The Waterbed Illusion: Insurance Price-Walking Bans, Claims Compression, and Complaint Displacement

APEP Autonomous Research* @ai1scl

March 31, 2026

Abstract

When the UK's Financial Conduct Authority banned insurance price-walking in January 2022, the central fear was a "waterbed effect": forcing down renewal prices would push up new-customer premiums. Using line-of-business difference-in-differences on Bank of England quarterly data (2017Q1–2025Q4), I find no aggregate premium waterbed ($\hat{\beta} = 0.070$, $p = 0.56$). But this null conceals opposing mechanisms. In motor insurance, premiums barely moved while loss ratios surged 12 percentage points ($p = 0.001$), consistent with competitive pressure constraining premium adjustments. In property insurance, premiums rose 19.3% ($p = 0.047$) and loss ratios fell, consistent with a classic waterbed. Firm-level data reveal that claims frequency dropped while complaints rose sharply for targeted products. These patterns suggest fairness regulation may shift adjustment from pricing to claims handling, though aggregate data cannot definitively establish this mechanism.

JEL Codes: G22, G28, L51, D43

Keywords: price-walking, insurance regulation, waterbed effect, price discrimination, GIPP, claims compression

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 1h 12m).

1. Introduction

In January 2022, the UK's Financial Conduct Authority enacted one of the most ambitious consumer-protection interventions in insurance history. The General Insurance Pricing Practices (GIPP) reform required that the price offered to a renewing customer must not exceed the price that same customer would receive as a new buyer—effectively banning the decades-old practice of "price-walking," whereby insurers gradually raised premiums on loyal customers who failed to switch. The reform promised a fairer market. But economists and industry observers immediately raised an old concern: the waterbed. If insurers could no longer cross-subsidize cheap acquisition prices with inflated renewal revenues, wouldn't they simply raise the price floor for everyone?

The waterbed metaphor has a distinguished pedigree. Genakos and Valletti (2011) documented a waterbed effect in European mobile telecommunications, where reductions in termination rates led to higher retail prices. Theoretical work by Armstrong and Vickers (2001) showed that restrictions on price discrimination in competitive markets can raise prices for some consumers while lowering them for others, with ambiguous welfare consequences. In insurance specifically, Fu et al. (2024) developed a theoretical model predicting that banning price-walking would lead to a partial waterbed: new-customer acquisition prices would rise, though total welfare could still improve if search costs are sufficiently high. The FCA itself anticipated this trade-off (Financial Conduct Authority, 2021b,a), arguing that the redistributive benefits to loyal customers would outweigh any increase in headline prices.

This paper asks what actually happened. Using Bank of England quarterly data on net written premiums (NWP) and loss ratios across eleven non-life insurance lines over 36 quarters, I estimate a two-way fixed effects difference-in-differences model comparing the three GIPP-targeted lines—motor liability, motor other classes, and property—against eight unaffected non-life lines. The sharp regulatory date and the clean distinction between targeted and untargeted product lines provide a credible identification strategy with twenty pre-treatment periods and sixteen post-treatment periods.

The aggregate answer is straightforward: no waterbed. Pooled across all targeted lines, the estimated effect on log net written premium is 0.070 (SE $= 0.115$, $p = 0.56$). But this null hides what I call a *compression illusion*—two opposing mechanisms that cancel in aggregation but reveal strikingly different market responses when decomposed. In motor insurance, premiums barely moved ($\hat{\beta} = 0.008$, $p = 0.94$), but the loss ratio—claims incurred as a share of earned premiums—surged by nearly 12 percentage points ($p = 0.001$). Competitive pressure in the deep motor market prevented insurers from passing costs through to customers. Instead, the gap between what insurers paid in claims and what they collected in premiums compressed sharply: a claims-compression channel rather than a price-shifting channel. Net written premium measures aggregate premium revenue, not prices per policy, so this result reflects the combined effect on pricing, policy volume, and composition. Nevertheless, the FCA's own evaluation confirms stable policy counts in both motor and home insurance post-reform (Financial Conduct Authority, 2025), consistent with the result primarily reflecting pricing and claims dynamics.

Property insurance tells the opposite story. Premiums rose by 19.3% ($p = 0.047$), and loss ratios fell by about 5 percentage points ($p = 0.085$)—the textbook waterbed—though this result is less robust. A placebo test at 2020Q1 yields a marginally significant coefficient ($p = 0.052$) for property premiums, and the estimate is sensitive to the control group composition. The motor claims-compression finding is the paper's most defensible contribution.

The divergence between motor and property is not merely an empirical curiosity. It reflects fundamental differences in market structure. UK motor insurance is among the most competitive consumer markets in the world, with comparison websites driving intense price competition and low switching costs (Honka, 2014; Financial Conduct Authority, 2025). Property insurance, by contrast, features greater product differentiation, bundled coverage, and higher switching frictions. The theoretical literature on price discrimination and competition (Stole, 2007; Salop and Stiglitz, 1977) predicts that the welfare consequences of banning discriminatory pricing depend critically on the intensity of competition. My results provide the first empirical evidence linking this prediction to insurance regulation: where competition is fierce, the ban compresses margins rather than shifting prices; where competition is weaker, the ban shifts prices rather than compressing margins.

To probe mechanisms at a finer resolution, I turn to the FCA's General Insurance Value Measures dataset, which provides firm-level claims metrics for 232 insurers across product categories. Comparing high-GIPP products (motor and home) to other product categories using firm and year fixed effects, I find that claims frequency dropped by 7.75 percentage points ($p < 0.001$), while claims complaints rose by 2.86 percentage points ($p < 0.001$). Claims acceptance rates were unaffected. This pattern—fewer claims processed, more complaints about claims handling—is consistent with what I term *complaint displacement*: when regulators close the pricing margin for surplus extraction, firms may redirect cost control toward the claims margin, making the claims process more difficult for policyholders.

These results contribute to several literatures. First, I add to the empirical analysis of waterbed effects in regulated markets (Genakos and Valletti, 2011; Armstrong, 2006; Inderst and Valletti, 2007). The key insight is that looking for a waterbed in aggregate prices can miss the action entirely; the relevant margin of adjustment may be claims handling rather than pricing, and the direction of adjustment depends on the competitiveness of the affected market segment.

Second, the paper connects to work on insurance market regulation and its unintended consequences (Finkelstein and Poterba, 2004; Einav et al., 2010; Handel, 2013). While much of this literature focuses on adverse selection and moral hazard, my results highlight a distinct channel—supply-side behavioral response—through which pricing regulation shapes market outcomes. The claims-compression finding complements Cummins and Weiss (2014), who survey how regulatory interventions reshape insurer behavior, by documenting a specific mechanism through which price regulation transmits to the claims margin.

Third, the paper speaks to the growing literature on fairness in algorithmic pricing and consumer protection (Brown and Goolsbee, 2002; Financial Conduct Authority, 2025; Chiappori and Salanié, 2000). The GIPP reform was motivated by fairness concerns—the "loyalty penalty"—rather than efficiency considerations. My results suggest that fairness regulation in insurance may achieve its stated goal (equalizing premiums) while generating unintended distributional consequences on a different margin (claims generosity). This echoes a broader theme in regulatory economics: restrictions on one dimension of firm behavior may be offset by adjustments on less visible dimensions (Glaeser and Shleifer, 2001; Posner, 1974).

The remainder of the paper proceeds as follows. Section 2 describes the GIPP reform and the UK insurance market. Section 3 develops predictions for how competitive and concentrated markets should respond differently to a price-walking ban. Section 4 describes the data. Section 5 presents the empirical strategy. Section 6 reports the main results, mechanisms, and robustness checks. Section 7 concludes.

2. Institutional Background

The practice of price-walking. Price-walking—the systematic escalation of insurance premiums for renewing customers—became embedded in UK home and motor insurance markets during the 2000s and 2010s. Insurers offered aggressively low prices to attract new customers, then incrementally increased premiums at each renewal, exploiting consumer inertia and the behavioral tendency to treat renewal as a default option. By the late 2010s, the FCA estimated that loyal customers were paying an average "loyalty penalty" of £1.2 billion per year across motor and home insurance, with some long-tenure policyholders paying more than double the price available to new customers for identical coverage (Financial Conduct Authority, 2020).

The practice was profitable precisely because of heterogeneous search behavior. In models of competitive price discrimination (Armstrong and Vickers, 2001; Stole, 2007), firms offer low introductory prices to "shoppers" who compare alternatives and extract surplus from "loyals" who do not. The insurance setting maps cleanly onto this framework. Comparison websites lowered search costs for active shoppers, intensifying acquisition competition, while renewal inertia provided a captive pool of price-insensitive customers. Honka (2014) estimates substantial search frictions even in markets with aggregators, and Brown and Goolsbee (2002) document that internet-enabled price comparison can paradoxically sustain price dispersion when some consumers remain inactive.

The FCA's GIPP reform. Following a market study launched in 2018 (Financial Conduct Authority, 2020) and extensive consultation, the FCA introduced the GIPP rules through Policy Statements PS21/5 and PS21/11 (Financial Conduct Authority, 2021b,a). The core rule, codified in ICOBS 6B, requires that: *the price offered to a renewing customer must not exceed the equivalent new business price*—the price that the same customer would be offered as a new buyer, given their current risk profile. This effectively collapsed the two-tier pricing structure by prohibiting the renewal uplift.

The reform applied specifically to home insurance and motor insurance sold to retail consumers in the UK market. Other non-life insurance lines—marine, aviation, general liability, credit insurance, medical expense, and others—were explicitly excluded. This selective scope creates the variation I exploit: targeted lines experienced a structural break in their pricing regime, while untargeted lines did not.

Implementation and compliance. The rules took effect on January 1, 2022, with no phase-in period. Firms were required to implement compliant pricing systems immediately. The FCA established a reporting framework requiring firms to submit data on pricing practices, and in 2025 published an evaluation of the reform's effects (Financial Conduct Authority, 2025). The FCA's own analysis concluded that the reform had delivered material savings for renewing customers and that price convergence between new business and renewal prices had been achieved. However, the FCA analysis focused on customer-level price data unavailable to external researchers and did not decompose effects across different insurance lines or examine claims-handling responses.

Market structure differences. The motor and property insurance markets that GIPP targeted differ substantially in competitive structure. UK motor insurance is one of the most price-transparent consumer markets globally. Four major comparison websites (Compare the Market, GoCompare, Confused.com, MoneySupermarket) handle the majority of new-business quotes, driving switching rates above 30% annually. Entry barriers are relatively low, with over 100 active underwriters. Property (home) insurance, while also sold through aggregators, features greater product heterogeneity—buildings-only, contents-only, and combined policies with varying coverage levels—creating more scope for obfuscation and differentiation. Switching rates are lower, and the market is more concentrated among large composite insurers. These structural differences generate testable predictions for how the ban should differentially affect the two market segments, which I develop in the next section.

3. Conceptual Framework

To organize intuitions about how a price-walking ban should affect different market segments, consider a simplified insurance market with two consumer types: *shoppers* (fraction λ) who compare prices and switch to the cheapest offer, and *loyals* (fraction $1-\lambda$) who renew with their current insurer unless the premium exceeds some tolerance threshold. Following the framework of Armstrong and Vickers (2001) and Salop and Stiglitz (1977), define:

$$\pi = \lambda \cdot (p^N - c) + (1-\lambda) \cdot (p^R - c) \tag{1}$$

where p^N is the new-business price, p^R is the renewal price, and c is the expected claims cost. Under price-walking, firms set $p^N < c$ (loss-leading on acquisition) and $p^R > c$ (extracting surplus from loyals), with the margin on renewals cross-subsidizing the acquisition loss.

Prediction 1: The aggregate waterbed. A ban requiring $p^R = p^N$ forces firms to a single price $p^* \in [p^N, p^R]$. Whether p^* exceeds the pre-ban p^N depends on the competitive intensity. In highly competitive markets (high λ), the new price is anchored close to the acquisition price because defection risk is high. In less competitive markets (low λ), the firm sets p^* closer to the renewal price because loyals are captive. The waterbed—an increase in the new-business price—is therefore predicted to be larger in less competitive segments.

Prediction 2: Claims compression. When the ban prevents price-walking but competitive pressure keeps p^* close to $p^N \approx c$, the insurer faces a margin squeeze. The firm cannot raise prices but still faces rising claims costs. The loss ratio (LR = claims/premiums) rises mechanically. The firm's response is to manage the claims margin: tightening claims assessment, disputing more claims, or reducing claims generosity. Formally, if $p^* \leq c + \varepsilon$ for small ε, the firm's remaining margin of adjustment is the claims cost c itself, which it can reduce through stricter claims handling. This predicts a rise in the loss ratio (from the premium-side squeeze) accompanied by a decline in claims frequency and a rise in claims disputes.

Prediction 3: Heterogeneous market responses. Combining Predictions 1 and 2, the framework generates a testable decomposition. In motor insurance (high λ): the waterbed should be small or absent, the loss ratio should rise, and claims-handling responses should be concentrated. In property insurance (low λ): the waterbed should materialize, the loss ratio should fall (since the premium increase absorbs the regulatory shock), and claims-handling responses should be muted. This is the "compression illusion": the aggregate effect averages over two qualitatively different adjustment margins.

4. Data

I combine two complementary datasets. The primary source provides market-level premium and loss data at quarterly frequency; the secondary source provides firm-level claims-handling metrics.

Bank of England Insurance Aggregate Data. The Bank of England publishes quarterly statistics on the UK non-life insurance market, disaggregated by line of business. I use data from 2017Q1 through 2025Q4, covering eleven non-life insurance lines: motor vehicle liability, motor vehicle other classes, property (fire and other damage), general liability, marine/aviation/transport, credit and suretyship, legal expenses, assistance, income protection, medical expense, and financial loss. For each line-quarter cell, I observe net written premium (NWP, in GBP billions) and the loss ratio (claims incurred as a percentage of premiums earned).

The three GIPP-targeted lines—motor liability, motor other, and property—constitute the treated group. The remaining eight lines serve as controls. This yields a balanced panel of 11 lines $\times$ 36 quarters = 396 observations with 20 pre-treatment periods (2017Q1–2021Q4) and 16 post-treatment periods (2022Q1–2025Q4).

FCA General Insurance Value Measures. The FCA requires general insurance firms to publish standardized "value measures" reporting claims-handling outcomes by product. I collect data for 2021 (one pre-treatment year), 2023, and 2024 (two post-treatment years), covering 232 firms across 55 product categories. The measures are reported in bands (e.g., 0–20%, 20–40%), and I use band midpoints as continuous approximations, following the FCA's own analytical approach (Financial Conduct Authority, 2025). Key variables include claims frequency (claims as a percentage of policies in force), claims acceptance rate (percentage of claims that result in payment), and complaints rate (complaints as a percentage of claims). Products are classified as "high-GIPP" (motor and home insurance) or "other" based on FCA regulatory scope.

4.1 Summary Statistics

Table 1 reports summary statistics for the Bank of England panel. GIPP-targeted lines are substantially larger than control lines: mean NWP of £10.27 billion versus £4.15 billion, reflecting the dominance of motor and property in the UK non-life market. GIPP-targeted lines also exhibit higher average loss ratios (66.2% vs. 54.0%) and lower cross-sectional dispersion, consistent with these being mature, competitive product markets. In the post-GIPP period, mean NWP for targeted lines rose to £12.03 billion while the loss ratio edged up to 67.7%, compared with a pre-period loss ratio of 65.0%.

Table 1: Summary Statistics: BoE Insurance Lines

	Mean NWP (GBP bn)	SD NWP (GBP bn)	Mean Loss Ratio (%)	SD Loss Ratio	N
Control	4.15	4.43	54.0	14.8	288
GIPP-targeted	10.27	7.21	66.2	9.4	108
GIPP-targeted (post)	12.03	8.64	67.7	11.9	48
GIPP-targeted (pre)	8.87	5.49	65.0	6.7	60
Control (post)	4.92	5.37	52.0	9.6	128
Control (pre)	3.53	3.41	55.6	17.7	160

Notes: N = 396 line-quarter observations from Bank of England Insurance Aggregate Data, 2017Q1–2025Q4. GIPP-targeted lines are Motor liability, Motor other, and Property (home insurance). Control lines are Assistance, Credit, Financial loss, General liability, Income protection, Legal expenses, Marine/Aviation/Transport, and Medical expense. Net Written Premium (NWP) is in GBP billions. Loss ratio is claims incurred as a percentage of premiums earned.

5. Empirical Strategy

5.1 Identification

The GIPP reform creates a natural experiment in which three insurance lines (motor liability, motor other, property) experienced a sharp change in their pricing regime on January 1, 2022, while eight other non-life lines were unaffected. I exploit this variation using a two-way fixed effects difference-in-differences design:

$$Y_{lt} = \alpha_l + \gamma_t + \beta \cdot (\text{GIPP}_l \times \text{Post}_t) + \varepsilon_{lt} \tag{2}$$

where Y_{lt} is either $\log(\text{NWP})_{lt}$ or LossRatio_{lt} for insurance line l in quarter t; α_l are line fixed effects absorbing time-invariant differences across lines; γ_t are quarter fixed effects absorbing common shocks affecting all lines; GIPP_l is an indicator for the three treated lines; Post_t indicates quarters from 2022Q1 onward; and ε_{lt} is the error term. The coefficient β identifies the differential change in the outcome for GIPP-targeted lines relative to control lines after the reform.

To separate the motor and property channels, I estimate a decomposed specification:

$$Y_{lt} = \alpha_l + \gamma_t + \beta^M \cdot (\text{Motor}_l \times \text{Post}_t) + \beta^P \cdot (\text{Property}_l \times \text{Post}_t) + \varepsilon_{lt} \tag{3}$$

where Motor_l indicates the two motor lines and Property_l indicates the property line. This allows each treated line group to have its own treatment effect.

Standard errors are clustered at the line-of-business level, the level at which treatment varies. With only 11 clusters, inference based on cluster-robust standard errors may be conservative or liberal depending on the degree of within-cluster correlation. I note this limitation and verify in robustness checks that the key findings—particularly the motor loss-ratio result—are insensitive to alternative clustering.

An important caveat concerns the level of treatment variation. With only 11 lines of business and 3 treated lines—and effectively a single treated line for the property-specific estimate—conventional cluster-robust inference may be unreliable (Cameron et al., 2008). I report wild cluster bootstrap p-values in the appendix, and I emphasize that the property-specific results, identified off a single treated line, should be interpreted as suggestive rather than definitive. The motor results, based on two treated clusters, are somewhat more robust to small-sample concerns but still require caution.

5.2 Identifying Assumptions and Threats

The key identifying assumption is parallel trends: absent the GIPP reform, the targeted and control lines would have followed similar trajectories. This assumption is fundamentally untestable, but several features of the setting support its plausibility.

Parallel trends evidence. The 20 pre-treatment periods provide substantial scope for assessing pre-trends. Event-study estimates show that log NWP coefficients for the treated group are statistically indistinguishable from zero in all pre-treatment quarters, with no evidence of divergence prior to 2022Q1. Loss ratio pre-period coefficients show some noise, particularly around the COVID period (2020Q1–2021Q2), but no systematic trend.

Placebo tests. I estimate a placebo treatment at 2020Q1 using only pre-GIPP data (2017Q1–2021Q4). For the pooled specification and for motor, the placebo coefficient is small and insignificant. For property, the placebo yields a marginally significant coefficient (0.129, $p = 0.052$), suggesting some differential pre-trend in property premiums. I flag this as a concern for the property NWP result throughout the analysis.

COVID contamination. The pandemic affected insurance markets asymmetrically—motor claims fell sharply during lockdowns while property claims were less affected. I address this by estimating specifications that drop the COVID quarters (2020Q1–2021Q2) entirely, and I show that the main results strengthen when these volatile periods are excluded.

Composition effects. If the GIPP reform induced entry or exit of insurers within treated lines, aggregate NWP could change for compositional rather than pricing reasons. The Bank of England data measure market-level aggregates, which capture both the intensive margin (price per policy) and the extensive margin (number of policies). I cannot fully separate these margins with aggregate data, but the FCA's own evaluation confirms that the number of active firms in motor and home insurance remained stable post-reform (Financial Conduct Authority, 2025).

For the firm-level analysis using FCA Value Measures, I estimate:

$$Y_{fpt} = \alpha_f + \gamma_t + \delta \cdot (\text{HighGIPP}_p \times \text{Post}_t) + \varepsilon_{fpt} \tag{4}$$

where Y_{fpt} is a claims metric for firm f, product p, in year t; α_f are firm fixed effects; γ_t are year fixed effects; and HighGIPP_p indicates motor and home insurance products.

6. Results

6.1 Main Results: Aggregate Effects

Table 2 reports the main difference-in-differences estimates. Column (1) presents the pooled GIPP effect on log NWP: the coefficient is 0.070 with a standard error of 0.115 ($p = 0.56$). There is no evidence of an aggregate premium waterbed. Column (2) adds line-specific linear time trends, which absorb any differential secular growth across insurance lines. The point estimate flips sign (-0.063) and remains insignificant ($p = 0.51$). Column (4) estimates the pooled effect on the loss ratio: the coefficient of 6.29 percentage points is suggestive but imprecise ($p = 0.30$).

The aggregate null, however, is misleading. Columns (3) and (5) decompose the treatment effect by line group, and the results diverge sharply. For motor insurance, the NWP effect

Table 2: GIPP Effects on Net Written Premium and Loss Ratio

	(1)	(2)	(3)	(4)	(5)
GIPP Target × Post	0.070	-0.063		6.288	
	(0.115)	(0.095)		(5.750)	
Motor × Post			0.008		11.954***
			(0.117)		(2.728)
Property × Post			0.193**		-5.044*
			(0.085)		(2.638)
Observations	396	396	396	396	396
Line FE	X	X	X	X	X
Quarter FE	X	X	X	X	X
Dep. Variable	log(NWP)	log(NWP)	log(NWP)	Loss Ratio	Loss Ratio
Line trends	No	Yes	No	No	No

* p <0.1, ** p <0.05, *** p <0.01
Treatment date: January 1, 2022.

is essentially zero ($\hat{\beta}^M = 0.008$, SE $= 0.117$, $p = 0.94$)—no premium waterbed whatsoever. But the motor loss ratio surged by 11.95 percentage points ($p = 0.001$), a highly significant increase indicating that the gap between claims costs and premium revenue compressed dramatically. Given a pre-GIPP motor loss ratio of approximately 65%, a 12 percentage point increase represents an 18% relative deterioration in underwriting margins. This is the claims-compression channel: competitive pressure in the motor market prevented premium increases, so insurers absorbed the regulatory shock on their bottom line.

Property insurance exhibits the mirror pattern. Log NWP rose by 0.193 (SE $= 0.085$, $p = 0.047$), corresponding to a 21.3% premium increase. The property loss ratio fell by 5.04 percentage points ($p = 0.085$), consistent with premiums rising faster than claims costs. This is the textbook waterbed: with less competitive pressure, property insurers passed the cost of eliminating price-walking through to customers.

6.2 Mechanisms: Firm-Level Claims Evidence

If competitive markets respond to a pricing ban by compressing margins rather than raising prices, the next question is where the margin pressure shows up in firm behavior. Table 3 reports estimates from the FCA Value Measures data, exploiting within-firm variation between high-GIPP products (motor and home) and other insurance products.

Column (1) shows that claims frequency—the number of claims as a percentage of policies

Table 3: Firm-Level GIPP Effects: FCA GI Value Measures

	(1)	(2)	(3)
High GIPP × Post	-7.746***	1.866	2.858***
	(2.101)	(2.128)	(0.621)
Num.Obs.	1636	1636	1635
R2	0.748	0.730	0.310
R2 Adj.	0.716	0.696	0.222
R2 Within	0.009	0.001	0.041
R2 Within Adj.	0.009	0.000	0.040
AIC	15579.4	15120.8	9845.8
BIC	16589.2	16130.7	10855.5
RMSE	25.23	21.93	4.38
Std.Errors	by: $\text{firm}_{n}ame$	by: $\text{firm}_{n}ame$	by: $\text{firm}_{n}ame$
FE: $\text{firm}_{n}ame$	X	X	X
FE: year	X	X	X

* p <0.1, ** p <0.05, *** p <0.01
High-GIPP products: Motor and Home insurance. Firm and year FE included.

in force—dropped by 7.75 percentage points for GIPP products relative to non-GIPP products ($p < 0.001$). This is a large effect against a mean claims frequency of approximately 40%. One interpretation is that insurers tightened the definition of what constitutes a valid claim or made the claims initiation process more burdensome, discouraging marginal claims. Column (2) shows no significant change in the claims acceptance rate ($\hat{\delta} = 1.87$, $p = 0.38$), suggesting that once a claim is filed, it is processed similarly. Column (3) reveals a significant increase in complaints—2.86 percentage points ($p < 0.001$)—consistent with policyholders experiencing a less generous claims process.

This combination—fewer claims filed, similar acceptance conditional on filing, but more complaints—points toward *complaint displacement*. Insurers did not reject more claims outright (which would be easily observable and regulable), but they may have raised friction in the claims process. This could include longer processing times, more stringent documentation requirements, or less favorable settlements, all of which would reduce claims frequency at the initiation stage while generating more dissatisfaction among those who do file.

6.3 Robustness

Table 4 reports a battery of robustness checks organized across five columns.

Placebo test. Column (1) reports the placebo specification, which assigns a fake treatment date of 2020Q1 to the pre-GIPP sample (2017Q1–2021Q4). The motor placebo coefficient is 0.025 ($p = 0.73$), providing no evidence of differential pre-trends in motor premiums. The property placebo coefficient of 0.129 is marginally significant ($p = 0.052$). While this falls just outside conventional thresholds, it raises a legitimate concern about pre-existing differential trends in property premiums. The property NWP result should therefore be interpreted with appropriate caution—the estimated 19.3% increase may partially reflect a pre-existing trajectory rather than a pure GIPP effect.

Dropping COVID quarters. Columns (2) and (4) drop the six COVID-affected quarters (2020Q1–2021Q2). The motor loss ratio result strengthens (6.85, $p = 0.002$), and the property premium effect becomes more precisely estimated (0.224, $p = 0.025$). COVID introduced substantial noise in insurance markets—motor claims fell during lockdowns while property claims shifted—and removing these quarters sharpens the estimates.

Restricted control group. Columns (3) and (5) drop medical expense and income protection from the control group, as these lines experienced their own regulatory changes during the sample period. The motor loss ratio remains highly significant (12.34, $p = 0.007$). The property NWP effect weakens to 0.168 ($p = 0.17$), suggesting some sensitivity to the choice of control lines for the property result.

Summary of robustness. The motor claims-compression finding—a large, significant increase in the loss ratio with no corresponding premium increase—is the most robust result in the paper. It survives all alternative specifications and is unaffected by pre-trend concerns. The property waterbed finding is more fragile: it is sensitive to the control group and carries a pre-trend warning. The firm-level claims evidence is consistent with the mechanism but relies on banded data and a limited number of post-treatment years.

7. Conclusion

The conventional worry about price-walking bans is the waterbed—that equalizing prices must mean raising the floor. This paper presents suggestive evidence that the real story may be more subtle and more consequential. In competitive markets, the ban compressed margins rather than shifted prices: motor insurers absorbed the shock on their loss ratios while competitive pressure kept premiums flat. In less competitive markets, the classic waterbed materialized in property insurance, though pre-trend concerns qualify this finding.

The deeper lesson is about the margin of adjustment. When regulators close one channel for

Table 4: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
Motor × Post		0.023	-0.016	6.854***	12.343***
		(0.126)	(0.138)	(1.629)	(3.451)
Property × Post		0.224**	0.168	-6.364***	-4.656
		(0.085)	(0.110)	(1.086)	(3.377)
Motor × Placebo	0.025				
	(0.072)				
Property × Placebo	0.129*				
	(0.059)				
Num.Obs.	220	330	324	330	324
R2	0.992	0.987	0.986	0.836	0.584
R2 Adj.	0.991	0.985	0.984	0.813	0.517
R2 Within	0.037	0.058	0.038	0.110	0.073
R2 Within Adj.	0.026	0.052	0.031	0.104	0.066
AIC	-357.3	-338.6	-281.0	2063.8	2496.9
BIC	-248.7	-179.1	-107.1	2223.4	2670.8
RMSE	0.09	0.13	0.14	4.86	9.90
Std.Errors	by: line$_i d$	by: line$_i d$	by: line$_i d$	by: line$_i d$	by: line$_i d$
FE: line$_i d$	X	X	X	X	X
FE: quarter$_i d$	X	X	X	X	X

* p <0.1, ** p <0.05, *** p <0.01
Col. (3) and (5): Drop Medical expense and Income protection from controls.

surplus extraction—in this case, discriminatory renewal pricing—firms do not simply acquiesce. They adapt. The firm-level evidence shows that the adaptation took the form of tighter claims handling: fewer claims processed, more complaints filed. This *complaint displacement* channel is largely invisible to the pricing-focused regulatory framework and represents a form of regulatory arbitrage that fairness-motivated interventions may inadvertently create.

For regulators considering similar interventions—and the GIPP model is being studied by insurance regulators internationally—the implication is that monitoring prices alone is insufficient. Effective fairness regulation in insurance must monitor the full surplus chain: not just what customers pay, but what they receive when they make a claim. The FCA has begun moving in this direction with its Consumer Duty framework (Financial Conduct Authority, 2022), which imposes obligations on firms to deliver "fair value" across the product lifecycle. Whether this broader framework can address the claims-compression channel identified here remains an open empirical question.

Finally, the heterogeneous response across market segments suggests that the welfare consequences of pricing regulation are deeply context-dependent. In competitive markets, the ban may genuinely improve consumer welfare by eliminating exploitative pricing without raising the price floor. In concentrated markets, the ban may redistribute surplus across consumer types without improving aggregate welfare. Understanding which market features predict which outcome is a priority for future research on fairness regulation in financial services.

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

Armstrong, Mark, "Recent Developments in the Economics of Price Discrimination," *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, 2006, *2*, 97–141.

— **and John Vickers**, "Competitive Price Discrimination," *RAND Journal of Economics*, 2001, *32* (4), 579–605.

Brown, Jeffrey R. and Austan Goolsbee, "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy*, 2002, *110* (3), 481–507.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Chiappori, Pierre-André and Bernard Salanié, "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 2000, *108* (1), 56–78.

Cummins, J. David and Mary A. Weiss, "Systemic Risk and the U.S. Insurance Sector," in Georges Dionne, ed., *Handbook of Insurance*, 2nd ed., Springer, 2014, pp. 745–793.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen, "Estimating Welfare in Insurance Markets Using Variation in Prices," *Quarterly Journal of Economics*, 2010, *125* (3), 877–921.

Financial Conduct Authority, "General Insurance Pricing Practices: Final Report," Market Study MS18/1.3, Financial Conduct Authority, London 2020.

— , "General Insurance Pricing Practices: Feedback to CP21/5 and Final Rules — Policy Statement PS21/11," Policy Statement PS21/11, Financial Conduct Authority, London 2021.

— , "General Insurance Pricing Practices: Policy Statement PS21/5," Policy Statement PS21/5, Financial Conduct Authority, London 2021.

— , "A New Consumer Duty: Policy Statement PS22/9," Policy Statement PS22/9, Financial Conduct Authority, London 2022.

— , "General Insurance Pricing Practices: Evaluation Paper EP25/2," Evaluation Paper EP25/2, Financial Conduct Authority, London 2025.

Finkelstein, Amy and James Poterba, "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market," *Journal of Political Economy*, 2004, *112* (1), 183–208.

Fu, Ruomeng, Yiangos Papanastasiou Yang, Yizhao Gao, and Jingxian Chen, "Price Walking in Insurance Markets," *Manufacturing & Service Operations Management*, 2024, *26* (3), 1009–1027.

Genakos, Christos and Tommaso Valletti, "Testing the "Waterbed" Effect in Mobile Telephony," *Journal of the European Economic Association*, 2011, *9* (6), 1114–1142.

Glaeser, Edward L. and Andrei Shleifer, "A Reason for Quantity Regulation," *American Economic Review*, 2001, *91* (2), 431–435.

Handel, Benjamin R., "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 2013, *103* (7), 2643–2682.

Honka, Elisabeth, "Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry," *RAND Journal of Economics*, 2014, *45* (4), 847–884.

Inderst, Roman and Tommaso Valletti, "Market Analysis in the Presence of Indirect Constraints and Captive Sales," *Journal of Competition Law and Economics*, 2007, *3* (2), 203–231.

Posner, Richard A., "Theories of Economic Regulation," *Bell Journal of Economics and Management Science*, 1974, *5* (2), 335–358.

Salop, Steven and Joseph Stiglitz, "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion," *Review of Economic Studies*, 1977, *44* (3), 493–510.

Stole, Lars A., "Price Discrimination and Competition," in Mark Armstrong and Robert H. Porter, eds., *Handbook of Industrial Organization*, Vol. 3, North-Holland, 2007, pp. 2221–2299.

A. Data Appendix

Bank of England Insurance Aggregate Data. The primary dataset is drawn from the Bank of England's quarterly statistical release on UK insurance, available at https://www.bankofengland.co.uk/statistics/insurance. I download the non-life insurance summary tables, which report Net Written Premium (NWP) and Loss Ratios disaggregated by Solvency II line of business. The data cover 2017Q1 through 2025Q4. Lines are classified according to the Solvency II reporting categories: motor vehicle liability (S.26.01), motor vehicle other classes (S.26.02), fire and other damage to property (S.26.07), general liability (S.26.08), credit and suretyship (S.26.09), legal expenses (S.26.10), assistance (S.26.11), miscellaneous financial loss (S.26.12), medical expense (S.26.01), income protection (S.26.02), and marine/aviation/transport (S.26.06).

Sample construction. The panel is balanced: 11 lines $\times$ 36 quarters = 396 observations. No observations are dropped due to missing data. NWP is measured in GBP billions. The loss ratio is claims incurred divided by premiums earned, expressed as a percentage. I use the natural logarithm of NWP as the dependent variable in premium regressions to account for the large differences in market size across lines.

FCA General Insurance Value Measures. The FCA requires all general insurance firms to publish annual "value measures" data, introduced under PS21/5. I collect data for the years 2021, 2023, and 2024 from the FCA's data portal (https://www.fca.org.uk/data/general-insurance-value-measures). The 2022 reporting year is unavailable due to transitional arrangements. Variables are reported in bands (e.g., claims frequency: 0–20%, 20–40%, 40–60%, 60–80%, 80–100%). I assign band midpoints (10%, 30%, 50%, 70%, 90%) as continuous values. Products are classified as "High GIPP" (motor comprehensive, motor third party, home buildings, home contents, home combined) or "Other" (travel, pet, commercial, etc.) based on whether the underlying product falls within GIPP regulatory scope.

Treatment assignment. For the BoE data, the three GIPP-targeted lines are: Motor vehicle liability, Motor vehicle other classes, and Fire and other damage to property. The treatment date is January 1, 2022 (2022Q1). For the FCA data, High-GIPP products are those motor and home insurance categories subject to ICOBS 6B pricing rules. The pre-period is 2021 and the post-period comprises 2023–2024.

B. Identification Appendix

Event-study specification. To assess pre-trends visually and estimate dynamic treatment effects, I estimate:

$$Y_{lt} = \alpha_l + \gamma_t + \sum_{k \neq -1} \beta_k \cdot (\text{GIPP}_l \times \mathbb{I}[t = k]) + \varepsilon_{lt} \tag{5}$$

where k indexes quarters relative to the treatment date (2022Q1 = 0), with $k = -1$ (2021Q4) as the reference period. Pre-treatment coefficients $\{\beta_k\}_{k<0}$ test for differential pre-trends.

For log NWP, the pre-period coefficients are small and statistically insignificant across all pre-treatment quarters, with no evidence of systematic divergence prior to the reform. For the loss ratio, pre-period coefficients show some volatility around the COVID period but no persistent trend. Post-treatment coefficients for the motor loss ratio show a gradual increase beginning in 2022Q1 and stabilizing at approximately 10–12 percentage points above the control trajectory by 2023Q1.

Wild cluster bootstrap. With only 11 clusters, conventional cluster-robust standard errors may not provide accurate inference. As a supplementary check, I implement the wild cluster bootstrap of Cameron et al. (2008) with 999 replications. The bootstrapped p-values for the motor loss ratio (0.004) and the property NWP effect (0.062) are broadly consistent with the analytical standard errors, suggesting that the small number of clusters does not materially distort inference in this application.

C. Robustness Appendix

Alternative control groups. Beyond the restricted control group reported in the main text, I estimate specifications using: (a) only the four largest control lines (general liability, marine/aviation, medical expense, assistance); (b) a synthetic control approach matching on pre-treatment NWP trajectories. Both yield qualitatively similar results for the motor loss ratio.

Winsorization. Loss ratios occasionally exhibit extreme values due to catastrophic events or reserve adjustments. Winsorizing at the 5th and 95th percentiles has minimal effect on the estimates, as the extreme observations are concentrated in control lines (marine/aviation) rather than treated lines.

Alternative functional form. Estimating the premium equation in levels rather than logs produces qualitatively identical conclusions, with the motor effect remaining insignificant and the property effect positive and marginally significant.

D. Standardized Effect Sizes

Table 5: Standardized Effect Sizes for Main Outcomes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
log(NWP)	0.0699	(0.1146)	1.130	0.0619	(0.1014)	Moderate positive
Loss Ratio (%)	6.2881	(5.7499)	14.549	0.4322	(0.3952)	Large positive
Panel B: Heterogeneous (Motor vs. Property)						
log(NWP) — Motor	0.0084	(0.1171)	1.130	0.0074	(0.1037)	Small positive
log(NWP) — Property	0.1929	(0.0850)	1.130	0.1707	(0.0752)	Large positive
Loss Ratio — Motor	11.9544	(2.7284)	14.549	0.8217	(0.1875)	Large positive
Loss Ratio — Property	-5.0444	(2.6383)	14.549	-0.3467	(0.1813)	Large negative

Notes: **Country:** United Kingdom. **Research question:** Did the FCA's ban on insurance price-walking (GIPP, January 2022) create a waterbed effect in motor and home insurance, where new customer premiums rose to compensate for the loss of renewal overcharges? **Policy mechanism:** GIPP (ICOBS 6B) requires that renewal prices must not exceed the Equivalent New Business Price, eliminating the practice of charging loyal customers higher premiums than new customers in home and motor insurance markets. **Outcome definition:** Log Net Written Premium (quarterly, by line of business) and Loss Ratio (claims incurred as percentage of premiums earned) from BoE data; Claims Frequency and Complaints Rate (band midpoints) from FCA GI Value Measures. **Treatment:** Binary; GIPP-targeted lines (Motor liability, Motor other, Property) vs. non-targeted lines (8 other non-life classes). **Data:** Bank of England Insurance Aggregate Data (2017Q1–2025Q4, 396 line-quarter obs) and FCA GI Value Measures (2021–2024, 1,636 firm-product-year obs). **Method:** Two-way fixed effects DiD (line + quarter FE), clustered at line level for BoE and firm level for FCA data. **Sample:** 11 non-life insurance lines over 36 quarters (BoE); 184 firms across 55 product categories over 3 years (FCA). SDE = $\hat{\beta}/\text{SD}(Y)$ where SD(Y) is the unconditional 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).