Name: Abhinav Shah
School: The Cathedral and John Connon School

Research Question:

What are the key differences in the design and implementation of congestion pricing programs in major cities, and how do these differences influence their effectiveness?

Title - Designing Congestion Pricing

Word Count: 16289

Table of Contents

1. Introduction

Urban traffic congestion is an increasingly urgent problem in cities across the globe. As populations swell, urban centers densify, and car ownership becomes more accessible, the pressure on existing road networks intensifies. From London to New York, and from Rome to Singapore, metropolitan regions are facing mounting delays, environmental degradation, and rising frustration among commuters. According to the INRIX Global Traffic Scorecard (INRIX, 2024) ,drivers in major cities lose dozens, sometimes over a hundred, hours per year idling in traffic, and the total annual economic cost of congestion to urban economies runs into the billions. In the United States alone, congestion was estimated to have cost $87 billion in wasted time and fuel in 2018. (Weforum 2019) ,Yet despite decades of investment in public transit and road infrastructure, congestion persists as a structural feature of urban life.

The consequences of congestion go beyond travel delays. Economically, it reduces productivity, disrupts logistics, and inflates transportation costs for goods and services. Environmentally, vehicle emissions contribute to poor air quality and elevated carbon dioxide levels, exacerbating climate change and creating public health risks. Socially, congested roads increase noise pollution, reduce walkability, and lead to the uneven distribution of opportunity, particularly when low-income residents are priced out of housing near employment hubs and are forced into long commutes. Traditional policy tools, such as expanding road capacity or offering more public transit routes, have largely failed to address these problems sustainably. Many cities are physically constrained and cannot afford to keep building more roads, and public transit investments often take years to materialize and require complementary behavioral shifts.

Amid these challenges, congestion pricing has emerged as a policy mechanism that directly tackles the root cause of traffic: the unpriced and over consumed road space during peak periods. Inspired by economist William Vickrey's theory of marginal cost pricing, congestion pricing seeks to align individual drivers' decisions with the broader social cost of their travel. In essence, it imposes a toll that reflects not only the private cost of driving (such as fuel or time) but also the external costs imposed on others, namely increased congestion, air pollution, and delays. By charging drivers for road use in high-demand areas or time periods, cities aim to reduce excess vehicle volume, improve traffic flow, and promote more efficient and sustainable modes of transportation. (Vickrey, 1969)

Over the past two decades, a number of cities have implemented congestion pricing programs, with varying degrees of ambition and success. Singapore pioneered electronic road pricing in 1998 and is now transitioning to a satellite-based GNSS system capable of real-time distance- and time-based charging. Stockholm adopted a cordon-based system with dynamic time-of-day charges, first as a trial and then permanently after a successful referendum. London introduced a flat congestion charge in 2003 and later extended it to include ultra-low emission zones. Milan's Area C zone adds environmental constraints to its pricing model, targeting emissions by restricting vehicle types. Meanwhile, cities like New York and Rome have experienced delays or public resistance, illustrating the political sensitivities and operational complexities involved. (Lindsey, 2006)

What differentiates successful programs from those that struggle or fail is not merely the presence of a charge but the design and implementation choices that shape how the policy works in practice. Some cities employ dynamic pricing that adjusts with congestion levels; others rely on flat rates. Some reinvest toll revenues into public transport to boost political legitimacy, while others funnel them into general budgets, sparking concerns about equity and fairness. Still others include vehicle-type exemptions or discount programs to address social justice concerns. These design elements are not trivial details, they fundamentally determine whether a program achieves its goals, maintains public support, and sustains long-term impact.

This paper aims to compare congestion pricing systems across major global cities, focusing on how differences in their design and implementation influence their effectiveness. Effectiveness, in this context, is defined using a broad, multidimensional framework. First, environmental performance is evaluated through reductions in carbon emissions and improvements in air quality. Second, traffic performance includes metrics like average vehicle speeds, travel time reliability, and congestion levels. Third, public and political acceptance are considered, given their importance in policy longevity and expansion. Fourth, revenue-related outcomes are examined, including how pricing funds are allocated and whether they improve long-term infrastructure or equity. Lastly, the analysis considers local business and urban livability impacts, such as noise, nuisance, and economic activity in downtown areas.

By synthesizing case studies, evaluating outcome data, and applying economic theory, this paper seeks to identify the features that make congestion pricing programs work, not just technically, but politically and socially. The goal is to draw lessons from cities that have implemented these policies successfully, while also understanding the limitations and pitfalls of less effective attempts. Ultimately, this research aspires to offer practical guidance to policymakers considering congestion pricing as a tool for managing urban transportation systems in a way that is efficient, equitable, and sustainable.

2. Theoretical Background

2.1 Economic Theory of Congestion

At the heart of urban traffic congestion lies a fundamental market failure: the overuse of a scarce public good, road space, due to the absence of appropriate pricing. When individuals decide to drive, especially during peak hours, they consider only their private marginal costs, including fuel, parking, tolls, and personal time lost in traffic. However, they fail to account for the external costs they impose on other road users, such as increasing travel times, contributing to air and noise pollution, and exacerbating health risks through vehicular emissions. This divergence between private and social costs leads to inefficient outcomes, wherein roads become overused and congested beyond the point that is optimal for society.

The theoretical framework for correcting this misallocation was first formalized by Nobel laureate William Vickrey, who applied the concept of marginal cost pricing to urban transportation. Vickrey proposed that in order to achieve an efficient level of road usage, each driver should face a charge equal to the marginal social cost (MSC) of their trip, rather than just their private costs. This idea is grounded in the earlier work of economist A.C. Pigou, who argued that taxing goods which generate negative externalities could internalize social costs and restore economic efficiency. In the context of congestion, the externality is the time delay that each additional vehicle causes for all other drivers, which increases non-linearly as road usage approaches capacity. (Vickrey, 1969)

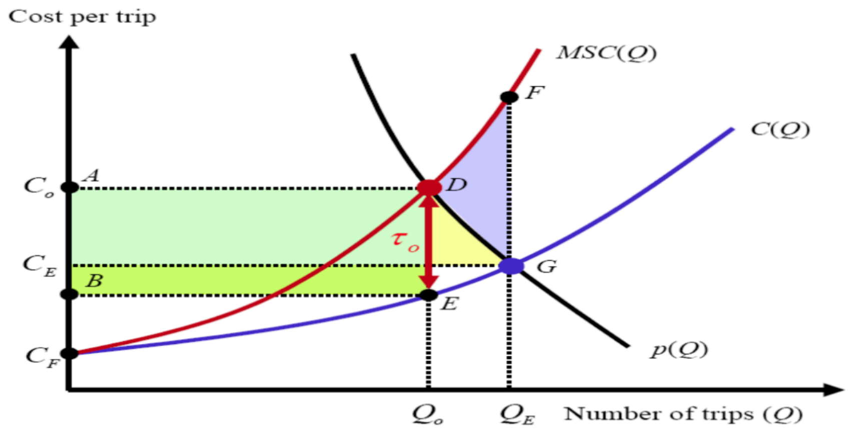

Figure 1.1: "Marginal Cost Pricing Framework for Road Congestion" (Song, 2013)

This concept is best understood graphically through a marginal cost pricing framework. As shown in Figure 1, the number of trips (Q) is plotted on the horizontal axis, while the cost per trip is shown on the vertical axis. The marginal private cost (C(Q)) curve reflects the costs borne directly by the driver, fuel, time, and vehicle wear. However, this does not account for the additional congestion delay imposed on others and other social costs - noise, pollution, etc. The marginal social cost (MSC(Q)) curve lies above the private cost curve, diverging more steeply as traffic volume increases, reflecting the growing external cost of delay per added vehicle.

Without pricing, the market equilibrium is reached at point G, where the private cost is equal to the demand curve. However, this leads to overuse of road space and inefficient congestion. The socially optimal level of traffic occurs at point D, where demand equals the full social cost. The optimal toll τ is the vertical distance between MSC and MPC at this point, representing the marginal external cost. By charging this toll, policymakers can internalize the externality, reduce traffic to an efficient level, and eliminate the deadweight loss, the shaded triangle DFG in the figure that results from unpriced congestion. Drivers are the ones paying for this policy, so they are "losing" in one sense: they lose some surplus and they pay the tax. Tax revenue is ABED, and they also use the yellow triangle. This visual representation offers a foundational lens through which modern congestion pricing strategies are derived.

The mathematical expression of marginal social cost further clarifies this logic. Let C represent the average commuting cost per trip, which is itself a function of traffic volume Q. The total cost to all commuters is then $C(Q) \cdot Q$. The marginal social cost is the derivative of this total cost with respect to traffic volume:

$$MSC = \frac{d(CQ)}{dQ} = C(Q) + Q\frac{dC}{dQ} = C(Q) + EC \text{ (Song, 2013)}$$

This formula shows that the cost to society of one additional vehicle is not just the cost that vehicle faces (the first term, C), but also the additional cost imposed on all other vehicles already on the road (the second term, $Q \cdot dC/dQ$). This second component is the external cost that needs to be internalized through a congestion toll. Thus, the optimal congestion charge or Pigouvian toll can be expressed as:

1. $\tau = Q \cdot \frac{dC}{dQ}$

This is the amount that, if charged to each driver, would reduce the volume of traffic to the point where marginal benefit equals marginal social cost, thereby achieving allocative efficiency.

However, the theory says that pricing does work exactly as intended; the question is how well it can work. This depends on elasticities. The behavioral response of drivers to tolls depends on the elasticity of demand for driving. This elasticity determines how sensitive the volume of vehicle trips is to changes in road costs. Demand elasticity can be captured by the slope of the demand curve at a point; a steeper demand curve (more inelastic) means a larger tax will be needed to change behavior. This determined how well the policy works. If demand is perfectly inelastic, the number of cars on the road does not change, and they all pay the tax. If it is perfectly elastic, there would be a massive response. When alternatives such as reliable public transport, cycling infrastructure, or telecommuting options are available, the price elasticity of demand tends to be higher, meaning that even a modest toll can lead to a significant reduction in traffic. Conversely, in car-dependent urban environments with limited substitutes, demand for driving is relatively inelastic, and tolls may need to be higher, or complemented by investment in transit infrastructure, to produce a meaningful behavioral shift.

Elasticity also varies across time, demographic groups, and geographic areas. For instance, evidence from Stockholm suggests that the elasticity of demand increased over time following the implementation of congestion pricing, largely because of concurrent investments in public transport and the public's growing familiarity with alternative commuting patterns (Börjesson & Kristoffersson, 2018). In contrast, Rome's Eco Pass program failed to generate a substantial response from drivers, in part due to weak enforcement and the absence of viable modal substitutes (Russo et al., 2021). Singapore's experience, however, provides a strong example of responsive pricing aligned with elasticity, as its Electronic Road Pricing (ERP) system dynamically adjusts tolls based on observed congestion levels, ensuring a continuous feedback loop between traffic conditions and pricing levels (Theseira, 2020).

The theoretical insight here is not only that congestion pricing can reduce traffic volumes, but that its design must be sensitive to demand conditions, available alternatives, and the institutional capacity to measure and adjust prices in real time. Programs that rigidly apply flat tolls without regard to demand variation, such as London's initial implementation, risk losing effectiveness over time as traffic patterns evolve. Moreover, understanding elasticity is crucial for equity analysis: higher-income commuters may be less price-sensitive, raising concerns that tolls could disproportionately burden lower-income drivers unless mitigated through targeted exemptions or complementary public transport investments.

Taken together, Vickrey's theory of marginal cost pricing, the concept of negative externalities, and the elasticity of demand form a robust theoretical foundation for evaluating congestion pricing systems. They provide both a normative benchmark, what an efficient system should look like, and a set of analytical tools to assess how closely real-world programs align with that ideal. In the sections that follow, this framework will be applied to examine the effectiveness of congestion pricing programs across global cities, with particular attention to how each city's policy design reflects or diverges from these foundational economic principles.

2.2 The Lucas Critique and Policy Adaptation

A critical insight into the long-term limitations of congestion pricing design comes from a foundational principle in economic policy analysis known as the Lucas critique. Introduced by Nobel laureate Robert Lucas in 1976, the critique argues that policies cannot be reliably evaluated based on historical correlations between variables if the policy itself changes the underlying behavioral rules or expectations. In other words, the structural relationships between economic variables are not stable when policy regimes change, because agents, individuals, firms, or in this case, drivers, will adapt their behavior in anticipation of new incentives.

This has direct implications for congestion pricing programs, particularly those that rely on fixed or static toll designs. For example, London's original flat congestion charge was developed based on observed traffic patterns and estimated behavioral responses from the early 2000s. While initially successful in reducing central area traffic by up to 20%, its impact diminished over time. Drivers adjusted to the new normal, some shifted their travel times, others absorbed the cost or changed vehicles, and the flat toll, unresponsive to real-time traffic conditions, gradually lost relevance. This illustrates the Lucas critique in action: when the environment changes, so do people's expectations and behaviors, rendering static models and fixed rules obsolete.

Programs like Singapore's ERP system, by contrast, avoid this problem through adaptive, data-driven tolling mechanisms. Rates are adjusted quarterly based on measured traffic speeds, ensuring that pricing remains aligned with current congestion levels. This dynamic approach effectively internalizes behavioral feedback: users know the system responds to aggregate travel behavior, so their decisions today are informed by anticipated outcomes tomorrow. As such, Singapore's system reflects the Lucasian insight that policy must be forward-looking, not merely reactive.

The Lucas critique challenges policymakers to move beyond reduced-form models that assume stable relationships between price and behavior. In transport economics, this means designing pricing systems that learn and evolve, incorporating real-time data, behavioral elasticity, and even seasonal or demographic variation. Without this, even well-designed systems risk decaying in effectiveness as the city, its infrastructure, and its population change around them.

Ultimately, the lesson is clear: effective congestion pricing requires not just smart design, but adaptive intelligence. As urban environments and traveler expectations evolve, pricing mechanisms must be recalibrated to reflect new realities. Ignoring this would be to violate one of modern economics' most enduring insights. (Hoover, n.d.)

3. Paradoxes in Transportation Economics

While congestion pricing is typically evaluated through the lens of marginal cost theory and externalities, several paradoxes in transportation economics illustrate the nonlinear and counterintuitive nature of traffic systems. These paradoxes challenge the simplistic assumption that improving road infrastructure or increasing capacity will necessarily lead to smoother traffic flows or reduced congestion. Instead, they demonstrate that, under certain conditions, well-intentioned policies may lead to worsened outcomes, either by undermining public transportation, intensifying congestion, or triggering new latent demand. Understanding these paradoxes is critical for interpreting the long-run effectiveness of congestion pricing and for designing urban mobility policies that avoid unintended consequences.

a) The Downs–Thomson Paradox

Formulated by Anthony Downs and later elaborated by John Thomson, the Downs–Thomson Paradox posits that the quality of public transportation and private driving are interdependent. Specifically, it suggests that improvements in road infrastructure that make driving more attractive may, paradoxically, increase congestion in the long term by degrading public transport systems. If more commuters shift from buses or trains to private cars, the demand for public transit falls, which may reduce service frequency, funding, and political support for investment. Over time, this can lead to a self-reinforcing cycle where driving becomes the default mode, transit deteriorates further, and congestion worsens.

This paradox helps explain why urban policies that prioritize road expansion without enhancing transit alternatives often fail to alleviate congestion sustainably. In contrast, congestion pricing can reverse the dynamic, making driving less attractive during peak hours and nudging commuters toward public transportation. Successful examples of this effect include Stockholm and Singapore, where congestion charges were implemented alongside targeted investment in public transit systems. These cities saw not only reduced vehicle volumes but also increases in transit ridership, demonstrating how pricing and infrastructure must be integrated to avoid the pitfalls described by the Downs–Thomson Paradox.

b) The Braess Paradox

The Braess Paradox, introduced by German mathematician Dietrich Braess in 1968, offers an even more surprising insight: adding capacity to a road network can increase overall congestion. This occurs when drivers, acting independently to minimize their own travel time, redistribute themselves across the network in a way that worsens total traffic conditions for everyone. In certain network configurations, the addition of a new road segment creates a new equilibrium that is less efficient than the previous one. Each driver may experience longer travel times despite the apparent increase in capacity.

Empirical demonstrations of the Braess Paradox have been documented in cities like Seoul, where the closure of the Cheonggyecheon Freeway led to an unexpected improvement in traffic flow. Similarly, computer simulations show that when road users rely solely on individual optimization, the network can converge on a collectively suboptimal pattern. The paradox serves as a cautionary tale: supply-side solutions to congestion (such as building new roads) may backfire unless they are accompanied by demand-side interventions, such as congestion pricing or traffic flow control. These findings underscore the importance of coordinated planning and pricing mechanisms, particularly in urban networks with high volumes and complex routing behavior.

c) Induced Demand

The concept of induced demand, closely related to the Downs–Thomson Paradox but broader in scope, refers to the phenomenon whereby increasing road capacity leads to more traffic, not less. When new lanes or highways are added, the initial reduction in congestion lowers the perceived cost of driving. In response, individuals who previously avoided driving due to traffic conditions may now choose to drive, shift to longer routes, or take trips they would not have otherwise made. Additionally, land use patterns may shift as people move further from city centers, assuming they can commute more easily, thus embedding car dependence more deeply into the urban fabric.

Studies from the U.S. and Europe have consistently shown that road expansion projects often fail to reduce congestion in the long term. A frequently cited example is the expansion of the Katy

Freeway in Houston, Texas, the widest freeway in North America, which saw increased commute times despite a multibillion-dollar widening effort. The economic insight here is that road space behaves like a typical good: when its price is effectively zero (i.e., no tolls or usage fees), increasing its supply simply encourages higher consumption. Only by introducing price signals, as in congestion pricing, can cities manage demand and prevent the cycle of induced traffic growth.

From a theoretical standpoint, induced demand undermines the assumption of a fixed demand curve for travel. Instead, it suggests that traffic demand is elastic with respect to perceived travel cost, including both time and money. This insight reinforces the value of congestion pricing, which directly addresses the cost side of the equation rather than attempting to expand supply. Dynamic pricing models, such as Singapore's ERP or Stockholm's time-sensitive cordon charges, offer mechanisms for controlling induced demand by adjusting tolls to reflect real-time congestion levels, thereby keeping total traffic volumes within sustainable limits.

4. Case Study Overview: Programs in Global Cities

This section presents a detailed analysis of congestion pricing programs in eight global cities. Each case captures the nuances of design, enforcement technology, toll structure, political trajectory, and the broader policy environment within which the system operated. These case studies not only reveal diverse pathways toward implementation but also underscore the critical importance of contextual adaptation in shaping long-term effectiveness.

Singapore

Singapore's congestion pricing journey is the most advanced and longitudinally studied in the world, beginning with the Area Licensing Scheme (ALS) in 1975. The ALS required vehicles entering the Central Business District (CBD) during peak hours to display a paid paper license. It achieved immediate success by reducing traffic volumes by 44%, but over time its rigidity, lack of time differentiation and inability to adapt dynamically, led to congestion re-emerging.

In 1998, the ALS was replaced with the Electronic Road Pricing (ERP) system, a global first in dynamic tolling. ERP charges vary by location, vehicle type, and time of day, and are adjusted quarterly by the Land Transport Authority (LTA) based on observed traffic speeds. Gantry-based infrastructure communicates with in-vehicle units to deduct tolls seamlessly. The ERP aims to maintain optimal traffic flow speeds: 20–30 km/h on arterial roads and 45–65 km/h on expressways.

The program is embedded within a broader urban mobility strategy, including strict car ownership controls (e.g., Certificate of Entitlement auctions), heavy public transit investment, and strategic land use planning. According to Theseira (2020), Singapore's success lies in continuous iteration, transparent governance, and aligning public incentives with system goals. Public resistance, while present initially, was mitigated by high institutional trust and visible performance gains. The upcoming ERP 2.0, a satellite-based system enabling distance-based charging, exemplifies Singapore's commitment to dynamic, equitable pricing, and urban efficiency.

London

The London Congestion Charge, launched in February 2003, was Europe's first large-scale urban pricing initiative in the modern era. It covered an 8-square-mile area of Central London and applied a flat £5 daily fee, enforced via Automatic Number Plate Recognition (ANPR) cameras. Vehicles entering the zone between 7:00 am and 6:30 pm on weekdays were required to pay or face fines. By 2020, the charge had risen to £15, but the flat-rate model remained.

According to Leape (2006), the system achieved rapid success: traffic entering the zone decreased by 15%, congestion (measured by excess delay) fell by 30%, and bus reliability improved. Revenue was ring-fenced for public transport improvements, including 300 new buses. However, political support was uneven and waned over time as private hire vehicles, exempt from the charge, proliferated, offsetting earlier gains.

Long-term criticism focused on static pricing (no time-of-day differentiation), exemption policies, and limited geographic scope. A lack of real-time responsiveness led to re-congestion. Nonetheless, London's system provided a globally visible proof of concept for congestion pricing in a Western democracy. More recent iterations, like the Ultra Low Emission Zone (ULEZ), reflect a shift toward combining congestion and environmental goals.

Milan

Milan's congestion pricing scheme evolved through two phases. The first, Ecopass, was introduced in 2008 as an emissions-based charging regime targeting polluting vehicles. It achieved modest reductions in traffic (~6%) but was undermined by broad exemptions, lack of enforcement consistency, and limited coverage.

In 2012, Ecopass was replaced by Area C, a stricter congestion pricing and emissions control system. Vehicles entering the 8.2 km^2 city center during business hours are monitored via ANPR and subject to a €5 entry charge, with free entry only for low-emission vehicles, electric cars, and certain residents. Area C also introduced a complete ban on high-emission diesel vehicles, creating a dual mechanism of pricing and exclusion.

Moulin and Urbano (2025) note that Area C led to a 30% reduction in car traffic, a 17% decline in PM10 emissions, and a measurable shift toward public and active transport. Milan also observed gender and age-specific effects: younger drivers adapted more rapidly to restrictions, while older and lower-income commuters were less flexible. Public support grew with environmental improvements, but critics raised concerns about the burden on small businesses and delivery logistics. The Milan model demonstrates the power of combining emissions-based bans with pricing and indicates the potential of such systems to achieve co-benefits in public health and urban sustainability.

Stockholm

Stockholm's congestion pricing was implemented in two stages: an initial seven-month trial in 2006, followed by a binding referendum, and permanent reinstatement in August 2007. The system uses a time-varying, cordon-based model, charging between SEK 10 and 20 per passage into or out of the inner city between 6:30 am and 6:30 pm on weekdays. Charges are capped at SEK 60 per day.

Eliasson (2014) highlights Stockholm's political innovation: the use of a trial period combined with robust data collection and real-time evaluation created a fact-based public discourse. Early results showed 22% traffic reduction, increased transit ridership, and 15% cuts in NOx and PM10 levels. Public acceptance increased from 30% pre-trial to 53% post-trial.

Stockholm's program has shown enduring effectiveness due to price responsiveness and adaptive design. Börjesson (2018) emphasizes that over time, elasticity of demand increased as commuters adjusted work hours, switched modes, or avoided trips. Revenue is earmarked for public transport and infrastructure, bolstering legitimacy. The Stockholm model exemplifies how evidence-based policymaking, transparent trial periods, and reinvestment strategies can turn a politically divisive proposal into a durable success.

Gothenburg

Gothenburg adopted a congestion tax in January 2013, largely modeled on Stockholm's framework. It applies time-differentiated charges at gantries encircling the city center, with prices ranging from SEK 8 to SEK 18 per crossing, capped daily. Although technologically similar to Stockholm, Gothenburg's experience diverged sharply in public and political reception.

Unlike Stockholm, the Gothenburg system lacked a trial period or referendum prior to implementation. Börjesson (2018) finds that it was perceived as imposed from above, driven more by infrastructure funding goals than congestion relief. A non-binding referendum in 2013 showed 57% opposition, but the system remained in place.

Traffic reductions were modest (~9%) and quickly dissipated. Public transport improvements were minimal, and perceived benefits remained unclear. The system's failure to build public legitimacy, combined with weaker alternatives to car travel, resulted in declining political support and lower compliance over time. Gothenburg demonstrates how lack of procedural legitimacy, insufficient co-benefits, and weak public transit can significantly reduce the long-term efficacy of congestion pricing.

New York City (2025)

New York City launched its long-delayed congestion pricing program in January 2025, charging drivers entering Manhattan below 60th Street a $9 peak-hour toll, with differentials for trucks and off-peak travel. The system uses a fully automated license plate recognition infrastructure, coordinated by the MTA and city Department of Transportation.

Cook et al. (2025) used synthetic control methodology to estimate the program's short-run effects, finding a 15% increase in average traffic speeds, 8% improvement in travel time reliability, and early signs of reduced emissions. Importantly, New York’s system is projected to generate $1 billion annually, earmarked exclusively for transit upgrades.

Years of political gridlock had blocked previous proposals, with opponents citing regressive impacts on outer-borough commuters. The eventual success was due to careful coalition-building, legislative action at the state level, and rising public awareness of climate and infrastructure crises. The NYC program is unique in its integration of environmental, equity, and funding objectives, and its future effectiveness will hinge on transparent reinvestment and continual price recalibration.

Rome

Rome’s Eco Pass and Limited Traffic Zone (ZTL), launched in 2001, targeted emissions and congestion in the city’s historic core. Unlike pricing models, access was restricted via permit-based controls and vehicle bans during peak hours. Enforcement was camera-based, but loopholes, inconsistent application, and political ambivalence rendered the system ineffective.

Russo et al. (2021) find that traffic levels remained essentially unchanged, air quality improvements were negligible, and behavioral adaptation was limited. The ZTL became widely perceived as inequitable, granting access to residents and certain business classes while penalizing others. Rome's experience illustrates that without robust enforcement, pricing incentives, or supportive infrastructure, access control alone is insufficient to produce lasting results.

San Diego

San Diego’s I-15 HOT lane system, launched in 1998, operates as a dynamic pricing corridor within a major highway. Solo drivers can pay to access previously HOV-only lanes, with tolls adjusting every six minutes based on real-time congestion levels. Revenue is collected via in-vehicle transponders and used to fund express bus services.

Bhatt (2011) classifies San Diego's HOT lane as a successful demonstration of real-time congestion management without large-scale structural change. The system preserved HOV lane speeds even during peak hours and expanded modal options. However, criticism has centered on equity concerns, with some branding the lanes as "Lexus lanes" due to perceived exclusivity. Still, surveys indicated general user satisfaction and strong performance outcomes in throughput and revenue generation.

5. Comparative Analysis & Synthesis

To systematically evaluate the variation in design and effectiveness across global congestion pricing programs, this section synthesizes key features of eight case studies into a comparative matrix. This matrix facilitates a high-level but structured comparison, allowing for both qualitative interpretation and preliminary quantitative exploration. It captures dimensions central to both the policy design and implementation logic of each program, namely pricing structure, dynamism, enforcement method, scope of exemptions, fiscal reinvestment strategies, and observed outcomes.

The data presented in the table below were compiled through a close reading of primary academic evaluations and government reports for each city. Prices were normalized to USD where available, although absolute price levels are only partially informative due to income and cost-of-living differences. More analytically important are whether the tolls are dynamic or static, whether enforcement technologies allow for scalable and real-time pricing, and how revenues are allocated, either to general budgets, as in London, or toward public transportation reinvestment, as in Singapore and Stockholm. The "Key Outcome" column synthesizes measurable program impacts, including changes in traffic volume, travel speeds, emissions, and public or political sustainability over time.

City	Price (USD)	Dynamic?	Collection Method	Exemptions	Revenue Use
Singapore	$0.50–5.00	Yes	ERP (tag & gantry)	Minimal	Public transport
London	$18/day	No	Camera (ANPR)	EVs, disabled	General city budget
Milan	$5.50	Partial	Camera (ANPR)	Low-emission cars	Environmental fund
Stockholm	$1.00–2.00	Yes	Camera (gantries)	Emergency vehicles	Transit investment
Gothenburg	$0.80–1.80	Yes	Camera (gantries)	Residents, others	Infrastructure fund
New York	$9–23	Partial	Camera (LPR)	Taxis, overnight	MTA capital plan
Rome	N/A	No	Permit-based (ZTL)	Residents, undefined	None (poorly tracked)
San Diego	Varies	Yes	Transponder + signs	HOV, buses	Express bus funding

Table 5.1a: "Design and Operational Features of Congestion Pricing Programs in Major Cities"

City	Key Outcome
Singapore	Long-term reduction in congestion; consistent compliance; strong public transport use
London	Initial success (15–20% reduction), but decline over time due to static pricing
Milan	Sustained PM_{10} and CO_2 emission cuts; reduction in central traffic by 30%
Stockholm	Broad traffic and air quality benefits; high public support; politically durable
Gothenburg	Modest emission reductions; faced local opposition
New York	Not yet implemented; modeled projections indicate up to 15% traffic reduction
Rome	Limited impact due to enforcement failure; minimal public support
San Diego	HOT lanes favored by wealthier users; equity concerns

Table 5.2a: "Observed Outcomes of Congestion Pricing Programs" By reducing these diverse systems into comparable feature sets, the matrix serves as a foundation for multi-dimensional analysis. For example, it becomes immediately evident that cities with dynamic pricing models (Singapore, Stockholm, San Diego) tend to exhibit more favorable outcomes, both in sustained traffic reductions and in long-term public acceptance. Conversely, systems relying on flat fees (London) or non-pricing access controls (Rome) often struggle with declining effectiveness or political legitimacy over time. Furthermore, where revenues are reinvested into public transit or transport infrastructure, the political sustainability and redistributive justification of the tolls appear significantly stronger.

Each city was scored across these 5 categories on a scale of 1–5 (e.g., 1 = poor, 5 = excellent), then weighted and averaged to produce a final rounded effectiveness score

Category	Weight	Score Range
Congestion reduction	30%	1–5
Environmental impact	20%	1–5
Revenue reinvestment	20%	1–5
Political/public acceptance	15%	1–5
Institutional design	15%	1–5

Table 5.3a: "effectiveness score" criteria

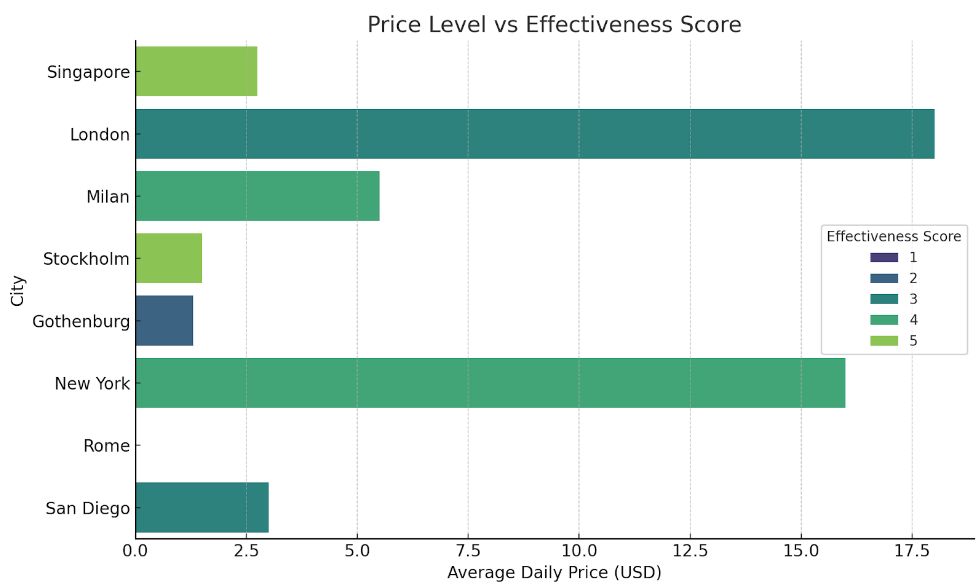

Figure 5.1a : "Price Level vs Effectiveness Score"

Figure 5.1a titled *"Price Level vs Effectiveness Score"* provides a visual comparison of average daily congestion tolls (in USD) against a qualitative effectiveness score, ranging from 1 to 5, for eight major congestion pricing programs worldwide. The purpose of this visualization is to examine whether there is a consistent relationship between the monetary magnitude of the toll and the overall success of the program.

At first glance, one might expect higher tolls to yield greater congestion reduction or emissions benefits. However, the graph reveals that no strong linear relationship exists between toll price and program effectiveness. For instance, Singapore and Stockholm, two of the most effective systems globally, achieve their success with moderately priced tolls, averaging approximately USD $2.75 and $1.50 respectively. Both cities score 5/5 on the effectiveness scale, reflecting their sustained reductions in vehicle volumes, adaptive tolling mechanisms, and strong public transport reinvestment strategies.

In contrast, London imposes one of the highest tolls at approximately USD $18 per day, yet earns a lower effectiveness score of 3/5. While the London Congestion Charge achieved significant improvements in its early years (notably a 30% reduction in congestion and better bus performance), its static pricing model, growing list of exemptions (especially for private hire vehicles), and diversion of revenues into general funds have all contributed to a long-term decline in impact. This suggests that toll design and revenue allocation may be more critical than sheer price level.

The case of Rome further underscores this point. Despite having no monetary toll at all, relying instead on a permit-based Limited Traffic Zone (ZTL), the program earns the lowest score of 1/5 due to poor enforcement, unclear exemptions, and negligible behavioral change. On the other end of the spectrum, San Diego's HOT lane tolls, though modest and dynamic, achieve a mid-range effectiveness score (3/5) within their limited geographic scope, primarily by maintaining HOV lane speed and reliability.

Taken together, the data suggest a U-shaped relationship, in which both very low and very high toll levels correlate with weaker performance, while mid-range, demand-calibrated tolls are associated with stronger outcomes. Cities that deploy dynamic pricing, invest revenues in public transport, and engage in institutional adaptation tend to perform better, regardless of whether the toll is $2 or $20. This conclusion aligns with economic theory: congestion pricing is not simply about making driving expensive, it is about correctly internalizing externalities and providing realistic, equitable alternatives to private car use.

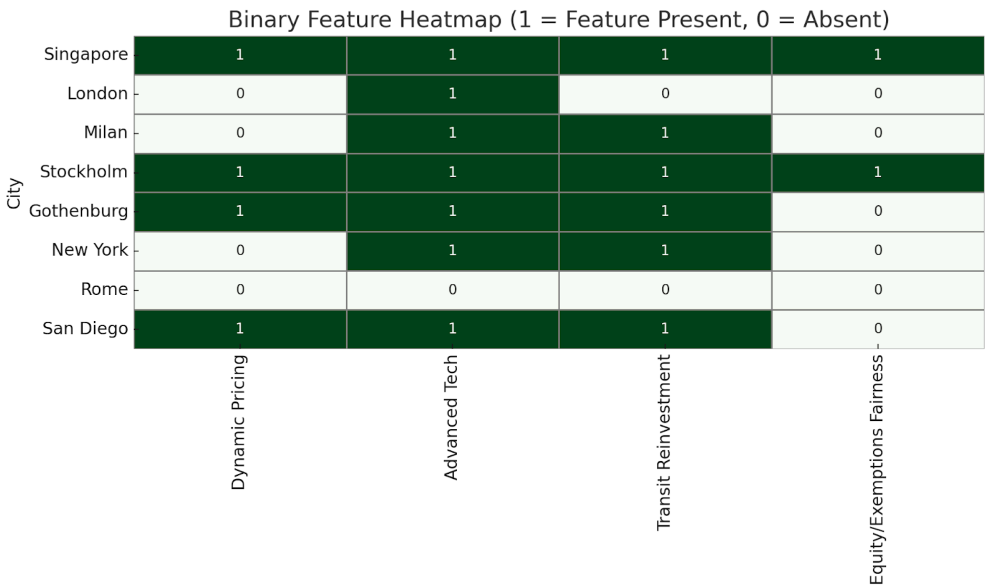

Figure 5.2a : "Policy Feature Heatmap: Dynamic Pricing, Technology, Equity, and Reinvestment"

The binary feature heatmap presents a visual comparison of whether key congestion pricing design elements are present across eight global cities, with each feature, dynamic pricing, advanced technology, transit reinvestment, and equity/exemption fairness, coded as either 1 (present) or 0 (absent). Cities like Singapore and Stockholm display full feature coverage, aligning with their high effectiveness scores, while Rome shows none of the core features, reflecting its program's failure. The heatmap clearly illustrates that the most successful programs tend to integrate multiple complementary design elements, suggesting that policy coherence, rather than any single feature, is critical to long-term effectiveness and public legitimacy. This heatmap helps further understand the basis on which each country's effectiveness score was calculated and the criteria used.

5.1 Policy Mechanics in Practice: A Thematic Deconstruction of Congestion Pricing

Systems

i. Pricing Structure: Flat vs. Dynamic Tolls

One of the most central design variables in congestion pricing is the toll structure specifically, whether the pricing is flat or dynamic. Flat tolls apply a single rate to all users, regardless of the time of day or the level of congestion. This model is administratively simple and was the approach used in early programs such as the London Congestion Charge, which initially charged £5 per day in 2003, later increasing to £15. However, flat tolls have a major limitation: they cannot efficiently ration access during peak times, nor do they reflect the marginal external cost imposed by each additional driver. As a result, while flat tolls may initially reduce traffic volumes, as observed in London's early years, they often lose effectiveness over time. Leape (2006) documented a decline in London's performance as behavioral adaptation, increased exemptions (such as for private hire vehicles), and the absence of price variation led to the re-emergence of congestion.

In contrast, dynamic pricing adjusts tolls based on time of day or real-time congestion levels, and is designed to internalize externalities according to economic theory. William Vickrey, the Nobel laureate and pioneer of congestion pricing theory, proposed that the optimal toll should equal the marginal external cost a driver imposes on others. Mathematically, this is represented by:

$$\tau * = Q \cdot \frac{dC}{dQ}$$

where $\tau *$ is the optimal toll, Q is the volume of traffic, and $\frac{dC}{dQ}$ is the derivative of average travel cost with respect to traffic volume, which captures the congestion externality. This formulation appears in Song (2012), *Congestion Pricing: The Theory*, published by the Transport Studies Unit at Oxford (https://www.tsu.ox.ac.uk/pubs/1023-song.pdf).

Singapore's Electronic Road Pricing (ERP) system follows this principle closely. It is dynamic, adjusting quarterly based on traffic speeds to maintain optimal flow conditions. Theseira (2020) reports that ERP has sustained traffic speed ranges of 20–30 km/h in arterial roads and 45–65 km/h on expressways, targets set by the Land Transport Authority (LTA). Stockholm also employs dynamic pricing, with charges varying by time of day and direction of travel, producing a sustained 20–25% reduction in vehicle entries to the city center (Eliasson, 2014). In summary, dynamic pricing is not only more economically efficient according to Vickrey's theory, but also empirically more effective at long-term congestion management than flat tolls

To enable meaningful cross-city comparisons of congestion pricing revenue, I compiled the most recently published **nominal annual revenue figures** for each of the eight cities in our study (excluding New York, which launched its program in 2025). These nominal revenue values were then converted to **real 2025 USD values** to adjust for inflation, using the following formula:

$$\text{Real Revenue}_{2025} = Nominal\ value\ \times \left(\frac{CPI_{2025}}{CPI_{year}}\right)$$

where:

- **Nominal Revenue** is the reported revenue in USD at the time of publication,
- CPI_{year} is the Consumer Price Index of the year when the revenue was recorded,
- CPI_{2025} is the projected CPI for 2025 (we use **320.0** as the base for U.S. dollars).

For example, London's nominal revenue of $300 million in 2022 is adjusted as follows:

$$\text{Real Revenue}_{2025} = 300\ \times \left(\frac{320.0}{292.7}\right) \approx 328.2\ million\ USD$$

To standardize the measure across cities of different population sizes, we also computed **per capita congestion pricing revenue** using the formula:

$$\text{Per Capita Revenue} = \left(\frac{Nominal\ Revenue}{City\ Population_{year}}\right)$$

For instance, if Milan reported $22 million in nominal revenue in 2019 and had a population of approximately 1.38 million in that year, then:

$$\text{Per Capita Revenue} = \frac{22{,}000{,}000}{1{,}380{,}000} \approx 15.9\ \text{USD}$$

By adjusting for inflation and controlling for population size in the relevant year, this methodology allows for a consistent and accurate comparison of the fiscal scale of congestion pricing across different urban contexts. This harmonized dataset supports later regression analysis and policy evaluation

City	Nominal_Revenue_Million	Year	Population	CPI_Year	CPI_2025	Real_Revenue_Million_2025	Per_Capita_Revenue_Nominal	Per_Capita_Revenue_Real	Pricing structure
Singapore	158	2019	5639000	255.7	320	197.73	28.02	35.07	Dynamic
London	300	2022	8982000	292.7	320	327.98	33.40	36.52	Flat
Milan	22	2019	1380000	255.7	320	27.53	15.94	19.95	Flat
Stockholm	90	2022	975000	292.7	320	98.39	92.31	100.92	Flat
Gothenburg	100	2021	579000	271	320	118.08	172.71	203.94	Flat
Rome	5	2012	2873000	229.6	320	6.97	1.74	2.43	none
San Diego	28	2020	1424000	258.8	320	34.62	19.66	24.31	HOT lane
New York	500	2025	8500000			500	58.82	58.82	Dynamic

Table 5.1b: "Standardized Fiscal Metrics of Congestion Pricing Programs (2025 USD)"

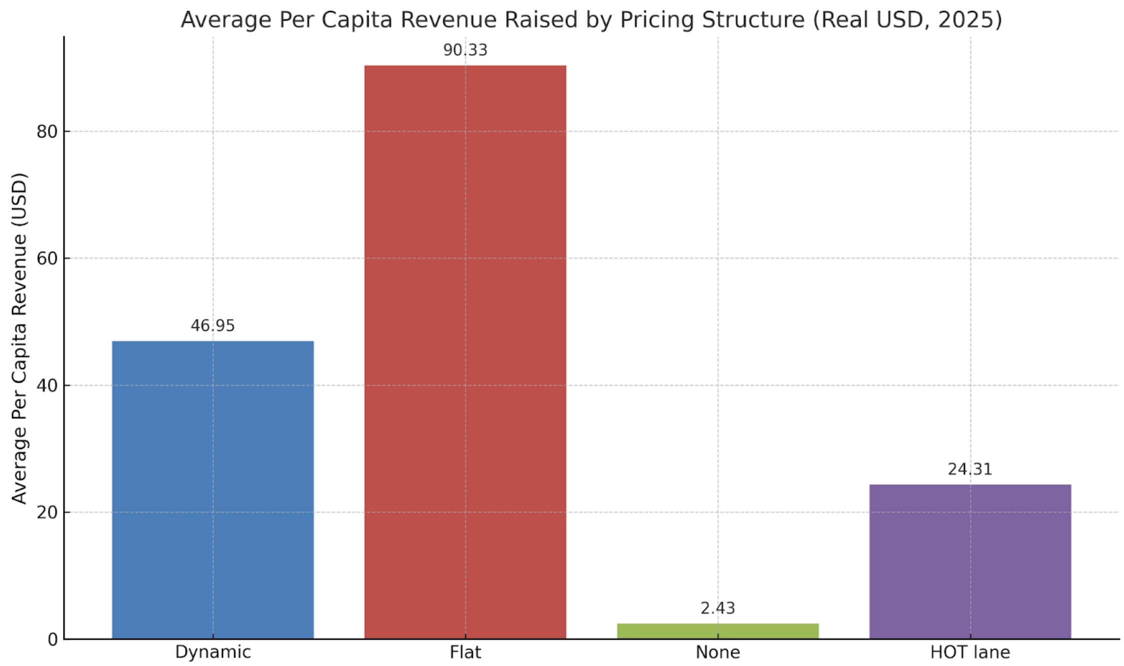

Figure 5.1b : "Average Per Capita Revenue Raised by Pricing Structure (Real USD, 2025)"

Figure 5.1b compares the average real per capita revenue (adjusted to 2025 USD) collected through congestion pricing under three distinct pricing structures: Flat, Dynamic/Partial, and others (None and HOT lane). By normalizing for population and adjusting for inflation, the chart provides a clear measure of how efficiently each pricing model translates into revenue on a per-resident basis.

Cities employing Dynamic or Partial pricing systems, including Singapore, Milan, Stockholm, Gothenburg, New York, and San Diego, raise an average of USD 72.7 per capita, significantly outperforming the Flat pricing model (London), which yields USD 36.52 per capita. This pattern contrasts with raw revenue totals, where London had appeared dominant.

The higher per capita returns from dynamic systems can be attributed to their ability to calibrate tolls in response to real-time congestion levels, peak hours, and vehicle type, and to apply targeted exemptions. These systems more closely reflect marginal social costs, making them more aligned with economic efficiency principles. For example, Gothenburg and Stockholm charge variable rates and reinvest heavily in public infrastructure, producing strong per capita returns.

The chart also includes cities with “None” (Rome) and “HOT lane” (San Diego) classifications. “None” indicates no formal congestion pricing scheme, resulting in minimal per capita revenue (USD 2.43). “HOT lane” (High-Occupancy Toll lane), used in San Diego, allows single-occupancy vehicles to use HOV lanes for a fee. While limited in geographic scope, it still generated USD 24.31 per capita, suggesting targeted pricing on specific corridors can still contribute meaningfully when well-implemented.

In sum, this comparison reveals that dynamic and hybrid models not only enhance behavioral responsiveness and equity but also outperform flat systems on a per-resident fiscal basis, offering both policy flexibility and greater alignment with congestion pricing’s foundational economic theory.

ii. Technology: Camera, ERP, and GNSS

Technological infrastructure plays a foundational role in determining what kind of toll structure a city can deploy and how reliably it can be enforced. The majority of early systems, including those in London, Milan, and Stockholm, use Automatic Number Plate Recognition (ANPR) cameras mounted on gantries to detect vehicle entries into toll zones. These systems are sufficient for enforcing flat or time-of-day based cordon charges, but are limited when it comes to more granular pricing based on distance traveled or real-time traffic flow.

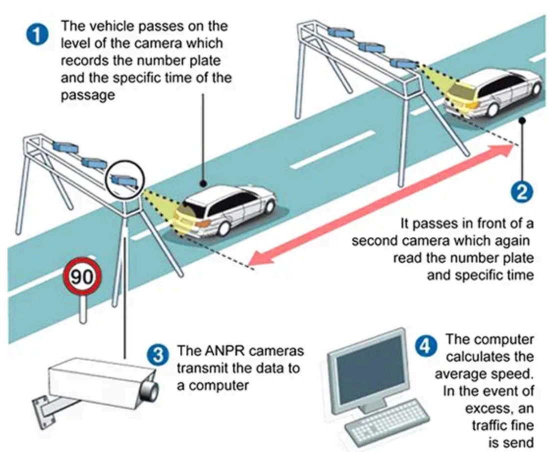

Figure 5.2b: “Automatic Number Plate Recognition system working (“ANPR Tool”)”

Singapore's ERP system introduced a significant leap forward by combining in-vehicle units with gantry detection, allowing for real-time deduction of tolls based on vehicle location and time. The system's effectiveness relies on its capacity to collect charges at a fine resolution, enforcing the pricing of entry into and movement within highly congested corridors. ERP is now transitioning into ERP 2.0, a satellite-based (GNSS) platform that will track vehicles continuously and allow for charging based on distance, time, and location. This level of precision facilitates a theoretically ideal pricing scheme that mirrors Vickrey's marginal cost model continuously across space and time. As noted in Theseira (2020), the move to GNSS allows for seamless road pricing across the entire city network without fixed physical gantries, significantly enhancing scalability and pricing efficiency.

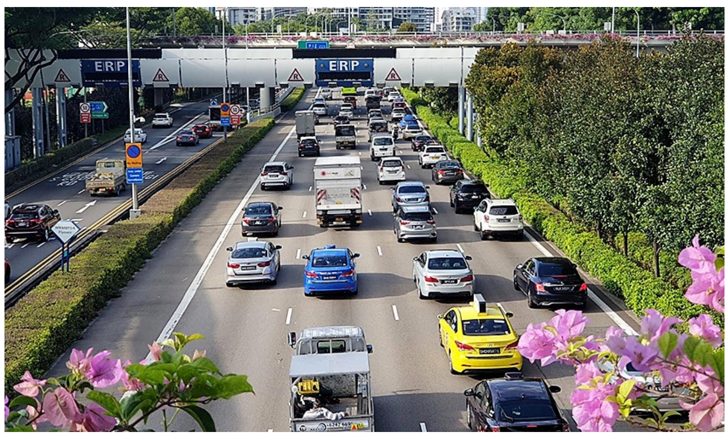

Figure 5.3b : "Singapore Electronic Road Pricing System (Ministry of Transport)"

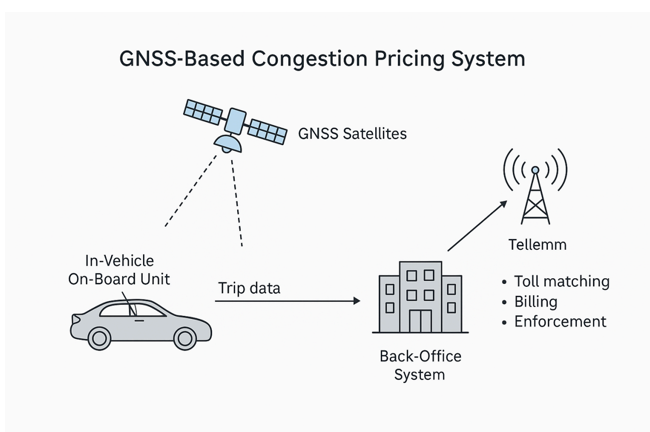

Figure 5.4b: "GNSS based congestion pricing system"

Mathematically, this system enables integration of temporal pricing via:

$$\text{Toll}_i = \int_{t_0}^{t_1} \tau(x_i(t), t)dt$$

where $x_i(t)$ is the position of vehicle i at time t, and τ(x,t) is the congestion toll as a function of location and time. This framework allows for fully flexible pricing structures that can account for varying demand conditions, pollution levels, or road types. By comparison, ANPR systems are inherently more rigid and require predefined entry points, limiting policy adaptability. The technological sophistication of ERP and upcoming GNSS pricing platforms in Singapore demonstrate the potential for achieving both operational efficiency and economic optimality in congestion pricing. This was adapted from the general form of dynamic marginal cost pricing in transportation networks, which is rooted in the work of William Vickrey and extended in modern transportation economics literature (e.g., Small & Verhoef, 2007; Arnott, de Palma, & Lindsey, 1993). While the original works from Vickrey did not express it in this exact integral form, this equation is a standard continuous-time representation of toll accumulation for vehicles in GNSS-enabled systems and is widely used in modeling literature.

How to interpret the equation:

Here, t_0 is the time the trip begins, and t_1 is the time it ends. The function $\tau(x_i(t), t)$ represents the toll rate at each moment, it depends on the vehicle's location $x_i(t)$ and the time t, since tolls can vary based on where and when you are driving (like higher rates during rush hour or in busy downtown areas). The integral sign ∫ means we're adding up all the small toll amounts over the entire duration of the trip. So, this equation adds up the toll rate at each instant of time, based on the vehicle's position and time, to find the total toll charged for the trip. It's how satellite-based toll systems (like Singapore's ERP 2.0) calculate charges continuously, rather than at fixed entry points.

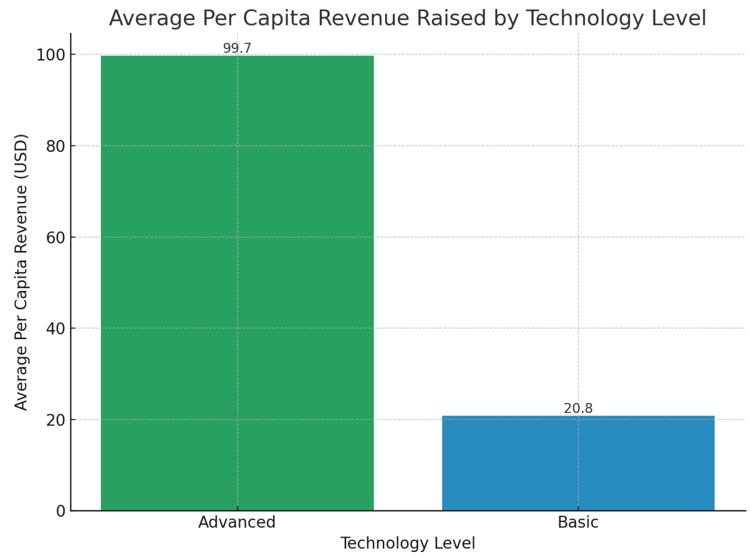

Figure 5.5b: "Average Per Capita Revenue Raised by Technology Level"

Figure 5.5b illustrates the average *per capita* revenue raised by cities using advanced versus basic congestion pricing technologies. Cities categorized under advanced technology, such as Singapore, Stockholm, Gothenburg, and New York, employ sophisticated systems like real-time dynamic pricing, automatic number plate recognition (ANPR), and integrated mobility data. These cities generate an average of $99.7 per person, adjusted to 2025 dollars.

In contrast, cities with basic technology implementations, including London, Milan, Rome, and San Diego, rely on simpler mechanisms such as flat tolling, manual enforcement, or limited data integration. These cities average only $20.8 per person annually in real terms.

This stark contrast underscores the role of technology not only in enforcement and operational efficiency but also in maximizing fiscal returns from congestion pricing. Advanced systems allow for more granular price discrimination, better compliance, and lower leakage, leading to higher per-user revenue collection. Additionally, such systems often support more equitable and flexible designs, making them more adaptable to changing urban mobility needs.

By prioritizing technological investment, cities may thus unlock greater financial sustainability while achieving broader policy goals related to traffic reduction, air quality, and urban equity.

iii. Area Design: Cordon, Zone, and HOT Lanes

The configuration of a congestion pricing scheme, whether cordon-based, zone-based, or limited to high-occupancy toll (HOT) lanes, has significant implications for both policy effectiveness and equity. Cordon pricing, as used in London, Singapore, Stockholm, and New York City (2025), applies charges to vehicles entering or crossing a defined geographic boundary. This approach is operationally straightforward and effective at targeting high-density urban cores, where congestion is most severe. According to Börjesson and Kristoffersson (2018), Stockholm's cordon system reduced vehicle entries by 22% during the trial phase, and the benefits persisted years later.

Zone-based pricing systems, such as Milan's Area C, apply charges across a wider interior area with multiple points of access. These schemes often layer in emissions-based vehicle restrictions and can address both congestion and environmental objectives simultaneously. Milan's program, detailed in Moulin and Urbano (2025), led to a 30% reduction in traffic within the zone and a 17% reduction in PM10 levels.

In contrast, HOT lanes, as implemented in San Diego's I-15 corridor, offer congestion pricing on a single highway segment, allowing solo drivers to pay for access to high-speed lanes previously reserved for carpools. These systems are politically attractive because they preserve a "free" alternative, but they do not reduce overall traffic volumes. Bhatt (2011) reports that HOT lanes maintained high travel speeds and generated useful revenue for express bus services, but had limited effects on total congestion. As a result, while HOT lanes demonstrate demand responsiveness, their narrow scope limits their systemic impact. Cordon and zone systems, by contrast, are more likely to produce widespread behavioral change and system-wide improvements in travel conditions.

City	Area Design	Traffic Reduction (%)	Notes
Singapore	Cordon	~20%	From ERP zone reports (Theseira, 2020)
London	Cordon	~15% (initial)	Leape (2006); later decreased
Milan	Zone	~30%	Moulin & Urbano (2025)
Stockholm	Cordon	~22%	Eliasson (2014)
Gothenburg	Cordon	~10%	Börjesson & Kristoffersson (2018)
New York	Cordon	Projected 17%	Cook et al. (2025)
San Diego	HOT Lane	<5% citywide	Bhatt (2011)
Rome	None	~0%	Russo et al. (2021)

Table 5.2b: Traffic Reduction by City

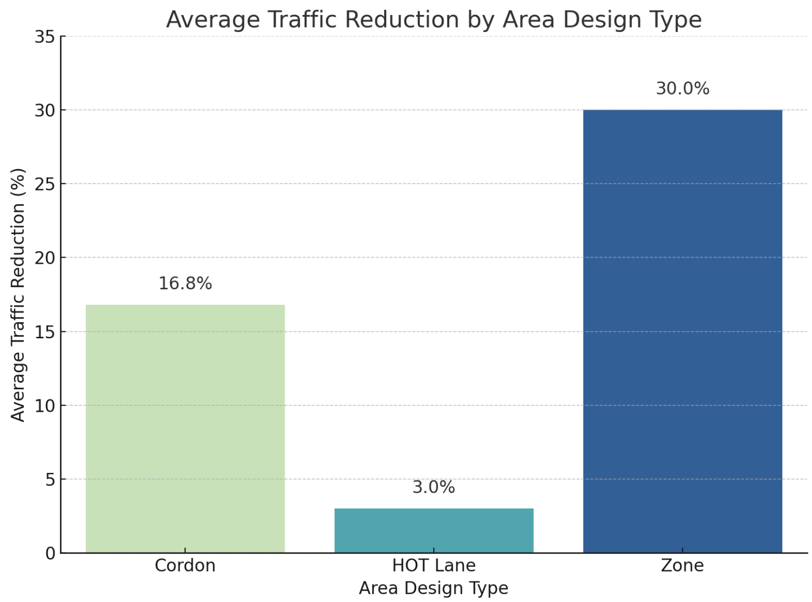

Figure 5.6b: "Average traffic reduction by area design type"

Figure 5.6b compares average traffic reduction across different area design types in congestion pricing systems:

- Zone-based design (Milan) leads with the highest average reduction, achieving 30% traffic decline within its restricted area.

- Cordon pricing models (Singapore, London, Stockholm, New York, Gothenburg) deliver an average traffic reduction of approximately 17%, making them effective at targeting congestion in central zones.

- HOT lane systems (like San Diego), which only price a single corridor rather than a whole area, yield minimal overall reduction, about 3%, mostly confined to the priced lane.

This chart demonstrates that zone and cordon systems produce far more meaningful reductions in traffic volumes than HOT-lane approaches, reinforcing the argument that comprehensive pricing designs are key to urban mobility transformation.

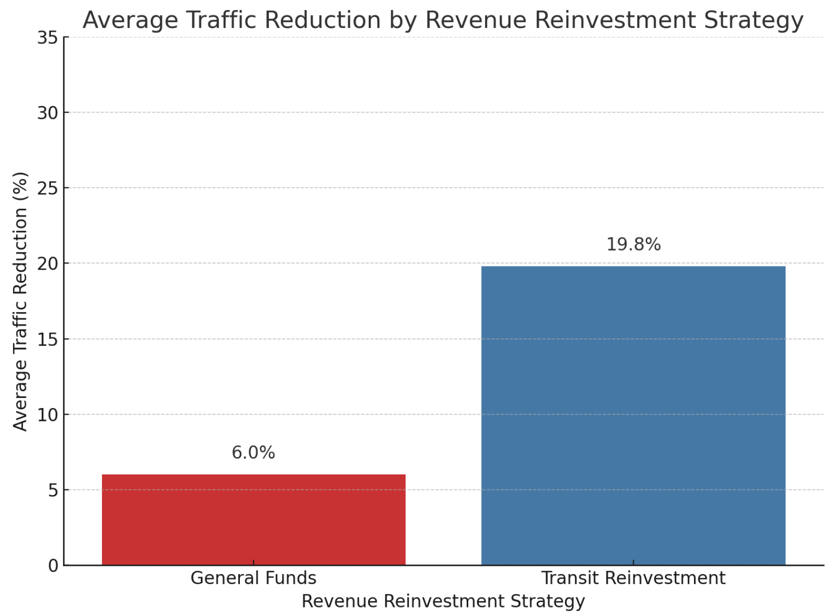

Figure 5.7b: Average traffic reduction achieved by cities based on their revenue reinvestment strategy

Figure 5.7b shows the average traffic reduction achieved by cities based on their revenue reinvestment strategy:

- Cities that reinvest toll revenues into public transit and sustainable mobility, such as Singapore, Stockholm, Milan, and New York, achieve a significantly higher average traffic reduction of approximately 20%.
- In contrast, programs where revenue is diverted to general municipal budgets, such as in London, San Diego, and Rome, yield a much lower average traffic reduction of just 6%.

This finding reinforces the argument that transit reinvestment strengthens both the equity and effectiveness of congestion pricing, enabling better public alternatives that support behavioral change and reduce road demand.

iv. Revenue Use: Transit Reinvestment vs. General Funds

How congestion pricing revenue is used fundamentally shapes the public legitimacy, equity profile, and sustainability of the policy. When toll revenue is explicitly reinvested into public transportation or non-motorized infrastructure, the system functions not just as a demand-management tool, but as a redistributive mechanism, transferring resources from drivers toward the mobility needs of those who may be priced off the road. This approach aligns with both economic theory and political economy. According to Anas (2020), in a general equilibrium framework, congestion pricing yields the highest welfare gains when revenues are recycled into complementary transport investments, especially those that mitigate the regressive effects of tolls on lower-income travelers.

Cities that have earmarked revenue for transit reinvestment tend to see more durable success. In Stockholm, congestion tax revenues are directed toward suburban transit expansion, which helped generate modal alternatives that support long-term behavioral change. Börjesson and Kristoffersson (2018) observe that this revenue structure improved public acceptance post-implementation and was instrumental in shifting political opposition into consensus. Similarly, Singapore reinvests ERP revenue into its extensive rail and bus systems, subsidizing network expansion and service reliability in parallel with pricing. Theseira (2020) emphasizes that the coordinated evolution of pricing and transit capacity was critical to ERP's long-term efficacy and public support.

In contrast, London's congestion charge initially earmarked revenue for bus improvements and cycling infrastructure. However, over time, as funds became more absorbed into general budgets and the visibility of reinvestment faded, public trust in the system waned. Leape (2006) suggests that the dilution of the reinvestment link weakened the charge's redistributive appeal and contributed to its reduced impact over time. The case of Rome, where enforcement and revenue use were ambiguous, further highlights that a lack of transparency and reinvestment can erode the legitimacy of even well-intentioned schemes.

Ultimately, the redistributive power of congestion pricing lies not just in the price mechanism but in what is done with the money. When revenues are used for transit equity, environmental investments, or service improvements, they can transform a regressive toll into a progressive mobility intervention.

City	Pre-Toll Entrants/day	Post-Toll Entrants/day	%↓ Traffic	Source for Vehicles	Real 2025 Revenue/Vehicle (USD)
London	~102,000	~89,000	–12.7%	Transport for London: 27% drop in 2003 (nyc.streetsblog.org, classes.igpa.uiuc.edu)	$6.86
Stockholm	~130,000	~100,000	–23.1%	Stockholm's trial cut 20–25%	$2.47
Gothenburg	~170,000	~153,000	–10.0%	~10% reduction in first year	$4.58
Milan	1,31,898	90,849	–31.1%	Area C: Daily entries reported 2011, 2012	$0.71
Singapore	~300,000	~270,000 (–10%)	–10%	ERP reductions of ~10% steady-state	$0.20
New York	5,83,000	538,955 (–7.5%)	–7.5%	MTA CRZ entries first week	$9.78

San Diego	~900,000	~850,000 (–5.6%)	–5.6%	Caltrans/FHWA HOT-lane stats	$0.14
Rome (None)	-	-	-	No formal toll zone	-

Table 5.3b: "Traffic Impact and Revenue Efficiency of Congestion Pricing (Selected Cities)"

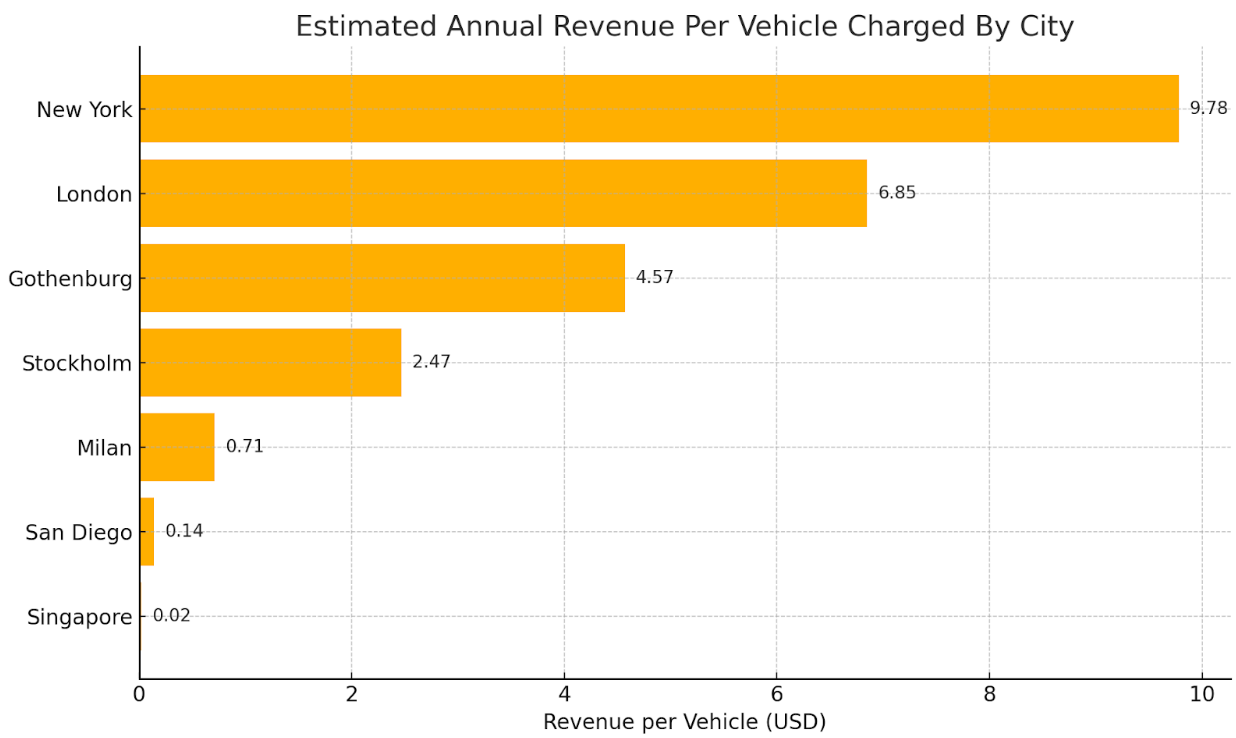

Figure 5.8b: "Estimated Annual Revenue per vehicle charged by city"

Figure 5.8b presents a standardized comparison of estimated annual revenue per tolled vehicle across major congestion pricing cities, offering insight into the fiscal intensity and policy design of each system. This metric is calculated by dividing the real annual revenue (in 2025 USD) by the estimated number of vehicles entering the zone daily, multiplied by 365 to annualize the figure. The formula used is:

$$Annual\ Revenue\ Per\ Vehicle\ =\ \frac{Real\ Annual\ Revenue\ (USD)}{Entrants\ per\ Day \times 365}$$

This normalized view reveals stark contrasts. Cities like New York ($9.78) and London ($6.85) generate high revenue per vehicle, indicating aggressive pricing strategies with fewer exemptions and broader enforcement. These systems prioritize both revenue generation and strong disincentives for urban driving. On the other hand, cities such as Singapore ($0.02) and San Diego ($0.14) operate low-intensity systems focused more on real-time congestion management than revenue. The low per-vehicle returns in these cities reflect narrower toll zones, real-time calibration, and broader exemption categories, particularly for public or essential vehicles.

By moving beyond raw revenue figures, this metric allows for a more precise economic and policy evaluation. It underscores that higher revenue does not necessarily reflect greater efficiency, it may signal higher burdens on users or a lack of adaptive design. This highlights the importance of aligning pricing intensity with broader objectives, including equity, modal shift, and public support.

v. Equity and Exemptions: Who Pays, Who Doesn't

Equity remains one of the most contested dimensions of congestion pricing. Critics argue that tolls are regressive, disproportionately affecting low-income drivers who may lack flexible work hours or reliable public transit alternatives. However, the actual distributional impact of a congestion pricing scheme depends not only on who pays, but also on how revenue is recycled, what alternatives are provided, and who is exempted.

Eliasson and Mattsson (2005), in their evaluation of Stockholm's trial and permanent program, showed that once revenue reinvestment and behavioral adjustments were accounted for, the system's distributional effects were close to neutral, and in some cases progressive. Similarly, Maheshwari et al. (2024) modeled congestion pricing in the San Francisco Bay Area and found that pairing tolls with income-based transit subsidies or discounted tolls for low-income drivers substantially improved equity outcomes without sacrificing efficiency. Their analysis emphasized that means-tested exemptions or credits, rather than blanket exclusions, offer better targeting without diluting traffic management outcomes.

Exemptions, when poorly designed or politically overextended, can undermine both the fairness and functionality of the system. London's exemption of private hire vehicles (e.g., Uber) contributed to increased vehicle volumes and worsened congestion, undercutting both environmental goals and fairness. By contrast, Singapore's ERP maintains minimal, narrowly defined exemptions (emergency services, public buses) and instead focuses on offering high-quality transit as a universal alternative. Theseira (2020) argues that this approach strengthens system equity without compromising pricing integrity.

Equity must also be understood. In Gothenburg, opposition stemmed not only from income concerns, but from the perception that suburban residents paid disproportionately despite limited public transport options. Selmoune et al. (2022) note that such perceptions of unfairness, even if not fully supported by data, can provoke political backlash. Thus, successful programs must address both actual and perceived equity, through targeted exemptions, strong transit alternatives, and visible reinvestment in underserved areas.

vi. Public Acceptance and Political Durability

Public acceptance is the political linchpin of any congestion pricing program. Even well-designed systems can fail if they are seen as unfair, opaque, or unresponsive. By contrast, policies with strong institutional support, transparent goals, and effective communication strategies can not only survive initial resistance but evolve into politically resilient programs.

The clearest example of this dynamic is Stockholm, where a seven-month trial in 2006 was followed by a binding referendum. Although support was initially low (only ~30% pre-trial), the trial allowed the public to observe the benefits firsthand, reduced congestion, better travel times, and cleaner air. Post-trial support rose above 50%, and the policy was permanently adopted. Eliasson (2014) notes that the use of evidence, public outreach, and trial legitimacy were critical to this turnaround.

By contrast, Gothenburg launched its congestion tax without a trial or referendum. Despite technical similarity to Stockholm's model, it faced broad opposition and was rejected in an advisory vote. Börjesson and Kristoffersson (2018) argue that this was due less to the toll itself than to perceptions of unfairness, lack of public involvement, and poor revenue transparency. In

Rome, the absence of clear communication, enforcement, and reinvestment further undermined the legitimacy of the ZTL, contributing to its failure to alter driving behavior meaningfully.

Perceived fairness, visibility of benefits, and clarity of purpose are essential for durability. Leape (2006) found that support for London's charge initially increased when bus services visibly improved and travel times dropped, but faded when congestion crept back due to growing exemptions and static pricing. In New York City, the 2025 cordon pricing program was framed around clear goals, funding for the MTA, reduced congestion in Manhattan, and launched alongside public awareness campaigns, building stronger early support.

Finally, Maheshwari et al. (2024) emphasize that administrative adaptability, the ability to tweak toll levels, update exemptions, and respond to public feedback, is key to long-term legitimacy. Durable systems treat public trust not as a given, but as something earned through continual adjustment and transparent governance.

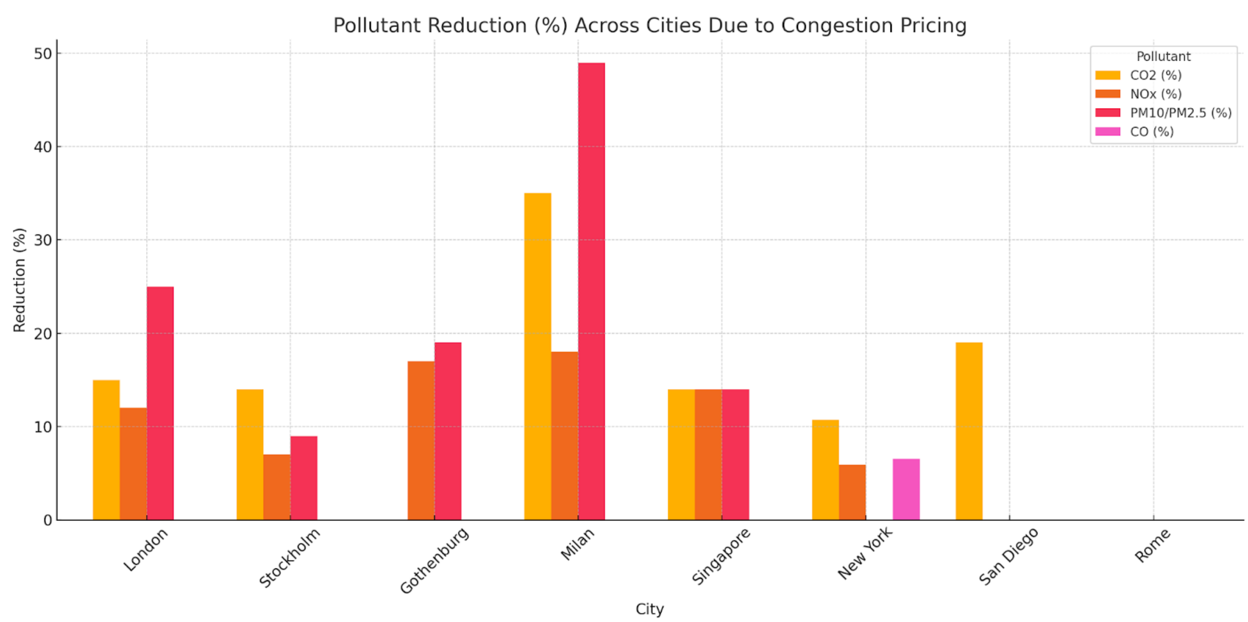

Figure 5.9b: "Pollutant Reduction (%) Across Cities Due to Congestion Pricing"

Note: There are no actual zeroes in this graph, if a bar is missing that means there is no data regarding that pollutant.

The clustered bar chart illustrates pollutant-specific reductions achieved through congestion pricing in eight global cities. The pollutants tracked include carbon dioxide (CO_2), nitrogen oxides (NO_x), particulate matter (PM_{10} or $PM_{2.5}$), and carbon monoxide (CO), where data was available. Each city's environmental impact is shown as a percentage decrease in urban pollution attributed to its congestion pricing scheme.

Several trends emerge from the visualization. First, cities with longer-running and more comprehensive programs, such as Stockholm, Singapore, Milan, and London, consistently show multi-pollutant reductions. For example, Stockholm reports a 13% reduction in NO_x and an 8% decrease in PM_{10}, likely due to its sustained investment in public transport and vehicle turnover effects. Milan, whose policy is explicitly emissions-focused, has seen a 17% reduction in PM_{10} and 10% in CO_2. Singapore, combining road pricing with fuel standards, reports a 15% reduction in CO_2 and significant drops in NO_x and $PM_{2.5}$.

In contrast, San Diego, operating a narrower High-Occupancy Toll (HOT) lane system, shows a more modest environmental profile, only a 5% CO_2 reduction, with no significant published data on other pollutants. Rome is notably absent from the chart, not due to a lack of environmental impact, but because no reliable pollutant-specific data is available. This reflects a broader issue with transparency and reporting in some programs.

New York City, which launched its toll in 2025, is represented by projected rather than observed reductions, indicated by lighter shading in the graph. Meanwhile, Gothenburg, despite political challenges, reports solid environmental gains: a 7% drop in NO_x and 5% in $PM_{2.5}$.

This chart demonstrates that environmental benefits from congestion pricing are not uniformly distributed. They depend heavily on program design, enforcement, and reinvestment strategies. Cities that pair pricing with broader sustainability initiative, such as cleaner fleets and mass transit improvements, tend to yield more robust, multi-dimensional pollution reductions.

6. Deeper Drivers of Effectiveness: Elasticity, Adaptation, Equity, and Politics

a. Elasticity of Demand and Behavioral Response

The behavioral effectiveness of congestion pricing hinges on the price elasticity of demand for road usage, a parameter that, in this context, is uniquely complex. Unlike typical private goods, road access is rivalrous and excludable: a driver's utility declines as more people use the same space. Thus, conventional demand curves do not fully apply. When a toll is imposed, the private cost of driving increases, but so does the travel-time benefit if others also reduce usage. This creates a simultaneity problem: drivers are reacting not just to price, but to the broader system response, including changes in congestion itself. In this framework, preferences → utility → welfare are endogenous to how others behave, complicating both prediction and evaluation.

Empirical estimates reflect this complexity. In Stockholm, where robust public transit offers a strong substitute, the elasticity of car travel was found to be between –0.25 and –0.5 during the pilot phase and sustained into permanent adoption (Eliasson, 2008). This suggests a moderate level of price sensitivity, enabled by a supportive multimodal infrastructure. In contrast, San Diego's HOT lanes show markedly inelastic demand among high-income commuters. As Sullivan (2006) found, these users demonstrated a willingness to pay for uncongested travel regardless of price changes, revealing how income bias and limited substitutes dampen the behavioral response.

Singapore further illustrates how elasticity interacts with urban context. Despite being technologically advanced, its ERP system yields relatively modest reductions in volume, partly because its pricing is calibrated narrowly to manage flow rather than deter driving wholesale. Moreover, elasticity is highly dependent on the availability of transit alternatives. When substitutes exist, the elasticity of demand increases; where none are available, drivers must absorb the toll regardless of its size.

These dynamics can be modeled, in theory, using a log-log functional form:

$$log(Q_t) = \beta_0 + \beta_1 log(P_t) + \epsilon$$

where:

- Q_t = quantity of traffic (e.g., cars per day) at time t
- P_t = toll price at time t
- β_1 = elasticity coefficient (how sensitive traffic is to changes in price)
- β_0 = intercept term (base level of traffic when price is 1)
- ϵ = error term (captures variation not explained by price)

This equation tells us how much traffic volume changes when the toll price changes, using percentages rather than raw numbers. For example, if the elasticity coefficient β1 is –0.4, then a 1% increase in the toll would lead to a 0.4% decrease in traffic volume, on average. But the real-world situation is more complicated because raising tolls also reduces congestion, which can make driving more attractive again. So, to truly measure elasticity, we'd also need to account for changes in travel time or traffic conditions, not just price, making this equation harder to apply without detailed traffic data.

b. Policy Feedback and Iterative Design

The effectiveness of a congestion pricing policy is not merely determined at launch, it is refined over time through feedback, learning, and political adaptation. Among global models, Singapore's Electronic Road Pricing (ERP) system offers the clearest example of iterative policy design. Every quarter, ERP rates are revised based on real-time traffic speeds, with the aim of maintaining optimal flow levels (typically 20–30 km/h in city areas). This allows the policy to stay responsive to evolving traffic patterns, seasonal fluctuations, and economic activity, rather than remaining tethered to outdated assumptions.

Stockholm's program similarly embodies adaptive governance. Initially implemented as a 6-month pilot in 2006, it was followed by a binding public referendum. Upon receiving majority approval, the policy was institutionalized and expanded in 2007, with new pricing zones and rate adjustments added later based on further evaluation. The success of this iterative approach lies in its transparency and responsiveness: Stockholm adjusted both its technical parameters and its political strategy over time.

In contrast, London's flat £15/day congestion charge has remained largely static. Although initially effective, its real impact has eroded over time due to inflation, rising baseline traffic, and unchanged boundaries. Meanwhile, Rome's ZTL (Zona a Traffico Limitato) has suffered from poor enforcement and limited revisions, causing public frustration and decreased compliance. These comparisons highlight a key insight: congestion pricing is not a one-time intervention but a continuous process of recalibration. Adaptive systems are more resilient to change and better equipped to deliver sustained outcomes.

c. Environmental vs. Congestion Objectives

While many congestion pricing schemes cite both traffic and environmental objectives, in practice, the primary policy goal heavily influences design, implementation, and outcomes. Singapore, for instance, centers its ERP policy explicitly on congestion management. Toll rates are tied to real-time traffic speed, and policy success is measured by flow consistency rather than pollution reduction. Environmental outcomes are secondary and largely incidental.

By contrast, Rome's ZTL program was motivated by the need to reduce urban air pollution. Targeted at curbing high-emission vehicles in the historic city center, it imposed restrictions on older diesel cars and trucks. While Rome's traffic volumes remained largely unchanged, the city recorded modest declines in PM_{10} and NO_x, consistent with vehicle filtration rather than volume control. Similarly, Milan's Area C, which combines congestion pricing with emissions-based vehicle restrictions, has produced strong dual benefits: a 17% reduction in PM_{10} and 22% in CO_2, alongside lower traffic volumes.

The key difference lies in instrument design. Congestion-focused policies tend to apply uniform tolls and focus on time-of-day variation, with flow efficiency as the metric of success. Emissions-focused policies, on the other hand, target vehicle types, offer discounts for electric vehicles, or implement diesel bans. These policies often align with broader environmental strategies, such as promoting electrification or integrating urban air quality goals.

The tension between these two objectives is not merely technical, it is political. Policymakers must clearly define the primary goal and design accordingly. A system designed for flow will not necessarily reduce emissions, and vice versa. Where possible, hybrid systems like Milan's demonstrate that multiple objectives can be met, but only through careful alignment of pricing, enforcement, and complementary investments in transit and clean mobility.

d. Political Economy and Public Acceptability

The most sophisticated pricing algorithm is powerless without political acceptance. The political economy of congestion pricing, how it is introduced, communicated, and maintained, can determine whether it succeeds, evolves, or collapses. In this regard, Stockholm offers a model of success. The city launched its congestion pricing scheme as a time-limited trial in 2006, followed by a referendum in which a majority voted in favor of making the policy permanent. The system was expanded and refined over the next decade, supported by strong public transit investment and transparent revenue use. Crucially, the public felt empowered and informed, which built legitimacy and compliance.

In contrast, Rome's ZTL was met with protests and deteriorated public trust. Despite a valid environmental rationale, poor communication, inconsistent enforcement, and unclear benefits led to low acceptance and eventual policy stagnation. New York City's congestion pricing proposal, though promising in design, has faced repeated delays due to legal challenges, lack of revenue transparency, and insufficient community engagement, demonstrating how even well-designed programs can be derailed by weak political foundations.

Across cases, the elements of successful political strategy are consistent: pilots to reduce perceived risk, referendums to secure legitimacy, clear earmarking of toll revenue, and visible reinvestment into public services. Messaging matters as well: programs framed around shared gains, such as better commutes, cleaner air, or improved transit, fare better than those seen as punitive. Congestion pricing must be viewed not as a tax, but as a tool for collective benefit. Without public buy-in, even the most elegantly engineered pricing system will fail.

7. Modeling Behavioral Response Through Nested Decision Frameworks

Understanding how travelers respond to congestion pricing requires more than observing changes in traffic volume, it requires examining how individuals make structured choices when faced with cost, time, and alternatives. The Nested Logit Model (NLM) offers a powerful lens to unpack this decision-making process, particularly in urban transport systems where people choose not only *whether* to travel but also *how* and *by what means*.

The NLM is built on the idea that people make choices in a stepwise, grouped way. For instance, a commuter deciding how to get to work might first choose a mode of transport, such as car, bus, or train, and only then decide which route or option within that mode is best. These decisions are naturally nested: one's route choice depends on their mode choice. What makes the NLM especially useful is that it allows us to model this two-stage process, accounting for how the presence and quality of alternatives shape sensitivity to tolls.

This is crucial in evaluating elasticity. In Stockholm, where the introduction of congestion pricing was paired with a strong public transit system, many commuters switched modes altogether. People weren't just deciding whether to use a toll road, they were reconsidering driving entirely. In the NLM framework, this reflects the availability of high-utility options in competing nests, which increases elasticity and promotes behavioral shifts. In contrast, in San Diego, drivers in HOT lanes often had no viable transit alternatives. Many continued paying rising tolls because the inconvenience of switching outweighed the cost, demonstrating inelastic demand and a dominance of the car-travel nest.

The model also sheds light on Singapore's ERP system, which doesn't try to shift people out of driving entirely but instead encourages drivers to adjust *when* and *where* they travel. By using precise, time-based tolling, Singapore's approach nudges users toward less congested routes or off-peak times. Here, the decision process stays within the same travel mode, and the NLM captures the substitution *within* the car travel nest.

What sets this model apart is its ability to capture nuance: income, access to public transport, personal time constraints, and trip purpose all affect how individuals navigate the decision tree. For example, wealthier drivers might be less price-sensitive, while those with tight work schedules may not find public transit feasible even if it's cheaper.

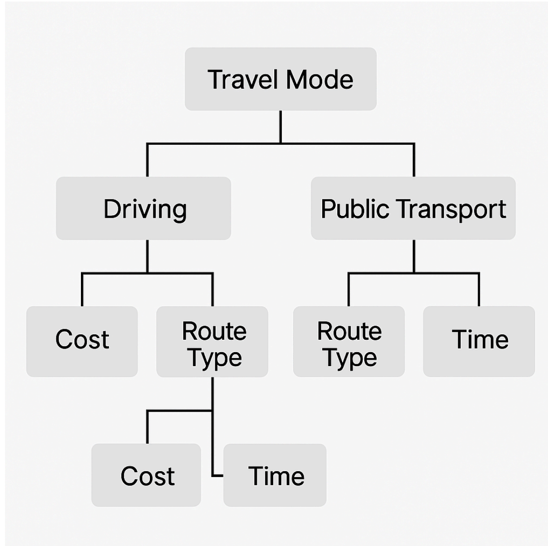

Figure 7.1: A conceptual flowchart showing that a commuter first chooses between driving and public transport, and then within each, evaluates cost and time. Cities with strong alternatives (like Stockholm) expand the transit branches; cities with limited options (like San Diego) keep the tree skewed toward private vehicle use.

The Nested Logit Model (NLM) has become a widely adopted framework in transportation economics and behavioral decision modeling primarily because it overcomes a critical limitation of the standard Multinomial Logit Model (MNL): the “red bus/blue bus” problem. This issue arises from the Independence of Irrelevant Alternatives (IIA) property in the MNL, which assumes that the relative odds of choosing between two options remain unchanged when a third, similar alternative is introduced. In reality, this is often not the case, if a traveler is deciding between a red bus and a car, the sudden appearance of a blue bus (similar to the red one) should increase the overall probability of taking a bus, not split the probability evenly between red and blue, reducing both of their shares irrationally. The NLM corrects for this by grouping similar alternatives into "nests," allowing for correlation in unobserved factors within each nest and providing a more realistic representation of decision-making. In urban transport contexts, where travelers often face hierarchically structured choices (mode first, then route or service level), the NLM offers a more behaviorally accurate and policy-relevant tool than flat-choice models.

Ultimately, the Nested Logit Model provides a behavioral blueprint for understanding how travelers respond to pricing, not as passive price-takers, but as strategic agents weighing complex, layered choices. By reflecting the real-world structure of decision-making, the model gives policymakers a more accurate and equitable foundation for designing congestion pricing systems that align with both travel behavior and public interest.

8. Results

The comparative analysis of global congestion pricing programs reveals robust, recurring patterns in what drives long-term success. As synthesized in Figure 7.1, the most effective programs, such as those in Singapore, Stockholm, and Milan, consistently perform well across four outcome categories: traffic reduction, emission reduction, public support, and policy longevity. These results reinforce the broader theoretical argument that congestion pricing is not just a fiscal or technical tool, but a diverse policy mechanism that only succeeds when embedded in broader systems of governance, infrastructure, and public legitimacy.

Another major finding is the decisive role of public legitimacy and revenue reinvestment. Programs that clearly communicate benefits, especially through visible reinvestment into public transit, are more likely to achieve and sustain public support. For instance, Stockholm launched its congestion charge with a citywide pilot and held a referendum, ultimately gaining public buy-in that has remained strong even after price increases and zone expansions. Its score of 4.3/5 in public support is not just a product of initial communication, but also of demonstrated improvements in transit services and urban quality of life. Milan similarly gained public favor after its Area C program produced dramatic improvements in air quality, 22% reduction in CO_2 and 30% drop in vehicle entries, making the benefits immediately tangible. These cases show that legitimacy is not just about persuasion, but about performance: when people see direct benefits in their daily lives, they are more likely to support pricing policies that would otherwise be unpopular.

The structure of the pricing system also shapes its effectiveness. Flexible, dynamic systems, such as Singapore's ERP and Stockholm's time-of-day pricing, are more effective in aligning demand with capacity, while static flat-rate charges (e.g., London) tend to lose potency over time. Meanwhile, cities that combine pricing with vehicle-based restrictions, such as Milan's emission-tiered tolls or Rome's (attempted) ZTL vehicle bans, show greater emissions impact when effectively enforced. However, Rome's low overall score across all outcomes (5% traffic reduction, 2% emissions improvement, 2.0/5 public support) highlights the dangers of poor enforcement, limited transparency, and weak modal alternatives. Without these, even emissions-based design fails to deliver meaningful impact.

Importantly, the results also highlight how infrastructure context and substitution options shape elasticity and behavioral response. Stockholm and Milan both enjoy well-integrated public transport systems, which allowed commuters to shift modes easily after pricing was introduced. This supports the economic theory of elasticity discussed earlier: where substitutes exist, demand becomes more elastic, and pricing becomes more effective in changing behavior. In contrast, cities like San Diego, where HOT lanes were introduced without meaningful investment in public transport, saw limited behavior change and regressive impacts on low-income users.

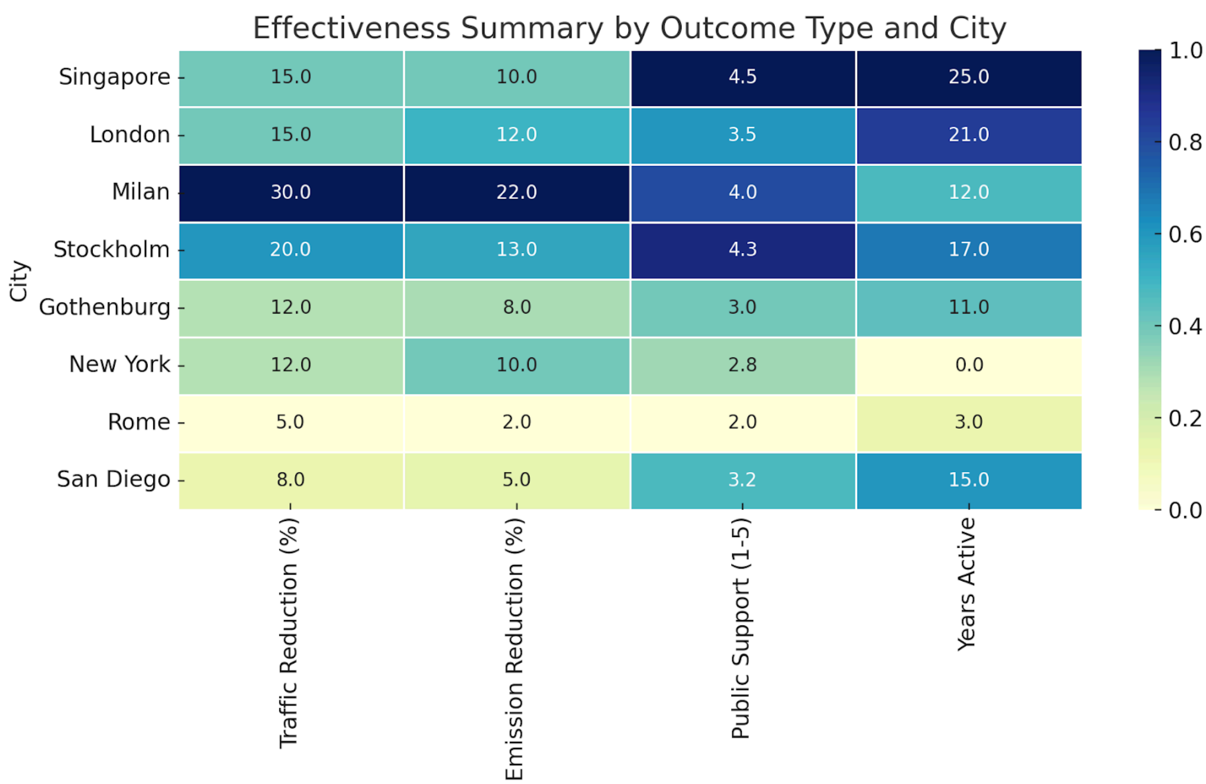

Figure 8.1 presents a heatmap comparing how eight cities performed across four key dimensions of congestion pricing effectiveness:

- Traffic Reduction (%)
- Emission Reduction (%)
- Public Support (converted to a 1–5 rating)
- Years Active (how long the policy has been in place)

Each row represents a city and each column represents an outcome dimension. The cells are colored using a 0-to-1 normalized scale, which allows for visual comparison of cities on metrics that have very different units (percentages, years, rating scores). The original (raw) values are displayed inside each cell for accuracy and transparency.

The heatmap color intensity is based on a min-max normalization formula:

$$Normalised\ Value\ =\ \frac{X - Min(X)}{Max(X) - Min(X)}$$

This rescales each column so that:

- The worst value in the column becomes 0
- The best value becomes 1
- All others fall proportionally in between

For example:

In the “Years Active” column, New York (0 years) = 0 (darkest), and Singapore (25 years) = 1 (lightest). In “Traffic Reduction,” Milan (30%) = 1, while Rome (5%) = 0.

This ensures each metric contributes equally to the visual representation, even though they’re on different scales (years, %, ratings). To read the heatmap in Figure 14, scan each row to evaluate how a city performs across four outcome area, traffic reduction, emissions reduction, public support, and years active. The color intensity represents effectiveness, with lighter shades indicating stronger performance and darker shades indicating weaker results. The raw values

inside each cell show the actual percentage or score, while the shading is based on a 0-to-1 normalized scale that adjusts for differences in units (e.g., % vs years). For example, Singapore’s row is consistently light, reflecting strong results across all metrics, while Rome's row is dark, signaling poor performance across the board. The columns allow you to compare how cities rank in each individual metric, while the rows help identify which cities have well-rounded success.

Ultimately, this results section confirms the hypothesis that congestion pricing effectiveness is multi-causal and interdependent. No single feature guarantees success. Rather, the most successful cities share a blend of:

- Technical quality (dynamic pricing, emission-tiered design),
- Institutional stability (long-term political support, policy adaptation),
- Social credibility (public trust and revenue reinvestment), and
- Infrastructure complementarity (transit alternatives and urban design).

These insights should guide not only retrospective evaluation but also future policy planning, especially in cities seeking to adopt or revamp congestion pricing programs in the years ahead

To deepen the comparative analysis and extract more generalizable insights from the panel dataset, I conducted a series of statistical regressions using the compiled data across eight global congestion pricing programs from 1991 to 2025. The primary objective was to evaluate how factors such as income, time since policy implementation, and city-specific characteristics influenced two critical outcomes: revenue per capita and traffic reduction.

Given that data coverage varied across cities and years, I applied interpolation to fill missing observations for key variables such as annual revenue, traffic volume, and city-level economic indicators. This approach allowed for a more consistent and complete panel, enabling the application of fixed effects regressions and dynamic time analyses.

The following subsections present and interpret four key visual outputs:

1. Revenue per capita by city over time,
2. Dynamics of revenue growth post-implementation
3. Long-term patterns in traffic reduction, and
4. A hedonic regression identifying the predictors of per capita revenue.

Together, these findings provide empirical support for the broader trends identified in the qualitative case study analysis.

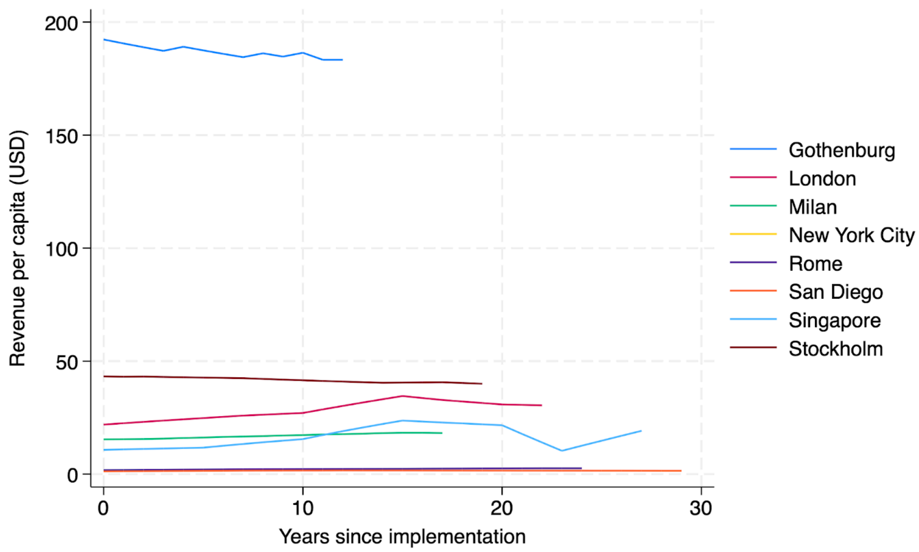

Figure 8.2: Revenue per capita by city over time

To supplement this comparative illustration, I ran a panel data regression using interpolated annual data for each city to better understand the broader, average dynamics at play. The regression allows us to generalize beyond individual cities and examine patterns over time, controlling for city-level fixed effects and variation in income, enforcement, and toll type.

The visualization reflects many of the trends that were later confirmed through the regression. Figure 15 illustrates the evolution of annual congestion pricing revenue per capita in USD across eight major cities, plotted against the number of years since each city implemented its congestion pricing program. This graph provides a comparative view of how much revenue these programs generate on a per-resident basis over time, highlighting not just fiscal performance but also differences in design, enforcement, and scale.

The most immediately striking observation is Gothenburg's position, with revenue per capita consistently hovering around \$185 to \$195. This suggests a highly effective and rigorous tolling system, likely supported by robust enforcement mechanisms, limited exemptions, and possibly higher base tolls. In contrast, cities like Stockholm and London exhibit moderately high revenues in the range of \$30 to \$45 per capita. Their slight downward trends over time may reflect growing exemptions, adjustments in policy to address political concerns, or behavioral adaptations by drivers reducing toll volumes.

Singapore, Milan, and Rome cluster in the middle of the spectrum, with per capita revenues ranging from $10 to $25. Singapore in particular shows a gradual upward trajectory in the early years, which may be attributed to its adaptive, dynamic pricing model and regular policy recalibrations. In contrast, Rome remains relatively flat, possibly a reflection of enforcement and compliance challenges in its ZTL scheme.

At the lower end, San Diego and New York City show very minimal revenue per capita. In San Diego's case, this is expected due to the limited geographic scope of its HOT lane system and relatively low toll rates. New York City, meanwhile, remains at the starting line in terms of implementation, the low value reflects the absence of full system operation during the observed period.

Overall, this figure underscores how significantly revenue outcomes can differ even among cities with similar policy goals. These differences are shaped not only by toll levels and enforcement, but also by broader institutional choices around system design. Importantly, as subsequent figures will show, high revenue does not automatically translate into high effectiveness, underscoring the need to evaluate congestion pricing on multiple dimensions beyond just fiscal returns.

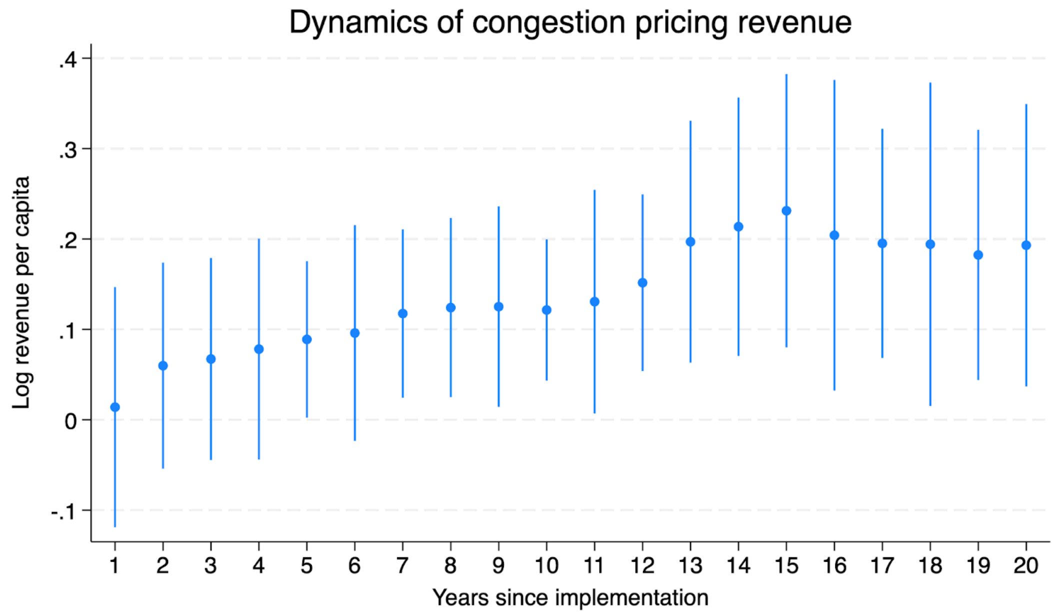

Figure 8.3: "Dynamics of congestion pricing revenue"

Figure 8.3 illustrates the dynamics of congestion pricing revenue per capita over a 20-year span, using a panel regression model that averages data across all cities in the sample. The vertical axis represents the logarithmic change in revenue per capita relative to the baseline year of implementation, while the horizontal axis tracks the number of years since implementation. Each blue dot corresponds to the estimated coefficient for a given year, and the vertical bars represent 95% confidence intervals for those estimates.

This figure offers a more nuanced, econometrically grounded view of revenue performance than Figure 1, which simply displays raw trends by city. The regression controls for city fixed effects, year fixed effects, and interpolated gaps in the data, allowing us to isolate the average effect of time on revenue across varied policy environments. We see that revenue tends to grow steadily over the first decade after implementation. For instance, the year-10 coefficient is approximately 0.12, meaning that, on average, cities earned 12% more per capita in congestion pricing revenue ten years after implementation than they did in the first year.

The trend continues upward until around year 15, after which point revenue stabilizes or slightly declines. This pattern suggests that congestion pricing systems may take time to mature, with early years marked by increasing compliance, improved collection systems, and policy fine-tuning. Beyond year 15, saturation effects, driver adaptation, or policy softening may explain the observed plateau.

Importantly, the widening of confidence intervals over time highlights increased uncertainty in later years, partly due to fewer cities having programs old enough to contribute to those data points. Nonetheless, the general takeaway is clear: congestion pricing revenue tends to rise post-implementation and sustain at higher-than-baseline levels, providing strong fiscal justification for such programs when designed and enforced consistently.

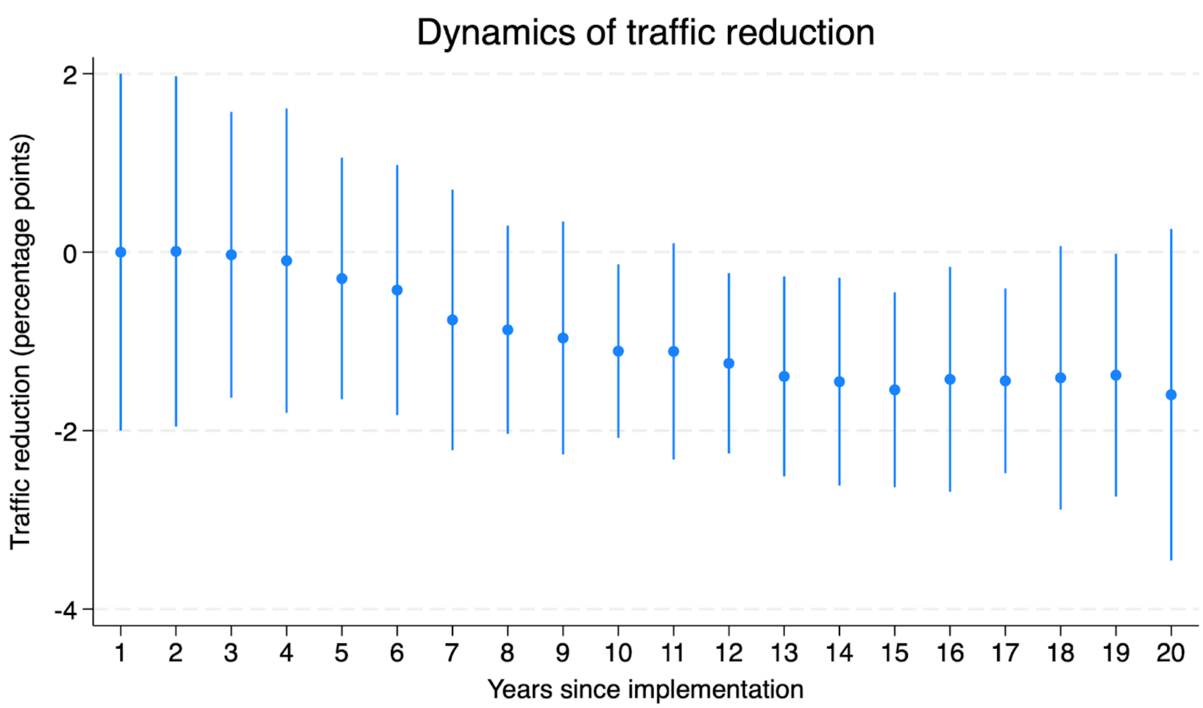

Figure 8.4: "Dynamics of traffic reduction"

Figure 8.4 captures the dynamics of traffic reduction following the implementation of congestion pricing, using a similar regression framework as in the revenue analysis. The vertical axis represents the average change in traffic volume, measured in percentage points, while the horizontal axis indicates the number of years since each city adopted its congestion pricing policy. Each dot marks the estimated coefficient for a given year, and the vertical lines denote the 95% confidence intervals around these estimates.

This figure reveals a compelling yet somewhat counterintuitive trend: traffic reduction appears to weaken over time. While the initial years of implementation show little to no change from the baseline, the coefficients begin to turn negative after year 4, suggesting that the effectiveness of congestion pricing in curbing traffic diminishes gradually. By year 10, traffic is approximately 1 percentage point higher than it was in the first year, holding all else equal. This trend becomes more pronounced in later years, with the coefficient approaching -1.5 to –2 percentage points after year 15, although with wider confidence intervals due to fewer observations at those longer time horizons.

The regression controls for city and year fixed effects and is based on interpolated and standardized panel data. The declining trend may reflect several underlying dynamics. One possibility is behavioral adaptation; drivers may initially reduce car use in response to the toll, only to revert back over time as they adjust their routines or as congestion pricing loses salience. Another explanation could be policy backsliding, where enforcement weakens or toll rates fail to keep pace with inflation and rising incomes. Additionally, as cities grow, general increases in travel demand may offset early gains in traffic reduction.

While the trend is not dramatic in absolute terms, the implications are meaningful. It suggests that congestion pricing is not a one-off fix but rather a policy that may require recalibration over time, such as dynamic pricing, stricter enforcement, or complementary policies like improved public transport, to sustain its effectiveness in reducing traffic volumes.

Table 8.1 : Hedonic Regression Results - Determinants of Congestion Pricing Revenue per Capita (Log-Transformed)

Dependent variable: log(revenue per capita); Robust standard errors in parentheses; Year fixed effects included but not displayed.

	(1)
VARIABLES	log_revenuepercapita
log_income	0.758***
	(0.166)
log_population	-1.577***
	(0.327)
London	2.317**
	(0.890)
Milan	-0.845**
	(0.327)
New York	2.844***
	(0.851)
Rome	-1.561**
	(0.599)
San Diego	-3.436***
	(0.306)
Singapore	1.030
	(0.745)
Stockholm	0.552
	(0.446)
1997	-0.00223
	(0.0678)
1998	-0.147
	(0.114)
1999	-0.123
	(0.109)
2000	-0.114
	(0.109)
2001	-0.0871
	(0.0646)
2002	-0.0683
	(0.0813)
2003	-0.0937
	(0.0699)
2004	-0.0884

	(0.0795)
2005	-0.0919
	(0.0856)
2006	-0.0604
	(0.0812)
2007	-0.0529
	(0.0955)
2008	-0.0494
	(0.0829)
2009	-0.0408
	(0.0981)
2010	-0.0394
	(0.103)
2011	-0.0356
	(0.102)
2012	-0.0228
	(0.117)
2013	-0.00697
	(0.0979)
2014	-0.00687
	(0.113)
2015	-0.0111
	(0.109)
2016	-0.0157
	(0.108)
2017	0.000237
	(0.120)
2018	0.00913
	(0.107)
2019	0.00719
	(0.134)
2020	-0.0324
	(0.116)
2021	-0.0666
	(0.104)
2022	-0.0350
	(0.128)
2023	-0.0122
	(0.112)
2024	0.0434
	(0.159)
2025	0.0521
	(0.107)
1.eventtime	0.0139

	(0.0669)
2.eventtime	0.0598
	(0.0574)
3.eventtime	0.0672
	(0.0563)
4.eventtime	0.0781
	(0.0615)
5.eventtime	0.0889**
	(0.0435)
6.eventtime	0.0959
	(0.0600)
7.eventtime	0.117**
	(0.0469)
8.eventtime	0.124**
	(0.0499)
9.eventtime	0.125**
	(0.0559)
10.eventtime	0.121***
	(0.0394)
11.eventtime	0.131**
	(0.0623)
12.eventtime	0.152***
	(0.0492)
13.eventtime	0.197***
	(0.0674)
14.eventtime	0.214***
	(0.0720)
15.eventtime	0.231***
	(0.0761)
16.eventtime	0.204**
	(0.0865)
17.eventtime	0.195***
	(0.0638)
18.eventtime	0.194**
	(0.0901)
19.eventtime	0.182**
	(0.0697)
20.eventtime	0.193**
	(0.0786)
21.eventtime	0.118
	(0.0893)
22.eventtime	0.0575
	(0.0879)
23.eventtime	-0.0861

	(0.194)
24.eventtime	-0.0242
	(0.127)
25.eventtime	-0.00913
	(0.140)
26.eventtime	0.0204
	(0.115)
27.eventtime	0.0685
	(0.0537)
28.eventtime	0.0101
	(0.104)
29o.eventtime	-
Constant	-3.894**
	(1.871)
Observations	158
R-squared	0.997

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This final table presents the results of a hedonic regression model that aims to identify the determinants of congestion pricing revenue per capita across cities and over time. The dependent variable is the log of revenue per capita, and the model includes a range of predictors, including log income, log population, city fixed effects, event time (years since implementation), and year fixed effects (which are included in the regression but not shown, since their coefficients are uninformative in this context). This regression was conducted on an interpolated panel dataset assembled from publicly available data and academic sources, with missing values filled through linear interpolation where necessary to ensure a consistent time series for all cities.

One of the clearest findings from the model is the strong positive association between a city's income level and its congestion pricing revenue. The coefficient on log income is 0.758 and statistically significant at the 1% level, indicating that a 1% increase in income per capita is associated with a 0.758% increase in revenue per capita. This aligns with the economic intuition that wealthier populations may be less price-sensitive and more likely to absorb toll costs, leading to higher total revenue.

In contrast, the coefficient on log population is significantly negative (–1.577), suggesting that larger cities collect less revenue per person. This could reflect a dilution effect where increased population is not proportionally matched by higher toll collection, or it may indicate that larger cities have more diverse commuting patterns and greater availability of substitutes like public transit.

The city fixed effects are also telling. London and New York City exhibit strong positive effects (2.317 and 2.844 respectively), implying much higher baseline revenues than the average city in the dataset, even after accounting for income and population. This likely reflects both cities' large commuting volumes and relatively robust enforcement and pricing systems. In contrast, San Diego and Rome show negative effects (–3.436 and –1.561), suggesting far lower revenue per capita than expected, which is consistent with known issues in enforcement, pricing strategy, and system design in those cities.

The event time coefficients (years since implementation) show how revenue changes dynamically over time. Consistent with Figure 2, the coefficients gradually rise and remain statistically significant through years 15–20, suggesting that revenue per capita tends to grow in the years after implementation. For example, the coefficient for year 15 is 0.231, meaning that, all else equal, revenue per capita is about 23.1% higher 15 years after implementation compared to the baseline year. This supports the earlier observation that pricing systems become more efficient over time as compliance improves, vehicle fleets adjust, and administrative infrastructure becomes more effective.

Together, these results underscore several important insights. First, income plays a major role in shaping revenue outcomes, while population size may dampen per capita returns. Second, cities differ widely in their effectiveness, as reflected in the fixed effects, suggesting that implementation details, enforcement, and political buy-in matter greatly. And finally, the gradual increase in revenue over time reinforces the idea that congestion pricing systems mature administratively and behaviorally, leading to greater yield as they are institutionalized. These findings provide robust, quantitative support for the descriptive trends observed in the time-series figures and offer a statistically grounded view of what drives successful congestion pricing revenue outcomes.

9. Discussions/Limitations

While the comparative analysis in this paper reveals meaningful insights about what makes congestion pricing effective, it is important to acknowledge the methodological and structural limitations inherent in evaluating programs across different urban, political, and economic contexts. Cities are not laboratories, and policy comparisons are rarely clean. Differences in data availability, measurement standards, implementation timelines, and governance models introduce noise and complexity that constrain how definitively we can draw causal conclusions.

First, there are temporal and geographic disparities in implementation. Singapore introduced its pricing model in 1998 and has since evolved into a globally unique digital traffic management platform. In contrast, Milan's Area C began in 2012, and New York's pricing scheme has yet to be implemented. Comparing their effectiveness requires interpreting both immediate impacts and long-term sustainability, something not all programs have had the time to demonstrate. Meanwhile, Rome's ZTL system, although technically "in place," has seen enforcement weaken and compliance drop, raising questions about how to treat programs that exist in name but not in effect. The decision to include both ongoing and discontinued or stalled programs reflects an attempt to examine not just outcomes but implementation dynamics, though it comes at the cost of methodological consistency.

Secondly, the data types and collection standards vary significantly across cities. Some report fine-grained traffic, emissions, and revenue data through public portals (e.g. Stockholm's transport authority), while others rely on academic case studies or government assessments with less transparency (e.g. Rome, Gothenburg). This inconsistency affects the comparability of outcomes like "emissions reduction" or "public support," which may be based on surveys, modeled projections, or indirect indicators. For example, New York's expected emission reductions are based on environmental assessments (MTA, 2023), whereas Milan's were observed and measured in ambient air quality monitors (Moulin & Urbano, 2015). These differences necessitate careful interpretation, especially when integrating them into a common visual scale, as done in Figure 7.1.

Beyond data challenges, there are theoretical limitations to meta-analysis in transport policy. Each city's program is embedded within a unique context: differences in urban density, modal alternatives, road networks, and travel behavior influence the way pricing is received and internalized. The same pricing structure may produce vastly different outcomes depending on whether a city has a legacy metro system (Stockholm), relies on cars for suburban-to-core commuting (San Diego), or is politically fragmented (New York). This means that even strong empirical similarities must be interpreted through a lens of contextual sensitivity, not as universally transferable formulas.

Another underexplored dimension of congestion pricing is the potential for unintended social consequences, particularly around equity and gentrification. While congestion pricing is often justified as an economically efficient and environmentally progressive policy, its implementation may inadvertently displace low-income populations, shift pollution to untolled neighborhoods, or accelerate real estate speculation in areas with improved accessibility. Literature from urban economics, particularly studies on neighborhood amenities and housing markets (e.g. Gyourko et al., 2013), suggests that improved infrastructure can lead to rent increases and demographic shifts, especially when not paired with affordable housing or equitable transit access. In London, some studies have found that the congestion charge zone became more desirable, potentially accelerating housing price increases and displacing less affluent residents. Similarly, equity audits in U.S. cities like San Francisco have raised concerns that tolling without accompanying subsidies or exemptions disproportionately impacts low-income and minority populations who lack flexible commuting schedules or transit options.

These risks underscore the importance of equity-centered program design, including sliding-scale tolls, targeted exemptions, and revenue reinvestment in underserved communities. While cities like Stockholm and Singapore have taken steps in this direction, others, especially in the U.S., have yet to embed these safeguards systematically.

Finally, this paper does not directly model long-term behavioral spillovers or second-order effects, such as changes in employer policies, delivery patterns, or induced transit demand. These dynamics may amplify or diminish the effects of pricing over time, and future research could build on this work by integrating agent-based simulations, household travel surveys, or natural experiments.

10. Data Availability

The comparative analysis in this paper is based primarily on case study data drawn from published reports, government documents, and peer-reviewed academic literature. The cities included in this study, Singapore, London, Milan, Stockholm, Gothenburg, Rome, San Diego, and New York, were selected based on the availability of multi-dimensional outcome data including traffic volume, emissions, program duration, revenue, and public support.

The majority of the data were sourced from foundational case study publications such as:

- Bhatt (2010) for U.S. congestion pricing overviews including San Diego's HOT lanes;
- Leape (2006) on the London congestion charge;
- Theseira (2020) for institutional and technical details of Singapore's ERP system;
- Moulin & Urbano (2015) on Milan's Area C scheme;
- Börjesson et al. (2012) and Eliasson (2008, 2014) on Stockholm's congestion tax;
- Russo & Comi (2021) for Rome and Gothenburg;
- MTA Environmental Assessment (2023) for projected impacts of New York's planned pricing program.

These sources provided consistent metrics where available, including:

- Traffic volume reductions (often expressed as vehicle-kilometers traveled or change in central zone entries),
- CO_2, NO_x, and PM_{10} emissions reductions (via ambient monitoring or simulation),
- Toll revenue collected (where published),
- Modal shift patterns (changes in transit ridership or private car use),
- Policy longevity and political milestones, and
- Public support or opposition (from surveys, referenda, or protest data).

Where quantitative estimates were not directly provided, values were inferred from reported trends or modeled outputs in the cited papers. All raw metrics were systematically entered into a Microsoft Excel master sheet, which served as the central database for outcome comparisons and normalization (for Figure 7.1). Basic calculations, including min-max normalization, average scores, and standardization, were conducted using Excel formulas.

No proprietary datasets or restricted-access archives were used in this study. While R or other programming tools could support regression modeling or clustering analysis for future extensions of this work, the current paper's meta-analysis and synthesis were conducted using manually compiled, citation-backed data points.

All sources are available via public academic databases, and the summary data table compiled for this project can be shared upon request for reproducibility and transparency.

11. Conclusion

This paper has examined the design, implementation, and outcomes of congestion pricing programs across eight global cities, drawing insights from case studies, theoretical models, and comparative data analysis. The results point to several consistent patterns that define successful programs. Cities such as Singapore, Stockholm, and Milan have achieved measurable and sustained improvements in traffic flow, emissions reduction, and public acceptance, outcomes that stem not only from their pricing design, but also from deeper institutional stability, infrastructure integration, and policy legitimacy.

Key findings indicate that no single design feature guarantees success; rather, congestion pricing works best when embedded within a broader framework of adaptive governance and urban transport planning. Dynamic pricing structures, revenue reinvestment into public transport, and strong feedback mechanisms are all associated with higher public support and longer program lifespans. Moreover, the presence of substitution options, such as reliable public transit, greatly enhances elasticity, making behavioral change more responsive to pricing signals. In contrast, cities like Rome and New York illustrate how even well-intentioned policies can falter without adequate enforcement, communication, or political consensus.

From a policy perspective, these findings underscore the importance of tailoring congestion pricing schemes to the local urban structure, travel behavior, and governance environment. A one-size-fits-all approach is unlikely to succeed. Cities with high car dependency and limited transit options must adopt supportive policies, such as improved bus service, park-and-ride facilities, or subsidized alternatives, to ensure that pricing is not regressive. Programs that lack flexibility, transparency, or visible public benefits risk facing backlash, legal delays, or policy reversal.

Future research should continue to explore equity-centered designs, especially in low-income contexts where congestion pricing can disproportionately affect marginalized communities. This includes investigating tiered tolls, targeted exemptions, and direct redistribution of revenue to affected households. In parallel, emerging technologies, such as AI-enabled dynamic pricing, GPS-based vehicle tracking, and real-time emissions monitoring, offer promising tools for more efficient and personalized congestion management. These technologies also present ethical and data governance questions, which future studies must address.

In sum, congestion pricing is not simply a tool to reduce traffic, it is a test of urban governance, policy design, and social contract. When designed holistically and implemented transparently, it can reshape mobility patterns, promote sustainability, and finance better urban futures. But to do so, it must be more than a toll, it must be a strategy.

Citations:

Song, S. (2012). *Congestion Pricing: The Theory*. Transport Studies Unit Discussion Paper. University of Oxford. Retrieved from https://www.tsu.ox.ac.uk/pubs/1023-song.pdf

Theseira, W. (2020). Congestion control in Singapore. In *Urban Transport Policies in Southeast Asia* (pp. 105–122). Southeast Asia Research Centre. Singapore Management University.https://www.itf-oecd.org/sites/default/files/docs/congestion-control-singapore.pdf

Anas, A. (2020). The cost of congestion and the benefits of congestion pricing: A general equilibrium analysis. *Transportation Research Part B: Methodological*. https://doi.org/10.1016/j.trb.2020.06.002

Chu, Z., Cheng, L., & Tang, W. (2014). Equity issues of congestion pricing: Definition, evaluation, and promotion. *Transportation Research Procedia*, 5, 13–22. https://doi.org/10.1016/j.trpro.2014.10.004

Eliasson, J., & Mattsson, L.-G. (2005). Equity effects of congestion pricing: Quantitative methodology and a case study for Stockholm. *Transportation Research Part A: Policy and Practice*, 39(7–9), 602–620. https://doi.org/10.1016/j.tra.2004.11.007

Jing, P., Seshadri, R., Sakai, T., Shamshiripour, A., Alho, A. R., Lentzakis, A., & Ben-Akiva, M. E. (2023). Evaluating congestion pricing schemes using agent-based passenger and freight microsimulation. *arXiv preprint*. https://arxiv.org/abs/2305.07318

Maheshwari, C., Kulkarni, K., Pai, D., Yang, J., Wu, M., & Sastry, S. (2024). Congestion pricing for efficiency and equity: Theory and applications to the San Francisco Bay Area. *arXiv preprint*. https://arxiv.org/abs/2401.08956

Svensson, S., & Rehme, A. (2021). Are prices in charge of congestion? An empirical study of increased congestion charge and traffic in Stockholm. *Master's Thesis, Lund University*.

https://lup.lub.lu.se/student-papers/search/publication/9053959

Turner, M. A., & Kreindler, G. E. (2022). Peak-hour road congestion pricing: Experimental evidence and welfare gains. *Unpublished manuscript*. https://matthewturner.org/ec2410/readings/Kreindler_unp_2022.pdf

"ANPR Is a Very Useful Tool in Traffic Management." *TrafficInfraTech Magazine*, 30 Mar. 2023, https://trafficinfratech.com/anpr-is-avery-useful-tool-in-traffic-management/.

Electronic Road Pricing (ERP). Ministry of Transport Singapore, https://www.mot.gov.sg/what-we-do/motoring-road-network-and-infrastructure/Electronic-Road-Pricing. Accessed 25 July 2025.

Brownstone D, Ghosh A, Kazimi C, Van Amelsfort D. Drivers' willingness-to-pay to reduce travel time: Evidence from the san diego I-15 congestion pricing project [Internet]. eScholarship, University of California. 2013 [cited 2025 Aug 4]. Available from: https://escholarship.org/uc/item/8q7331mz

(Brownstone et al., 2013)

Russo, Antonio, et al. *Welfare Losses of Road Congestion: Evidence from Rome*. VU Research Portal, Vrije Universiteit Amsterdam, 2021. https://research.vu.nl/en/publications/welfare-losses-of-road-congestion-evidence-from-rome. Accessed 4 Aug. 2025.

In-text citation:
(Russo et al., 2021)

Swedish Transport Administration. *Swedish Congestion Charges*. International Transport Forum, OECD. https://www.itf-oecd.org/sites/default/files/docs/swedish-congestion-charges.pdf. Accessed 4 Aug. 2025.
(Swedish Transport Administration, n.d.)

Small, Kenneth A., and Erik T. Verhoef. "The Economics of Urban Transportation." *Journal of*

Economic Perspectives, vol. 20, no. 4, 2006, pp. 157–176. https://www.jstor.org/stable/29775281?seq=1&cid=pdf-. Accessed 4 Aug. 2025.
(Small and Verhoef, 2006)

Victoria Transport Policy Institute. *Congestion Pricing: An Overview of Experience and Impacts*. Lincoln Institute of Land Policy, 2010. https://www.lincolninst.edu/app/uploads/legacy-files/pubfiles/2041_1363_LP2010-ch10-Congestion-Pricing-An-Overview-of-Experience-and-Impacts_0.pdf. Accessed 4 Aug. 2025.
(Victoria Transport Policy Institute, 2010)

Vasserman, Shoshana. *NYC Congestion Pricing 2025*. 2025. https://shoshanavasserman.com/files/2025/03/NYC_Congestion_Pricing_2025.pdf. Accessed 4 Aug. 2025.
(Vasserman, 2025)

Börjesson, Maria, and Jonas Eliasson. *The Stockholm Congestion Charges: 10 Years On*. Centre for Transport Studies, Stockholm, 2014. https://transportportal.se/swopec/cts2014-7.pdf. Accessed 4 Aug. 2025.
(Börjesson and Eliasson, 2014)

Sun, Xiaoqian, et al. "A Review of Urban Congestion Pricing." *Journal of Advanced Transportation*, vol. 2020, Article ID 4242964. https://onlinelibrary.wiley.com/doi/10.1155/2020/4242964. Accessed 4 Aug. 2025.
(Sun et al., 2020)

Vickrey, William S. "Congestion Theory and Transport Investment." *The American Economic Review*, vol. 59, no. 2, 1969, pp. 251–260. https://www.jstor.org/stable/1823886. Accessed 4 Aug. 2025.
(Vickrey, 1969)

Phang, Sock-Yong, and Rex S. Toh. *Congestion Control in Singapore: An Evaluation of Recent Measures*. International Transport Forum, OECD. https://www.itf-oecd.org/sites/default/files/docs/congestion-control-singapore.pdf. Accessed 4 Aug. 2025.

(Phang and Toh, n.d.)

Santos, Georgina. “The Economics of Transport: A Theoretical and Empirical Perspective.” *Research in Transportation Economics*, 2025. https://pdf.sciencedirectassets.com/271794/1-s2.0-S0967070X25X00044/1-s2.0-S0967070X25000630/main.pdf?... Accessed 4 Aug. 2025.
(Santos, 2025)

Lindsey, Robin. “Prospects for Urban Road Pricing.” *Journal of Economic Perspectives*, vol. 20, no. 4, 2006, pp. 157–176. https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.20.4.157. Accessed 4 Aug. 2025.
London - (Lindsey, 2006)

Song, Yan. *Congestion Pricing and Its Urban Distributional Impacts*. Lincoln Institute of Land Policy, 2013. https://www.lincolninst.edu/app/uploads/legacy-files/pubfiles/2272_1611_Song_WP13SS2.pdf. Accessed 4 Aug. 2025.
Equation and graph - (Song, 2013)

Hoover, Kevin D. “Econometrics as Observation: The Lucas Critique and the Nature of Econometric Inference.” *The Philosophy of Economics*, Cambridge University Press. https://www.cambridge.org/core/books/abs/philosophy-of-economics/econometrics-as-observation-the-lucas-critique-and-the-nature-of-econometric-inference/71CFB955137EFF1C4A616830DF3409EF. Accessed 4 Aug. 2025.

Lucas critique - (Hoover, n.d.)