WORKING PAPER | July 2026

Columbia University, Department of Economics

Tipping the Tax: How Government Surcharges Become Gratuities*

Tyler Sotomayor§ Gabriel Uceda-Sosa¶

Columbia University

July 2026

Working Paper

Abstract

New York taxi screens suggest tips as percentages of the fee-inclusive total, so every government surcharge on the meter raises every suggested tip. We estimate the causal pass-through of statutory charges into gratuities: *eleven cents of each legislated surcharge dollar becomes tip*, about four-fifths of the rate at which riders tip the fare itself. Identification comes from six sources of quasi-experimental variation: surcharges that switch on and off at fixed clock times, the holiday calendar, a large repricing, a flat fare, and the spatial boundary of the 2025 congestion toll. The per-dollar estimate is stable across doses from $0.50 to $5.00 and in both directions, and riders absorb the charges rather than re-optimize. At the 2025 schedule, $141 million of government charges flow through the tip base each year, generating $15–21 million in gratuities that no legislature enacted; ignoring tips misstates the incidence of these charges.

JEL Codes: D91, H22, J33, R48

Keywords: default effects, tipping, tax salience, pass-through, congestion pricing

*We thank the NYC Taxi & Limousine Commission for the public trip records on which this paper relies. All errors are our own. § Corresponding author: tyler.sotomayor@columbia.edu. ¶ Correspondence: gabrielucedasosa.com. Please cite as: Sotomayor, Tyler and Gabriel Uceda-Sosa (2026). "Tipping the Tax: How Government Surcharges Become Gratuities." Working Paper, Columbia University.

Contents

1 Introduction

At 3:58 on a weekday afternoon, a Manhattan taxi ride ends with a \$17.70 metered fare. The backseat screen assembles the bill—fare, fixed fees, congestion surcharge—into a \$21.70 total and offers three buttons: 20, 25, or 30 percent, computed on that total. Four minutes later, the identical ride ends differently. The rush-hour surcharge has switched on, the total reads \$24.20, and every button is higher—the 20 percent button by exactly fifty cents, which is 20 percent of the \$2.50 the fare schedule just added (Figure 2). Nothing about the trip, the driver, or the rider's regard for the driver has changed. A scheduled line item changed, and the suggestion arithmetic carried it into the tip.

The mechanism on the screen. Figure 1 shows what the buttons multiply: every component of the fare stack, from the metered fare the driver keeps to the congestion fees the driver merely collects and remits, lands in one total, and the percentages are computed on the sum. This is not one vendor's quirk. All three payment vendors match tips to the fee-inclusive base at four to five times the rate of any narrower base (Table 1), and the base is visible in behavior alone: 57 percent of positive credit-card tips sit in the half-point bin at *exactly* 20 percent of the all-in total, and the same tips, recomputed against the metered fare, show no spikes at all (Figure 3). Haggag and Paci (2014) established that suggested tips cause tips. A suggestion, however, is a percentage *of something*. This paper is about the something—specifically, about what happens when public policy moves it.

The question is the pass-through rate: when a statute puts one more dollar on the meter, how many cents cross into the tip? Call it ρ. Answering requires surcharge variation that is not correlated with who is riding, and New York's fare schedule supplies it daily: the rush-hour surcharge applies to weekday pickups from 4:00pm, assigned by the pickup minute. We build date-by-minute cells from the universe of standard-rate credit-card trips, September 2021 through September 2024, and compare tips minutes apart across the cutoffs.

A daily experiment, and a trap in it. Tips jump at 4:00pm. But the naive comparison is a trap: the metered fare jumps too, by \$0.49, because rush-hour traffic makes trips slower and dearer (Table 4), and attributing the whole tip step to the surcharge inflates it by a third (\$0.389

against \$0.292). The repair is in the fare schedule itself. On weekends the same clock carries no surcharge, so the weekday–weekend *difference in discontinuities* at 4:00pm nets out whatever rush hour does to trips on any day of the week (Grembi et al., 2016); our analysis plan locked weekends as controls before estimation. Conditioning on the non-surcharge fare, the weekend placebo step falls from \$0.088 to \$0.024; differencing weekday against weekend gives 0.108 per statutory dollar in the \$2.50 regime and 0.105 in the \$1.00 regime, and dropping the fare control entirely moves those by about a cent. Identification does not lean on a functional-form assumption about fares; the weekend difference does the work.

One parameter. The headline result is that about *eleven cents of every statutory surcharge dollar becomes tip.* Across the six primary within-day changes—the rush surcharge switching on at two statutory doses, the overnight surcharge switching on at two doses, and two downward changes where the net surcharge falls—the intention-to-treat effect per statutory dollar spans 0.090 to 0.117, with a precision-weighted mean of 0.108 (SE 0.001). Scaled instead by the surcharge dollars the meter actually records, ρ^{IV} runs 0.122 to 0.150. The average tip in this sample is 17.4 percent of the base, so riders tip a surcharge dollar at roughly four-fifths the rate they tip a fare dollar: the surcharge is not distinguished from the fare because, on the screen, it is not distinguishable from the fare.

Figure 7 assembles every locked estimate on one axis, and it is the paper's argument in one frame. Two statutory doses, because the legislature repriced both surcharges by 150 percent in December 2022. Surcharges switching on and switching off. Three clocks. Weekday and weekend cutoffs. A second control series—legal holidays, working days with no surcharge—replicates the weekend difference at 0.095–0.096, and its exception proves the rule: Juneteenth, which the meters do not exempt, shows the full step. A \$5.00 surcharge on JFK *flat* fares—where the metered fare cannot jump at 4:00pm by construction—gives 0.133–0.149 per statutory dollar with no controls at all, twice our largest dose on a base three times larger. And—with no clock in it at all—the January 2025 congestion toll, identified from a boundary in space: the spatial estimate is 0.107 per statutory dollar, against the clock designs' pooled 0.108. Two estimates fail: both at 6:00am on weekdays, where the dawn shift toward airport-bound trips breaks the design's conditioning. We report them in grey rather than discard them, because their weekend twins at the same clock and dose are normal—which localizes the failure to weekday-dawn composition, not to the mechanism, and shows the identifying assumption is checkable (Section 5).

Riders do not undo it. Three findings say the pass-through is a property of the interface rather than of attentive generosity. First, the button is not the mechanism: the two major vendors' riders take an exact default on 19 versus 65 percent of tips, yet pass-through is statistically identical at both (0.147 and 0.152 per recorded dollar)—riders who type custom amounts compute percentages on the same inclusive total the screen displays. Second, riders barely react to repricing: when the rush surcharge rose from \$1.00 to \$2.50 overnight in December 2022, keeping tip dollars fixed required the rush-window tip rate to fall 1.1 points; it fell 0.22, immediately, and never drifted further over the following six months. Third, the quantity margins are silent: pickup density is smooth through the cutoff, card use does not fall, and zero-tipping rises 0.15 percentage points. Riders absorb the fee on every margin except the one the software moves for them.

What is new. That payment interfaces differ in the base they multiply, and that add-ons inside the base raise tips, is established: Thakral and Tô (2026) document both in this market, using the 2012 fare change and the two archival vendors' differing treatment of tax and tolls, and our archival evidence confirms their account—the base difference explains about three-quarters of the vendor tip gap that Haggag and Paci (2014) used to identify default effects (Appendix B). What has not been measured is what *statutory policy* does when it rides that base. A legislated surcharge that switches at a known minute, at a dose the legislature set and then changed, is variation of a different kind than a fare change or a route-determined toll, and it identifies a policy object: the pass-through of a public fee into a private gratuity, per statutory dollar. We estimate it, test its stability across doses, directions, clocks, and two congestion-pricing programs, and aggregate the resulting transfer. Prior work shows that interfaces construct the tipping norm; this paper shows that fiscal policy conscripts it.

Incidence, stakes, and a policy dial. The driver remits the congestion fee and keeps the tip, so the toll's true incidence splits: on a \$0.75 CBD toll, the rider pays about \$0.83, the MTA receives \$0.75, and the driver nets the induced tip. Aggregated against fees actually charged, the 2019 congestion surcharge's \$127.2 million on card trips generated roughly \$18.7 million a year in unlegislated gratuities, and the 2025 toll adds about $2.2 million a year (Section 6). These are transfers, not losses—and they run from riders to drivers, plausibly progressively. The point is not that the transfer is bad; it is that no one chose it. The dial that sets it is one line of payment-terminal arithmetic—the definition of the base—and the 2012 market, in which one vendor computed on a narrower base, shows the dial is real engineering, not hypothesis. What a default is computed *on* matters as much as what it is set *to*.

The rest of the paper proceeds as follows. Section 2 situates the paper. Section 3 presents the fare stack, the data, and the base facts. Section 4 states the designs. Section 5 reports the estimates, the mechanism evidence, and the congestion-fee designs. Section 6 prices the transfers and the counterfactuals, and Section 7 concludes.

2 Related Literature

Defaults and choice architecture. That pre-set options move consequential choices is among the most replicated results in economics, from retirement saving (Carroll et al., 2009; Madrian & Shea, 2001) and organ donation (Johnson & Goldstein, 2003) to the broad program of Thaler and Sunstein (2008); DellaVigna (2009) surveys the field. Haggag and Paci (2014) brought the result to tipping using the same New York taxi system we study, and Alexander et al. (2021) showed in the same setting that raising the suggested-tip menu raises tips—about seventeen cents per percentage point of suggestion—at the cost of more zero tips. Our contribution is orthogonal to the menu *level*: we hold the menu fixed and study the *base* it multiplies.

Thakral and Tô (2026) are closest to us and should be read alongside this paper. Using the September 2012 fare change and the two vendors' differing treatment of tax and tolls, they establish that a larger base raises tips even among riders who do not press a default button, and that add-ons crowd tips in or out according to whether the interface includes them—facts our vendor evidence in Section 5 and Appendix B reproduce. Their question is what tipping *is*: they build a model of a default-anchored norm with a convex pain of paying. Ours is what a *tax* does: their identifying variation is a fare change and a route-determined toll, while ours is a legislated surcharge that switches at a known minute, at a dose the legislature set and then changed by 150 percent. That difference is what lets us report a pass-through rate per statutory dollar, test its stability across doses and directions, extend it to two congestion-pricing programs, and aggregate the transfer. Choice architecture here is not an anchor but a transmission mechanism from fiscal policy to a private transfer.

Tax salience, attention, and shrouding. Chetty et al. (2009) and Finkelstein (2009) established that how a tax is displayed changes how agents respond to it; Taubinsky and Rees-Jones (2018) show attention to sales taxes is endogenous and heterogeneous, Feldman and Ruffle (2015) and Goldin and Homonoff (2013) develop the toolkit for partially attentive consumers, and Morrison and Taubinsky (2023) document rule-of-thumb responses to price changes. Gabaix and Laibson (2006) supply the market logic of add-on prices that firms do not compete away. Our channel is distinct from all of these in one respect: the surcharge is fully disclosed, itemized on the screen, and riders still tip on it—not because the tax is hidden, but because the interface has already folded it into the arithmetic of an unrelated voluntary payment. It is shrouding by *composition* rather than by concealment, and the December 2022 test shows riders undo only a fifth of a 150 percent surcharge increase.

Tipping. Tipping is a large voluntary transfer whose variation service quality explains only partly (Azar, 2020; Lynn & McCall, 2000); Conlin et al. (2003) model it as norm enforcement, and Chandar et al. (2019), in Uber's national rollout, find rider fixed traits dominate trip attributes—most riders never tip, a few always do. Field evidence on point-of-sale prompts (Warren et al., 2021) documents consumer irritation with suggestion screens. The channel is not specific to taxis: Lynn and Kwortnik (2023), in 68 million restaurant checks, find tipping rates rise with the local sales-tax rate—which is what tipping on a tax-inclusive check total implies, in an industry with no taxi meter in sight. Against this backdrop our estimates show that a substantial share of observed tip variation is set by payment-terminal arithmetic: a dollar of government surcharge moves the average tip by fifteen cents regardless of which vendor's screen, which dose, or which direction.

Congestion pricing. Evaluations of cordon pricing measure traffic, revenue, and environmental margins, from London (Leape, 2006) to the January 2025 New York program, whose short-run traffic effects Cook et al. (2025) measure with network speed data. Fiscal incidence in these evaluations ends at the statutory payment. We document a channel none of them counts: the toll enters the tip base, so part of the rider's payment continues past the MTA to the driver. Roughly \$2.2 million a year of New York's congestion toll becomes gratuities—small beside program revenue, but a transfer created entirely by interface design, and one that generalizes to every percentage-suggestion payment system that sits downstream of a government fee.

3 Institutional Setting and Data

The fare stack. A metered yellow-cab fare is a stack, and its components differ in one respect that matters throughout this paper: who ends up with the money. Figure 1 sets them side by side.

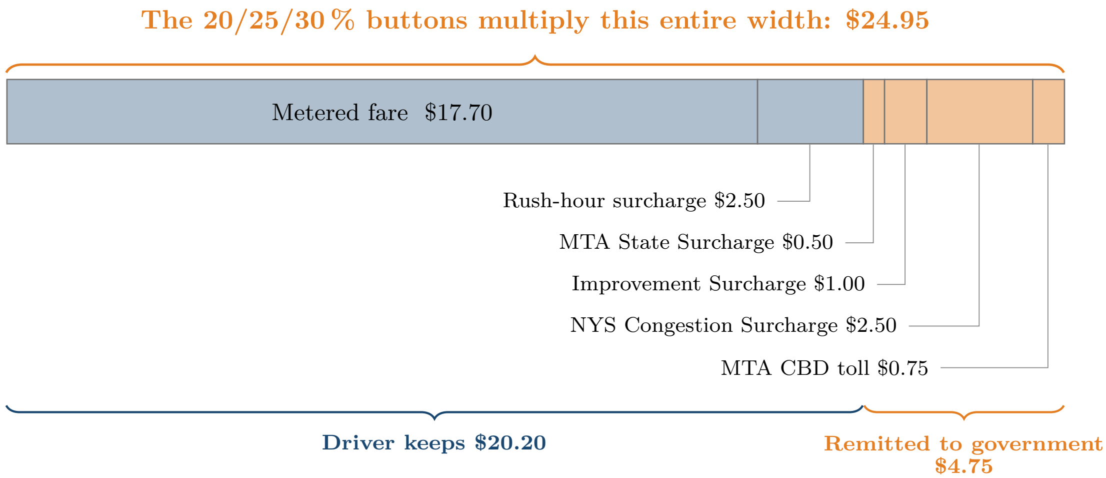
Two components are *fare revenue the driver keeps*: a rush-hour surcharge on weekday pickups between 4:00pm and 8:00pm, and an overnight surcharge from 8:00pm to 6:00am. Both are set by the TLC, and both were \$1.00 and \$0.50 respectively until December 18, 2022, and \$2.50 and \$1.00 thereafter (New York City Taxi and Limousine Commission, 2022). They are prices, not taxes.

Four components are *collected from the rider and remitted*: a \$0.50 MTA State Surcharge, a \$1.00 improvement surcharge, a \$2.50 New York State congestion surcharge on trips south of 96th Street (since 2019), and a \$0.75 MTA congestion toll south of 60th Street (since January 5, 2025). For these the driver is a collection agent and nets nothing; on medallion taxis the technology service provider withholds the congestion surcharge before funds reach the owner.

The distinction organizes the empirical strategy. The within-day clock designs identify pass-through from the two *fare* surcharges, which switch at fixed minutes and so deliver sharp, high-powered, repeatedly replicated variation. The spatial designs identify it from the two *remitted* congestion fees, which is the fiscal case the paper is ultimately about but which offers weaker variation. That the two give the same answer is the point: the payment interface does not distinguish them, because on the screen they are indistinguishable. Every component in Figure 1, whoever receives it, is added to the metered fare before the payment screen appears, and the percentage buttons multiply the sum.

Figure 2 shows the mechanism as a rider encounters it, before any econometrics: the same trip four minutes apart, with every suggested tip higher after 4:00 PM because a government surcharge has entered the base.

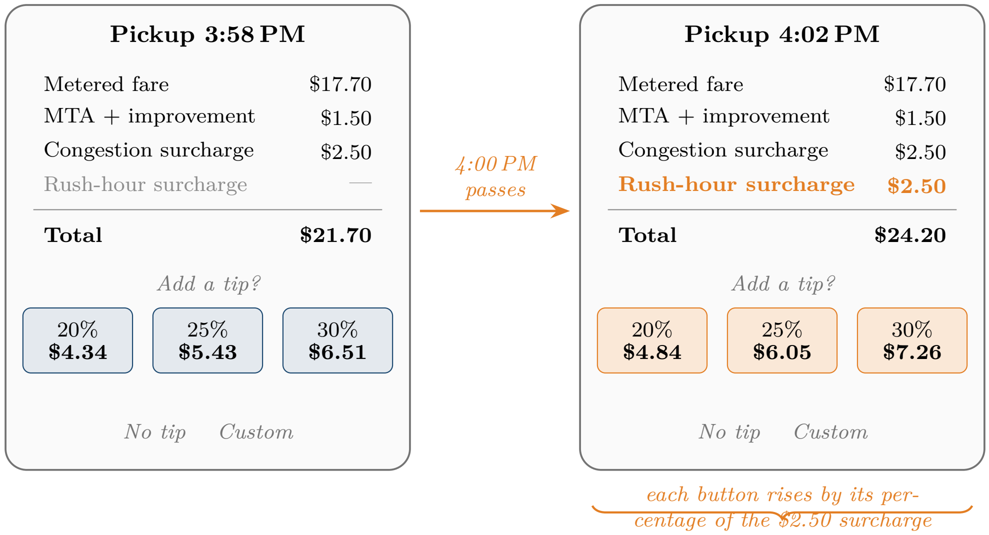
Figure 1: The Fare Stack: Who Receives Each Component, and What the Tip Buttons Multiply



Component	Amount	When it applies	Who receives it
Metered fare	varies	always	**Driver**
Rush-hour surcharge	$2.50	weekdays 4–8pm	**Driver**
Overnight surcharge	$1.00	8pm–6am	**Driver**
MTA State Surcharge	$0.50	always	MTA
Improvement Surcharge	$1.00	always	Taxi improvement fund
NYS Congestion Surcharge	$2.50	south of 96th St	New York State
MTA CBD toll	$0.75	south of 60th St	MTA
Bridge and tunnel tolls	varies	by route	Bridge authority

Notes: Illustrative weekday 4:30pm pickup south of 60th Street in 2025; the metered fare is the estimation sample's pre-4pm mean. Segment widths are proportional to dollars. Navy components are fare revenue the driver keeps; orange components are collected from the rider and remitted—the driver is a collection agent for them and nets nothing. The rush-hour and overnight surcharges are set by the TLC but are *not* government revenue. Every component, whoever receives it, sits inside the base the percentage buttons multiply, which is the fact this paper measures: the $4.75 of remitted fees on this trip raises the suggested tip by $0.95 at the 20 percent button, and raises the tip actually paid by about $0.52 at our estimated pass-through— money that reaches the driver rather than the agency levying the fee.

Figure 2: The Mechanism on the Screen: One Trip, Four Minutes Apart



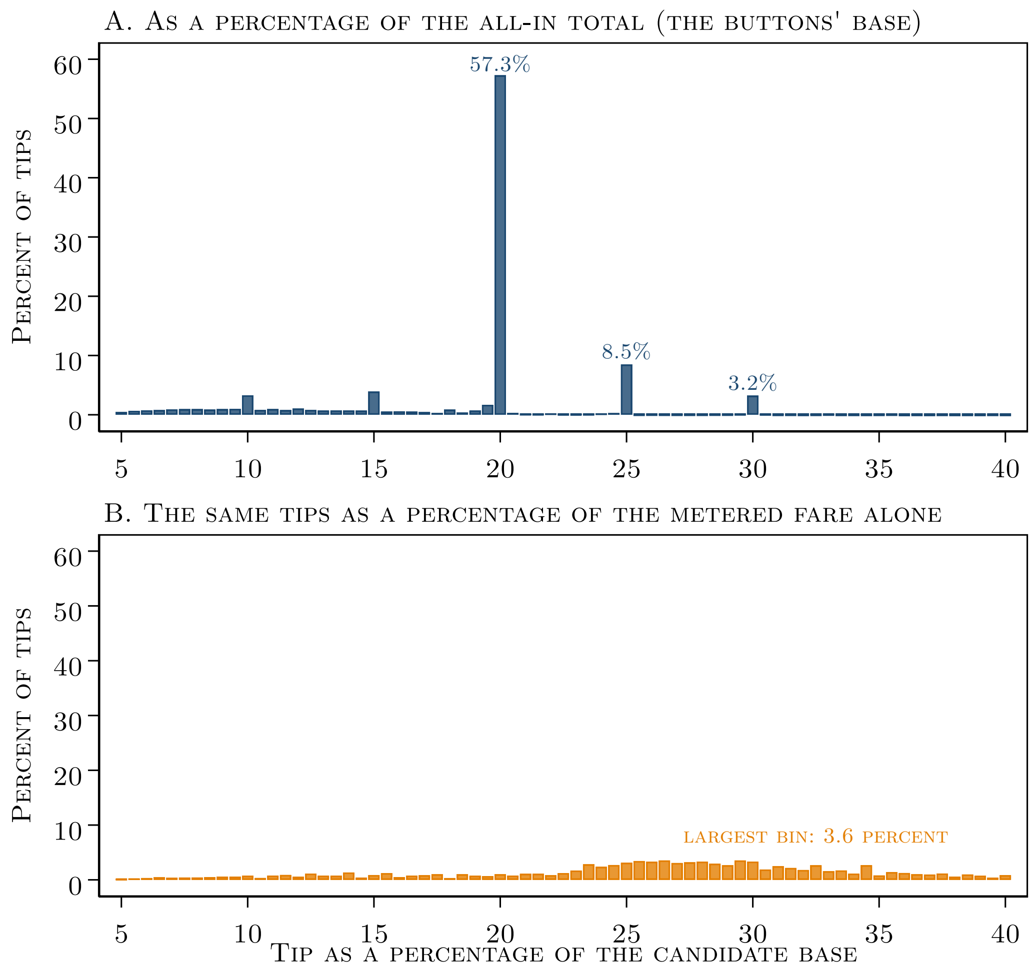
Notes: Illustration of the identifying variation, using the estimation sample's pre-4pm mean metered fare. The two screens show the same trip picked up four minutes apart on a weekday. At 4:00 PM the $2.50 rush-hour surcharge (in orange) enters the total, and because the suggested tips are percentages of that surcharge-inclusive total, every button rises with it—the 20 percent button by $0.50. Nothing about the ride, the driver, or the rider's intended generosity has changed. The paper measures how much of the surcharge reaches the driver through this channel: about fifteen cents per surcharge dollar, averaged over riders who press a button, tip a custom amount, or skip the tip.

What base do the buttons use?. Figure 3 shows the answer before any matching exercise. Expressed as a percentage of the all-in pre-tip total, 57.3 percent of positive credit-card tips fall in the half-point bin at *exactly* 20 percent, with further spikes at exactly 25 and 30—the three buttons. Expressed as a percentage of the metered fare alone, the same tips have no spikes anywhere: no bin holds more than 3.6 percent. The smaller spikes at 10 and 15 percent of the all-in total are custom round percentages computed on the same inclusive base—a first sign, confirmed in Section 5, that the base governs even riders who never press a button.

Table 1 sharpens the fact by vendor. For each payment vendor we compute the share of credit-card tips that equal 20, 25, or 30 percent of a candidate base to within half a cent. In July 2024, before the CBD toll existed, the fee-inclusive and fee-exclusive bases are numerically identical and the match rates coincide exactly, as they must. In June 2025 they diverge sharply: the fee-inclusive total matches 17 to 74 percent of tips depending on vendor, the fee-exclusive total only 4 to 17 percent. All three vendors compute suggested tips on the surcharge-inclusive total.

The table also documents striking vendor heterogeneity in adherence: at one vendor 17 percent

Figure 3: The Base, Drawn: The Same Tips Against Two Candidate Bases



Notes: June 2025, standard-rate credit-card trips with positive tips and no tolls (2.18 million trips), in bins of half a percentage point over 5–40 percent. Panel A expresses each tip as a percentage of the all-in pre-tip total—the base the buttons multiply; annotations mark the bins at exactly 20, 25, and 30 percent. Panel B expresses the same tips as a percentage of the metered fare alone, on identical axes. Only the buttons' base produces spikes.

of tips are exact button presses, at the others 66 and 74 percent. Our estimates aggregate over this heterogeneity, so ρ should be read as a market-wide average rather than a vendor-specific behavioral parameter.

Table 1: Which Base Do the Percentage Buttons Use?

		Share of tips matching 20/25/30% of:		
Vendor, month	Card tips	All-in total	Excl. CBD fee	Fare only
Vendor 1, 2024-07	440,099	0.199	0.199	0.026
Vendor 2, 2024-07	1477491	0.671	0.671	0.017
Vendor 1, 2025-06	461,413	0.166	0.045	0.015
Vendor 2, 2025-06	1709442	0.663	0.133	0.014
Vendor 7, 2025-06	53,040	0.736	0.170	0.012

Notes: Card trips with positive tips and no tolls. July 2024 precedes the CBD congestion fee, so the first two bases are identical by construction; June 2025 follows it. A match requires equality to within half a cent.

Data and estimation cells. Trip records for every yellow-cab trip—pickup timestamp to the second, fare components, tip (recorded for card payments only), payment type, vendor, and pickup zone—are published monthly by the TLC (New York City Taxi and Limousine Commission, 2026). We aggregate to date-by-minute cells in windows around the three clock cutoffs, recording trip counts, mean metered surcharge, mean pre-tip total, mean card tip, the share of tips exactly at a default button, the zero-tip share, and the card share. Separate zone-by-fare-bin cells support the congestion-fee designs. Cells, not raw trips, are the estimation input; the build is deterministic and versioned with the code.

Table 2 describes the sample. The two surcharge regimes are similar in composition—card share near 77 percent, mean tips of \$3.44 and \$4.49, mean pre-tip totals of \$19 and \$26—with the difference in the mean metered surcharge reflecting the December 2022 increase. Roughly half of all credit-card tips are exactly equal to a default button, which is the behavioral fact the rest of the paper builds on.

The motivating pattern. Figure 4 shows the treatment. Mean metered surcharge by minute of day rises in a single step at 4:00pm—by \$0.72 in the \$1.00 regime and \$1.95 in the \$2.50 regime. The step is sharp because the surcharge is assessed on the pickup timestamp; were it assessed at dropoff, the trip-duration distribution would smear it across half an hour.

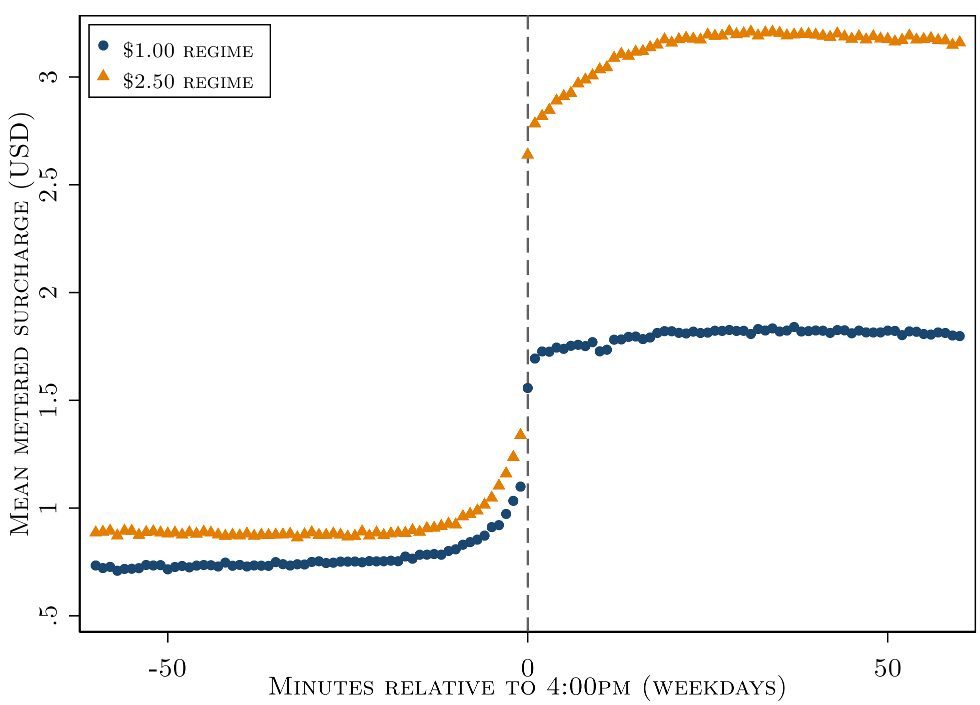
Table 2: Descriptive Statistics by Surcharge Regime

	Value
Lower-dose regime (Sep 2021–Dec 2022)	
Date-by-minute cells	205,753
Trips (millions)	18.0
Card trips (millions)	13.9
Card share of trips (%)	77.0
Mean credit-card tip (USD)	3.44
Mean pre-tip total (USD)	18.96
Mean metered surcharge (USD)	1.18
Tips exactly at a default button (%)	49.4
Zero tips (%)	3.6
Higher-dose regime (Dec 2022–Sep 2024)	
Date-by-minute cells	426,031
Trips (millions)	38.3
Card trips (millions)	29.1
Card share of trips (%)	76.2
Mean credit-card tip (USD)	4.49
Mean pre-tip total (USD)	26.08
Mean metered surcharge (USD)	1.81
Tips exactly at a default button (%)	51.3
Zero tips (%)	5.1

Notes: Date-by-minute cells in the estimation windows around the 4:00pm, 8:00pm, and 6:00am cutoffs, September 2021–September 2024. Trip-weighted means. Tips are observed for credit-card payments only.

The steps fall short of the statutory \$1.00 and \$2.50, and the reason is a property of the data rather than of the fare rules. The TLC's `extra` field records miscellaneous extras, not the rush surcharge alone, and it is contaminated on both sides of the cutoff. In the \$2.50 regime, 23.0 percent of standard-rate card trips in the hour *before* 4:00pm already carry exactly \$2.50 in that field, while 4.5 percent of trips in the hour after carry nothing; the balance sits at values such as \$5.00 and \$7.50 that no single surcharge explains (Table 3). Because the measured step is therefore not a clean statutory dose, we report two scalings throughout: an intention-to-treat effect per *statutory* dollar, which divides by a known constant and requires no assumption about the field, and an instrumental-variables ρ that scales by the measured step. The first is the conservative number and the second requires the discontinuity in `extra` to be entirely rush surcharge—an assumption the composition table does not establish. We report both and flag which is which at every point.

Figure 4: The First Stage: Metered Surcharge by Minute of Day



Notes: Trip-weighted mean of the metered surcharge field by minute of pickup, weekdays excluding holidays. Navy: \$1.00 regime (September 2021–December 2022). Orange: \$2.50 regime (December 2022–September 2024). The dashed line marks 4:00pm.

TABLE 3: WHAT THE SURCHARGE FIELD IS MADE OF

	Share of card trips with `extra` equal to:				Mean
Cutoff, side	\$0	\$1.00	\$2.50	other	`extra`
4:00pm, before	0.685	0.000	0.230	0.084	1.04
4:00pm, after	0.045	0.001	0.659	0.296	3.31
8:00pm, before	0.006	0.007	0.719	0.268	3.25
8:00pm, after	0.001	0.689	0.045	0.265	1.88
6:00am, before	0.024	0.602	0.027	0.347	2.28
6:00am, after	0.651	0.023	0.258	0.068	1.01

Notes: Standard-rate card trips, weekdays excluding holidays, \$2.50 regime, within 60 minutes of each cutoff. The field is contaminated on both sides of every cutoff: a substantial share of trips carries the higher value before the surcharge switches on, and a nonzero share carries none after. This is why the measured step falls short of the statutory dose, and why estimates scaled by the measured step (ρ^{IV}) are reported alongside estimates scaled by the statutory dose (ITT).

4 Empirical Strategy

Pass-through from within-day surcharge discontinuities. Let T_{dm} be the mean credit-card tip on trips picked up on date d in minute-of-day m, S_{dm} the mean metered surcharge, and B_{dm} the mean pre-tip total *net* of the surcharge—the metered fare and other components. Around a clock cutoff c where a surcharge switches, with running variable $r = m - c$ and $D = \mathbf{1}\{r \geq 0\}$, we estimate

$$T_{dm} \;=\; \alpha + \pi D + \gamma B_{dm} + f(r) + \varepsilon_{dm}, \tag{1}$$

with $f(r)$ local linear on each side, cells weighted by card-trip counts, and standard errors clustered by date. Two scalings of π appear throughout. Dividing by the *statutory* dose gives the intention-to-treat pass-through per legislated dollar, the paper's headline estimand. Instrumenting the measured surcharge S_{dm} with D gives $\rho^{IV} = \pi/\phi$, with ϕ the first-stage step; with a continuous treatment this is an average causal response (Angrist & Imbens, 1995), and it carries an exclusion restriction discussed in Section 5.

The preferred specification differences the discontinuity of equation (1) against the weekend one at the same clock, where no surcharge changes (Grembi et al., 2016). With W_d an indicator for weekday dates and slopes interacted with W_d,

$$T_{dm} \;=\; \alpha + \pi D + \omega W_d + \delta\,(D \times W_d) + \gamma B_{dm} + f(r) \times W_d + \varepsilon_{dm}, \tag{2}$$

and δ—the difference in discontinuities—divided by the statutory dose is the headline number. Differencing removes any 4:00pm discontinuity common to weekdays and weekends.

Why the non-surcharge base must be controlled. The term γB_{dm} is not cosmetic; it is the difference between a credible estimate and a biased one. At 4:00pm the tip base rises by \$2.49, of which \$1.95 is the surcharge. The remaining \$0.53 is the non-surcharge base—chiefly the metered fare, which jumps \$0.49 because rush-hour trips are slower and dearer. Omitting B_{dm} attributes those fare-driven tip increases to the surcharge and inflates the estimate by roughly a third. Two diagnostics confirm the correction. First, the weekend placebo—the same clock on days when no rush surcharge exists—falls from \$0.088 to \$0.024 once B_{dm} is included, which is what a correctly specified design should do to a placebo. Second, the corrected estimates agree across designs that have no reason to share a bias, including one identified purely in space (Section 5). Under the weekday–weekend difference the control is nearly redundant: dropping it moves the estimate by about a cent (Section 5).

Identifying assumption. For the single difference, identification requires rider composition and tipping norms to evolve smoothly through the clock cutoff conditional on the fare. For the preferred difference-in- discontinuities the requirement is weaker: whatever jumps at 4:00pm for reasons other than the surcharge—shift changes, light, the rhythm of the working day—must jump equally on weekends, where the clock carries no surcharge. The residual threat is a weekday-specific, non-surcharge discontinuity at exactly the statutory minute; placebo cutoffs at surcharge-free minutes and a permutation over candidate cutoffs bound it. The assumption is also checkable, and Section 5 shows where it fails: at 6:00am on weekdays the dawn compositional shift is discontinuous in precisely the prohibited way, and the design falls apart there while its weekend twin does not. We take the failure as evidence the diagnostics bind, not as grounds to discard them.

Dose, direction, and transportability. The December 2022 increase supplies a second dose at each clock, so comparing discontinuities across regimes tests whether one ρ governs both. The within-day structure also signs the shocks: surcharges switch *on* at 4:00pm (weekdays) and 8:00pm (weekends), and the net surcharge *falls* at 8:00pm on weekdays when the rush surcharge ends and the smaller overnight surcharge begins. Comparing upward with downward changes tests whether pass-through is symmetric.

Finally, the 2019 New York State congestion surcharge and the January 2025 MTA congestion toll moved the same tip base at boundaries in space. Because the fee columns exist only after each policy begins, treated areas must be defined geographically: we learn the set of pickup zones where each fee is actually charged from post-period incidence and apply those zone sets to the pre-period, then estimate difference-in-differences on zone-by-fare-bin cells with zone-fare and month fixed effects, clustering by pickup zone. This design shares no variation with the clock designs, so agreement between them is a genuine out-of-sample test.

5 Results

How to read the estimates. Three numbers appear throughout this section, and they answer different questions.

The reduced form (RF) is the raw causal quantity: how many dollars the average credit-card tip changes at the cutoff, holding the non-surcharge fare constant. It is the coefficient a rider would recognize—"tips went up 29 cents at 4:00pm."

ITT per statutory dollar divides the reduced form by the *statutory* surcharge change ($2.50, $1.00, and so on). Because the statute is a known constant, no first stage is estimated, no exclusion restriction is invoked, and the standard error needs no adjustment. It answers the policy question: for each dollar the rule adds, how much do tips rise on average across *all* standard-rate trips—including any trip the meter failed to charge? It is the design-based number, and the one we headline.

ρ^{IV} instruments the measured surcharge with the cutoff indicator and estimates both stages jointly, so its standard error accounts for the first stage being estimated rather than known. With a continuous treatment it is an average causal response (Angrist & Imbens, 1995), and it is causal only under an exclusion restriction: that the entire discontinuity in the measured surcharge field is rush surcharge. Table 3 is why that assumption cannot be verified—the field is contaminated on both sides of every cutoff—so we read ρ^{IV} as the behavioral rate among trips whose recorded surcharge moved, corroborated externally by the mechanical benchmark (Appendix A). Which scaling matters depends on the question: pricing a fee *before* enactment uses ITT per statutory dollar; aggregating the transfer from fees already collected uses ρ^{IV}, because collections are recorded dollars. ITT sits below ρ^{IV} exactly insofar as some trips' recorded surcharge does not move at the cutoff.

One further habit of reading: every specification conditions on the non-surcharge base (the metered fare and fixed fees). Section 5 shows why this is not optional.

What else changes at 4:00pm: balance. Table 4 reports discontinuities in trip characteristics at the weekday 4:00pm cutoff. The honest news first: trips themselves change. The metered fare jumps \$0.49, trips get 0.58 miles shorter and 0.69 minutes slower—the signature of rush-hour congestion. Continuity of trip characteristics fails, which is why we describe the design as a *fare-conditioned* within-day discontinuity rather than a conventional regression discontinuity: identification leans on conditioning on the fare, and on the weekend contrast, not on an assumption that nothing else moves at 4:00pm.

What does *not* change is the composition of payers: the card share moves 0.2 percentage points and the vendor mix 0.03 points. The people tipping on either side of the cutoff are paying the same way, through the same machines.

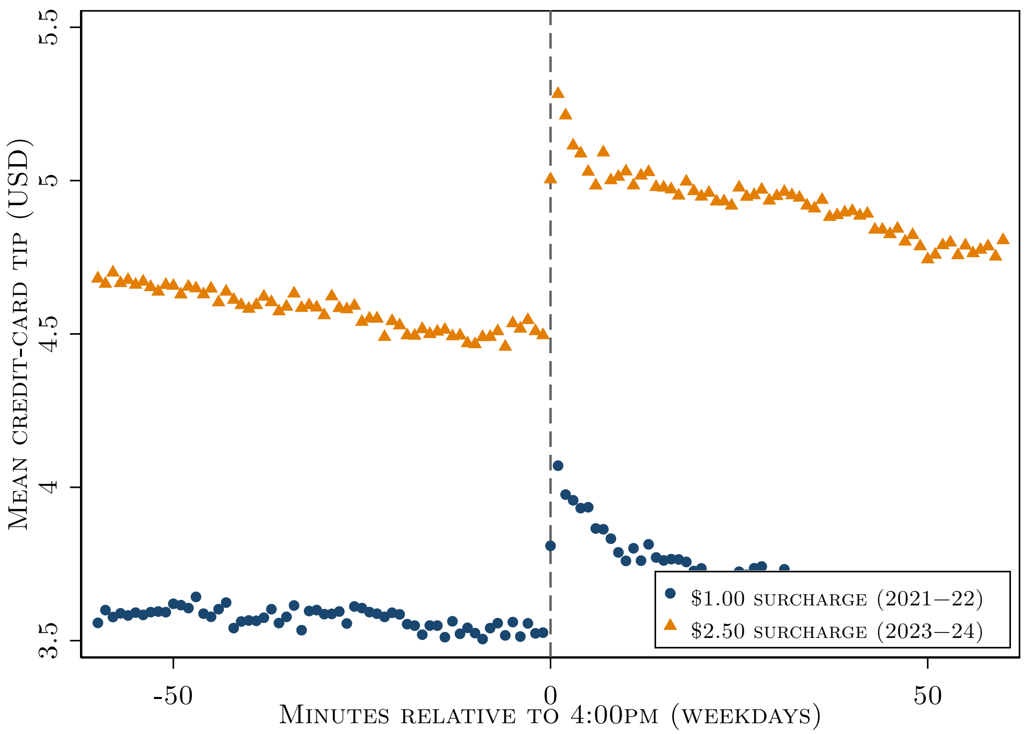
Table 4: Balance at the 4:00pm Cutoff

Trip characteristic	Pre-cutoff mean	Jump at 4:00pm	(SE)
Metered fare (USD)	17.695	0.4862***	(0.0353)
Trip distance (miles)	2.962	-0.5827***	(0.0985)
Trip duration (minutes)	18.076	0.6919***	(0.0819)
Card share of trips	0.804	0.0019***	(0.0007)
Vendor 2 share of card trips	0.739	0.0003	(0.0008)

Notes: Weekday standard-rate card trips, \$2.50 regime, bandwidth ±60 minutes, local linear, clustered by date. The fare, distance, and duration jumps are why every estimate in this paper conditions on the non-surcharge base; the null card-share and vendor-share jumps are why payment composition is not a confound. ***, **, *: significance at 1, 5, 10 percent.

The main estimates. Figure 5 shows the raw pattern the estimates formalize: tips step up at 4:00pm in both regimes, by visibly more when the surcharge is larger, then drift down through the evening as post-rush trips shorten and cheapen—the fare dynamics the specification conditions away. Figure 6 shows the same clock on weekends, where no surcharge exists: drift without a step.

Figure 5: Credit-Card Tips by Minute Around 4:00pm

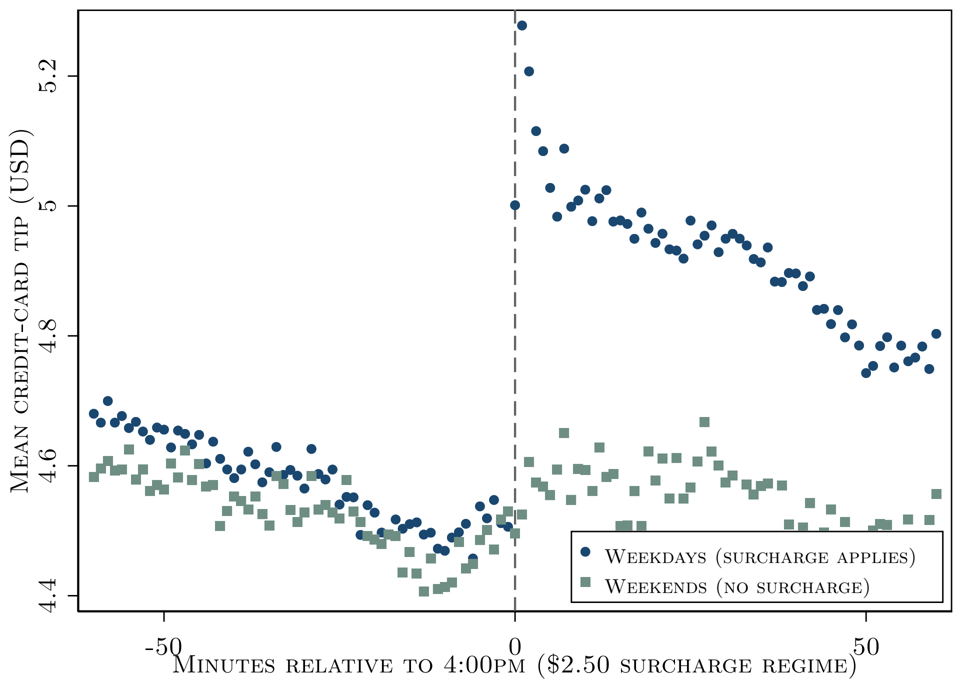


Notes: Standard-rate card trips, weekdays excluding holidays, trip-weighted means by minute of pickup. Navy: \$1.00 regime. Orange: \$2.50 regime. The step at the dashed line is the treatment; the gradual decline afterward is rush-hour fare dynamics, absorbed by the fare control in estimation.

Table 5 is the paper's central table: the same specification at all six within-day surcharge changes, with the two weekend placebos beneath. Reading across the top row: at the weekday 4:00pm cutoff in the \$2.50 regime, tips rise \$0.292 (SE \$0.005); per statutory dollar that is 0.117; instrumenting the measured surcharge, $\rho^{IV} = 0.150$ (SE 0.002). Reading down the ρ^{IV} column is the paper's result: 0.150, 0.141, 0.134, 0.134, 0.122, 0.128. Two doses, two clocks, both directions, one parameter near 0.14. The placebo rows show what the same machinery finds when no surcharge changes: \$0.024 and $-$\$0.002.

Figure 7 is the replication argument in one frame: every locked change, expressed per statutory

Figure 6: Weekdays Versus Weekends at the Same Clock



Notes: Mean credit-card tips by minute around 4:00pm, $2.50 regime, standard-rate card trips. Weekdays (navy) receive the rush surcharge from 4:00pm; weekends (teal) never do.

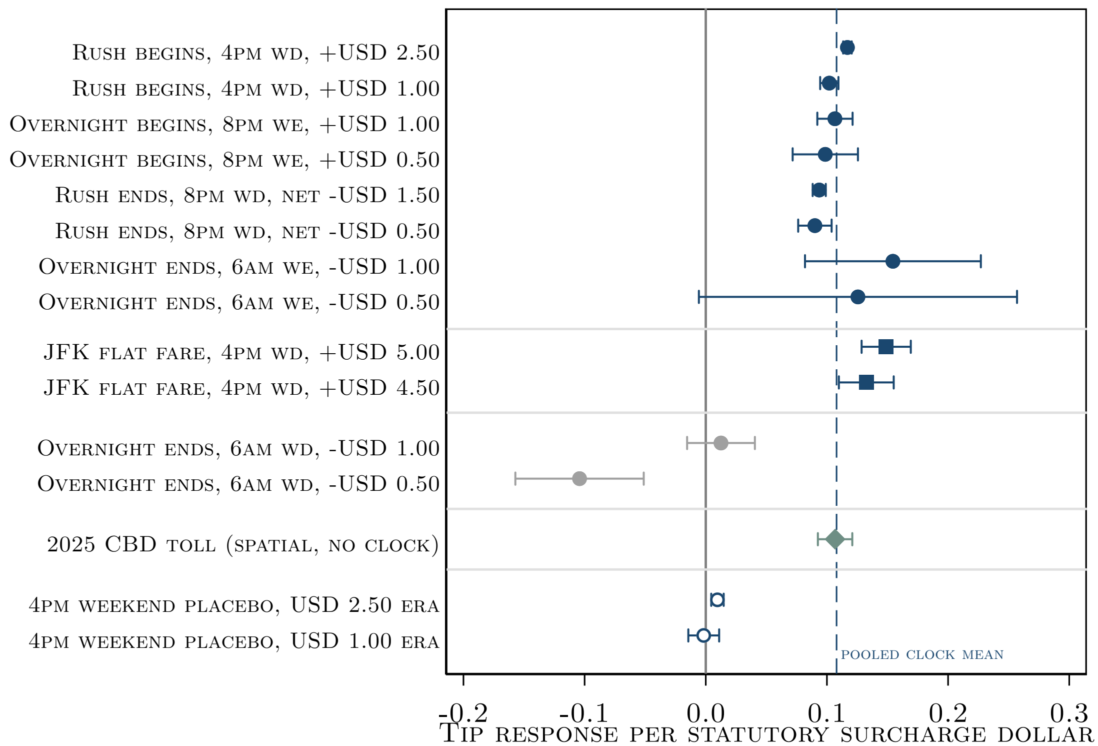
Table 5: Pass-Through at Six Surcharge Changes

	Reduced form (USD)		ITT per	ρ^{IV}	
Surcharge change	Tip step	(SE)	statutory $	Estimate	(SE)
Rush begins 4pm, wk, USD 2.50	0.292***	(0.005)	0.117	0.150	(0.002)
Rush begins 4pm, wk, USD 1.00	0.102***	(0.004)	0.102	0.141	(0.005)
Overnight begins 8pm, wke, USD 1.00	0.107***	(0.007)	0.107	0.134	(0.009)
Overnight begins 8pm, wke, USD 0.50	0.049***	(0.007)	0.099	0.134	(0.018)
Rush ends 8pm, wk, net minus USD 1.50	-0.140***	(0.004)	0.094	0.122	(0.003)
Rush ends 8pm, wk, net minus USD 0.50	-0.045***	(0.004)	0.090	0.128	(0.010)
4pm, weekend placebo, USD 2.50 era	0.024***	(0.007)	—	—	
4pm, weekend placebo, USD 1.00 era	-0.002	(0.006)	—	—	

Notes: Standard-rate card trips, September 2021–September 2024, bandwidth ±60 minutes, local linear in pickup minute, conditioning on the non-surcharge base, weighted by card trips, clustered by date. ITT divides the reduced form by the statutory dose (a known constant, so its SE is the RF's SE rescaled). ρ^{IV} instruments the measured card-trip surcharge with the cutoff indicator; its SE reflects joint estimation of both stages. Weekend placebo rows apply the identical specification where no surcharge changes. ***, **, *: significance at 1, 5, 10 percent.

dollar, on one axis. The six primary changes span 0.090–0.117 with a precision-weighted mean of 0.108 (SE 0.001). The 6:00am weekend changes sit at 0.126 and 0.155 with wide intervals. The JFK flat-fare surcharge (squares; Section 5) contributes the \$4.50 and \$5.00 doses at 0.133 and 0.149. The 2025 congestion toll, estimated with no clock at all, lands at 0.107. The weekend placebos sit at zero, and the two 6:00am weekday failures appear in grey rather than disappearing (Section 5).

Figure 7: One Parameter: Pass-Through per Statutory Dollar at Every Locked Change



Notes: Intention-to-treat tip response per statutory surcharge dollar, with 95 percent confidence intervals; specification of Table 5 throughout. Navy circles: within-day surcharge changes on the standard-rate sample. Navy squares: the JFK flat-fare surcharge (Section 5), whose higher ITT reflects near-complete treatment delivery (94 versus 78 percent of the statutory dose), not different behavior. Grey circles: the 6:00am weekday changes, where the dawn compositional shift breaks the fare conditioning (Section 5). Teal diamond: the January 2025 congestion toll, estimated from the spatial origin–destination design with no clock involved, divided by the \$0.75 statutory toll. Hollow circles: weekend placebos at 4:00pm, scaled by the matched regime's dose. The dashed line marks the precision-weighted mean of the six primary standard-rate changes, 0.108 (SE 0.001). Minimum detectable effects (2.8 × SE) for the six primary changes range from 0.005 to 0.039 per statutory dollar.

Estimated as an explicit difference-in-discontinuities—the weekday step minus the weekend step at the same clock (Grembi et al., 2016), the specification of equation (2)—the effect is \$0.271 (SE \$0.008) in the \$2.50 regime and \$0.105 (\$0.008) in the \$1.00 regime: 0.108 and 0.105 per statutory dollar. Dropping the fare control entirely moves these to 0.120 and 0.092. The conclusion does not lean on the functional form of the fare adjustment; the weekend difference carries it.

The fuzzy version of the same design reports all three quantities of the standard presentation. The first stage—the weekday–weekend difference in the *surcharge* discontinuity—is \$1.934 (SE \$0.011) in the \$2.50 regime; the reduced form is the \$0.271 above; and the just-identified 2SLS coefficient, instrumenting the measured surcharge with the cutoff-by-weekday interaction, is $\hat{\tau} =$ 0.140 (SE 0.004). In the \$1.00 regime the same three numbers are \$0.713 (\$0.007), \$0.105 (\$0.008), and 0.147 (0.011): the fuzzy estimate is dose-stable, and it sits inside the ρ^{IV} family, as it must, since it differs from the single-cutoff IV only in differencing the weekend counterfactual through both stages.

A working-day control, and a \$5 dose on a fare that cannot move. Two further designs close the remaining doors on the 4:00pm result, and both were run after the main estimates as out-of-sample checks.

Holidays. Legal holidays are calendar weekdays on which the rush surcharge does not apply—a working-day counterfactual that weekends cannot supply. The federal holiday list is not the TLC list, so we classify each holiday by its own first stage: the traditional holidays are cleanly exempt, one observed date is partially charged and dropped, and Juneteenth is *fully charged* in all three years (first stages of \$0.90, \$2.29, and \$2.28 against statutory doses of \$1.00 and \$2.50)—the meters do not observe the newest federal holiday. Differencing the weekday step against the exempt holidays gives 0.096 per statutory dollar in the \$2.50 regime and 0.095 in the \$1.00 regime (wild-cluster bootstrap $p < 10^{-4}$ and $p = 0.035$; holiday clusters are few, so the bootstrap is the relevant inference)—the same answer the weekend difference gives. And Juneteenth supplies the converse check: a day with holiday *composition* but the full *dose* shows a tip step of \$0.241 (SE \$0.067), 0.096 per statutory dollar. Composition off, dose on: the step follows the dose.

JFK flat fares. The analysis plan quarantines special rate codes from the main sample; analysed separately, the JFK flat fare is the cleanest design in the paper. Its rush-hour surcharge is \$4.50 in the early regime and \$5.00 after December 2022—twice the largest dose in the main design—and it sits on a *flat* fare, so the metered fare cannot jump at 4:00pm by construction: the confound that motivates the fare conditioning everywhere else does not exist here. The estimates behave accordingly. The tip step is \$0.745 (SE \$0.052) at the \$5.00 dose and \$0.596 (\$0.052) at \$4.50—0.149 and 0.133 per statutory dollar—and dropping every control changes the answer by three hundredths of a cent (0.1486 against 0.1489). Weekend placebos are null on both sides (tip −\$0.03 and −\$0.05, first stages −0.01 and −0.04). Per *recorded* surcharge dollar the JFK estimates are 0.159 and 0.150, on top of the main design's ρ^{IV}; the higher ITT relative to the pooled 0.108 is treatment delivery, not behavior—JFK meters deliver 94 percent of the statutory dose against 78 percent on standard rates. The mechanical benchmark tells the same story at five times the stakes: the pre-4:00pm JFK tip rate of 17.5 percent predicts a step of \$0.875 under complete pass-through, and the estimate is 85 percent of it, the same fraction as the main design. Riders tipping on an \$85 airport fare pass a \$5 surcharge into the tip at the same rate riders tipping on a \$25 Manhattan fare pass \$2.50.

Interpreting the magnitude: the average tip in this sample is 17–18 percent of the base. If riders treated a surcharge dollar exactly like a fare dollar, complete default-mediated pass-through would put ρ near 0.17; if riders fully distinguished the surcharge—tipping on the service, not the stack—ρ would be 0. The estimate of 0.14–0.15 says riders pass surcharges into tips at about four-fifths of the rate they tip the fare itself: not total blindness to the base, but close.

Robustness. Table 6 shows ρ^{IV} across bandwidths of 30 to 90 minutes (0.147–0.156) and donuts excluding 3, 5, or 10 minutes around the cutoff (0.148–0.152). Because the surcharge is assessed on the pickup time, the donut is not needed for assignment reasons; we report it because cells adjacent to the cutoff mix minutes unevenly within trips, and it makes no difference. Placebo cutoffs at 3:15pm and 5:00pm give tip-side effects of −\$0.0003 (SE \$0.007) and −\$0.008 (SE \$0.007)—both sides of the placebo, first stage and tip, are now reported and both are null. A permutation over 60 candidate cutoffs, each estimated at the same ±30 minute bandwidth as the true cutoff and drawn only from windows that never cross 4:00pm, finds none with an effect as large as the true cutoff's, and a wild cluster bootstrap on the corrected reduced form rejects zero at $p < 10^{-6}$. Minimum detectable effects—2.8 times the standard error, the analysis plan's locked criterion—range from 0.005 to 0.039 per statutory dollar across the six primary changes, so even the least-powered design could detect an effect one-third the size of the pooled estimate.

Table 6: Bandwidth and Donut Robustness, ρ^{IV}

Specification	ρ^{IV}	(SE)
Bandwidth $\pm$30 min	0.147***	(0.004)
Bandwidth $\pm$45 min	0.147***	(0.003)
Bandwidth $\pm$60 min	0.150***	(0.002)
Bandwidth $\pm$90 min	0.156***	(0.002)
Donut $\pm$3 min	0.148***	(0.002)
Donut $\pm$5 min	0.150***	(0.002)
Donut $\pm$10 min	0.152***	(0.003)

Notes: Weekday 4:00pm cutoff, \$2.50 regime, corrected specification, clustered by date. ***, **, *: significance at 1, 5, 10 percent.

The third clock: 6:00am, and where the design breaks. The analysis plan locks a third clock. At 6:00am the overnight surcharge *ends*, so the statutory step is downward—\$0.50 before December 2022, \$1.00 after—and because the overnight surcharge applies on every day of the week, there is no weekend placebo at this cutoff. That absence turns out to be informative.

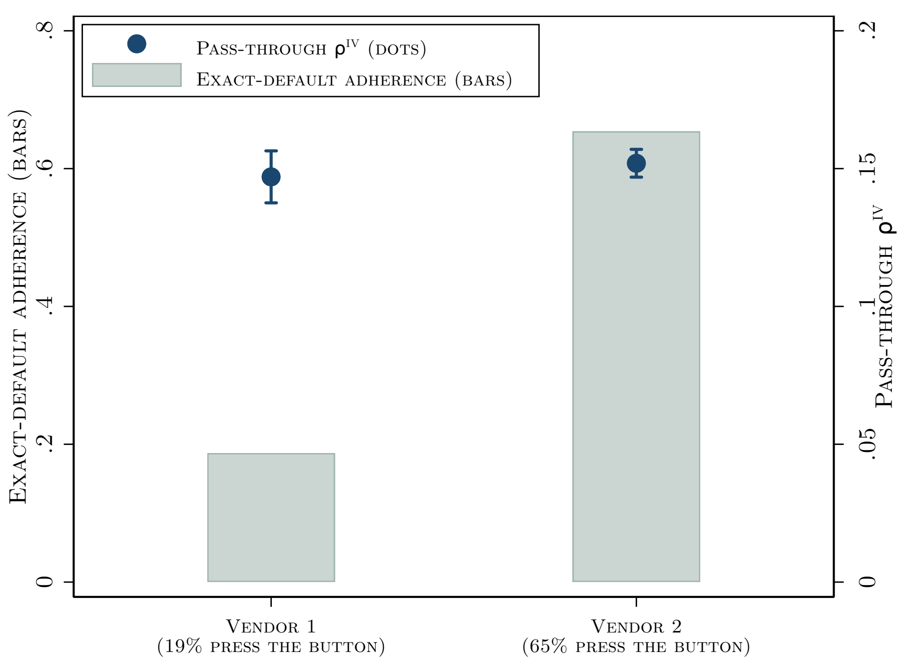
On weekends the estimates behave: $\rho^{IV} = 0.175$ (SE 0.041) at the \$1.00 dose and 0.165 (0.088) at \$0.50, a little above the main cluster but within its neighborhood. On weekdays they do not: the \$1.00 dose gives 0.016 (0.018) and the \$0.50 dose gives -0.143 (0.037)—null and wrong-signed respectively.

We report this because the plan locked it, and because the pattern diagnoses itself. The same clock and the same statutory dose produce sensible estimates on weekends and broken ones on weekdays. What differs between them is what happens at dawn on a working day: the pickup mix shifts sharply toward airport-bound and commuter trips, which are longer and dearer in a way that conditioning on the mean non-surcharge fare does not absorb. On weekends that shift is absent and the design recovers the same parameter as the 4pm and 8pm cutoffs. The honest reading is that this design identifies pass-through wherever rider composition evolves smoothly through the cutoff, and that the 6:00am weekday window is the one place in our data where it plainly does not. We therefore exclude the two weekday 6am estimates from the headline range and report them here rather than in Table 5.

Do riders undo it?. At the 4:00pm cutoff the share of tips exactly at a default button falls 0.85 percentage points, zero-tipping rises 0.15 points, and card use *rises* 0.19 points. Against a mechanical benchmark—the surcharge step times the prevailing 17.4 percent tip rate—the estimated pass-through is about nine-tenths of complete. Riders offset almost none of the base change.

The button is not the mechanism: vendor evidence. The two payment vendors offer a built-in test of *how* the pass-through operates, and it refutes our own initial hypothesis. If pass-through worked purely through riders pressing default buttons, it should scale with button use: Vendor 1's riders take an exact default on 19 percent of tips, Vendor 2's on 65 percent, so ρ should differ by a factor of three. It does not. Estimated separately on the same 4pm design, $\rho^{IV} = 0.147$ (SE 0.005) at Vendor 1 and 0.152 (0.003) at Vendor 2—statistically indistinguishable. Riders who type custom amounts evidently compute their percentages on the same fee-inclusive total the screen displays. The operative mechanism is the *base the interface presents*, not the button that presents it, which broadens the result: redesigning buttons would not undo the fiscal channel; redesigning the base would. Figure 8 shows the contrast in one frame: adherence bars that differ by a factor of 3.5, pass-through dots that coincide.

Figure 8: Same Pass-Through, Very Different Button Use



Notes: Bars (right axis): share of card tips exactly at a default button, pre-cutoff, by payment vendor. Dots (left axis, 95 percent CIs): ρ^{IV} from the weekday 4:00pm design estimated separately by vendor, \$2.50 regime. If pass-through operated only through button presses, the dots would differ as much as the bars.

Who passes it through. The mechanism has a sharper test than vendor comparisons: classify every tip by *how it was chosen*, and ask whose tips move at the cutoff. We split card tips into four mutually exclusive types—exactly a 20/25/30 percent button, a whole-dollar amount, other custom amounts, and zero—and estimate the weekday–weekend difference-in- discontinuities separately for each type's mean tip and its share of trips (Table 7).

If pass-through were mechanical rounding—riders nudging totals to the nearest dollar—whole-dollar tips would carry it. They do not. In the $2.50 regime, button tips rise $0.421 (SE $0.011) and custom non-round tips $0.247 ($0.027), while whole-dollar tips move $0.043 ($0.025). The dose comparison is more damning still: when the statutory dose was $1.00, button tips rose $0.182—the response scales one-for-one with the legislated dose—while the whole-dollar response is $0.04 in *both* regimes, insensitive to a 150 percent dose change. Riders who think in percentages of the displayed total pass the surcharge through; riders who think in round dollar amounts are inert. And nobody switches type at the cutoff: every share moves by less than half a percentage point. The aggregate offset from complete mechanical pass-through is not partial attention spread thinly—it is the whole-dollar fifth of riders sitting still.

Table 7: Who Passes It Through: DiDisc by Tipping Type

	DiDisc	(SE)	Per statutory $
Exact-button tips	0.421	(0.011)	0.168
Whole-dollar tips	0.043	(0.024)	0.017
Custom (non-round) tips	0.247	(0.027)	0.099
Share: button	-0.0047	(0.0018)	—
Share: whole-dollar	0.0048	(0.0015)	—
Share: custom	-0.0027	(0.0016)	—
Share: zero tip	0.0026	(0.0007)	—

Notes: Weekday–weekend difference-in-discontinuities at 4:00pm, $2.50 regime, fare-conditioned, clustered by date; tip rows weight by the category's own trips, share rows by card trips. Types are mutually exclusive: button = tip equals 20/25/30 percent of the pre-tip total to within half a cent; whole-dollar = round-dollar tips not at a button. In the $1.00 regime the button response is $0.182 (0.009) and the whole-dollar response $0.043 (0.027): button pass-through scales with the dose, whole-dollar tipping is inert at both doses.

Demand: riders do not dodge, and barely notice. Three further margins complete the behavioral picture, and all three are near-zero in the same direction.

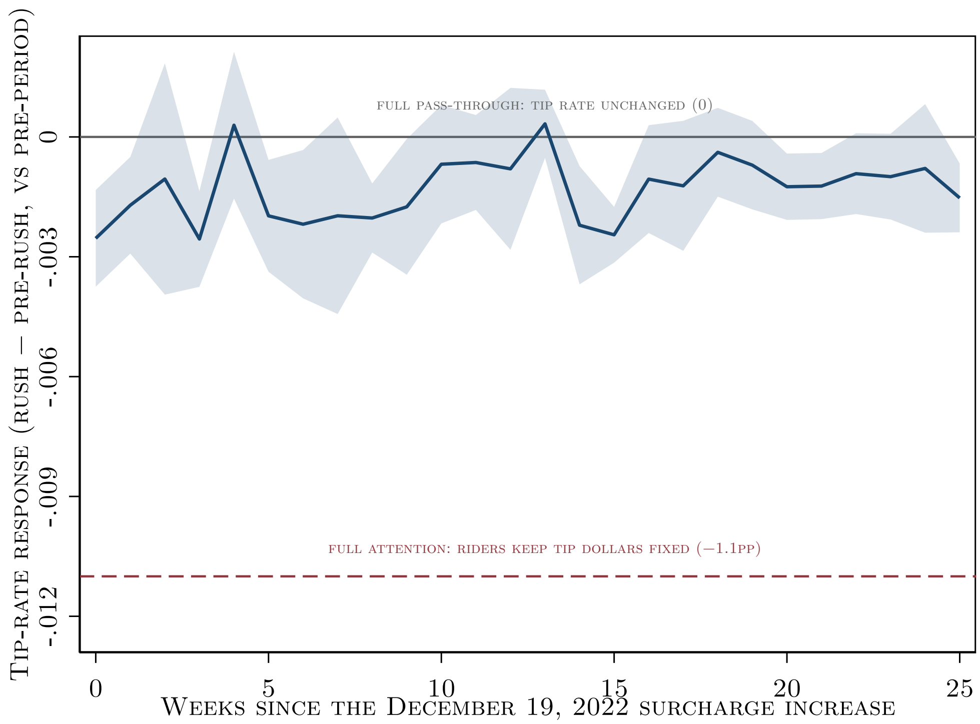
No retiming. If riders timed hails to dodge the surcharge, pickup density would pile up before 4:00pm and hollow out after. It does not: the share of window pickups per minute is essentially identical across the cutoff (ten-minute ratio 0.997 at the \$1.00 dose, 0.987 at \$2.50; Figure 10). A 150 percent increase in the fee produced a one percent change in timing. This doubles as the design's sorting diagnostic: the running variable is not manipulated.

No demand response at the toll. Log eligible trips at the January 2025 congestion boundary move −1.0 percent (SE 0.9)—statistically and economically indistinguishable from zero for a \$0.75 fee on a \$27 ride. Combined with the tip pass-through, the driver-side ledger of the toll is striking: drivers gain roughly \$2.2 million a year in induced tips on eligible card trips and lose no detectable fare revenue. At the point estimate, the congestion toll made drivers net *better* off; the confidence interval cannot exclude a net loss if the true demand decline sits at its lower bound.

No reaction to a 150 percent surcharge increase. On December 19, 2022 the rush surcharge jumped from \$1.00 to \$2.50 overnight. If riders monitored their totals, percentage tipping should have dropped by about 1.1 points in the rush window (the fall required to keep tip dollars fixed). Netting seasonal drift with the surcharge-free pre-rush window, the tip rate fell 0.22 points (SE 0.07)—riders undid about a fifth of the increase and passed the rest through (Figure 11). That one-fifth offset independently implies pass-through of roughly four-fifths of the prevailing tip rate, which is precisely where the clock designs put ρ. And the offset is not the start of slow learning: Figure 9 traces the response horizon by horizon, local-projection style, for six months after the repricing. The path settles immediately at about −0.2 points and never drifts toward the −1.1-point full-attention benchmark. Riders who were going to notice noticed at once; the rest never did.

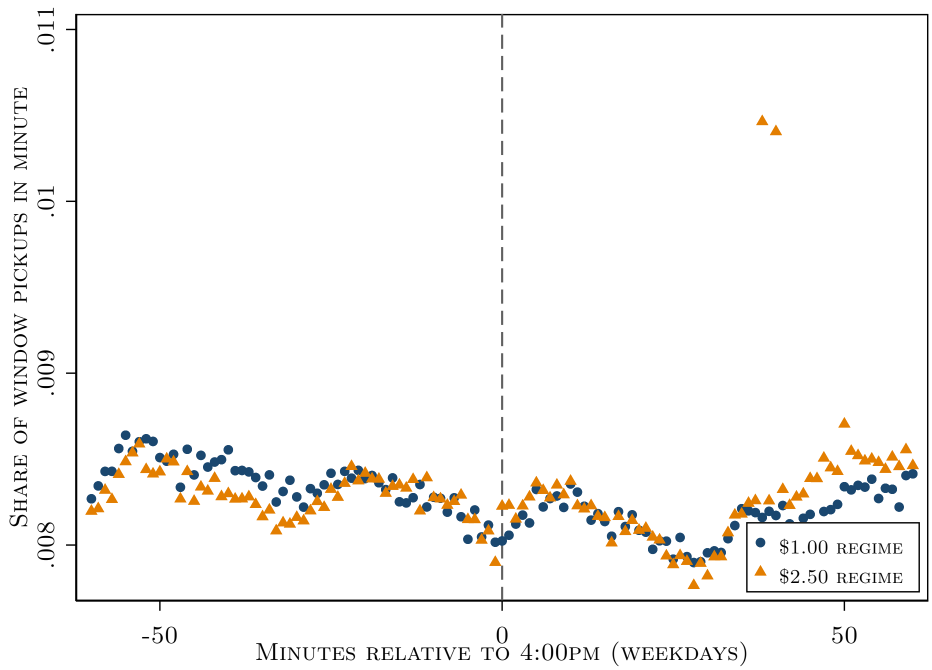
Transportability: congestion fees. The clock designs identify ρ from surcharges that switch on a schedule; congestion policy switches them on at boundaries in space. The re-estimated design assigns treatment by the statutory rule—a trip is eligible if its pickup *or* dropoff lies in the fee zone—on standard-rate card trips, and reports joint pre-trend tests. At the January 2025 toll the origin–destination design gives a pooled tip effect of \$0.080 (SE \$0.005) against a first stage of \$0.602 of fee per card trip: $\rho = 0.133$ (SE 0.009), within a cent of the clock designs. Correcting the eligibility matters exactly as contamination predicts: the old pickup-only assignment gave \$0.058, because treated dropoffs sat in its control group, and its apparent post-event fade disappears under the correct rule—post coefficients are stable at \$0.041–\$0.052 across all six months. Two refinements

Figure 9: No Delayed Learning: The Adaptation Path After December 2022



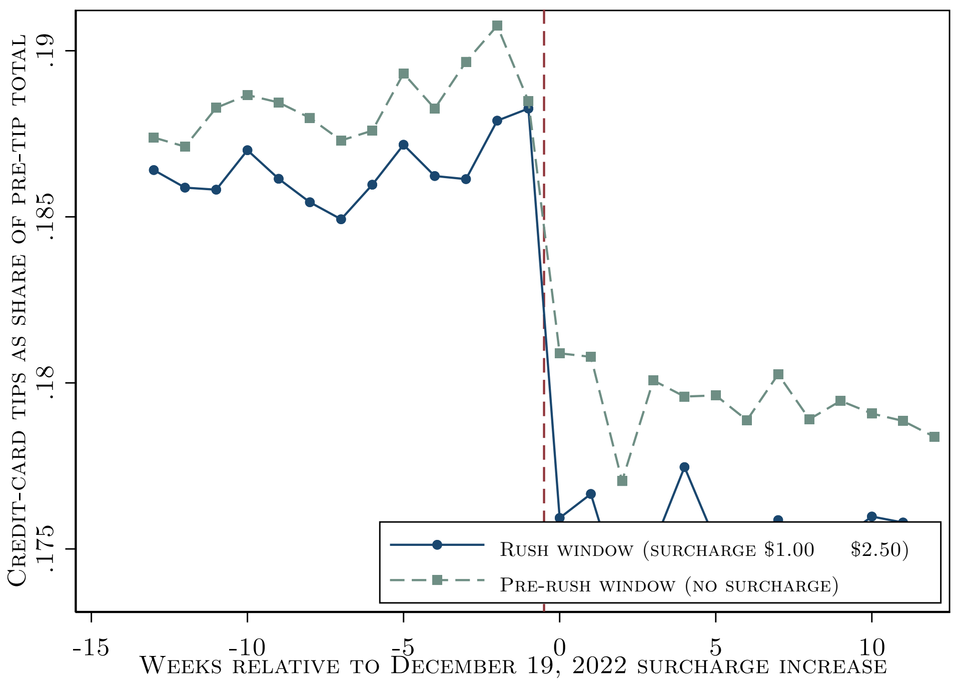
Notes: Weekly horizon-by-horizon responses of the rush-window tip rate, netted against the surcharge-free pre-rush window and measured relative to the 26-week pre-period mean, for 26 weeks after the $1.00-to-$2.50 repricing. The shaded band is a 95 percent interval from daily variation within each week. The maroon line marks the response that would keep tip dollars unchanged; zero marks full pass-through.

FIGURE 10: PICKUP DENSITY THROUGH THE 4:00PM CUTOFF



Notes: Each date's share of window pickups in each minute, averaged by minute, weekdays. Smooth density through the cutoff means riders do not retime hails around the surcharge and the running variable is unmanipulated.

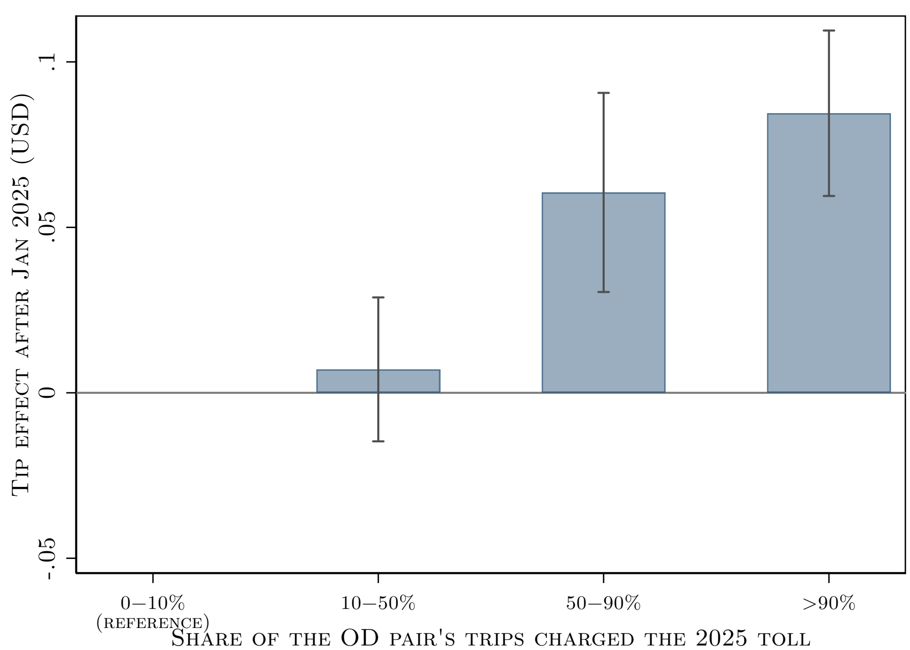
FIGURE 11: THE DECEMBER 2022 SURCHARGE INCREASE: DID ANYONE NOTICE?



Notes: Weekly mean credit-card tips as a share of the pre-tip total, weekdays, in the rush window (navy; surcharge rose $1.00 to $2.50 at the dashed line) and the surcharge-free pre-rush window (teal). Keeping tip dollars fixed would require the navy series to fall about 1.1 points relative to teal; it falls 0.22.

discipline the pre-period. First, seasonality: December's holiday tipping makes any single reference month misleading, so we add treated-by-calendar-month fixed effects, identified from a second pre-year (2023), and measure post effects against all pre months. The effects strengthen and stabilize—\$0.074 to \$0.089 in every one of the six post months ($t > 8$; Figure 13)—but a joint Wald test on the 2024 pre-year against 2023 still rejects ($p = 0.0007$): treated zones genuinely drifted during 2024's recovery, beyond seasonality. Second, and more decisively, the eligibility rule itself supplies a *continuous dose*: the post-period share of each pickup–dropoff pair's trips actually charged the fee, which nets out the exempt highway corridors no geometric rule can see. Conditioning on the non-surcharge base—the same correction the clock designs required—the dose design yields a monotone staircase: pairs charged up to half the time show \$0.007 (SE \$0.011), pairs charged 50–90 percent show \$0.061 (0.015), and pairs charged above 90 percent show \$0.085 (0.013), for an implied $\rho = 0.121$. Effects that rise step-for-step with measured treatment intensity (Figure 12) are difficult to attribute to zone-level drift, which would have to reproduce the dose gradient by coincidence.

Figure 12: The Dose Staircase: Tip Effects Rise with Charging Intensity



Notes: Post-toll tip effects by the share of each pickup–dropoff pair's trips actually charged the 2025 congestion fee (measured from post-period fee incidence), relative to pairs charged up to 10 percent of the time. Fare-conditioned, OD-pair and month fixed effects, clustered by pickup zone; bars are 95 percent confidence intervals.

Stacking both fee adoptions into one design—each cohort with its own unit and time effects, tips regressed on the observed fee instrumented by eligibility-times-post, fare-conditioned throughout—yields the *average rollout effect* of putting a fee on the meter:

$$\bar{\rho}_{\text{rollout}} = 0.163 \text{ (SE 0.021)}.$$

Within the same specification the cohorts are 0.171 (0.033) for 2019 and 0.138 (0.007) for 2025, statistically indistinguishable ($p = 0.36$). The uniform fare conditioning also resolves this section's one outlier: the earlier 2019 estimate of 0.218 carried the same omitted-fare bias the clock designs taught us to remove, and falls into line once treated identically. We still keep the spatial evidence in a corroborating role—the binary event studies do not pass a strict pre-trend test and we say so—while noting that the dose gradient and the pooled rollout estimate land within two cents of the clock designs.

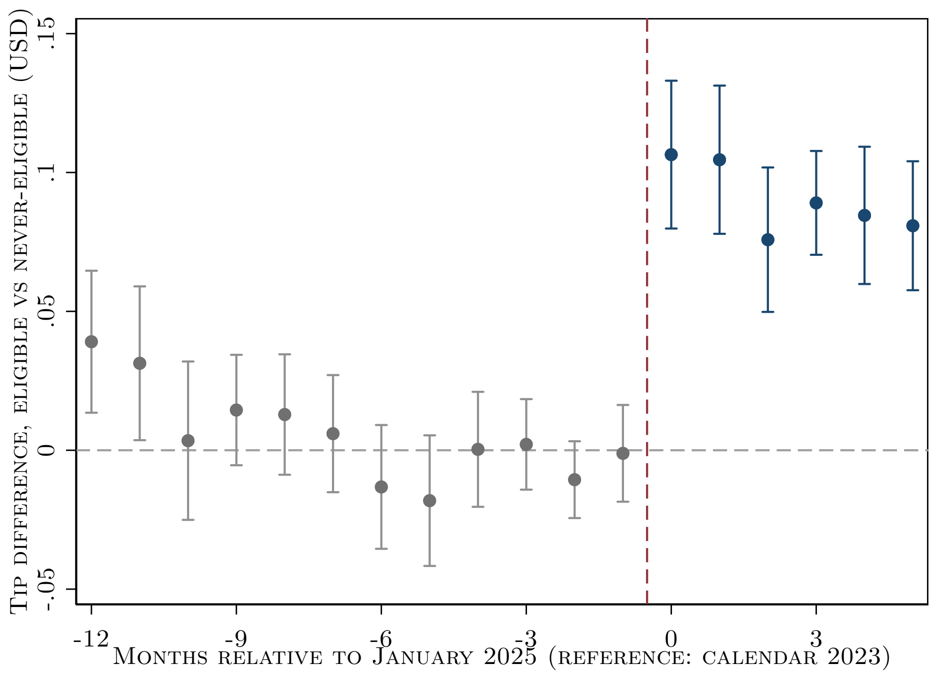
Table 8: Pass-Through at Congestion Fees: The Initial Pickup-Only Design

Policy	First stage	Tip effect	(SE)	$\hat{\rho}$
2019 congestion surcharge	2.180	0.555	(0.104)	0.254
2025 CBD congestion toll	0.426	0.062	(0.007)	0.146

Notes: Zone-by-fare-bin difference-in-differences, zone-fare and month fixed effects, clustered by pickup zone. Windows: 2018m7–2019m12 and 2024m1–2025m6. *This is the initial pickup-only specification and is superseded*: it assigns treatment by pickup zone alone, so trips whose dropoff was in the fee zone sit in its control group. It is retained to show what the contamination does. The preferred estimates—origin–destination eligibility, fare conditioning, and the continuous dose—are reported in the text and give ρ of 0.133, 0.121, and 0.163 rather than the 0.146 and 0.254 shown here.

The size and geography of the transfer. At $\rho = 0.15$, the \$127.2 million of 2019 congestion surcharges on card trips implies roughly \$18.7 million of additional gratuities, and the 2025 toll's \$17.4 million annualized implies about \$2.2 million a year. Figure 14 is the spatial design's summary exhibit: the per-zone, fare-conditioned change in tips—measured relative to never-eligible zones—estimated at a placebo date (July 2024) and at the real toll (January 2025), with zones muted unless significant at five percent. At the placebo date, 6 percent of non-CBD zones and 10 percent of CBD zones show anything, scattered at random. At the real date, **82 percent of congestion-district zones turn significantly positive** (mean +\$0.084) against 18 percent elsewhere: the

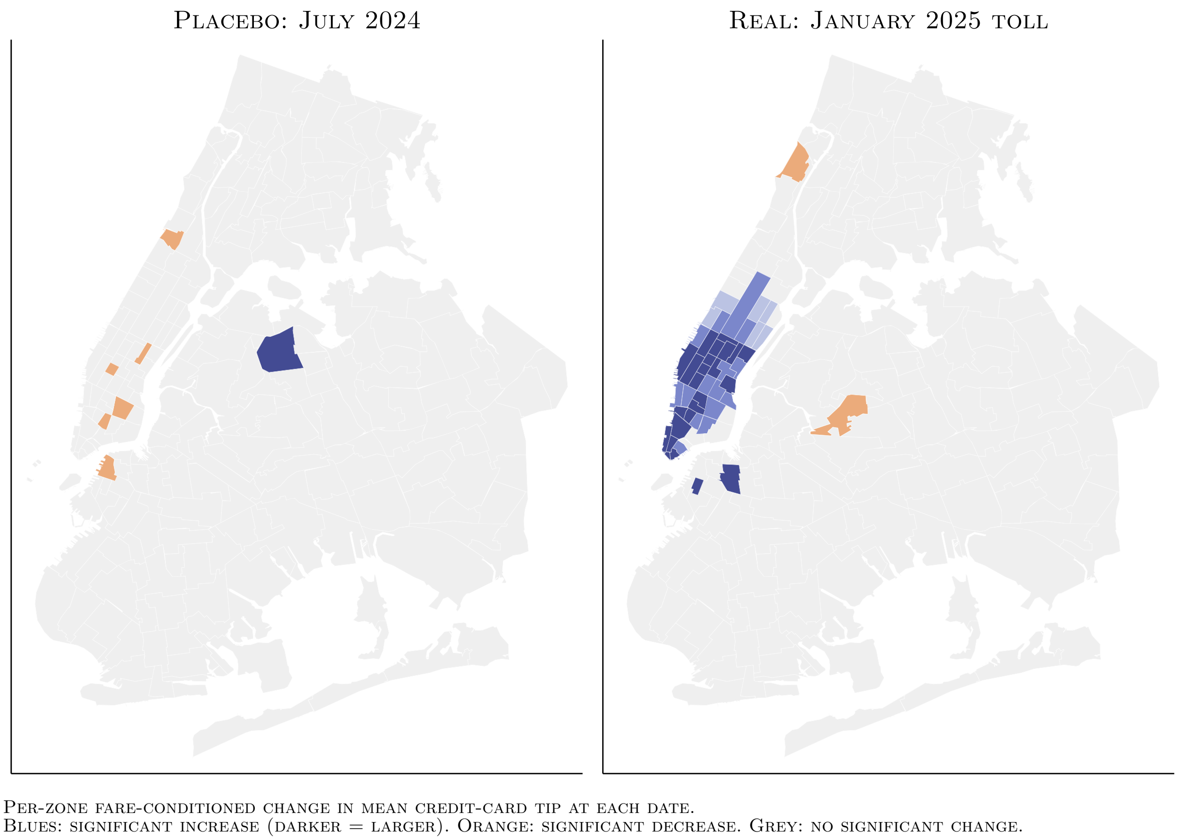
Figure 13: Event Study: The January 2025 Congestion Toll



Notes: Corrected design: origin–destination eligibility with treated-by-calendar-month fixed effects identified from calendar 2023; the twelve 2024 pre-months are shown as deviations from that seasonal baseline, and post coefficients are stable at \$0.074–\$0.089 with $t > 8$ throughout. The pre-period looks flat, and no single 2024 coefficient is large—but a joint Wald test on all twelve rejects ($p = 0.0007$): treated routes drifted slightly during 2024's recovery. That is why the spatial evidence is read as corroboration, anchored by the dose gradient of Figure 12, rather than as the headline. Bars: 95 percent confidence intervals.

tolled district materializes from the estimates alone, with no boundary drawn. Figure 15 then maps where the resulting transfer is generated: congestion-district pickup zones average $46,000 of induced tips a year each, against $900 outside, so the policy boundary is visible in the map without being drawn.

Figure 14: The Toll Appears in the Map: Placebo Versus Real Date

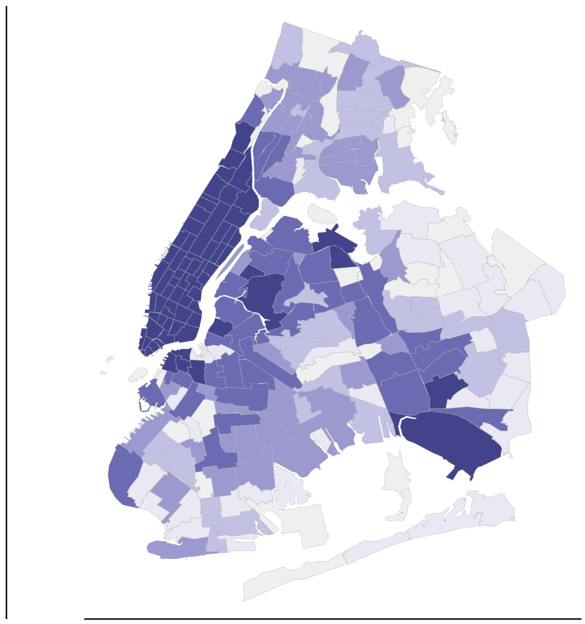


Notes: Per-pickup-zone change in mean credit-card tips, conditioned on fare bins and dropoff eligibility and measured relative to never-eligible zones, at a placebo split (July 2024, left) and the actual toll (January 2025, right). Blues: significant increases (darker larger); orange: significant decreases; grey: no significant change or insufficient trips. The congestion district emerges only at the real date.

6 Transfer Accounting and Policy Counterfactuals

The estimates support a compact transfer accounting, and its striking feature is how little of it is assumption. Table 9 collects the rows.

FIGURE 15: WHERE THE TRANSFER IS GENERATED



DARKER = MORE ADDITIONAL TIP DOLLARS GENERATED BY THE 2025 CONGESTION TOLL (FEE DOLLARS CHARGED ON CARD TRIPS X ESTIMATED PASS-THROUGH, ANNUALIZED). GREY = ZONE GENERATES NONE. MANHATTAN AND INNER BOROUGHS.

Notes: Annualized additional tip dollars induced by the 2025 congestion toll, by pickup taxi zone: fee dollars charged on card trips times the estimated pass-through. Darker = larger; grey = none.

TABLE 9: THE TIP-TAX LEDGER (USD MILLIONS PER YEAR)

	USD m/yr
Annual transfers (measured)	
Riders → MTA, 2025 toll on card trips	16.6
Riders → drivers via tips, 2025 toll ($\rho = 0.133$)	2.2
Riders → drivers via tips, 2019 surcharge ($\rho = 0.150$)	18.7
Counterfactuals	
Fare-only tip base (both fees): induced tips	0.0
Toll at USD 2.50 instead of USD 0.75: induced tips	7.4
Scale gauges (not welfare bounds)	
Harberger triangle at demand-CI edge (-2.9%; point est. ≈ 0)	≤ 0.34
Retiming, cash-switching, zero-tip margins	≈ 0

Notes: Transfers are sums of observed fee payments on standard-rate card trips (2025 toll: 2025H1 annualized; 2019 surcharge: calendar 2019) multiplied by the design-estimated pass-through. The counterfactual rows apply the dose linearity established across the \$0.50–\$2.50 range. The deadweight-loss bound is one-half times the toll times the trip decline at the lower edge of the demand confidence interval; the point estimate of that decline is indistinguishable from zero.

Transfers. The toll moves \$16.6 million a year from riders to the MTA on card trips—the intended, legislated flow—and \$2.2 million a year from riders to drivers through the tip base, the unintended flow this paper measures. The 2019 congestion surcharge, at its larger dose and volume, moved about \$18.7 million a year through the same channel. These are transfers, not losses: dollars change hands at face value.

Counterfactual designs. Because pass-through is stable in the dose from \$0.50 to \$5.00—the JFK flat-fare surcharge supplies the \$4.50 and \$5.00 points (Section 5)—the formula $\rho \times$ fee $\times$ volume prices design alternatives directly. A toll set at \$2.50 would have induced roughly \$7.4 million a year in tips. A mandated fare-only tip base—the design one payment vendor used voluntarily in 2012 (Appendix B)—sets the induced flow to zero at no collection cost, returning the full \$21 million a year across both fees to riders. The tip-side of any future meter fee can be priced the same way before enactment.

The base-rule menu. Because pass-through applies to whatever the base contains, every candidate base rule has a price, and Table 10 prices the 2025 fee environment. On card trips, \$210 million a year of non-fare dollars sit inside the tip base, generating \$23–32 million a year in tips at the paper's two scalings. The government-remitted components alone contribute \$141 million of base and \$15–21 million of tips— and the largest of them is not the new CBD toll but the 2019 New York State congestion surcharge, still running at \$2.50 per trip: \$73 million of base, \$8–11 million a year of induced tips. A regulator choosing the base is choosing a row of this menu. Excluding the CBD toll alone returns \$2–3 million a year to riders; excluding every government fee returns \$15–21 million; the fare-only base that one vendor used in 2012 returns \$23–32 million, at zero collection cost. Tips induced by the driver-kept surcharges are a different object—a gratuity on a price rather than on a tax—and the table labels them separately.

Efficiency: the margins we measured are silent. A transfer becomes a welfare problem when it distorts behavior, and on every margin we can measure, it does not. Riders do not retime hails around the 4:00pm surcharge (density ratio 0.99), do not reduce congestion-zone ridership detectably (-1.0 percent, SE 0.9), do not flee to cash (the card share rises at the cutoff), and barely adjust tipping behavior itself (exact-default use falls under one point). As a gauge of scale

Table 10: The Base-Rule Menu (USD Millions per Year, 2025 Fee Environment)

Component / base rule	In tip base (USD m/yr)	Induced tips at ρ of: 0.108	0.150
Components the driver keeps			
Rush-hour and overnight surcharges	51.4	5.6	7.7
Components collected and remitted			
MTA State Surcharge (USD 0.50)	15.4	1.7	2.3
Improvement surcharge (USD 1.00)	30.7	3.3	4.6
NYS congestion surcharge (USD 2.50)	72.7	7.9	10.9
MTA CBD toll (USD 0.75)	17.4	1.9	2.6
Airport access fee	4.4	0.5	0.7
Bridge and tunnel tolls (third party)	17.9	1.9	2.7
Base rules (tips returned to riders)			
Exclude the CBD toll only	17.4	1.9	2.6
Exclude all government fees	140.6	15.2	21.1
Fare-only base (exclude everything)	209.9	22.7	31.5

Notes: Card trips, all rate codes, 2025H1 annualized. "In tip base" sums each component's dollars on card trips. Induced tips apply the pooled ITT per statutory dollar (0.108) and ρ^{IV} per recorded dollar (0.150); collections are recorded dollars, so the higher column is the aggregation-consistent one and the lower is conservative. Base rules give the tips returned to riders if the percentage buttons excluded the listed components.

rather than a welfare bound: a Harberger triangle computed at the unfavorable edge of the demand confidence interval—one-half times the toll times a 2.9 percent trip decline—comes to $0.34 million a year, an order of magnitude below the transfer, and the point estimate of the decline is zero. Margins we cannot measure—whether inattentive riders would endorse the tip on reflection, and any driver labor-supply response—are discussed below.

What kind of transfer this is. Two clarifications about its nature. First, on the receiving side, a tip is the most driver-kept dollar in the fare stack: unlike metered revenue it does not offset lease payments or commissions, so an induced tip dollar is close to marginal driver income (net of card-processing fees, and taxable, since card tips are recorded). The driver remits the congestion fee itself and nets nothing on it; the induced tip is the only part of the flow that stays. Second, the transfer is not regulatory capture in the classic sense of an interested industry bending rules in its favor: no beneficiary chose the base. The TLC certifies the payment systems, so the base rule sits within the regulable surface, but the record shows convergence by payment vendors, not advocacy by drivers—closer to regulatory neglect than capture. "Unlegislated" describes a decision nobody made.

What this accounting cannot say. Whether the transfer harms riders depends on whether the inattentive four-fifths would endorse the tip on reflection—a preferences question our design-based evidence cannot answer, so we report transfers as transfers. Driver labor supply, medallion values, and general-equilibrium effects are likewise out of scope. The distributional geography of the burden appears in Section 5; we label it neighborhood exposure, not individual incidence.

7 Conclusion

New York taxi payment screens compute suggested tips as percentages of a total that includes every government surcharge on the meter. We show that this design choice has a measurable fiscal consequence: a statutory dollar of surcharge raises the average credit-card tip by about eleven cents—about fifteen cents per surcharge dollar the meter actually records.

The estimate is well identified for a behavioral parameter of this kind. It comes from surcharges that switch on a clock, assigned to the minute of pickup, replicated at statutory doses from $0.50 to $5.00, in both directions, at three separate times of day, against two independent control series (weekends and exempt legal holidays), on a flat fare where the confound cannot exist by construction (JFK), and—crucially—reproduced by a design with no clock in it at all, the spatial boundary of the 2025 congestion toll. Across six within-day changes the effect spans 0.090 to 0.117 per statutory dollar (ρ^{IV} of 0.122–0.150 per recorded dollar); the spatial estimate is 0.107 per statutory dollar against the clock designs' pooled 0.108. Bandwidth and donut variations move it by less than a cent. Two estimates, both at 6:00am on weekdays, fail, and Section 5 shows why: the weekend estimates at that same clock and dose are normal, so what breaks the weekday window is the dawn compositional shift, not the mechanism.

Getting there required one correction we want to flag for anyone replicating this design. The obvious specification—regress the tip on an indicator for being past the cutoff—is badly biased, because the metered fare jumps at 4:00pm along with the surcharge. Rush-hour trips are slower and dearer, and part of the observed tip step is that fare increase rather than the surcharge. Conditioning on the non-surcharge base removes it, and the weekend placebo falls by roughly three-quarters when it does. Designs that use time-of-day discontinuities in prices should expect the same problem wherever the underlying price moves with the time of day.

The interpretation is straightforward. Riders tip on a surcharge dollar at essentially the rate they tip on a fare dollar, because the interface gives them no way to distinguish the two. They do not switch to cash, stop tipping, or abandon the default buttons in economically meaningful numbers. What looks like generosity toward drivers is, at the margin, arithmetic performed by a payment terminal.

The demand margins sharpen the picture into a single sentence: riders do not retime hails around the 4:00pm fee, do not ride less when the congestion toll arrives, do not switch to cash, and offset only a fifth of a 150 percent surcharge increase—the fee is absorbed on every margin except the one the software moves for them. It also yields a concrete policy dial. One 2012 vendor already computed suggestions on the fare alone, so the counterfactual is not hypothetical: mandating a fare-only tip base would return roughly \$18.7 million a year of the 2019 surcharge's induced gratuities and \$2.2 million of the toll's to riders, at zero collection cost, by changing one line of payment-terminal arithmetic—about thirty cents per surcharge-paying card trip.

Two implications follow. For fiscal policy, the incidence of a fee levied on a metered transaction does not stop at the statutory payment: part of it continues past the agency to the worker, unlegislated and uncounted—roughly \$18.7 million from the 2019 congestion surcharge and \$2.2 million a year from the 2025 toll. For the growing literature on interface-mediated payments, the result identifies a channel by which policy and choice architecture interact: what a default is computed *on* matters as much as what it is set *to*.

Three limitations bound the claims. Tips are observed only for card payments, so the pass-through we estimate is conditional on card use, which the design shows does not itself respond at the cutoff. Our estimates aggregate across payment vendors whose adherence rates differ by a factor of four, so ρ is a market average rather than a behavioral primitive. And the 2019 congestion-surcharge window predates our minute-cell sample and abuts the pandemic; we treat it as corroborative. Appendix B ties the modern mechanism to the canonical result: replicating Haggag and Paci (2014) on 2012–2013 data shows the two payment vendors of that era disagreed about the base itself—one stopped at the fare plus surcharge, the other went on to include tax and tolls—and that this base difference accounts for about three-quarters of the vendor tip gap they documented. By 2025 every vendor had converged on the inclusive base, which is what turns a vendor-level quirk into a market-wide fiscal channel.

References

Alexander, D., Boone, C., & Lynn, M. (2021). The effects of tip recommendations on customer tipping, satisfaction, repatronage, and spending. *Management Science*, *67*(1), 146–165.

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, *90*(430), 431–442.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, *111*(12), 4088–4118.

Azar, O. H. (2020). The economics of tipping. *Journal of Economic Perspectives*, *34*(2), 215–236.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, *90*(3), 414–427.

Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2009). Optimal defaults and active decisions. *Quarterly Journal of Economics*, *124*(4), 1639–1674.

Chandar, B., Gneezy, U., List, J. A., & Muir, I. (2019). The drivers of social preferences: Evidence from a nationwide tipping field experiment [National Bureau of Economic Research Working Paper 26380]. *Working Paper*.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, *99*(4), 1145–1177.

Conlin, M., Lynn, M., & O'Donoghue, T. (2003). The norm of restaurant tipping. *Journal of Economic Behavior & Organization*, *52*(3), 297–321.

Cook, C., Kreidieh, A., Vasserman, S., Allcott, H., Arora, N., van Sambeek, F., Tomkins, A., & Turkel, E. (2025). The short-run effects of congestion pricing in New York City [National Bureau of Economic Research Working Paper 33584; revised July 2026 as "The Network-Wide Effects of Congestion Pricing: Evidence from New York City"]. *Working Paper*.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), 315–372.

Feldman, N. E., & Ruffle, B. J. (2015). The impact of including, adding, and subtracting a tax on demand. *American Economic Journal: Economic Policy*, *7*(1), 95–118.

Finkelstein, A. (2009). E-ZTax: Tax salience and tax rates. *Quarterly Journal of Economics*, *124*(3), 969–1010.

Gabaix, X., & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, *121*(2), 505–540.

Goldin, J., & Homonoff, T. (2013). Smoke gets in your eyes: Cigarette tax salience and regressivity. *American Economic Journal: Economic Policy*, *5*(1), 302–336.

Grembi, V., Nannicini, T., & Troiano, U. (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics*, *8*(3), 1–30.

Haggag, K., & Paci, G. (2014). Default tips. *American Economic Journal: Applied Economics*, *6*(3), 1–19.

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, *302*(5649), 1338–1339.

Leape, J. (2006). The London congestion charge. *Journal of Economic Perspectives*, *20*(4), 157–176.

Lynn, M., & Kwortnik, R. J. (2023). Large sample evidence on tipping rates in the restaurant industry. *International Journal of Hospitality Management*, *112*, 103400.

Lynn, M., & McCall, M. (2000). Gratitude and gratuity: A meta-analysis of research on the service-tipping relationship. *Journal of Socio-Economics*, *29*(2), 203–214.

Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics*, *116*(4), 1149–1187.

Morrison, W., & Taubinsky, D. (2023). Rules of thumb and attention elasticities: Evidence from under- and overreaction to taxes. *Review of Economics and Statistics*, *105*(5), 1110–1127.

New York City Taxi and Limousine Commission. (2022). Industry notice 22-02: Taxi fare increase [Rush-hour surcharge $1.00 to $2.50 and overnight surcharge $0.50 to $1.00, effective December 19, 2022].

New York City Taxi and Limousine Commission. (2026). TLC trip record data, yellow taxi [Monthly parquet files; accessed July 2026].

Taubinsky, D., & Rees-Jones, A. (2018). Attention variation and welfare: Theory and evidence from a tax salience experiment. *Review of Economic Studies*, *85*(4), 2462–2496.

Thakral, N., & Tô, L. T. (2026, January). *Five facts about tipping* [Working paper, Brown University and Boston University]. https://linh.to/files/papers/tips.pdf

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Warren, N., Hanson, S., & Yuan, H. (2021). Feeling manipulated: How tip request sequence impacts customers and service providers? *Journal of Service Research*, *24*(1), 66–83.

A Additional Robustness

Wild cluster bootstrap. With 449 date clusters, cluster-robust inference on the headline reduced form is unlikely to be distorted, but the analysis plan locks a wild cluster bootstrap as a check. Using Webb weights and 9,999 replications on the *corrected* weekday step in the \$2.50 regime—the preferred specification, not the uncorrected one—the bootstrap rejects the null of no effect at $p < 10^{-6}$, with a 95 percent confidence set of $[0.2834, 0.3012]$ around a point estimate of \$0.2923 (Cameron et al., 2008).

Permutation over candidate cutoffs. The clock design has a natural randomization test: re-estimate the discontinuity at candidate minutes where no surcharge changes and locate the true cutoff in the resulting distribution. Every candidate—true and placebo alike—is estimated with the same specification and the same ± 30 minute bandwidth, so the placebo distribution is comparable to the actual statistic; candidates are drawn only from windows that never cross 4:00pm, so no placebo inherits part of the true discontinuity. Of 60 such candidates, *none* produces an absolute estimate as large as the true cutoff's 0.2401.

Placebo cutoffs. Placebo cutoffs at 3:15pm and 5:00pm—one before the surcharge begins, one after it is already in force—produce first stages of -0.001 and -0.023 respectively, against 1.951 at the true cutoff, and tip-side effects of $-\$0.0003$ (SE \$0.0068) and $-\$0.0079$ (SE \$0.0071). Both sides of each placebo are null. Whatever generates the 4:00pm discontinuity is specific to the minute at which the fare schedule changes.

Mechanical benchmark. A purely mechanical account of pass-through predicts a tip step equal to the surcharge step times the prevailing tip rate on the base. In the \$2.50 regime the surcharge step is \$1.951 and the pre-cutoff tip rate is 0.174, for a mechanical prediction of \$0.339. The corrected estimate is \$0.292—about 86 percent of the mechanical benchmark. The remaining fourteen percent is the behavioral offset: the small declines in exact-default adherence and the small rise in zero-tipping documented in Section 5. Pass-through is therefore close to, but slightly below, complete.

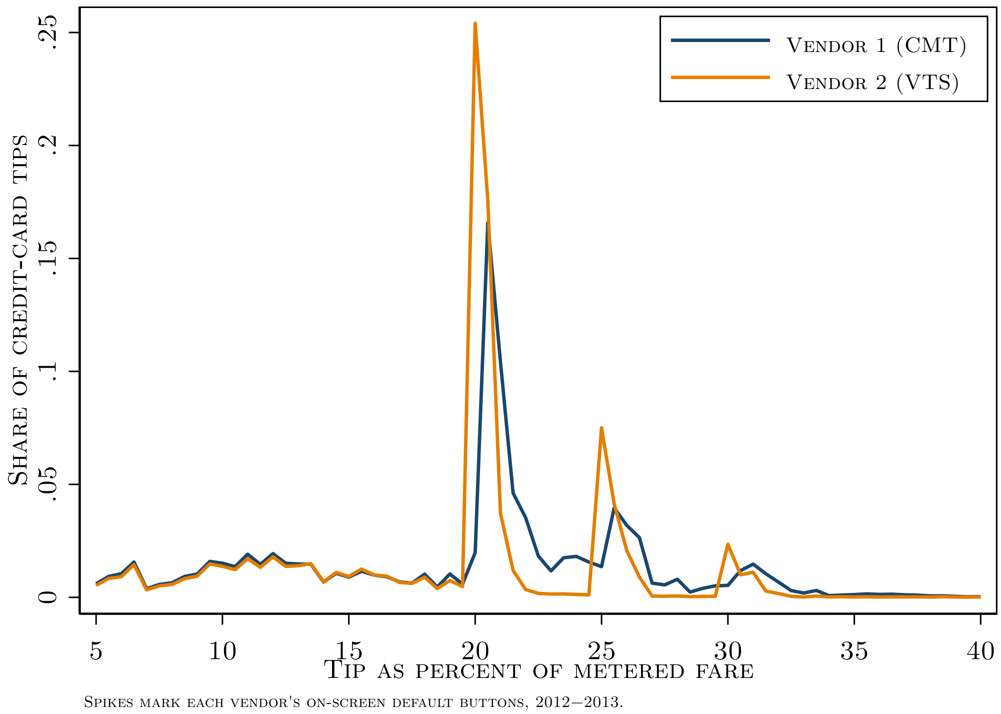
Synthetic difference-in-differences. Because the binary toll event study fails a strict pre-trend test, we re-estimate the adoption effect with synthetic difference-in-differences (Arkhangelsky et al., 2021), which reweights both control zones and pre-periods to match the treated group's pre-treatment trajectory—the estimator built for exactly this failure mode. On a balanced zone-month panel of fare-conditioned tips (treated: the 52 CBD-pickup zones; donors: never-eligible zone pools), the SDID average treatment effect is \$0.061 (bootstrap SE \$0.050), an implied $\rho = 0.102$—inside the family of estimates but underpowered at zone-level aggregation, where the design surrenders the cell-level precision of the main specifications. An in-time placebo assigning a fake July 2024 adoption returns $-$\$0.125 (SE 0.069), reflecting the same 2024 drift the Wald test flags. We read SDID as corroborating the magnitude while confirming, rather than repairing, the pre-period's imperfection; the dose staircase remains the spatial design's strongest evidence.

B Replication: Haggag and Paci (2014) and the Origins of the Base

Haggag and Paci (2014) established that default menus cause tips using New York taxi data from 2009–2013, exploiting the fact that the city's two payment vendors displayed different suggestions. We replicate that vendor comparison on six archival months (April–June and October 2012, April–May 2013; 42.7 million credit-card trips). The two vendors differed not only in the *level* of their suggestions but in the *base* they computed them on—a fact documented independently by Thakral and Tô (2026), whose description of the two bases our match rates confirm. What we add here is the arithmetic linking that base difference to the tip gap Haggag and Paci (2014) measured.

Recovering the menus. Figure A1 plots the distribution of tips as a percentage of the metered fare, by vendor, for fares of \$15 and above. Vendor 2 (VTS) shows sharp spikes at exactly 20.0 and 25.0 percent. Vendor 1 (CMT) shows spikes displaced slightly to the right, at 20.5 and 25.5 percent. The displacement is not noise, and it is not a different menu level: it is what a 20 percent button produces when the percentage is applied to a base that includes the tax and tolls sitting on top of the fare. Vendor 2's spikes land on round percentages of the fare because its base is the fare plus the time-of-day surcharge, and that surcharge is zero on the majority of daytime trips—so for most of the distribution its base and the metered fare coincide. This is exactly why the next subsection tests the candidate bases directly rather than reading them off spike locations: the spikes alone cannot separate "fare" from "fare plus a surcharge that is usually zero."

Figure A1: Vendor Default Menus in the Archival Data, 2012–2013



Notes: Distribution of credit-card tips as a percentage of the metered fare, in half-point bins, for trips with fares of $15 or more. Spikes mark each vendor's on-screen default buttons. Vendor 2's spikes fall exactly on 20 and 25 percent of the fare; Vendor 1's are displaced upward because its percentages are computed on a surcharge-inclusive base.

Testing the base directly. Table A1 makes the point without reference to spike locations. For each vendor we compute the share of positive card tips that equal a 20, 25, or 30 percent button applied to each candidate base, walking up the fare stack one component at a time. Testing the full ladder matters: the surcharge field is zero on most daytime trips, so "fare" and "fare plus surcharge" coincide there, and a test that compares only the fare against the all-in total will appear to match the fare for a vendor that in fact uses fare-plus-surcharge.

The pattern is categorical and identical in all six months. Vendor 2 (VTS) matches *fare plus surcharge*—60.7 percent, against 30.4 percent for the fare alone and 1.4 percent for the all-in total.

Vendor 1 (CMT) matches the *all-in total including tax and tolls*—53.8 percent, against essentially nothing for either narrower base. So the vendors did not disagree about whether to include the time-of-day surcharge; both included it. They disagreed about the sales tax and tolls sitting above it. This is the base description in Thakral and Tô (2026), and our match rates confirm it.

Table A1: Which Base Did Each Vendor Use in 2012–2013?

		Share of tips matching a 20/25/30% button on:			
Vendor	Card trips	Fare	+ surcharge	+ tax	+ tolls
Vendor 1 (CMT)	22,143,216	0.024	0.015	0.519	0.538
Vendor 2 (VTS)	20,511,439	0.304	0.607	0.014	0.014

Notes: Six archival months (April–June and October 2012, April–May 2013), credit-card trips with metered fares of $4–60 and positive tips. A match requires equality to within half a cent. Each column adds one component of the fare stack to the base of the preceding column. Vendor 2 computes on fare plus surcharge; Vendor 1 computes on the all-in total.

Does the base difference explain the tip difference?. Vendor 1's mean tip rate is 20.6 percent of fare; Vendor 2's is 19.7 percent, an observed gap of 0.94 percentage points. Conditioning on month and fare bin, the gap is 0.0089 (SE 0.0011), reported in Table A2. That gap has a prediction attached to it, and the correct prediction follows from the correct base difference: what Vendor 1 includes and Vendor 2 does not is the sales tax and tolls, which average 6.51 percent of the metered fare. A rider pressing a 20 percent button therefore tips $0.20 \times 0.0651 = 1.30$ percentage points of fare more at Vendor 1 than at Vendor 2.

Not every rider presses a button. Scaling by Vendor 1's exact-default share of 53.8 percent gives an expected gap of 0.70 percentage points, against an observed 0.94. Base inclusion accounts for roughly three-quarters of the vendor gap; the residual quarter is the level difference in the menus themselves and the difference in zero-tipping (2.0 percent of Vendor 1's trips against 3.6 percent of Vendor 2's). The vendor difference that Haggag and Paci (2014) used to identify default effects is therefore substantially—though not wholly—attributable to the base-inclusion mechanism this paper studies.

Executing their deposited code. The authors' replication archive (openICPSR 113897) contains their cleaned 2009 analysis file and estimation code. Running their Table 2 specification verbatim on their data—the $15 menu-threshold discontinuity in CMT trips, third-order fare poly-

Table A2: Vendor Gap in Mean Tip Rate, 2012–2013

	Tip as share of fare
CMT mean tip rate	0.2060
VTS mean tip rate	0.1966
Difference (VTS − CMT)	-0.0089***
	(0.0011)

Notes: Six archival months, credit-card trips with fares of \$4–60. The difference conditions on month-by-fare-bin fixed effects and clusters by fare bin. ***, **, *: significance at 1, 5, 10 percent.

nomial on each side, day-of-week, hour, and borough controls, driver fixed effects, standard errors clustered by fare bin, fares \$5–25—yields a tip jump at the threshold of \$0.2757 (SE \$0.0060) on 6.22 million trips, with a mean tip of \$2.22 just below the cutoff: their headline default effect, regenerated from their archive on our machines. The lineage of the present paper is then exact: their experiment moves the *menu* at a fare threshold and tips jump 12 percent of the base; ours moves the *base* at a clock threshold and tips move by 80 percent of the prevailing tip rate. Both are the same object—suggested-gratuity arithmetic doing fiscal work riders do not undo.

Convergence. The historical comparison closes the paper's argument. In 2012 the market was split: one vendor stopped its base at the fare plus the time-of-day surcharge, the other carried it all the way to the all-in total including tax and tolls. By 2025, as Table 1 shows, all three vendors compute on the inclusive total. The design choice that was once a point of disagreement between payment systems has become the industry standard, and with it the channel through which government fees flow into gratuities. That convergence is what makes the channel a market-wide fiscal fact rather than one vendor's quirk, and it is the feature of the modern setting that the archival literature could not have observed.

C Data Appendix

Sources. All trip data come from the NYC Taxi and Limousine Commission's monthly yellow-taxi parquet files (New York City Taxi and Limousine Commission, 2026), which cover the universe of metered yellow-cab trips. Taxi-zone geometry comes from the TLC's published shapefile. No restricted or proprietary data are used; the entire analysis is reproducible from public sources with the code in this repository.

Cell construction. Two builds produce the estimation inputs. The first streams each monthly file and aggregates to date-by-minute cells within windows around the three clock cutoffs (14:30–17:30, 18:30–21:30, and 05:00–07:00), recording for each cell the trip count, card-trip count, mean metered surcharge, sum of card tips, sum of card pre-tip totals, the count of tips exactly at a 20/25/30 percent button, and the zero-tip count. The second aggregates by month, pickup zone, and $5 fare bin over the congestion-fee windows (2018m7–2019m12 and 2024m1–2025m6), recording counts, tip and base sums, and mean fee incidence. Four further builders complete the inputs: the 6:00am window cells (the main minute build collects only the afternoon and evening windows; the 6:00am builder is schema-identical), the JFK flat-fare minute cells (RatecodeID 2, all three windows, same schema), the 2012–13 vendor base-ladder cells, and the June-2025 tip-rate histogram cells behind Figure 3.

Sample restrictions are applied identically in both builds: metered fares between $3 and $200, non-negative tips. Tips are recorded only for credit-card payments, so all tip means are conditional on card use; the card share is reported as a locked outcome at every cutoff. Date-by-minute cells are restricted to dates within each file's own month, which removes a small number of misdated records.

Regime and holiday coding. The rush-hour surcharge rose from $1.00 to $2.50 and the overnight surcharge from $0.50 to $1.00 on December 19, 2022 (New York City Taxi and Limousine Commission, 2022); cells are assigned to regimes on that date. Federal and municipal holidays, on which the rush surcharge does not apply, are excluded from weekday samples and are listed in the cleaning code.

Geographic treatment assignment. The congestion-fee columns exist only after each policy takes effect, so a treated indicator built from them would place every pre-period observation in the control group. We therefore learn the treated zone set from post-period fee incidence—pickup zones where the mean fee exceeds $0.30, giving 52 zones for the 2025 CBD toll and 115 for the 2019 surcharge—and apply that fixed zone set to the whole window.