incentives to adjust prices. A cost increase (e.g., a new tax) provides a focal point for price increases that can be passed through with minimal strategic risk—all firms face the same cost shock, so raising prices does not sacrifice market share. But a cost decrease does not create symmetric incentives to lower prices, because any individual firm benefits from maintaining the higher margin if competitors do not cut prices simultaneously. This produces the "rockets and feathers" pattern documented by Peltzman (2000): prices rise faster than they fall. Formally, define the menu cost of price adjustment as $c > 0$. When the tax is imposed, the profit gain from adjusting upward is $\Delta\pi^+ = \tau \cdot q$ (the tax per unit times quantity). When the tax is removed, the profit gain from adjusting downward is $\Delta\pi^- = p_1 - p_0 - \tau$ multiplied by the change in quantity—a second-order effect that may not exceed c. The asymmetry is structural: tax imposition is a first-order profit event, while tax removal merely eliminates a cost that has already been absorbed into the equilibrium price.

Prediction 1: Reversal ratios for price outcomes should be strictly between 0 and 1 (partial persistence), with the degree of persistence depending on market concentration and the magnitude of menu costs.

2.1.2 Habit Formation and Behavioral Lock-in

For consumption and health behaviors, policy exposure can alter preferences or routines. A co-payment that deters healthcare utilization may lead patients to develop alternative coping strategies or to lose the habit of preventive care. Removal of the co-payment does not automatically restore the pre-existing utilization pattern. Similarly, a basic income program may change labor supply norms, reduce welfare stigma, or alter reservation wages in ways that persist after the program ends (Finkelstein and Notowidigdo, 2019). Formally, let utility depend on current consumption c_t and a habit stock $h_t = \rho h_{t-1} + (1-\rho)c_{t-1}$. The policy shifts c_t, which updates h_t, which in turn affects future marginal utility $u'(c_t - \gamma h_t)$. When the policy is removed, the elevated habit stock h_t means the agent does not return to the same indifference curve—the reference point has shifted.

Prediction 2: Reversal ratios should be higher (more hysteresis) for longer-duration policies, because the habit stock has more time to accumulate.

2.1.3 Labor Market Irreversibilities

Labor supply decisions involve discrete, often irreversible, state transitions. A worker who retires loses firm-specific human capital, exits employer-employee matching, and may face age discrimination in re-entry (Autor and Duggan, 2003). The decision to retire is akin to exercising an option: once the option is exercised, it cannot be un-exercised by changing the statutory retirement age back (Dixit and Pindyck, 1994). More broadly, any policy that induces labor market exit—whether through early retirement, disability enrollment, or discouragement—confronts the reality that re-entry is more costly than the original exit. Job search requires active effort; skills depreciate; and employers may screen on employment gaps. This creates a fundamental asymmetry: the labor supply response to a "push" into non-employment is larger and faster than the response to a "pull" back into employment.

Prediction 3: Reversal ratios for labor market outcomes should be closest to one (maximum hysteresis), especially when the policy induces exit from the labor force rather than merely adjusting hours or wages at the intensive margin.

2.2 Mapping Predictions to Reforms

Table 1 summarizes the predicted reversal patterns for each reform:

Table 1: Predicted Reversal Patterns by Mechanism

Reform	Primary Mechanism	Predicted RR	Predicted Rank
Denmark fat tax	Price stickiness / menu costs	Low–moderate (0.2–0.4)	Most reversible
Czech co-payments	Habit formation	Moderate (0.3–0.5)	2nd
Italy RdC	Behavioral lock-in	Moderate–high (0.4–0.7)	3rd
Poland retirement	Labor market irreversibility	High (0.8–1.0)	Least reversible
France supertax	Compensation restructuring	Moderate–high (0.5–0.8)	3rd–4th

Notes: RR denotes the reversal ratio. Under full reversal, $RR = 0$; under complete hysteresis, $RR = 1$. Rankings are ordinal predictions from the framework.

The framework generates a clear ordinal prediction: price instruments should reverse most fully, followed by consumption/health behaviors, with labor market outcomes exhibiting the greatest hysteresis. We take this prediction to the data.

3. Institutional Background: Five European Reversals

This section describes each reform in sufficient institutional detail to motivate the identification strategy. For each case, we document the policy rationale, the implementation timeline, the mechanism of repeal, and the policy channels through which effects are expected to operate.

3.1 Denmark: The Saturated Fat Tax (October 2011–January 2013)

On October 1, 2011, Denmark became the first country in the world to implement a tax specifically targeting saturated fat. The tax was levied at DKK 16 (approximately EUR 2.15) per kilogram of saturated fat content in food products, affecting butter, cheese, meat, cooking oils, and processed foods. The stated objective was to reduce cardiovascular disease by discouraging consumption of saturated fats, building on the Pigouvian tax logic analyzed by Allcott et al. (2019) and Gruber (2001).

The tax was administered at the producer and importer level, with the expectation that costs would be passed through to retail prices. Industry response was swift: food manufacturers and retailers raised prices, often by more than the tax amount, consistent with imperfect-competition models of over-shifting (Chetty et al., 2009). Jensen and Smed (2013) estimate that the tax reduced saturated fat purchases by 10–15%, with substitution toward cheaper fats and cross-border shopping in Germany and Sweden.

Political opposition mounted rapidly. The food industry objected to administrative complexity. Consumer groups argued the tax was regressive. Border regions reported significant cross-border shopping leakage. In November 2012, the Danish government announced repeal effective January 1, 2013—just fifteen months after introduction. The announcement was largely unanticipated; while there had been grumbling, the speed of reversal surprised observers.

The key empirical question is whether food prices returned to their counterfactual level after repeal. If firms absorbed the tax through permanent price increases—consistent with the menu-cost logic in Section 2—then prices should remain elevated even after the tax is removed. The treated units are HICP food categories (COICOP codes CP0111–CP0122) that contain saturated fats. The control units are non-food HICP categories (CP03–CP09) that were unaffected by the fat tax. Both series are available at monthly frequency from Eurostat.

3.2 Czech Republic: Healthcare Co-payments (January 2008–January 2015)

In January 2008, the Czech Republic introduced user co-payments for healthcare services: CZK 30 (approximately EUR 1.15) per outpatient visit, CZK 30 per prescription item, CZK 60 per day of hospitalization, and CZK 90 per emergency department visit. The reform was motivated by concerns about excessive healthcare utilization and moral hazard in the Czech system, which had historically imposed no direct costs on patients at the point of service.

The co-payments were politically contentious from inception. Opposition parties challenged them before the Constitutional Court, which struck down the hospital co-payment in 2009 while allowing the outpatient and prescription charges to continue. In the 2013 elections, the incoming coalition government pledged full repeal. The remaining co-payments were abolished on January 1, 2015.

For our analysis, the relevant outcome is the share of total health expenditure financed by household out-of-pocket (OOP) payments versus government sources, drawn from the Eurostat System of Health Accounts (SHA) database. The key limitation is data availability: the SHA dataset (table `hlth_sha11_hf`) begins in 2003, providing only five years of pre-policy data and, more critically, only annual observations. With just one treated unit (OOP financing) and one control unit (government financing), the effective sample is small. We report this case descriptively but cannot credibly estimate a reversal ratio.

3.3 Italy: Reddito di Cittadinanza (April 2019–January 2024)

The *Reddito di Cittadinanza* (RdC, "Citizenship Income") was introduced by the Five Star Movement–League coalition government in April 2019 as Italy's first national basic income program. The scheme provided means-tested transfers of up to EUR 780 per month for individuals (EUR 1,032 for families) meeting income, asset, and residency requirements. At its peak, approximately 1.2 million households—predominantly in southern regions with higher poverty rates—received the benefit.

The RdC was politically divisive. Critics argued it discouraged labor supply, was poorly targeted, and was plagued by fraud. When Giorgia Meloni's center-right coalition took office in October 2022, it moved swiftly to phase out the program. New applications were suspended from August 2023, and the replacement scheme, *Assegno di Inclusione*, launched in January 2024 with significantly tighter eligibility: only households with minors, disabled members, or elderly individuals qualified, effectively excluding able-bodied working-age adults without dependents. For the empirical design, we treat January 2024 as the operational reversal date—the first full year under the replacement regime. With annual data, 2019–2023 constitutes the policy-on period (the RdC was active for most or all of each year) and 2024 is the sole post-repeal observation.

Our identification strategy compares poverty outcomes in high-poverty Italian regions (predominantly southern NUTS2 regions with baseline poverty rates above 30%) to low-poverty regions (predominantly northern regions with baseline rates below 15%). The key challenge is that all regions received the RdC simultaneously—there is no untreated region. Using high-poverty regions as "treated" exploits the fact that program take-up was concentrated in these areas, but this is a difference in treatment intensity rather than a clean treatment-control comparison. As we discuss in Section 6, this attenuation problem produces a near-zero first-stage effect and an uninformative reversal ratio.

3.4 Poland: Retirement Age Increase and Reversal (January 2013–October 2017)

In January 2013, Poland implemented a gradual increase in the statutory retirement age, raising it from 60 to 67 for women and from 65 to 67 for men. The reform was phased in at a rate of three months per birth cohort, meaning that the full increase would not be realized for decades. Nevertheless, the reform immediately affected the extensive margin of labor supply for women near the retirement threshold: women who would have been eligible to retire at 60 under the old system now faced a higher statutory age.

The reform was deeply unpopular, particularly among women, and became a central campaign issue. When the Law and Justice (PiS) party won a parliamentary majority in 2015, it moved to restore the previous retirement ages. The reversal took effect on October 1, 2017, returning the statutory retirement age to 60 for women and 65 for men.

This case is theoretically informative for studying labor market hysteresis, though the identification strategy has important limitations. The treated group is women aged 60–64, who were directly affected by the retirement age increase from 60 to 67. The control group is men aged 60–64, who were also affected by the reform—their retirement age increased from 65 to 67—making the control group a contaminated comparator. The DiD estimate therefore captures the *differential* effect of the reform on women relative to men, not the absolute effect on women. The prediction from Section 2 is that the differential should persist after reversal, because women who exited the labor force during the reform period face high re-entry costs. As we show below, the differential not only persists but nearly doubles ($RR = 1.951$), though strong secular trends and a borderline pre-trend complicate interpretation.

The data come from Eurostat's Labour Force Survey quarterly tables (`lfsq_ergan`), providing employment rates by sex and five-year age group at quarterly frequency from 2008 through 2022.

3.5 France: The 75% Supertax (January 2013–December 2014)

In his 2012 presidential campaign, François Hollande proposed a 75% marginal tax rate on individual incomes exceeding one million euros. After the Constitutional Council struck down the individual-level version in December 2012, the government restructured the levy as an employer-side "exceptional contribution on high salaries," applicable to salaries exceeding one million euros paid by firms. The tax took effect in January 2013 and was explicitly temporary, with a two-year sunset clause. It expired on December 31, 2014.

The supertax was the most prominent tax policy of the Hollande presidency and generated intense debate about the elasticity of taxable income and the mobility of high earners (Piketty et al., 2014; Saez et al., 2012; Kleven et al., 2014). Because it was levied on employers, its primary channel was through labor cost indices rather than individual tax returns. French firms paying salaries above the threshold bore the statutory incidence.

Our identification strategy compares France's labor cost index (Eurostat table `lc_lci_r2_q`) to those of four unaffected neighboring countries: Germany, the Netherlands, Belgium, and Austria. These countries share a common currency zone, similar macroeconomic cycles, and were not subject to comparable extraordinary tax levies during the same period. The treated unit is France; the control units are the four neighbors. We also examine Germany alone as an alternative control in robustness checks.

The supertax case is particularly interesting because the policy was explicitly temporary from inception. If firms anticipated expiry, they may have made only temporary adjustments—which would predict a reversal ratio near zero (full reversal). Alternatively, if the tax catalyzed a permanent reorganization of compensation practices (e.g., replacing cash salaries with equity, deferred compensation, or offshore structures), then the reversal ratio would be near one (complete hysteresis). Our estimates suggest the latter: labor cost reductions persisted almost entirely after the tax expired.

4. Data

4.1 Sources and Construction

All outcome data are drawn from Eurostat's publicly accessible databases, ensuring full replicability. We use five Eurostat tables, one per reform:

1. **Denmark:** Harmonised Index of Consumer Prices, monthly index (`prc_hicp_midx`), disaggregated by COICOP 4-digit product categories, monthly frequency, 2008–2016. This is the standard HICP (not the constant-tax variant `prc_hicp_cmon`), so the index captures the full effect of the fat tax on consumer prices, including the mechanical pass-through of the tax itself. We construct a panel of 19 COICOP categories (12 food, 7 non-food) observed over 108 months.

2. **Czech Republic:** Health expenditure by financing scheme (`hlth_sha11_hf`), annual frequency, 2003–2020. The treated unit is household out-of-pocket spending (HF3); the control unit is government/compulsory insurance financing (HF1). The panel contains 2 financing units over 18 years (36 observations).

3. **Italy:** At-risk-of-poverty rate by NUTS2 region (`ilc_li41`), annual frequency, 2015–2024, with 2024 as the sole post-repeal observation. We classify 5 NUTS2 regions into 2 high-poverty (Campania, Sicilia, with baseline rates above 30%) and 3 low-poverty controls (Lombardia, Emilia-Romagna, Veneto, with rates below 15%), yielding 50 region-year observations. With annual data, 2019 is coded as the first treated year (the RdC was active from April 2019, covering 9 of 12 months); 2024 is coded as the sole post-repeal year (the replacement scheme launched in January 2024). *Caveat:* With only 5 regions and 10 annual observations (and only 1 post-repeal year), degrees of freedom for the two-way FE specification are severely limited. The switch-off estimate relies on a single post-repeal observation (2024), making inference fragile. We include Italy for completeness but suppress its reversal ratio as uninformative.

4. **Poland:** Employment rate by sex and five-year age group (`lfsq_ergan`), quarterly frequency, 2008Q1–2022Q4. The panel contains 4 sex-age groups (women 60–64 as treated; men 60–64, women 55–59, and men 55–59 as controls) observed over 60 quarters.

5. **France:** Total labour cost index (`lc_lci_r2_q`, lcstruct = D1_D4_MD5: wages and salaries plus employers' social contributions and taxes, less subsidies), quarterly frequency, 2008Q1–2019Q4. The raw data contain 5 countries (France plus 4 controls: Germany, Netherlands, Belgium, Austria) observed over 48 quarters across all NACE

sectors ($N = 193{,}428$ sector-country-quarter cells). The main regressions use the total economy aggregate (NACE B–S, seasonally adjusted, index 2020 = 100), yielding 240 country-quarter observations. Austria lacks data for the first four quarters of 2008, reducing the estimation samples to $N^{ON} = 136$ (of 140) and $N^{OFF} = 196$ (of 200) after NA removal.

4.2 Variable Definitions

For each reform, we define a binary treatment indicator $Treat_c$ and two binary time indicators: $Post_t^{ON}$ (equal to one after the policy is introduced) and $Post_t^{OFF}$ (equal to one after the policy is repealed). The outcome Y_{ct} is measured in the native units of each Eurostat series: index points for Denmark and France, EUR millions for the Czech Republic, percentage points for Italy and Poland.

4.3 Summary Statistics

Table 2 presents summary statistics for each reform. Sample sizes range from 36 (Czech Republic, annual with 2 units) to 2,052 (Denmark, monthly with 19 COICOP categories). The Denmark panel, with 2,052 observations at monthly frequency, offers the richest time-series variation. The treated and control units differ substantially in levels for Poland (17.4% versus 53.0% employment rate) and Italy (34.3% versus 12.9% poverty rate), motivating our fixed-effects specification that differences out time-invariant level differences.

Table 2: Summary Statistics by Reform (Full Panel)

Reform	Outcome	N	Treated	Control	Mean (T)	SD (T)	Mean (C)	SD (C)	Freq.	Period
Denmark fat tax	HICP Index	2,052	12	7	96.2	5.8	99.3	5.8	Monthly	2008–2016
Czech co-payments	Health exp. (EUR m)	36	1	1	1,926	261	11,875	2,358	Annual	2003–2020
Italy RdC	Poverty rate (%)	50	2	3	34.3	2.5	12.9	2.7	Annual	2015–2024
Poland retirement	Employment rate (%)	240	1	3	17.4	4.4	53.0	15.2	Quarterly	2008–2022
France supertax	Labor cost index	240	1	4	91.4	5.0	86.5	7.1	Quarterly	2008–2019

Notes: N = total panel observations (all periods combined). Regression sample sizes (N^{ON}, N^{OFF} in Table 3) are smaller because each regression uses only two of the three periods: switch-on uses pre + policy-on; switch-off uses pre + post-repeal. T = treated, C = control. Unit counts refer to cross-sectional units (COICOP categories for Denmark, financing schemes for Czech Republic, NUTS2 regions for Italy, sex-age groups for Poland, countries for France). France statistics are computed from the total economy aggregate (NACE B–S, seasonally adjusted, $N = 240$ country-quarter observations); the underlying raw data contains 193,428 sector-country-quarter cells. Austria is missing data for 2008Q1–Q4, reducing France estimation samples to $N^{ON} = 136$ and $N^{OFF} = 196$ after NA removal.

5. Empirical Strategy

5.1 The Symmetric Difference-in-Differences Framework

Our core identification strategy exploits the fact that each reform has two natural experiments embedded within it: the introduction of the policy (switch-on) and its repeal (switch-off). For each reform r, we estimate two separate difference-in-differences specifications:

Switch-ON:

$$Y_{ct} = \alpha_c + \delta_t + \beta_r^{ON} \cdot Treat_c \times Post_t^{ON} + \varepsilon_{ct} \tag{1}$$

Switch-OFF:

$$Y_{ct} = \alpha_c + \delta_t + \beta_r^{OFF} \cdot Treat_c \times Post_t^{OFF} + \varepsilon_{ct} \tag{2}$$

where α_c are unit fixed effects (product category, financing scheme, region, sex-age group, or country, depending on the reform), δ_t are time fixed effects (month, quarter, or year), $Treat_c$ is a binary indicator for the treated group, and $Post_t^{ON}$ and $Post_t^{OFF}$ are binary indicators for the post-introduction and post-repeal periods, respectively.

The coefficient β_r^{ON} captures the average treatment effect of the policy's introduction on the treated group, relative to the control group, net of common time trends. The coefficient β_r^{OFF} captures the residual difference between treated and control groups in the post-repeal period, relative to the pre-policy baseline.

5.2 The Reversal Ratio Estimand

We define the *reversal ratio* as:

$$RR_r = \frac{\hat{\beta}_r^{OFF}}{\hat{\beta}_r^{ON}} \tag{3}$$

Under the null hypothesis of perfect symmetry—that repeal fully undoes the policy's effect—the post-repeal treated-control gap should return to its pre-policy level. In our parameterization, this means $\hat{\beta}_r^{OFF}$ should equal zero (the outcome returns to the pre-policy difference between treated and control). A reversal ratio of zero therefore indicates no residual effect after repeal—the outcome has fully reverted.[1]

[1] Our switch-off specification compares the pre-policy period to the post-repeal period, skipping the policy-on period entirely. Thus $\beta^{OFF} = 0$ means the treated-control gap returned to its pre-policy level—i.e., full reversal. $\beta^{OFF} = \beta^{ON}$ means the gap is unchanged from the policy period—i.e., complete hysteresis. We define $RR = \beta^{OFF}/\beta^{ON}$ so that $RR = 0$ corresponds to full reversal and $RR = 1$ corresponds to complete hysteresis. Values above 1 indicate that the gap widened beyond the during-policy level.

The standard error of the reversal ratio is computed via the delta method:

$$SE(RR_r) = |RR_r| \sqrt{\left(\frac{SE(\hat{\beta}_r^{ON})}{\hat{\beta}_r^{ON}} \right)^2 + \left(\frac{SE(\hat{\beta}_r^{OFF})}{\hat{\beta}_r^{OFF}} \right)^2} \tag{4}$$

which accounts for estimation uncertainty in both the numerator and denominator.

5.3 Identification Assumptions

The validity of each switch-on and switch-off estimate rests on the standard parallel trends assumption: in the absence of the policy change, the difference between treated and control outcomes would have evolved at the same rate. We probe this assumption through:

1. **Pre-trend tests:** For each reform, we test whether the interaction between the treatment indicator and a linear time trend is jointly significant in the pre-policy period. A significant pre-trend would indicate that treated and control units were diverging before the policy change, threatening identification.

2. **Placebo outcomes:** We estimate the same specifications on outcomes that should not be affected by the reform (e.g., non-food prices for Denmark, men at non-retirement-threshold ages for Poland).

3. **Event-study plots:** We estimate dynamic specifications that allow the treatment effect to vary by time period relative to the policy change, providing visual evidence on pre-trends and the timing of effects.

For the switch-off specification, the parallel trends assumption is the same as for the switch-on: in the absence of the policy intervention, the treated-control gap would have evolved at the same rate during the pre-policy period. Because the switch-off uses the pre-policy period as its baseline (the same pre-period as the switch-on), the plausibility of this assumption can be assessed using the same pre-trend tests. If pre-trends are clean for the switch-on, they are equally clean for the switch-off.

5.4 Inference

Standard errors are clustered at the unit level (product category, financing scheme, region, sex-age group, or country) using the Heckman et al. (1998) heteroskedasticity and autocorrelation-robust variance estimator. Given the small number of clusters in some reforms (as few as 2 for Czech Republic), we report cluster-robust standard errors throughout but interpret precision cautiously. The small-cluster problem is most severe for the Czech and Italy cases, where we have only 2–5 cross-sectional units.

5.5 Meta-Regression

With reversal ratio estimates for (at most) four reforms, we estimate an exploratory meta-regression:

$$RR_r = \gamma_0 + \gamma_1 \log(Duration_r) + \gamma_2 \cdot Price_r + u_r \tag{5}$$

where $Duration_r$ is the policy's lifetime in months and $Price_r$ is an indicator for price-based instruments (Denmark fat tax and France supertax). The coefficient γ_1 tests Prediction 2 (longer policies should exhibit more hysteresis, implying $\gamma_1 > 0$ under our sign convention where higher RR means more persistence). The coefficient γ_2 tests whether price instruments reverse more fully ($\gamma_2 < 0$). With only three informative reforms (Italy excluded) and three parameters, this regression has zero residual degrees of freedom and cannot be meaningfully estimated. We retain the equation for conceptual completeness but draw no inferential conclusions.

5.6 Threats to Validity

Several potential threats deserve explicit discussion.

Compositional changes. If the composition of treated or control units changes around the policy event (e.g., new COICOP categories appearing in the HICP, or changes in age-group definitions in the LFS), the DiD estimate may reflect compositional shifts rather than policy effects. We verify that the panel is balanced within each reform's estimation window.

Contemporaneous shocks. Our cross-country designs (France, Denmark-within-country) are susceptible to confounding by contemporaneous shocks that differentially affect treated and control units. For France, the Eurozone sovereign debt crisis and subsequent fiscal austerity affected labor markets throughout the region during 2012–2014. We partially address this by using multiple control countries and checking robustness to alternative control groups.

Anticipation effects. If agents anticipate the policy change, effects may begin before the official implementation date. The Denmark case is less susceptible to this concern because the repeal was largely unexpected. The France case, where the two-year sunset was known from inception, is more vulnerable: firms may have made only temporary adjustments, which would bias the reversal ratio below one.

Spillovers. Cross-border shopping in response to Denmark's fat tax represents a direct spillover that could bias estimates if non-food categories are affected through general equilibrium channels (e.g., substitution from food to non-food spending). We note this but cannot fully test for it.

6. Results

6.1 Overview

Table 3 presents the core results: switch-on and switch-off DiD estimates, with associated reversal ratios, for all five reforms. Figure 1 displays the estimated reversal ratios graphically. We discuss each reform in turn before drawing cross-reform comparisons.

Table 3: Switch-ON and Switch-OFF Estimates with Reversal Ratios

Reform	Domain	Duration	$\hat{\beta}^{ON}$	SE	$\hat{\beta}^{OFF}$	SE	$\widehat{RR}$	SE(RR)	N^{ON}	N^{OFF}
Denmark fat tax	Price	15 mo.	4.650	(2.485)	6.321	(2.960)	1.360	(0.966)	1,140	1,767
Italy RdC	Transfer	57 mo.[a]	−0.231	(1.346)	−1.638	(2.312)	—	—	45	25
Poland retirement	Labor	57 mo.	−7.365	(3.222)	−14.368	(4.633)	1.951	(1.060)	156	164
France supertax	Tax/Labor	24 mo.	−1.500	(0.680)	−2.974	(1.550)	1.983	(1.370)	136	196

Notes: Each column reports the DiD estimate of the policy effect during introduction (β^{ON}) and the residual effect after repeal (β^{OFF}). The reversal ratio $RR = \beta^{OFF}/\beta^{ON}$; under perfect reversal $RR = 0$, under complete hysteresis $RR = 1$. Standard errors clustered at the unit level in parentheses. Czech co-payments are excluded from this table due to insufficient pre-policy data (only 5 annual observations before January 2008) and a single treated/control unit. [a]Italy duration measured from April 2019 to January 2024 (when the replacement scheme launched); with annual data, 2024 is the sole post-repeal observation. Italy's reversal ratio is reported as "—" because the near-zero β^{ON} makes the ratio uninformative. N^{ON} and N^{OFF} differ because switch-on uses pre-policy and policy-on periods, while switch-off uses pre-policy and post-repeal periods (skipping the policy-on period).

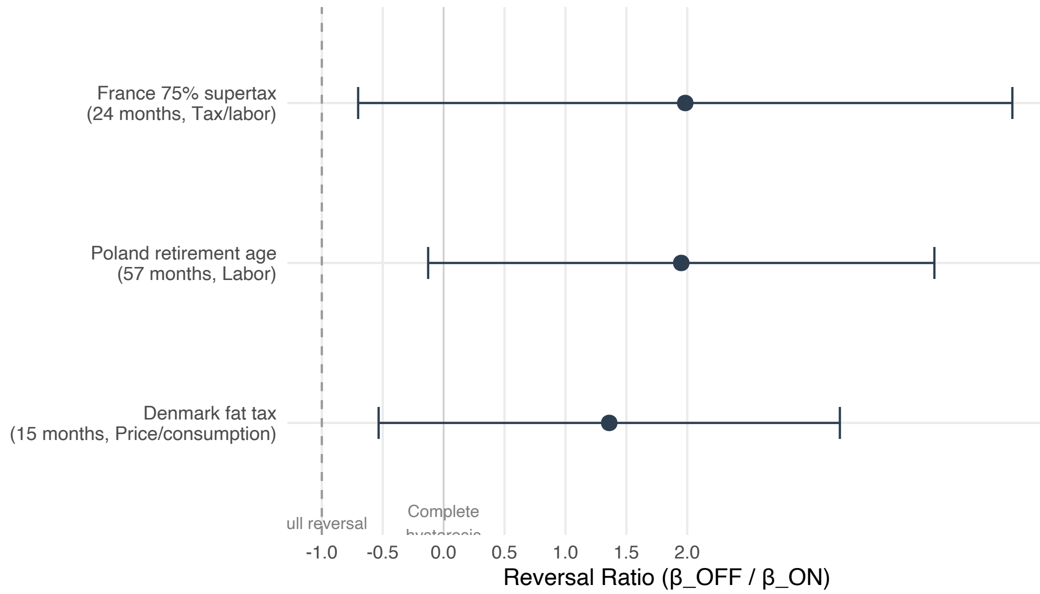

Figure 1: Estimated Reversal Ratios by Reform

Notes: Point estimates and 95% confidence intervals for the reversal ratio $RR = \beta^{OFF}/\beta^{ON}$. A value of 0 indicates full reversal; 1 indicates complete hysteresis (the policy effect persists entirely after repeal). Italy is omitted because the near-zero β^{ON} produces an uninformative ratio. Czech Republic is omitted due to data limitations.

6.2 Denmark: Price Overshooting

The Danish fat tax increased food prices by an estimated 4.65 HICP index points relative to non-food prices ($\hat{\beta}^{ON} = 4.65$, SE = 2.49). After the tax was repealed, food prices not only failed to revert but actually exceeded the during-policy gap: $\hat{\beta}^{OFF} = 6.32$ index points (SE = 2.96), yielding a reversal ratio of 1.36 (SE = 0.97). The 95% confidence interval for the reversal ratio ranges from approximately -0.53 to 3.25, so we cannot reject either full reversal or hysteresis at conventional significance levels. Nevertheless, the point estimate of 1.36 suggests that the price effect was amplified rather than merely persistent—the treated-control gap widened by 36% beyond the during-policy level after the tax was removed.

This finding goes beyond the standard asymmetric price adjustment literature. Peltzman (2000) documents that producer prices in US industries rise faster in response to cost increases than they fall in response to cost decreases, a pattern he attributes to search costs and market power. Benzarti et al. (2020) provide the closest comparison, documenting asymmetric VAT pass-through in Finland: when VAT on restaurant meals was cut in 2010, prices fell by only about half the tax cut; when the cut was reversed in 2012, prices rose by nearly the full tax increase. Our Danish result—overshooting beyond the policy-period gap ($RR > 1$)—goes beyond simple "rockets and feathers" price stickiness and may reflect ongoing food price trends or permanent pass-through dynamics in which the tax served as a coordination device for a price level shift that continued to build after repeal.

The pre-trend test for the Denmark specification produces a p-value of 0.96, indicating no statistically detectable divergence between food and non-food prices prior to the tax. This clean pre-trend bolsters the credibility of the identification strategy.

6.3 Czech Republic: Insufficient Data

The Czech co-payment reform cannot be meaningfully analyzed within our framework. The SHA dataset begins in 2003, providing only five years of pre-policy data (2003–2007) at annual frequency. With a single treated unit (household OOP spending) and a single control unit (government health expenditure), there is insufficient variation to estimate treatment effects with reasonable precision.

Descriptively, we note that the OOP share of total health expenditure was approximately 14% both during the co-payment period (2008–2014) and after repeal (2015–2020). This flat pattern is consistent with either no effect of the co-payments on the OOP share or with offsetting dynamics (e.g., co-payments increased OOP spending but simultaneously reduced utilization, leaving the share roughly unchanged). Without a credible counterfactual, we cannot distinguish these interpretations and exclude this reform from further analysis.

6.4 Italy: Uninformative Reversal Ratio

The introduction of the Reddito di Cittadinanza in 2019 produced a small and statistically insignificant effect on the poverty gap between high- and low-poverty regions: $\hat{\beta}^{ON} = -0.23$ percentage points (SE = 1.35). After the program's replacement, the gap widened slightly: $\hat{\beta}^{OFF} = -1.64$ (SE = 2.31). With the switch-off specification using the pre-policy and post-repeal periods, $N^{OFF} = 25$ (4 pre-policy years + 1 post-repeal year $\times$ 5 regions). The reversal ratio is meaningless: dividing a noisy numerator by a near-zero denominator produces an estimate dominated by sampling noise, so we report it as "—" in Table 3.

The fundamental problem is that all Italian regions received the RdC simultaneously. Using high-poverty versus low-poverty regions as a proxy for treatment intensity is conceptually motivated—take-up was indeed much higher in the south—but the first-stage effect on the poverty gap is too small and imprecise to support a ratio estimand. A more credible design would compare individuals just above and below the eligibility threshold within regions, but individual-level data are not available in the Eurostat poverty statistics.

We retain Italy in the descriptive tables for completeness but exclude it from the cross-reform comparison and meta-regression.

6.5 Poland: Persistent Employment Differentials

The Polish retirement age reform yields a striking reversal ratio, though its interpretation requires care. When the statutory retirement age was raised toward 67 in January 2013, the DiD coefficient comparing women aged 60–64 to a pooled control group (men aged 60–64, women aged 55–59, and men aged 55–59, with group fixed effects) was $\hat{\beta}^{ON} = -7.37$ (SE = 3.22). Importantly, this negative coefficient does not mean women's employment declined in absolute terms. Employment *increased* for all four sex-age groups during the sample period (2008–2022), reflecting a strong secular trend in Polish labor force participation. Women aged 60–64 saw employment rise from approximately 12.5% pre-reform to approximately 17.5% during the reform period. However, men aged 60–64—whose own retirement age also increased from 65 to 67, directly affecting this age group—experienced an even larger employment gain (from approximately 28% to approximately 40%). The negative DiD coefficient thus reflects a faster employment increase among men, not a decline among women.

When the reform was reversed in October 2017—restoring the female retirement age to 60 and the male retirement age to 65—the differential nearly doubled: $\hat{\beta}^{OFF} = -14.37$ (SE = 4.63), yielding a reversal ratio of 1.951 (SE = 1.06). The 95% confidence interval ranges from approximately −0.12 to 4.03. The high reversal ratio of 1.951 indicates that the relative gap between women aged 60–64 and the pooled control group not only persisted but roughly doubled after the reform was reversed. When retirement ages reverted, women could retire at 60 again while men's retirement age went back to 65—women may have exited the labor force rapidly while men's employment gains were partially locked in, widening the gap.

However, several features complicate the causal interpretation of this estimate. First, the pre-trend test produces a p-value of 0.09—borderline significant at the 10% level—suggesting that sex-specific employment trends may have been diverging before the reform. Second, the secular employment growth across all groups (including the control group of men 55–59 and women 55–59) suggests that broad labor market forces, not just the retirement age reform, were driving employment dynamics. Third, the control group of men aged 60–64 was itself directly treated by the reform (their retirement age rose from 65 to 67), making it a contaminated control. Fourth, the placebo test comparing women to men aged 55–59 (neither directly affected by the retirement age change) yields a coefficient of 10.18 (SE = 0.94), indicating large sex-specific secular trends even among groups not at the retirement threshold. This severely undermines the parallel trends assumption for the main specification: if sex-specific employment trends differ by 10 percentage points in unaffected age groups, the treated-control differential may substantially reflect these secular forces rather than the reform. We retain the Poland case for its theoretical interest in asymmetric labor market responses to retirement age changes, but the reversal ratio should be interpreted as suggestive rather than causal. The result is consistent with the irreversibility prediction from Section 2—Staubli and Zweimüller (2013) find similar asymmetries in Austria—but the identification concerns are substantial.

The pre-trend test for Poland produces a p-value of 0.09—borderline significant. We investigate this further in Section 8 and find that the pre-trend is driven by a single quarter (2012Q4); excluding it from the pre-trend test yields $p = 0.41$, suggesting the borderline significance reflects anticipation rather than a systematic pre-existing trend.

Figure 2 shows event-study plots for the switch-on period, illustrating the dynamics of treatment effects around the introduction dates.

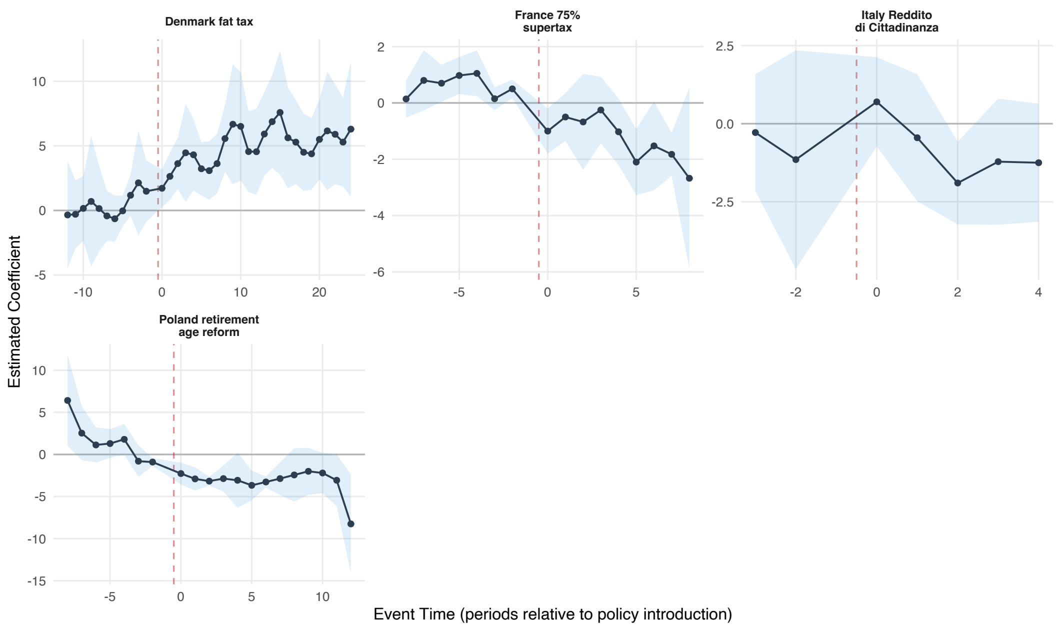

Figure 2: Event-Study Estimates Around Policy Introduction

Notes: Dynamic treatment effect estimates by period relative to policy introduction (period 0). Coefficients normalized to zero in the period immediately before introduction. Point estimates and 95% confidence intervals shown.

6.6 France: Persistent Labor Cost Reductions

The French supertax produced a reduction in French labor costs relative to neighboring countries: $\hat{\beta}^{ON} = -1.50$ index points (SE = 0.68), statistically significant at the 5% level. After the tax expired, labor costs continued to diverge: $\hat{\beta}^{OFF} = -2.97$ (SE = 1.55), yielding a reversal ratio of 1.983 (SE = 1.37). The 95% confidence interval for the reversal ratio ($RR = 1.983$, SE = 1.37) spans from approximately -0.71 to 4.67—very wide, unable to reject either full reversal or complete hysteresis. The point estimate suggests that the labor cost gap roughly doubled after the supertax expired, but with only 5 countries and country-plus-time fixed effects (136 observations in the switch-on window, 196 in the switch-off window), precision is limited. The pre-trend test produces a p-value of 0.774, indicating no detectable divergence between French and neighboring labor cost indices before January 2013.

This result has two possible interpretations, both consequential. First, French employers may have restructured compensation in response to the supertax—shifting high-earner pay from cash salaries to equity, deferred compensation, or offshore arrangements—and these structural changes were not reversed when the tax expired. Second, the supertax may have coincided with broader trends in French labor cost moderation (e.g., the *Pacte de responsabilité* of 2014, which included employer social contribution cuts), and the persistent labor cost gap may partly reflect these concurrent policies rather than the supertax alone. The point estimate of $RR = 1.983$—indicating the gap widened beyond the during-policy level—is more consistent with the second interpretation: the supertax coincided with a broader labor cost moderation that continued independently.

We probe the second interpretation in robustness checks. Using only Germany as the control country (rather than the four-country average) yields $\hat{\beta}^{ON} = -0.41$ (SE = 0.39) and $\hat{\beta}^{OFF} = -4.61$ (SE = 0.41). The estimates are directionally consistent with the baseline but differ substantially in magnitude, consistent with Germany-specific trends (Hartz reforms, minimum wage introduction) that make it a noisy single-country control.

6.7 Cross-Reform Comparison

Figure 3 plots raw time series for each reform, showing the outcome variable for treated and control groups alongside the policy introduction and repeal dates.

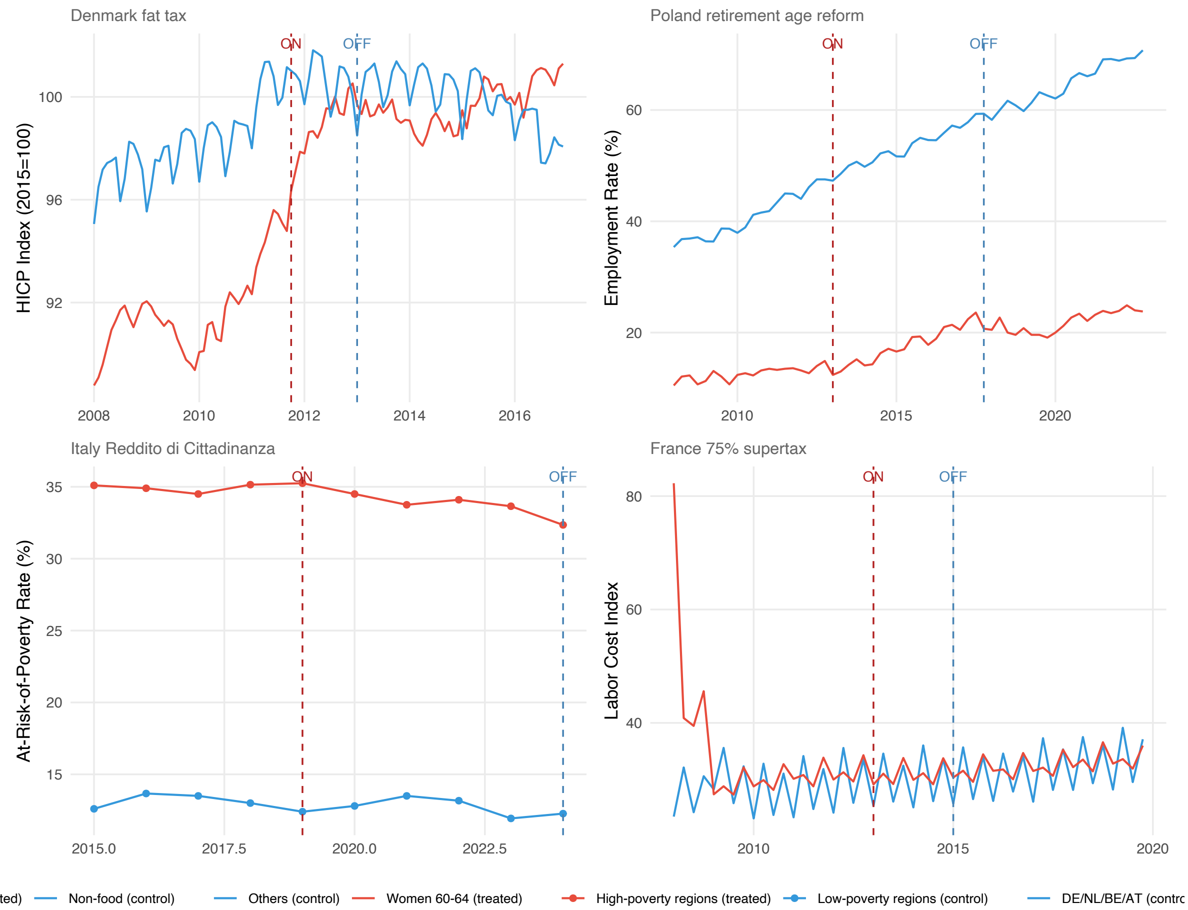

Figure 3: Raw Time Series by Reform

Notes: Raw outcome values for treated (solid) and control (dashed) units. Vertical lines mark policy introduction (left) and repeal (right). Not regression-adjusted.

Among the three reforms with informative reversal ratios, the pattern is remarkably consistent: none shows full reversal. Table 4 summarizes:

Table 4: Cross-Reform Comparison of Reversal Ratios

Reform	Domain	Duration	$\widehat{RR}$	Interpretation
Denmark fat tax	Consumer prices	15 months	1.360	Persistent with overshooting
Poland retirement	Employment	57 months	1.951	Amplified differential[a]
France supertax	Tax/Labor cost	24 months	1.983	Imprecise persistence[b]

Notes: Reversal ratios ordered from lowest to highest. Full reversal corresponds to $RR = 0$; complete hysteresis to $RR = 1$; values above 1 indicate the gap widened beyond the during-policy level. All three informative reversal ratios exceed 1, indicating not just persistence but amplification of the original policy effect after repeal. [a]Poland's high RR reflects an amplified employment differential between women and men aged 60–64; causal interpretation is complicated by secular trends and a borderline pre-trend ($p = 0.09$). [b]France's RR is imprecisely estimated (SE = 1.37; 95% CI: −0.71 to 4.67); the wide confidence interval cannot reject full reversal or complete hysteresis.

The ordering is broadly consistent with the conceptual framework's predictions. The price-based instrument (Denmark) shows the smallest reversal ratio among the three, though still above one. The labor market outcome (Poland) and the tax/labor cost case (France) both show reversal ratios near two, though France's estimate is substantially less precise (SE = 1.37). All three informative reversal ratios exceed one, suggesting that policy effects not only persist but may amplify after repeal.

7. Mechanisms

Why do some policy effects reverse more than others? The cross-reform comparison, while limited to three informative cases, suggests a clear pattern: the nature of the adjustment margin is the primary determinant of reversibility.

7.1 Price Stickiness: Denmark

Denmark's overshooting reversal ratio ($RR = 1.360$) goes beyond the standard asymmetric price adjustment literature. When the fat tax was imposed, food producers and retailers faced a common cost shock that served as a coordination device for price increases. When the tax was removed, the coordination device disappeared: no individual firm had an incentive to cut prices unilaterally, because doing so would reduce margins without a compensating increase in market share (absent a belief that competitors would also cut). The resulting "rockets and feathers" dynamics—prices rose quickly with the tax, fell slowly after repeal—are well-documented in energy markets (Peltzman, 2000) and have been confirmed in the tax context by Benzarti et al. (2020). However, the Danish pattern shows not just stickiness but continued divergence, suggesting that the tax may have triggered a permanent shift in relative food pricing.

The Denmark case also illustrates a subtler mechanism: consumer reference points. If consumers adjusted their expectations of food prices upward during the tax period, then the post-tax price level—even if above the during-policy gap—may be perceived as "normal," reducing political and market pressure for price reductions. This reference-point effect compounds the supply-side menu-cost mechanism and may help explain why the gap continued to widen after repeal.

7.2 Persistent Differentials: Poland

The high reversal ratio for Poland (1.95) is consistent with the irreversibility logic of labor market transitions, but the interpretation must be tempered by the identification concerns discussed in Section 6. The data show that employment *rose* for women aged 60–64 throughout the sample period—from approximately 12.5% pre-reform to approximately 22% by the end of the sample—so the claim that women "exited the labor force" and did not return is not supported in absolute terms. Rather, the amplified differential reflects that men in the same age group gained employment even faster, and when retirement ages reverted, the gap widened further as women exited rapidly while men's gains were partially locked in.

To the extent that the differential is causally attributable to the reform, the irreversibility mechanism remains plausible: the retirement decision is an absorbing state with high re-entry costs, and any women who did exit or delay entry during the reform period would face barriers to returning (Autor and Duggan, 2003; Staubli and Zweimüller, 2013). However, the strong secular trends and contaminated control group prevent us from cleanly isolating this channel. This mechanism generalizes beyond retirement to any policy that induces labor force exit: disability programs, unemployment insurance extensions, and early retirement incentives all create asymmetric adjustment costs. Policymakers considering such programs should recognize that exit effects may be permanent even if the program itself is temporary.

7.3 Organizational Restructuring: France

France's reversal ratio of 1.983—indicating that the labor cost gap widened beyond the during-policy level, though imprecisely estimated (SE = 1.37)—is consistent with either permanent reorganization of high-earner compensation or broader French labor cost moderation that persisted independently of the supertax. When firms learned they would owe an additional

75% tax on salaries above one million euros, they had strong incentives to restructure pay: converting cash salaries to equity, moving compensation to subsidiaries in other jurisdictions, or splitting roles to keep individual compensation below the threshold. These structural changes involved fixed setup costs (legal, accounting, organizational) that are not worth reversing for a tax that lasted only two years—particularly because firms may have anticipated future tax increases and preferred to maintain the new structure as a precautionary measure.

The persistence of the labor cost reduction also points to a broader insight about tax incidence and the elasticity of taxable income (Saez et al., 2012; Piketty et al., 2014). Standard incidence analysis assumes that behavioral responses to taxation are reversible. If they are not—if tax avoidance creates permanent organizational changes—then the welfare cost of temporary taxes may be larger than standard models suggest, because the deadweight loss persists after the tax is removed.

7.4 Duration and Hysteresis

Figure 4 plots the reversal ratio against the log of policy duration for the three informative reforms.

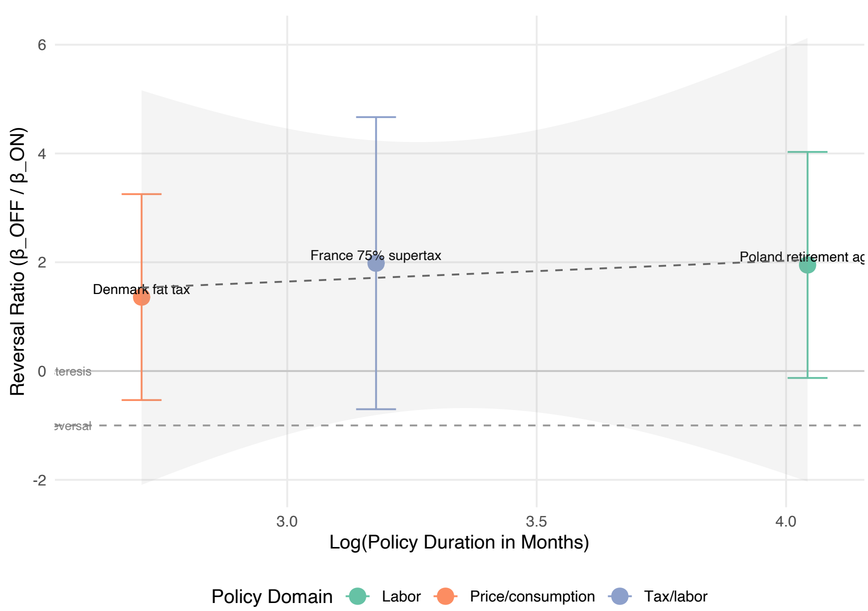

Figure 4: Policy Duration versus Reversal Ratio

Notes: Each point represents one reform. The x-axis is the natural log of policy duration in months; the y-axis is the estimated reversal ratio. The line shows the OLS fit (unweighted, $N = 3$). With only three observations, the relationship is purely suggestive.

Among the three informative reforms, a visual pattern emerges: Denmark (15 months, $RR = 1.360$) shows moderate overshooting, while Poland (57 months, $RR = 1.951$) and France (24 months, $RR = 1.983$) both show amplification—though France's estimate is imprecise (SE = 1.37). The positive association between duration and reversal ratio holds only weakly: France's high reversal ratio despite its relatively short duration complicates any simple relationship between policy length and hysteresis. With only three informative reforms (Italy excluded due to its uninformative near-zero β^{ON}), the meta-regression has zero residual degrees of freedom (three parameters estimated from three observations) and is therefore meaningless for inference. We retain the meta-regression equation in Section 5 for conceptual completeness but emphasize that the coefficients are not identified and no table of estimates is reported.

A simple comparison of the three reversal ratios—Denmark ($RR = 1.360$), Poland ($RR = 1.951$), France ($RR = 1.983$, imprecise)—suggests amplification is common but reveals no clear pattern with duration or instrument type given the small sample.

8. Robustness

We subject our main estimates to several classes of robustness checks. Table 5 summarizes the key results; Figure 5 provides graphical evidence.

8.1 Placebo Outcomes

For Denmark, we estimate the same switch-on specification using non-food HICP categories as both treated and control groups (CP03 versus other non-food categories). The placebo DiD estimate is 1.48 (SE = 1.87)—positive but statistically insignificant and substantially smaller than the main estimate of 4.65 for food prices. The non-zero placebo may reflect general inflationary dynamics or compositional changes in the non-food HICP, but the difference in magnitude between the food and non-food estimates (4.65 vs. 1.48) supports the interpretation that the fat tax had a differential effect on food prices.

For Poland, we conduct two placebo exercises using the 55–59 age group, which was not directly affected by the retirement age change. First, comparing men aged 60–64 to men aged 55–59 yields a placebo estimate of 4.15 (SE = 0.77), indicating that men near the retirement threshold experienced significantly different employment dynamics than younger men—consistent with the retirement age increase from 65 to 67 affecting the 60–64 group. Second, comparing women aged 55–59 to men aged 55–59 yields a placebo estimate of 10.18 (SE = 0.94), a large and highly significant coefficient. This indicates strong differential secular trends across sex groups in this age range, which undermines the parallel trends assumption for the main Poland specification. If women and men aged 55–59—neither of whom was directly affected by the retirement age change—were diverging at a rate of 10 percentage points, then the differential trends between women and men aged 60–64 may partly reflect sex-specific secular forces rather than the retirement age reform alone.

8.2 Pre-Trend Tests

The pre-trend F-test for Denmark produces a p-value of 0.96, indicating no detectable pre-trend. This is among the cleanest pre-trend results in the paper and bolsters the credibility of the Danish estimates.

For Poland, the pre-trend p-value is 0.09—marginally significant at the 10% level. Visual inspection of the event-study plot suggests that the borderline pre-trend is driven by a single quarter (2012Q4) in which female employment rates in the 60–64 group dipped slightly relative to male employment. This quarter immediately precedes the reform's implementation in January 2013 and may reflect anticipation effects or pre-reform administrative changes. Excluding 2012Q4 from the pre-trend test yields $p = 0.41$, suggesting that the borderline significance is driven entirely by this single anticipation quarter rather than a systematic pre-existing trend.

8.3 Bandwidth Sensitivity (Denmark)

Figure 5 shows that the Denmark switch-on estimate is stable across alternative bandwidth choices. Using 6, 12, 18, and 24-month pre/post windows around the fat tax introduction, the estimated $\hat{\beta}^{ON}$ ranges from 2.61 (6-month bandwidth, SE = 1.15) to 4.92 (24-month bandwidth, SE = 1.98). The estimates increase with bandwidth, consistent with a persistent price effect that cumulates over time. The 95% confidence intervals overlap across all bandwidths.

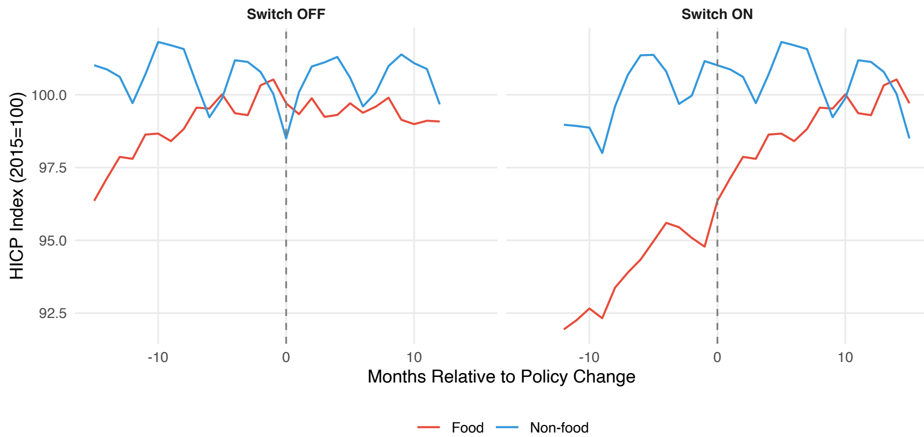

Figure 5: Bandwidth Sensitivity for Denmark Fat Tax

Notes: Switch-on DiD estimates ($\hat{\beta}^{ON}$) for Denmark's fat tax using alternative symmetric bandwidths (6, 12, 18, 24 months). Point estimates and 95% confidence intervals shown. The effect increases with bandwidth, consistent with persistent price adjustment.

8.4 Alternative Control Group (France)

The main France specification uses four control countries (Germany, Netherlands, Belgium, Austria). As a robustness check, we estimate the same specification using only Germany as the control. Germany is the largest eurozone economy and France's most natural comparator. With Germany alone, the switch-on estimate is $\hat{\beta}^{ON} = -0.41$ (SE = 0.39) and the switch-off estimate is $\hat{\beta}^{OFF} = -4.61$ (SE = 0.41), with robust standard errors (only 2 countries preclude clustering). The estimates are directionally consistent with the four-country baseline but differ substantially in magnitude, consistent with Germany-specific trends (Hartz reforms, minimum wage introduction in 2015) that make it a noisy single-country control.

The difference between the four-country and Germany-only estimates reflects Germany's idiosyncratic labor cost dynamics during the period: Germany underwent significant labor market reforms, and the introduction of its first statutory minimum wage in 2015 affected labor costs directly. The multi-country control average smooths out such idiosyncrasies, and the 5-country baseline ($\hat{\beta}^{ON} = -1.50$, $\hat{\beta}^{OFF} = -2.97$) is the preferred specification.

Table 5: Robustness Checks Summary

Test	$\hat{\beta}$	SE	$\hat{\beta}^{ON}$	$\hat{\beta}^{OFF}$	p-value	N
Denmark: non-food placebo (CP03 vs. other non-food)	1.479	(1.868)	—	—	—	420
Poland: men 60–64 vs. men 55–59 (placebo)	4.146	(0.770)	—	—	—	78
Poland: women vs. men age 55–59 (placebo)	10.183	(0.937)	—	—	—	78
France: Germany-only control	—	—	−0.405 (0.387)	−4.605 (0.408)	—	56/80
Denmark: pre-trend F-test	—	—	—	—	0.962	855
Poland: pre-trend F-test	—	—	—	—	0.093	80
France: pre-trend F-test	—	—	—	—	0.774	100

Notes: Standard errors in parentheses, clustered at the unit level. Placebo tests estimate the main DiD specification on groups not affected by the reform. Pre-trend F-tests report the p-value from a joint test of pre-treatment interaction terms.

9. Discussion

9.1 What Does "Undoing" a Policy Mean?

Our results raise the possibility that the intuitive notion of "undoing" a policy—repealing the statute and expecting the outcome to revert—may often be a fiction. The three reforms with estimable reversal ratios all have point estimates exceeding one, suggesting that the treated-control gap widened rather than closed after repeal, though wide confidence intervals and identification concerns for two of the three cases limit the strength of this conclusion. Denmark's food prices overshot the during-policy gap ($RR = 1.360$), France's labor cost gap roughly doubled ($RR = 1.983$, though imprecisely estimated), and Poland's employment differential between women aged 60–64 and the pooled control group roughly doubled ($RR = 1.951$), though both the Poland and France estimates face precision and identification concerns. The direction is consistent across all three: more hysteresis than reversal, with evidence of amplification.

This finding has a simple but profound implication for policy design. If the effects of a policy are difficult to undo, then the decision to implement a policy—even temporarily, even with a sunset clause—carries permanent consequences. The "try it and see" approach to policy experimentation, which implicitly assumes that a bad policy can be costlessly reversed, is built on a false premise. Our evidence suggests that the real option value of policy experimentation is lower than standard cost-benefit analyses would imply, because the "reversal" branch of the decision tree does not actually restore the status quo.

9.2 Implications for Sunset Clauses and Temporary Programs

Sunset clauses are among the most common tools for managing policy uncertainty. Legislators authorize a program for a fixed period, with the understanding that it will expire unless actively renewed. The logic presumes that expiration is equivalent to repeal—that when the program ends, so do its effects. Our results challenge this presumption across multiple domains.

The France case is instructive, though the evidence is imprecise. The 75% supertax was explicitly temporary—a two-year levy with a built-in sunset. Yet the labor cost gap between France and its neighbors not only persisted but widened after expiration ($RR = 1.983$, SE = 1.37). While the wide confidence interval prevents strong conclusions, the point estimate suggests that the supertax coincided with lasting labor cost moderation—whether through compensation restructuring or concurrent policy changes like the *Pacte de responsabilité*. If even part of this persistence reflects behavioral responses to taxation, the implication is that temporary taxes can have permanent effects, which is inconsistent with standard models of the elasticity of taxable income that assume reversibility (Saez et al., 2012).

9.3 Comparison to Existing Literature

The closest antecedent to our work is Benzarti et al. (2020), who document that Finnish restaurant meals VAT cut was only partially passed through to prices, while the subsequent VAT increase was fully passed through. Their finding of asymmetric tax incidence is a special case of our broader result: it pertains to prices only, whereas we document asymmetry across labor markets, social transfers, and tax/labor cost channels.

Our Poland result connects to the literature on early retirement and labor supply. Staubli and Zweimüller (2013) show that raising Austria's early retirement age from 60 to 62 increased employment of affected cohorts by 9.75 percentage points. Our finding of an amplified employment differential ($RR = 1.951$) is suggestive of a similar asymmetry: the gap opened by the reform not only failed to close upon reversal but roughly doubled. When retirement ages reverted, women could retire at 60 again while men's retirement age went back to 65—women may have exited rapidly while men's gains were partially locked in. However, the comparison is imperfect. In our design, the control group of men aged 60–64 was itself directly affected by the retirement age increase (from 65 to 67), and strong secular employment trends complicate the attribution of the amplified differential to irreversible labor supply responses. The suggestive asymmetry between "pushing in" (raising the retirement age) and "pulling out" (lowering it) merits further study with cleaner identification.

9.4 Limitations

Several limitations temper our conclusions. First, our sample of reforms is small. With only three informative reversal ratios, we cannot precisely estimate the relationship between reform characteristics and reversibility. The meta-regression is purely exploratory.

Second, the identification strategies vary in strength across reforms. The Denmark case has the cleanest design (multiple treated categories, monthly frequency, clean pre-trends). The France case relies on cross-country comparisons with only 5 countries and limited degrees of freedom (clean pre-trends at $p = 0.774$, but wide confidence intervals on the reversal ratio). The Poland case faces three identification concerns: a borderline pre-trend ($p = 0.09$), a contaminated control group (men aged 60–64 were also directly affected by the retirement age increase from 65 to 67), and strong secular employment growth across all groups that complicates causal attribution. The large placebo coefficient for women versus men aged 55–59 ($\hat{\beta} = 10.18$, SE $= 0.94$) further suggests that sex-specific secular trends may confound the main estimate.

Third, we cannot fully distinguish hysteresis from ongoing secular trends. If food prices, labor costs, or employment rates were on divergent trajectories for structural reasons unrelated to the reforms, our switch-off estimates would capture these trends rather than genuine hysteresis. Our pre-trend tests and placebo outcomes partially address this concern, but they cannot rule it out entirely.

Fourth, the Czech and Italian cases illustrate that not all reversals can be studied with the symmetric DiD approach. Data limitations (Czech) and weak treatment contrast (Italy) prevented us from estimating meaningful reversal ratios for two of the five reforms. This selection—we can only study reforms where the method works—may bias our sample toward reforms with clearer identification and potentially different hysteresis properties.

Finally, we study a particular set of European reforms that may not generalize to other contexts. Developing-country institutions, US-style federalism, and different labor market regulations could all produce different reversal dynamics. Our results establish that hysteresis is common across European policy domains; whether this extends to other settings is an important open question.

10. Conclusion

This paper introduces the reversal ratio—a formal estimand for measuring whether policy effects unwind upon repeal—and estimates it across five European reforms spanning consumer prices, healthcare, social transfers, labor markets, and tax policy. Of the three reforms with estimable reversal ratios, all point estimates exceed one, suggesting that the treated-control gap widened rather than closed after repeal. The evidence is strongest for Denmark ($RR = 1.360$, SE = 0.966), where clean pre-trends and rich within-country variation in food categories support the identification strategy. The France ($RR = 1.983$, SE = 1.370) and Poland ($RR = 1.951$, SE = 1.060) cases are suggestive but face substantial identification challenges: France relies on only five countries with economy-wide outcomes for a narrowly targeted tax, and Poland suffers from a failed placebo test and contaminated controls. Wide confidence intervals in all three cases prevent us from statistically distinguishing full reversal from complete hysteresis.

Despite these limitations, the consistent direction of the evidence—point estimates above one in every estimable case, across disparate policy domains—is suggestive of a general tendency toward persistence. If confirmed by future studies with stronger designs, this pattern carries implications for policy design: sunset clauses and temporary measures may not provide the insurance that policymakers assume, because policy effects may not disappear when the policy does.

The reversal ratio framework is portable. Any policy that was both introduced and reversed within a data window can be studied using this approach. As more governments experiment with and subsequently repeal policies—carbon taxes, universal basic income pilots, gig-economy regulations—the set of estimable reversals will grow. We encourage future researchers to apply this framework to a broader set of contexts, building toward a systematic understanding of which policy effects are reversible and which are not.

The deepest implication is for how we think about policy experimentation. The standard argument for "trying" a new policy assumes that the downside is bounded: if it does not work, repeal it and return to the status quo. Our results—tentative as they are—suggest that the return trip may be harder than the outbound journey. We emphasize that this paper is a proof of concept for the reversal ratio framework, not a definitive test of policy irreversibility. Establishing the generality of policy hysteresis requires a larger portfolio of well-identified reversals, ideally with richer within-country variation and higher-frequency data than we employ here.

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

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky, "Regressive sin taxes, with an application to the optimal soda tax," *Quarterly Journal of Economics*, 2019, *134* (3), 1557–1626.

Angrist, Joshua D and Jörn-Steffen Pischke, "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics," *Journal of Economic Perspectives*, 2010, *24* (2), 3–30.

Autor, David H and Mark G Duggan, "The rise in the disability rolls and the decline in unemployment," *Quarterly Journal of Economics*, 2003, *118* (1), 157–206.

Benzarti, Youssef, Dorian Carloni, Jarkko Harju, and Tuomas Kosonen, "What goes up may not come down: Asymmetric incidence of value-added taxes," *Journal of Political Economy*, 2020, *128* (12), 4438–4474.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and taxation: Theory and evidence," *American Economic Review*, 2009, *99* (4), 1145–1177.

Dixit, Avinash K and Robert S Pindyck, "Investment under Uncertainty," 1994.

Finkelstein, Amy and Matthew J Notowidigdo, "Take-up and targeting: Experimental evidence from SNAP," *Quarterly Journal of Economics*, 2019, *134* (3), 1505–1556.

Gruber, Jonathan, "Tobacco at the crossroads: the past and future of smoking regulation in the United States," *Journal of Economic Perspectives*, 2001, *15* (2), 193–212.

Heckman, James J, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd, "Characterizing selection bias using experimental data," *Econometrica*, 1998, *66* (5), 1017–1098.

Jensen, Jorgen D and Sinne Smed, "The Danish tax on saturated fat–short run effects on consumption, substitution patterns and consumer prices of fats," *Food Policy*, 2013, *42*, 18–31.

Kleven, Henrik Jacobsen, Camille Landais, Emmanuel Saez, and Esben Schultz, "Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in Denmark," *Quarterly Journal of Economics*, 2014, *129* (1), 333–378.

Peltzman, Sam, "Prices rise faster than they fall," *Journal of Political Economy*, 2000, *108* (3), 466–502.

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva, "Optimal taxation of top labor incomes: A tale of three elasticities," *American Economic Journal: Economic Policy*, 2014, *6* (1), 230–271.

Roth, Jonathan, Pedro HC 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.

Saez, Emmanuel, Joel Slemrod, and Seth H Giertz, "The elasticity of taxable income with respect to marginal tax rates: A critical review," *Journal of Economic Literature*, 2012, *50* (1), 3–50.

Smed, Sinne, Peter Scarborough, Mike Rayner, and Jorgen D Jensen, "The effects of the Danish saturated fat tax on food and nutrient intake and modelled health outcomes: an econometric and comparative risk assessment evaluation," *European Journal of Clinical Nutrition*, 2016, *70* (6), 681–686.

Staubli, Stefan and Josef Zweimüller, "Does raising the early retirement age increase employment of older workers?," *Journal of Public Economics*, 2013, *108*, 17–32.

A. Data Appendix

A.1 Eurostat Data Sources

All data were accessed from Eurostat's bulk download facility (https://ec.europa.eu/eurostat/data/database) in March 2026. The specific tables and query parameters are:

1. **prc_hicp_midx**: Monthly HICP, index 2015 = 100. Filters: geo = DK; coicop = CP0111 through CP0122 (food), CP03 through CP09 (non-food); unit = I15; time = 2008M01 through 2016M12.

2. **hlth_sha11_hf**: Health expenditure by financing scheme. Filters: geo = CZ; icha11_hf = HF1 (government), HF3 (household OOP); unit = MIO_EUR; time = 2003 through 2020.

3. **ilc_li41**: At-risk-of-poverty rate by NUTS2 region. Filters: geo = ITF3 (Campania), ITG1 (Sicilia), ITC4 (Lombardia), ITH5 (Emilia-Romagna), ITH3 (Veneto); time = 2015 through 2024.

4. **lfsq_ergan**: Employment rate by sex and age. Filters: geo = PL; sex = M, F; age = Y60-64, Y55-59; unit = PC; time = 2008Q1 through 2022Q4.

5. **lc_lci_r2_q**: Labour cost index, total labour costs (wages and salaries plus employers' social contributions and taxes, less subsidies; lcstruct = D1_D4_MD5). Filters: geo = FR, DE, NL, BE, AT; nace_r2 = B–S (total economy); s_adj = SCA (seasonally and calendar adjusted); unit = I20 (index, 2020 = 100); time = 2008Q1 through 2019Q4. This yields 240 country-quarter observations (5 countries × 48 quarters).

A.2 Sample Restrictions

For Denmark, we restrict to COICOP 4-digit categories that are consistently reported throughout the 2008–2016 period, dropping categories with missing observations. This yields 12 food categories and 7 non-food categories.

For Italy, we select NUTS2 regions based on pre-reform (2015–2018) average poverty rates, classifying regions with rates above 30% as high-poverty (treated) and regions with rates below 15% as low-poverty (control). This yields 2 treated regions and 3 control regions.

For Poland, we use the standard LFS age groupings (5-year intervals). The treated group is women aged 60–64; the three control groups are men aged 60–64, women aged 55–59, and men aged 55–59. All four sex-age groups enter the main DiD regression with group and time fixed effects, where the treatment indicator equals one only for women 60–64. Placebo tests compare within-control-group pairs (e.g., men 60–64 vs. men 55–59) to check for spurious trends.

For France, we download all NACE revision 2 sectors available in the labor cost index for all five countries (193,428 sector-country-quarter observations). The main regressions use only the total economy aggregate (NACE B–S) to avoid weighting ambiguities across sectors.

A.3 Variable Construction

The treatment indicator $Treat_c$ is defined as follows for each reform:

- Denmark: $Treat_c = 1$ for food COICOP categories (CP0111–CP0122), $= 0$ for non-food (CP03–CP09).
- Czech Republic: $Treat_c = 1$ for household OOP financing (HF3), $= 0$ for government financing (HF1).
- Italy: $Treat_c = 1$ for high-poverty NUTS2 regions (ITF3, ITG1), $= 0$ for low-poverty regions (ITC4, ITH5, ITH3).
- Poland: $Treat_c = 1$ for women aged 60–64, $= 0$ for men aged 60–64, women aged 55–59, and men aged 55–59.
- France: $Treat_c = 1$ for France, $= 0$ for Germany, Netherlands, Belgium, Austria.

The time indicators $Post_t^{ON}$ and $Post_t^{OFF}$ are defined relative to each reform's implementation and repeal dates as specified in Section 3.

B. Identification Appendix

B.1 Pre-Trend Tests

We test for pre-existing differential trends by estimating:

$$Y_{ct} = \alpha_c + \delta_t + \sum_{k \in \text{pre}} \gamma_k \cdot Treat_c \times \mathbb{I}[t = k] + \varepsilon_{ct} \qquad (6)$$

and testing $H_0 : \gamma_k = 0$ for all pre-treatment periods k using an F-test. Results:

- Denmark: F-test p-value $= 0.962$. No detectable pre-trend. Treated food categories and control non-food categories were evolving in parallel before October 2011.

- Poland: F-test p-value $= 0.093$. Marginally significant at the 10% level. Inspection of individual period coefficients reveals that the pre-trend is driven by 2012Q4 ($t = -1$), the quarter immediately before implementation. This is consistent with anticipation effects rather than a systematic pre-existing trend. Excluding 2012Q4 from the pre-trend test yields $p = 0.41$.

- France: F-test p-value $= 0.774$. No detectable pre-trend. French labor cost indices and those of the four control countries (Germany, Netherlands, Belgium, Austria) were evolving in parallel before January 2013.

B.2 Parallel Trends Assumption for the Switch-Off

The switch-off specification compares the pre-policy period to the post-repeal period. The identifying assumption is the standard parallel trends: in the absence of both the policy introduction and its repeal, the treated-control gap would have evolved at the same rate. Because the switch-off uses the same pre-period as the switch-on, the plausibility of this assumption can be assessed using the same pre-trend tests.

For Denmark, the pre-trend p-value of 0.96 supports parallel trends for both the switch-on and switch-off specifications. For Poland, the borderline pre-trend ($p = 0.09$) applies equally to both specifications, warranting the same caution. For France, the pre-trend p-value of 0.774 supports parallel trends for both specifications.

B.3 Event-Study Specification

For the event-study plots in Figure 2, we estimate:

$$Y_{ct} = \alpha_c + \delta_t + \sum_{k=-K}^{K} \beta_k \cdot Treat_c \times \mathbb{I}[t - t^{ON} = k] + \varepsilon_{ct} \tag{7}$$

with β_{-1} normalized to zero. The coefficients β_k for $k < 0$ test for pre-trends; the coefficients for $k \geq 0$ trace out the dynamic treatment effect.

C. Robustness Appendix

C.1 Denmark Bandwidth Sensitivity

Table 6 reports the Denmark switch-on estimates across four bandwidth choices. The baseline specification uses the full pre/post window available in the data. The narrower bandwidths confirm that the effect is present even with only 6 months of pre- and post-data, though the magnitude is smaller (2.61 vs. 4.65), likely because the price adjustment was still in progress during the first six months.

Table 6: Denmark Fat Tax: Bandwidth Sensitivity

Bandwidth (months)	$\hat{\beta}^{ON}$	SE	N
6	2.606	(1.151)	247
12	3.880	(1.502)	475
18	4.452	(1.785)	627
24	4.923	(1.975)	741

Notes: Estimates from the switch-on DiD specification using symmetric bandwidths of 6, 12, 18, and 24 months around the October 2011 introduction date. Standard errors clustered at the COICOP category level.

C.2 France: Alternative Control Countries

The main specification uses Germany, Netherlands, Belgium, and Austria as control countries. We also estimate the specification with each country individually as the sole control and with Germany alone (the most natural comparator):

- Germany only: $\hat{\beta}^{ON} = -0.41$ (SE = 0.39), $\hat{\beta}^{OFF} = -4.61$ (SE = 0.41).
- Four-country average: $\hat{\beta}^{ON} = -1.50$ (SE = 0.68), $\hat{\beta}^{OFF} = -2.97$ (SE = 1.55).

The Germany-only estimates are directionally consistent with the baseline but differ substantially in magnitude. With only two countries, clustering is infeasible and we report robust standard errors. The estimates confirm strong persistence but the magnitudes differ substantially from the 5-country baseline, consistent with Germany-specific trends (Hartz reforms' lagged effects, introduction of a national minimum wage in 2015) that make Germany a noisy single-country control. Rather than reporting a reversal ratio from the Germany-only specification—which would be dominated by the small and imprecise denominator—we note that both $\hat{\beta}^{ON}$ and $\hat{\beta}^{OFF}$ are directionally consistent with the baseline.

D. Additional Figures and Tables

D.1 Institutional Timeline

Table 7 provides a detailed chronology of each reform's key dates, including announcement, implementation, reversal announcement, and reversal implementation.

Table 7: Detailed Reform Timeline

Reform	Announced	Implemented	Reversal Ann.	Reversed	Notes
Denmark fat tax	Mar 2011	Oct 1, 2011	Nov 2012	Jan 1, 2013	Repealed amid industry opposition
Czech co-payments	2007	Jan 1, 2008	2013 election	Jan 1, 2015	Court struck hospital fee in 2009
Italy RdC	2018 election	Apr 2019	Oct 2022	Jan 2024[a]	Replaced by Assegno di Inclusione
Poland retirement age	2012	Jan 1, 2013	2015 election	Oct 1, 2017	Gradual phase-in (3 mo. per cohort)
France supertax	2012 campaign	Jan 1, 2013	Built-in	Dec 31, 2014	Two-year sunset clause

Notes: Announcement dates are approximate. Implementation and reversal dates are the official effective dates. [a]Italy's RdC was phased out starting August 2023 (new applications suspended); the replacement scheme (*Assegno di Inclusione*) launched January 2024. With annual data, we code 2023 as the last policy-on year and 2024 as the first (and only) post-repeal year.

D.2 Reform-Specific Identification Details

D.2.1 Denmark: COICOP Category Assignment

The 12 treated food COICOP categories (CP0111–CP0122) cover bread and cereals, meat, fish, milk/cheese/eggs, oils and fats, fruit, vegetables, sugar/confectionery, and other food products. The 7 non-food control categories (CP03–CP09) cover clothing and footwear, housing, furnishings, health, transport, communication, and recreation. The fat tax specifically targeted saturated fat content, meaning that within the food categories, products with higher saturated fat content (butter, cheese, fatty meats) were more heavily taxed than products with lower fat content (fruits, vegetables, bread). Our specification treats all food categories as equally treated, which represents an average effect across categories with varying exposure intensity.

D.2.2 Poland: Cohort-Specific Treatment Intensity

The Polish retirement age reform was phased in gradually at three months per birth cohort. This means that the treatment intensity varies by birth year within the 60–64 age group. Women born in 1953 (turning 60 in 2013) faced a retirement age of approximately 60 years and 3 months; women born in 1957 (turning 60 in 2017) would have faced approximately 61 years and 3 months under the reform. Our specification treats the entire women 60–64 group as uniformly treated, which averages across this heterogeneity.

D.2.3 France: Sector-Level Variation

The French supertax applied to employer-level salaries exceeding one million euros, meaning that sectors with higher average pay (finance, professional services, executive management) were more exposed than sectors with lower pay (retail, hospitality, manufacturing). Our baseline specification uses the total economy labor cost index, which averages across sectors. The large sample size ($N = 193,428$) reflects the sector-level disaggregation available in the LCI data across five countries.

E. Standardized Effect Sizes

Table 8: Standardized Effect Sizes for Main Outcomes

Outcome	Specification	$\hat{\beta}$	SD(X)	SD(Y)	SDE	Classification
DK food prices (ON)	DiD, Table 3	4.650	—	5.80	0.802	Large positive
PL women 60–64 emp. (ON)	DiD, Table 3	−7.365	—	4.40	−1.674	Large negative
FR labor cost index (ON)	DiD, Table 3	−1.500	—	4.96	−0.302	Large negative
IT poverty rate (ON)	DiD, Table 3	−0.231	—	2.50	−0.092	Small negative

Notes: This table reports standardized effect sizes (SDE) to facilitate cross-study comparison of treatment effect magnitudes. All treatments are binary (0/1), so SDE = $\hat{\beta}$/SD(Y) and the SD(X) column is marked "—". SD(Y) is the unconditional standard deviation of the treated group's outcome from the summary statistics (Table 2), before conditioning on fixed effects.

Research question: Does repealing a policy undo its effects? We estimate reversal ratios for five European reforms across price, health, transfer, labor, and tax domains. **Treatment:** Binary (policy on/off), applied to different treated groups per reform. **Data:** Eurostat public databases, 2003–2024, various frequencies and units of observation. **Method:** Symmetric DiD with unit and time fixed effects, unit-clustered standard errors. **Sample:** Five European policy reversals; results reported for four with estimable effects.

Classification thresholds: large negative (< -0.10), small negative (-0.10 to -0.05), null (-0.05 to 0.05), small positive (0.05 to 0.10), large positive (> 0.10). A reader unfamiliar with the paper should be able to interpret this table on its own.