

Distributional welfare impacts and compensatory transit strategies under NYC congestion pricing

Xiyuan Ren^a, Zhenglei Ji^a, Joseph Y. J. Chow^{a,*}

a. C2SMARTER University Transportation Center, New York University Tandon School of Engineering,
Brooklyn, USA

* Corresponding author: joseph.chow@nyu.edu

ABSTRACT

Early evaluations of NYC's congestion pricing program indicate overall improvements in vehicle speed and transit ridership. However, its distributional impacts remain understudied, as does the design of compensatory transit strategies to mitigate potential welfare losses. This study identifies population segments and regions most affected by congestion pricing, and evaluates how welfare losses can be compensated through transit improvements. We estimate joint mode and destination models using aggregated synthetic trips in New York and New Jersey and calibrate toll-related parameters with traffic counts reported by the MTA. Welfare impacts of congestion tolls are measured as changes in consumer surplus (CS) before and after program implementation. Compensatory transit strategies are evaluated by quantifying the reductions in transit wait time and fare discounts required to offset the welfare losses. The results show that the program leads to an accessibility-related welfare loss of approximately \$240 million per year, which is considerably lower than the gains from toll revenues: the gross revenue estimated by our models (\$1.077 billion per year) and the net revenue projected by the MTA (\$450 million per year). However, these benefits gains conceal significant disparities. Welfare losses are concentrated in Upper Manhattan, Brooklyn, and Hudson County, NJ, particularly among travelers less able to shift to transit or alternative destinations. For NYC residents, compensating aggregate welfare loss requires a 0.48-minute reduction in transit wait time or a \$135.59 million annual fare subsidy. Ensuring accessibility gains for all populations and counties (Pareto improving) requires a 1–2 minute reduction in wait time combined with an annual subsidy of about \$100–300 million. For New Jersey residents, achieving aggregate welfare gains primarily through fare discounts (requiring \$108.53 million per year) is more feasible and efficient; however, uniform discounts should be replaced by targeted mechanisms such as origin-based fare reductions or commuter pass bundles.

Keywords: congestion pricing, mode and destination choice, public transit, welfare analysis, synthetic data, New York City

1. Introduction

On January 5, 2025, New York City Metropolitan Transportation Authority (MTA) launched the Central Business District Tolling Program (CBDTP), which is the first cordon-based congestion pricing scheme in the United States (Cook et al., 2025; National Bureau of Economic Research, 2025; Nogueira, 2025). The program charges vehicles entering the Congestion Relief Zone (CRZ), with tolls varying by time of day, vehicle type, and payment method (MTA, 2025a). Toll rates start at \$9 for passenger cars and small commercial vehicles with E-ZPass, \$4.50 for motorcycles, and \$14.40–\$21.60 for trucks and buses, with a 75% overnight discount and surcharges of up to 50% for non-E-ZPass users billed by mail. Taxis and ride-hailing vehicles pay per-trip fees of \$0.75 and \$1.50, respectively. The program exempts emergency vehicles, provides partial credits for trips entering via tolled bridges and tunnels, and offers a 50 discount to low-income drivers after their first ten trips each month. Although its implementation follows decades of political debate and failed attempts dating back to proposals in the 1970s and Mayor Bloomberg’s PlaNYC initiative in 2007 (Bloomberg, 2007; Schaller, 2010; Schwartz et al., 2008), the program now serves as a critical case study for assessing the economic, behavioral, and environmental impacts of congestion pricing in the U.S. context.

Congestion pricing is designed to internalize the external cost of driving during peak periods, particularly the additional travel times imposed on others (De Palma & Lindsey, 2011; Downs, 2005). Landmark programs in London (Santos & Bhakar, 2006), Stockholm (Eliasson, 2009), and Singapore (Kockelman & Kalmanje, 2005) have demonstrated that charging drivers for access to constrained urban road space reduces traffic volumes, improves air quality, and generates stable funding for public transit. Short-run evaluations of NYC’s CBDTP point to similar benefits: average traffic speeds in the CBD have increased by 15%, alongside a 2–3% reduction in CO₂ emission rates (Cook et al., 2025). In addition, the program is projected to generate \$500 million in net revenue during its first year (MTA, 2025c). However, these gains have not quelled public opposition: at least ten lawsuits have been filed against the MTA and state officials by business coalitions, elected officials from New Jersey, and other stakeholders (Harris & Ley, 2024). Resistance is reinforced by longstanding narratives of necessity and fairness, as drivers question the feasibility of shifting away from car travel while lower-income groups and New Jersey commuters underscore the disproportionate financial burdens they face (Baghestani et al., 2022; Chen, 2025; Schaller, 2010). These debates highlight the tension between the program’s demonstrated benefits and the persistent anxieties over its distributional impacts.

Reinvesting toll revenues to improve public transit services is crucial to realizing both efficiency and equity goals under congestion pricing (Basso & Jara-Díaz, 2012; Chen & Nozick, 2016; Marazi et al., 2024). Effective compensatory strategies can not only ensure equitable opportunities for disadvantaged groups but also encourage a broader mode shift, yielding environmental benefits and long-term public trust (Isaksen & Johansen, 2025). However, this requires rigorous quantitative evaluation of how congestion pricing influences travel behavior and how revenues can be redistributed to offset increased monetary costs. Although welfare analysis based on logsum utilities from discrete choice models (DCMs) has been widely applied in transportation research (He et al., 2021; Ji, 2025; W. Li et al., 2021), the absence of consistent trip data for New York and New Jersey, as well as the lack of models that capture shifts in traveler preferences before and after the

program implementation, remain key barriers to assessing compensatory transit policies under NYC’s congestion pricing.

This study aims to measure the distributional impacts of NYC congestion toll on trip welfare and to evaluate the effectiveness of compensatory strategies such as increasing transit service frequency and providing fare discounts. We propose to estimate DCMs using synthetic and observed data collected before and after the implementation of the program.

For the pre-implementation period, we use synthetic population data for New York and New Jersey provided by Replica Inc., comprising over 60 million trips on a typical weekday in the second quarter of 2023. To consider behavioral heterogeneity, these trips are split into 16 segments defined by four population groups (NotLowIncome, LowIncome, Senior, and Student), two time periods (Peak and Overnight), and two trip purposes (Commute and Non-commute). Truck trips are excluded from the analysis, as truck drivers are less likely to benefit from improvements in public transit. For each trip segment, we estimate a market-level joint mode and destination choice model, in which trips originating from the same county are treated as a market, and each mode–destination pair is treated as an alternative. To better capture the substitution patterns across both modes and destinations, we employ the inverse product differentiation logit (IPDL) model (Fosgerau et al., 2024) in place of the traditional nested logit (NL) model (McFadden, 1977).

For the post-implementation period, we compile observed traffic counts and transit ridership published by MTA (MTA, 2025b). Traffic counts on major roads, bridges, and tunnels are used to calibrate toll-related taste parameters, while ridership data is employed to validate the post implementation model. We then apply the models to quantify changes in trip welfare attributable to the congestion toll. Trip welfare is measured using consumer surplus derived from the choice models, calculated as the logsum of utilities (Small & Rosen, 1981), and converted into monetary terms by dividing by the estimated travel cost parameter (Vij & Walker, 2016).

Finally, while congestion pricing should net welfare gains when the revenues are redistributed back to users, the redistribution mechanisms are not clear and outcomes can take many years to observe. Instead, we pose the question: where *would revenues need to be distributed* in the transit system to make improvements to compensate for the welfare losses associated with the toll. Specifically, we calculate the increase in transit service frequency and the level of fare discounts necessary to compensate for the reduction in consumer surplus. Moreover, we consider two goals: Kaldor–Hicks efficiency (Kaldor, 1939) and Pareto improving (Varian, 1992). The former seeks to compensate the aggregate welfare loss, while the latter requires that no traveler group is made worse off. To facilitate future research, we upload the processed datasets and estimated model parameters to a [GitHub repository](#).

The remainder of the paper is organized as follows. [Section 2](#) reviews implementations of congestion pricing and welfare analysis based on DCMs. [Section 3](#) outlines the methodology, including data processing, model estimation, and welfare measurement. [Section 4](#) presents the results of joint mode and destination choice models, impact analysis, and strategy evaluation. [Section 5](#) discusses the policy implementations of our findings. [Section 6](#) concludes with key takeaways and directions for future research.

2. Literature review

2.1 Congestion pricing: implementations, effectiveness, and concerns

The concept of congestion pricing stems from the idea of internalizing the external cost of peak-hour driving, particularly the additional delays imposed on other travelers (De Palma & Lindsey, 2011; Downs, 2005). Congestion pricing as a public policy was first introduced in Singapore in 1975 as the Area Licensing Scheme (ALS). The policy then evolved into the Electronic Road Pricing (ERP) system, which included dynamic, time-of-day toll charging (Olszewski & Xie, 2005; Phang & Toh, 2004), which reduced daily emissions by 80 metric tons of CO₂ (National Environment Agency, 2010). In 2003, London launched its “Congestion Charge” policy, which led to significant reductions in traffic congestion and improvements in bus speeds (Leape, 2006; Santos & Bhakar, 2006). The congestion charge also decreased London’s PM₁₀ and NO_x by about 12% and reduced CO₂ in the charged area by up to 20% in the first year (Beevers & Carslaw, 2005; Tonne et al., 2008). Stockholm made the congestion pricing permanent after the success of a seven-month trial of congestion taxes (Börjesson & Kristoffersson, 2015; Eliasson, 2009). The policy reduced NO_x by about 8% and PM₁₀ by around 13%, while children’s asthma attacks dropped significantly (Simeonova et al., 2021). Milan and Gothenburg also implemented congestion pricing policies, with the former evolving from an emissions-focused “Ecopass” to a comprehensive Area C charge (Beria, 2016), and the latter adopting a cordon tax to both manage congestion and finance infrastructure investments (West & Börjesson, 2020).

In addition to its immediate effect of lowering traffic volumes, congestion pricing encourages lasting shifts in traveler behavior. By imposing a toll on driving into the congestion zone during peak hours, congestion pricing may prompt commuters to change their departure times, adjust routes, or shift to public transit. For instance, Karlström et al. (2009) stated that cordon-based pricing in Stockholm shows that drivers are more likely to adjust either their departure time or mode when congestion tolls are applied, demonstrating the policy’s effectiveness in reallocating demand. Similarly, a study in Singapore revealed that even a small toll can significantly decrease peak-time demand by shifting traffic to less congested hours (Wongpiromsarn et al., 2012). Beyond temporal behavioral changes, congestion pricing can also lead to modal shifts for travelers. A stated-preference study in Beijing found that about 23% of habitual car users were willing to switch to public transit, biking, or walking under peak-hour tolls, indicating the potential for pricing policies to encourage more sustainable travel modes (Li et al., 2019).

New York City implemented as the first cordon-based congestion pricing program in the U.S. in January 2025, targeting one of the most congested areas in the world. The policy charges most vehicles about \$9 during peak hours to enter Manhattan’s Congestion Relief Zone (CRZ), with the revenue allocated to fund MTA infrastructure projects (MTA, 2025a). In its first six months, the program reduced vehicle entries into the CRZ by roughly 11%, while Manhattan traffic delays fell by 25% and average travel speeds improved by 5 to 15% during peak hours (Regional Planning Association, 2025). Compared to the prior year, bus performance improved significantly, with some express routes operating up to 20% faster, while overall transit ridership also increased: subway ridership rose by 7%, bus ridership by 12%, and commuter rail by 8% (Cook et al., 2025). The program is also projected to generate \$500 million in net revenue during its first year (MTA, 2025c).

Despite the promising outcomes, the program is not implemented without controversy and opposition. [Siena College Research Institute \(2024\)](#) showed that statewide opposition rates were as high as 63% before its implementation. At least ten lawsuits have been filed against the program by business coalitions, elected officials from New Jersey, and other stakeholders ([Harris & Ley, 2024](#)). A central criticism is its fairness as it imposes a flat fee for road use and its impacts on people from different income groups. Although low-income drivers can get discounts after 10 trips to the CRZ each month, higher-income motorists tend to benefit more from time savings ([Eliasson, 2016](#)). In addition, truck drivers and freight companies have voiced strong concerns over the toll's financial burden, arguing that it threatens freight-dependent businesses ([American Trucking Associations, 2025](#)). Small business owners have also reported financial strains, both from higher freight costs and from declines in customer foot traffic ([Shalma, 2025](#)).

With all of these challenges, the success of New York's congestion pricing depends on overcoming the dual challenge of demonstrating efficiency gains while ensuring that travelers perceive both the necessity and feasibility of shifting away from car use. This requires deeper insight into how congestion tolls reshape travel behavior and influence overall trip welfare.

2.2 Welfare analysis based on DCMs

DCMs have been widely applied in transportation research to forecast travel demand by assuming travelers make decisions by maximizing the overall utility they can expect to gain ([Bowman & Ben-Akiva, 2001](#)). These models enable researchers to examine how attributes such as travel time, monetary cost, and convenience of transfer influence the probability of selecting specific modes and destinations ([Hensher & Ho, 2016](#); [Ren & Chow, 2022](#)). Welfare impacts of transportation policies can be evaluated using consumer surplus (CS) derived from discrete choice models, where the logsum of utilities provides a measure of aggregate accessibility ([Small & Rosen, 1981](#); [Vij & Walker, 2016](#)). For instance, [Standen et al. \(2019\)](#) explored the use of the logsum measure of CS for valuing the user benefits of new separated cycleways in Sydney. [Ren et al. \(2024\)](#) proposed a choice-based decision support tool for determining optimal service regions for on-demand mobility that balances revenue generation and welfare gains. CS in joint mode and destination models is often linked to "accessibility", referring to the "ease" with which desired destinations may be reached ([Niemeier, 1997](#)). [Bills et al. \(2022\)](#) estimated a destination-mode (destination in the upper branch and mode in the lower branch) choice model to calculate logsum utilities and explored the equity impacts of a microtransit service in Metropolitan Detroit.

A growing body of literature have estimated DCMs and conducted welfare analysis in the context of congestion pricing. [He et al. \(2021\)](#) conducted a validated multi-agent simulation for NYC and demonstrated that congestion pricing produces heterogeneous demographic impacts, differing substantially across time periods and neighborhoods, highlighting the need for differentiated pricing strategies that account for varied commuter patterns and spatial traffic dynamics. [Li et al. \(2021\)](#) examined solutions for maximizing travelers' welfare by varying toll levels and locations across road network in Austin, Texas. They suggested that congestion tolls can do more harm than good unless travelers shift out of the peak periods or revenues are returned to travelers as credits. [Tarduno \(2022\)](#) proposed a departure time choice model to evaluate second-best congestion pricing schemes in proposed cordon zones across several U.S. cities, showing that spatial leakage and imperfect

pricing prevent the realization of welfare benefits. [Ji \(2025\)](#) estimated a mode choice model using NYC Citywide Mobility Survey data to examine the distributional impacts of NYC's CBDTP, highlighting that raising tolls without reinvestment to public transit delivers negligible mode shift. Together, these studies underscore the importance of systematically examining distributional impacts and evaluating compensatory strategies to ensure that congestion pricing policies balance efficiency gains with equity considerations.

Despite numerous innovative ideas and valuable empirical findings, several critical research gaps remain in applying DCMs to evaluate NYC's congestion pricing program. First, most existing studies rely on travel survey data from a single city, overlooking the broader regional impacts of New York City's congestion toll, which extend to upstate New York and New Jersey ([MTA, 2025b](#)). Second, relevant choice models either focus only on mode choice ([Ji, 2025](#)) or employ a nested structure by assuming unidimensional correlations ([Bills et al., 2022](#)), which may overlook the multidimensional substitution patterns across modes and destinations. Third, these studies typically treat travelers' taste parameters as stable across pre- and post-implementation conditions, neglecting potential preference shifts induced by the program. Last but not least, limited attention has been given to policy scenario analysis, particularly regarding how congestion pricing revenues might be redistributed to travelers through compensatory strategies that offset the additional monetary burden, as well as the trade-offs under aggregate efficiency and distributional equity goals.

2.3 Our contributions

This study contributes to the literature on congestion pricing by addressing several of the key gaps identified above. First, we use synthetic trip data to incorporate a broader spatial scope including New York City, upstate New York, and selected counties in New Jersey, thereby capturing the major regional spillover effects of the program. Second, unlike prior studies that consider joint mode and destination choice as a unidimensional nested structure (either mode–destination or destination–mode), we estimate IPDL models to consider correlations across both mode and destination groups. This is particularly important for congestion pricing studies, as tolls can simultaneously shift mode shares and destination distributions, reflecting substitution patterns across multiple dimensions. Third, we relax the common assumption of stable taste parameters by explicitly estimating traveler preferences before and after the implementation of congestion pricing, leveraging observed traffic counts to calibrate preference shifts induced by the toll.

In addition, this study advances policy evaluation by quantifying the role of compensatory transit strategies. We assess distributional impacts on accessibility through logsum utilities from the choice models disaggregated by demographic and trip segments. Building on this approach, we examine how increased transit service frequency and fare discounts can offset welfare losses attributable to congestion tolls. This approach provides a quantitative basis for designing revenue allocation policies that promote accessibility and equity, encourage sustainable mode shifts, and strengthen public trust. Collectively, these contributions position our study among the first to jointly assess the accessibility impacts and compensatory transit strategies of NYC's congestion pricing program within a utility-based framework.

3. Data and methodology

3.1 Study area and data collection

While the toll cordon applies only to Manhattan south of 60th Street (also known as Congestion Relief Zone, CRZ), its impacts extend well beyond the city's boundaries (MTA, 2025b). Our study area encompasses New York State and selected counties in New Jersey (Fig. 1). Specifically, we select the five New Jersey counties with largest number of commuting trips to Manhattan, as recorded in the Census Transportation Planning Products (CTPP) data (American Association of State Highway and Transportation Officials, 2025). This allows us to account for cross-state commuting and potential spillover effects that may reshape the distributional impacts of congestion pricing.

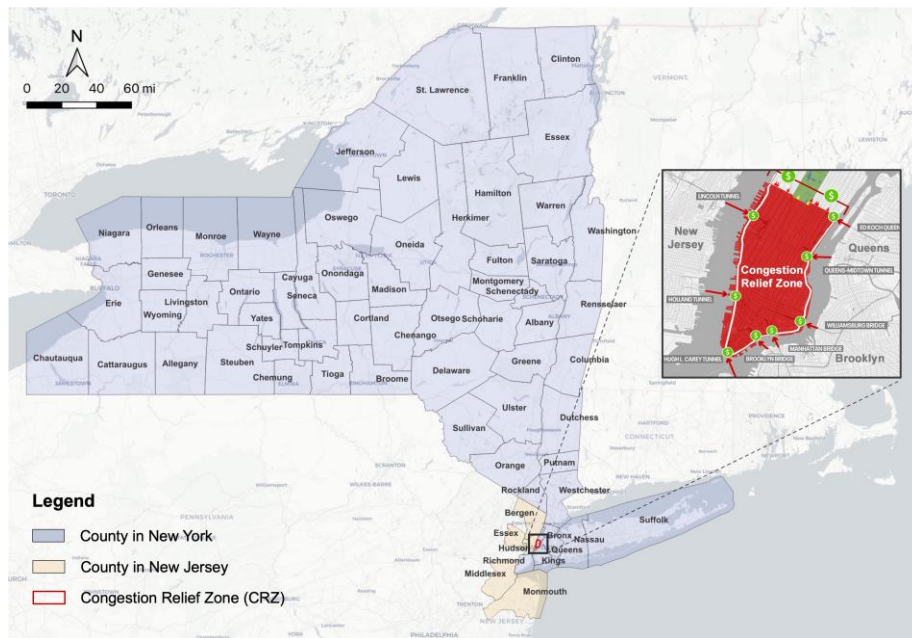


Fig. 1. Study area covering New York and New Jersey. The inset map is adapted from a figure in Meier (2024).

Consistent travel data remains scarce, particularly across different states and between urban and rural communities (Bachir et al., 2019; Parr et al., 2020). The lack of representative data can lead to biased identification of behavioral patterns. Synthetic population data help address this limitation by providing harmonized, large-scale trip records that capture travel behavior across regions in a consistent framework (Hörl & Balac, 2021).

We use synthetic population data provided by Replica Inc., containing synthetic residents and their trips in New York and New Jersey on a typical weekday in the second quarter of 2023. The dataset was generated through a combination of mobile phone data, census data, economic activity data, and built environment, representing large-scale travel behavior before the congestion pricing program. Each synthetic individual is assigned demographic attributes such as gender, age, income, and education level. Each synthetic trip includes information on origin, destination, departure and arrival time, and travel mode (driving, public transit, on-demand service, biking, walking, or carpool). According to Replica's data quality report (Replica, 2024), the largest error of demographic attributes for

a single census tract unit is within 5% compared to census data, and the largest error of commute mode share for a single census tract unit is within 10% compared to CTPP data. The number of trips to the CRZ, aggregated by origin, destination, and travel mode, is provided in [Appendix Table A1](#).

We further specify 16 trip segments based on four population groups, two trip purposes, and two time periods. For population groups, we consider low-income, not-low-income, senior, and student population. The senior population is defined as individuals aged 65 and older. The student population is defined as individuals still in schools, colleges, and universities. The low-income population is defined as individuals from households with an annual income below 200% of the 2023 Federal Poverty Guidelines¹. All other individuals are classified as the not-low-income population. Trip purposes are divided into commute and non-commute trips, where commute trips are identified as home-to-work and work-to-home itineraries, and all other trips are categorized as non-commute. Following the CBDTP pricing periods, we distinguish between peak and overnight trips, defining peak hours as 5 a.m.–9 p.m., and treating the remaining period as overnight. [Table 1](#) lists the number of trips and mode share by segment. Trips made by students to school locations are classified as commute trips. Since seniors may hold part-time jobs, they are also assigned commute trips in Replica’s data.

Table 1
Number of synthetic trips and driving/transit mode share by segment

Population group	Trip purpose	Time period	Num. trips (trips/day)	Driving mode share (%)	Transit mode share (%)
LowIncome	Commute	Peak	3,045,699	36.38%	29.33%
		Overnight	302,383	34.58%	30.34%
	Non-commute	Peak	4,531,991	43.64%	9.08%
		Overnight	859,007	42.11%	11.54%
NotLowIncome	Commute	Peak	14,050,299	47.58%	21.60%
		Overnight	1,185,598	45.82%	23.35%
	Non-commute	Peak	16,088,600	53.14%	6.13%
		Overnight	3,007,391	52.04%	7.55%
Senior	Commute	Peak	2,143,192	51.15%	15.85%
		Overnight	277,909	49.07%	15.24%
	Non-commute	Peak	3,556,543	57.59%	3.20%
		Overnight	727,529	56.37%	3.59%
Student	Commute	Peak	7,511,892	19.87%	6.20%
		Overnight	212,439	42.45%	18.38%
	Non-commute	Peak	2,920,304	50.52%	5.40%
		Overnight	532,461	49.35%	6.73%

Since Replica’s synthetic trip data do not distinguish transit travel time into access, egress, in-vehicle, and wait times, we use OpenTripPlanner (OTP) to obtain this information. OTP is an open-source tool that uses imported Open Street Map (OSM) data

¹ <https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines>

for routing on the street network and supports multi-agency public transport routing through imported General Transit Feed Specification (GTFS) data (Young, 2018). We first download OSM data for our study area. Then for the pre- and post-implementation periods, we obtain GTFS data for June 2023 and June 2025 from the Mobility Database (MobilityData, 2025). The OSM and GTFS datasets are imported into OTP to calculate transit service performance between any origins and destinations given a departure time. The average transit times and number of transfers aggregated by trips starting from various regions are provided in Appendix Table A2.

Moreover, we compile observed traffic counts, transit ridership, and vehicle entries published by MTA (MTA, 2025b). To calibrate the post-implementation model, we selected four tunnels with data available for both 2023 and 2025, including Queens Midtown Tunnel, Hugh Carey Tunnel, Lincoln Tunnel, and Holland Tunnel. To validate our model predictions, we draw on bus and subway passenger counts from April to June in both 2023 and 2025. The number of entries by vehicle class from January 5 to September 20 is used to estimate the gross toll revenue, which is then compared with the toll revenue predicted by our model. Further details are provided in Section 4.1.2.

3.2 Model specification

This section introduces joint mode and destination choice models for both pre- and post-implementation periods. Notations used in our models are summarized in Appendix Table A3.

3.2.1 Joint mode and destination model with market-level data

Given the large trip volume throughout New York and New Jersey, we aggregate trips in each segment by origin and destination county and compute the average travel time, average monetary cost, and total number of trips by each mode. The Manhattan county is divided into CRZ and non-CRZ areas to account for the congestion toll. This enables us to estimate joint mode and destination models at the market level, in which trips within the same segment and originating from the same county are treated as a market, and each mode–destination pair is treated as an alternative.

There are two reasons for aggregating the trips to the county level. First, synthetic trips are difficult to validate at the individual level but become more reliable when aggregated into larger spatial units. Second, the county is a suitable geographic unit for congestion pricing studies, as it aligns with common administrative boundaries where policies are implemented and equity concerns are raised.

We start from the utility function specified in market-level DCMs proposed by Berry et al. (1995). The utility of individual $n \in N$ in market $t \in T$ choosing alternative (or product) $j \in J$ is defined in Eq. (1).

$$U_{njt} = \bar{\delta}_{jt} + \mu_{njt} + \varepsilon_{njt}, \quad \forall n \in N, \forall j \in J, \forall t \in T \quad (1)$$

where $\bar{\delta}_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$ is the generic utility of alternative j in market t ; x_{jt} denotes a vector of alternative attributes besides price; p_{jt} denotes the price (monetary cost); ξ_{jt} denotes alternative specific constants; β and α are parameters to be estimated. μ_{njt} denotes the individual-specific unobserved utility and ε_{njt} is an i.i.d. Gumbel variate serving as the

random disturbance. Since trips were aggregated into counties and distinguished into 16 segments defined in Section 3.1, we assume individuals within the same market are homogeneous, which implies $\mu_{njt} = 0$ and allows us to drop the index n from U_{njt} and ε_{njt} . Using a compact form, the utility function can be rewritten as Eq. (2).

$$U_{jt} = V_{jt} + \varepsilon_{jt} = \theta^T X_{jt} + \varepsilon_{jt}, \quad \forall j \in J, \forall t \in T \quad (2)$$

where V_{jt} is a function of systematic utility; $X_{jt} = \{x_{jt}, p_{jt}, 1\}$ is a vector of all alternative attributes, and $\theta = \{\alpha, -\beta, \xi_{jt}\}$ is a vector of taste parameters. Accordingly, the market share of alternative j in market t is predicted as Eq. (3), and the logarithm form of a ratio between two market shares is presented as Eq. (4).

$$s_{jt} = \frac{e^{V_{jt}}}{\sum_{q \in J} e^{V_{qt}}}, \quad \forall j \in J, \forall t \in T \quad (3)$$

$$\ln\left(\frac{s_{jt}}{s_{qt}}\right) = \ln\left(\frac{e^{V_{jt}}}{e^{V_{qt}}}\right) = V_{jt} - V_{qt} = V_{jt}, \quad \forall j \in J, j \neq q, \forall t \in T \quad (4)$$

where s_{jt} denotes the market share or choice probability of alternative j in market t . Market-level models typically include an outside alternative ($j = 0$), representing the option of “not buying anything” (Berry, 1994). The systematic utility of the outside alternative is set to zero ($V_{0t} = 0$). The logarithm form of the ratio between the market share of alternative j and that of the outside alternative is called the inverse market share of j (Berry, 1994).

We focus on the joint choice of travel mode and destination, rather than mode choice alone, to better capture behavioral responses to cordon-based congestion pricing. The study area covers 63 counties in New York and 5 counties in New Jersey. Six travel modes are considered, including driving, public transit, for-hire vehicle (FHV), biking, walking, and carpool (trips made by several passengers in an auto vehicle). Travel to destinations outside New York State is defined as the outside alternative. Accordingly, there are $(63 + 5) \times 16 = 1,088$ markets and $63 \times 6 = 378$ alternatives in the model. Since each alternative is a combination of travel mode and destination, its systematic utility V_{jt} can be replaced with $V_{md,t}$ in which m is the index of mode and d is the index of destination. We include alternative attributes such as travel time, monetary cost, mode-specific constants, county-specific constants, and region-related interaction terms. The systematic utilities of the six modes given any $t \in T, d \in D$ are specified in Eqs. (5) – (10).

$$V_{driving,d,t} = (\theta_{autoTT} + \theta_{autoTT}^{NYC} IsNYC_t) TT_{driving,d,t} + (\theta_{cost} + \theta_{cost}^{NYC} IsNYC_t) CO_{driving,d,t} + \theta_{asc}^{driving} + \theta_{asc}^d \quad (5)$$

$$V_{transit,d,t} = (\theta_{AT} + \theta_{AT}^{NYC} IsNYC_t) AT_{transit,d,t} + (\theta_{ET} + \theta_{ET}^{NYC} IsNYC_t) ET_{transit,d,t} + (\theta_{WT} + \theta_{WT}^{NYC} IsNYC_t) WT_{transit,d,t} + (\theta_{IVT} + \theta_{IVT}^{NYC} IsNYC_t) IVT_{transit,d,t} + \theta_{trans} Trans_{transit,d,t} + (\theta_{cost} + \theta_{cost}^{NYC} IsNYC_t) CO_{transit,d,t} + \theta_{asc}^{transit} + \theta_{asc}^d \quad (6)$$

$$V_{fhv,d,t} = (\theta_{autoTT} + \theta_{autoTT}^{NYC} IsNYC_t) TT_{fhv,d,t} + (\theta_{cost} + \theta_{cost}^{NYC} IsNYC_t) CO_{fhv,d,t} + \theta_{asc}^{fhv} + \theta_{asc}^d \quad (7)$$

$$V_{biking,d,t} = (\theta_{nonautoTT} + \theta_{nonautoTT}^{NYC} IsNYC_t) TT_{biking,d,t} + \theta_{asc}^{biking} + \theta_{asc}^d \quad (8)$$

$$V_{walking,d,t} = (\theta_{nonautoTT} + \theta_{nonautoTT}^{NYC} IsNYC_t) TT_{walking,d,t} + \theta_{asc}^{walking} + \theta_{asc}^d \quad (9)$$

$$V_{carpool,d,t} = (\theta_{autoTT} + \theta_{autoTT}^{NYC} IsNYC_t) TT_{carpool,t} + \theta_{asc}^d \quad (10)$$

where $TT_{m,d,t}$ and $CO_{m,d,t}$ are the average travel time (minute) and cost (dollar) for market t traveling to destination d by mode m ; $AT_{transit,d,t}$, $ET_{transit,d,t}$, $WT_{transit,d,t}$, $IVT_{transit,d,t}$, $Trans_{transit,d,t}$ are the access time, egress time, wait time, in-vehicle time, and number of transfers for taking public transit. $IsNYC_t$ is an interaction variable that equals 1 if market t is located within NYC and 0 otherwise. This interaction variable allows us to distinguish travelers' sensitivities to travel time and cost within NYC from that in other regions. We use destination-specific constants instead of attributes such as facility proximity or employment density since these attributes are not directly affected by congestion pricing or its compensatory transit strategies. All terms associated with θ represent taste parameters to be estimated, including sensitivities to travel time and cost as well as mode- and destination-specific constants.

3.2.2 Pre-implementation model estimation

The assumed distributions of random disturbances ($\varepsilon_{md,t}$) determine how the choice model is estimated (McFadden & Train, 2000). In the joint mode mode and destination choice model, the random disturbances are not assumed to be i.i.d.; rather, they exhibit correlation both across modes to the same destination and across destinations when choosing the same mode. Previous studies typically address one if these correlation structures by estimating nested logit (NL) models, capturing dependence either across destinations with upper-branch destination choice and lower-branch mode nests, or the reverse (Bills et al., 2022; Newman & Bernardin, 2010). However, this is at the risk of misunderstanding substitution and complementarity patterns (Huo et al., 2024).

Fosgerau et al. (2024) proposed inverse product differentiation logit (IPDL) to address the limitation of hierarchical structures and provide faster estimation. IPDL allows alternatives to be nested across multiple hierarchical dimensions $h \in H$, with each alternative belonging to exactly one nest within each dimension. Huo et al., (2024) proved that IPDL is a general form of multinomial logit (MNL) and nested logit (NL). MNL is obtained when there is no hierarchical structure ($H = 0$). NL is obtained when there is only one hierarchical structure ($H = 1$). Fig. 2 illustrates a simplified case showing the difference between a NL with a destination–mode structure and the IPDL. In the NL model, a reduction in driving within the CRZ results in a general increase for all modes in the non-CRZ area, since the two branches are independent. By contrast, the IPDL model captures cross-dimensional substitution, where a reduction in driving within the CRZ may also lead to a decline in driving outside the CRZ, as driving overall becomes less attractive. This enables us to capture the broader spillover effects of congestion tolls across the entire study area.

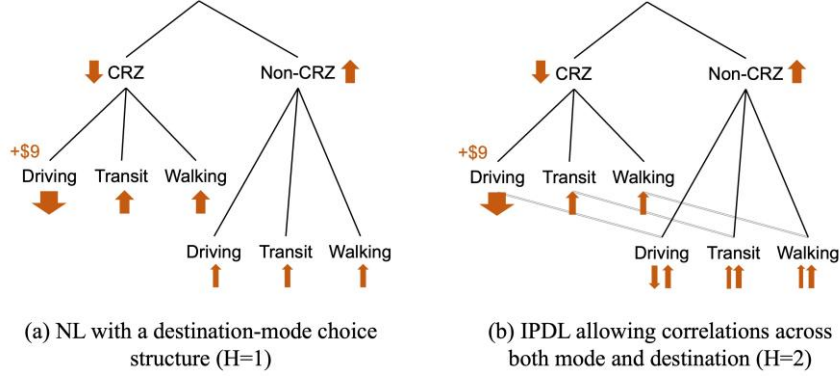


Fig. 2. A comparison between NL and IPDL.

In IPDL, the inverted market share is specified as Eqs. (11) – (12).

$$V_{jt} = \ln G_j(s_t; \varphi) + c_t \quad (11)$$

$$\ln G_j(s_t; \varphi) = \left(1 - \sum_{h=1}^H \rho_h\right) \ln(s_{jt}) + \sum_{h=1}^H \rho_h \ln\left(\sum_{q \in J_h} s_{qt}\right) \quad (12)$$

where $G_j(s_t; \varphi)$ is the invertible function of market share, c_t is a constant for market t , ρ_h is the grouping parameter for dimension h , and J_h is a set of alternatives grouped by dimension h . The higher value of ρ_h implies that alternatives in the same group are more similar in dimension d than other dimensions. To this end, correlation among multiple dimensions is captured by $\varphi = \{\rho_1, \rho_2, \dots, \rho_H\}$. Since the systematic utility of the outside alternative is assumed to be zero ($V_{0t} = 0$), we have $\ln G_j(s_t; \varphi) + c_t = \ln(s_{0t}) + c_t = V_{0t} = 0 \rightarrow c_t = -\ln(s_{0t})$. Linking this to Eqs. (11) – (12) we obtain Eq. (13) that relates the inverse market share to alternative attributes and nesting variables.

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \theta^T X_{jt} + \sum_{h=1}^D \rho_h \ln\left(\frac{s_{jt}}{\sum_{q \in J_h} s_{qt}}\right) \quad (13)$$

where $\ln\left(\frac{s_{jt}}{\sum_{q \in J_d} s_{qt}}\right)$ serve as a nesting variable associated with hierarchical dimension h , with parameter ρ_h to be estimated. Fosgerau et al. (2024) demonstrated that estimating IPDL reduces to a linear regression, where $\ln\left(\frac{s_{jt}}{s_{0t}}\right)$ is the dependent variable and $\left(X_{jt}, \ln\left(\frac{s_{jt}}{\sum_{q \in J_h} s_{qt}}\right)\right)$ are the independent variables. Since Eq. (13) holds for all alternatives and markets, the total number of regression observations is $|J| \times |T|$. Moreover, IPDL can be estimated using the two-stage least squares (2SLS) approach to handle endogeneity bias in market-level models (Angrist & Krueger, 2001). Following Krueger et al. (2023)'s work, we treat travel cost as an endogenous variable. We first group alternatives across two dimensions: mode and destination. Using the approach adopted by Fosgerau et al. (2024) and Ren et al. (2025), travel time variables of other alternatives in the same group are averaged. Since we have three travel time variables related to auto travel time, transit in-vehicle time, and non-auto travel time, we create $3 \times 2 = 6$ instrumental variables. Finally, we run instrumental regression to address the price endogeneity, which allows an unbiased estimation of θ_{cost} that is crucial for welfare analysis.

We estimated a separate model with deterministic parameters for each trip segment, assuming that taste parameters are homogenous within the same trip segment while heterogeneous across segments. Accordingly, taste parameters θ is indexed by $g \in G$, where each segment defined in [Section 3.1](#) has a unique set of parameters.

3.2.3 Post-implementation parameter calibration

Since only marginal data is available for the post-implementation period, it is not feasible to estimate a choice model with confidence intervals using maximum log-likelihood estimation (MLE). Instead, we introduce toll-related parameters to capture the general effects of congestion pricing on travelers' preferences. Additionally, we consider three types of changes in alternative attributes: (1) congestion tolls are applied to auto trips entering the CRZ, (2) a 15% increase in average driving speeds within the CRZ according to rates from [Cook et al. \(2025\)](#)'s report, and (3) changes in transit service performance between 2023 to 2025 captured by OTP. Together, these adjustments results in the utility function for the post-implementation period as shown in [Eq. \(14\)](#).

$$V_{jt}^{post} = \theta^T(X_{jt} + \Delta X_{jt}) + \sum_{m \in M_{toll}} \theta_{asc-toll}^m I_{jt} + \theta_{asc-toll}^{CRZ} L_{jt}, \quad \forall j \in J, \forall t \in T \quad (14)$$

where X_{jt} and θ^T denote the alternative attributes and taste parameters in the pre-implementation model; ΔX_{jt} represents changes in alternative attributes. I_{jt} equals 1 if a congestion toll is applied to trip jt , and 0 otherwise. L_{jt} equals 1 if the destination of trip jt is located within the CRZ, and 0 otherwise. $M_{toll} = \{driving, fhv, carpool\}$ is a set of modes charged by the program. $\theta_{asc-toll}^{driving}$, $\theta_{asc-toll}^{fhv}$, $\theta_{asc-toll}^{carpool}$, and $\theta_{asc-toll}^{CRZ}$ are four toll-related parameters to be calibrated.

To calibrate these parameters, we formulate an optimization problem that minimizes the gap between predicted and observed changes in traffic volume from 2023 to 2025, as shown in [Eqs. \(15\) – \(17\)](#).

$$\min_{\theta_{asc-toll}^{driving}, \dots, \theta_{asc-toll}^{CRZ}} (\Delta MTA_{NY} - \Delta Pred_{NY})^2 + (\Delta MTA_{NJ} - \Delta Pred_{NJ})^2 \quad (15)$$

subject to:

$$\Delta Pred_r = \frac{\sum_{j \in J_r} s_{jt}^{post} - \sum_{j \in J_r} s_{jt}^{pre}}{\sum_{j \in J_r} s_{jt}^{pre}}, \quad \forall r \in \{NY, NJ\} \quad (16)$$

$$s_{jt}^p = \frac{e^{V_{jt}^p}}{\sum_{q \in J} V_{qt}^p}, \quad \forall p \in \{pre, post\} \quad (17)$$

where ΔMTA_{NY} and ΔMTA_{NJ} denote the MTA-observed percentage changes in traffic volumes from New York and New Jersey to the CRZ. $\Delta Pred_{NY}$ and $\Delta Pred_{NJ}$ denote the corresponding predicted changes, which are computed from the utilities and market shares of the pre- and post-implementation models. Due to current data availability, we approximate traffic from New York to CRZ using the Queens Midtown Tunnel and Hugh Carey, and traffic from New Jersey to CRZ using the Lincoln Tunnel and Holland Tunnel. We solve the optimization problem using the Sequential Least Squares Programming

(SLSQP) algorithm in SciPy (Kraft, 1988), which efficiently handles nonlinear optimization problems using a quasi-Newton method.

Since our study places particular emphasis on transit strategies, we rely on MTA's transit ridership data to validate the model predictions. Specifically, we collect ridership records from April to June in both 2023 (pre-implementation) and 2025 (post-implementation). For each period, we calculate the average daily ridership and then compare these empirical values with the model-predicted number of public transit trips. We focus on April to June because this timeframe aligns with the synthetic population data used in our choice model estimation, ensuring consistency between observed and modeled travel patterns.

3.3 Metrics for welfare analysis

Based on the taste parameters from the choice models, we calculate several metrics for welfare analysis, including value of time, consumer surplus, compensating variation.

Value of time (VOT) measures traveler's trade-off between time savings and monetary costs, reflecting their willingness to pay for reduced travel time (Small, 2012). Following existing studies, we compute VOT as the marginal rate of substitution between travel time and travel cost. For instance, the value of auto travel time for trip segment $g \in G$ is defined in Eq. (18).

$$VOT_g = \frac{\theta_{autoTT,g}}{\theta_{cost,g}}, \quad \forall g \in G \quad (18)$$

where G is a set containing 16 trip segments; $\theta_{autoTT,g}$ is the parameter of auto travel time for segment g ; $\theta_{cost,g}$ is the cost parameter for segment g .

Consumer surplus (CS) is an economic concept that quantifies consumer welfare using the difference between the highest price a consumer is willing to pay for a good or service and the actual price they pay (Small & Rosen, 1981). Consistent with other choice models, CS in the IPDL framework can be computed as the logsum of alternative utilities (Fosgerau et al., 2024), as shown in Eq. (19).

$$CS_t = \ln \left(\sum_{j \in J} H_j(e^{V_{jt}}) \right) + C = c_t + C = -\ln(s_0) + C, \quad \forall t \in T \quad (19)$$

where $H_j(e^{V_{jt}}) = G_j^{-1}(e^{V_{jt}})$ denotes the utility function adjusted by considering alternative correlations, c_t is the constant for market t in Eq. (11), and C is an unknown constant.

CSs from different model specifications cannot be directly compared due to C . However, it can be converted to monetary units, and thus comparable units using compensating variation (CV), which is interpreted as the dollar amount an individual would have to be compensated to be as well off as before a policy change (Freeman et al., 2014). In our study, CV brought by the congestion toll is defined in Eq. (20).

$$CV_t = -\frac{1}{\theta_{cost}} (CS_t^{post} - CS_t^{pre}), \quad \forall t \in T \quad (20)$$

where θ_{cost} is the cost parameter, serving as a proxy for the marginal utility of income. CS_t^{pre} and CS_t^{post} denote consumer surplus before and after the implementation of

congestion pricing, where C drops out. Since transit service performance improved between 2023 and 2025 (see [Appendix Table A2](#)), we control for this effect by using 2023 transit travel times when calculating post-implementation CS. This avoids confounding improvements in transit service with the welfare impacts of congestion pricing.

3.4 Compensatory transit strategies and experimental design

How the toll revenue is reinvested to offset the potential negative welfare (welfare loss) will be critical to the long-term success of NYC's congestion pricing program. We evaluate two compensatory transit strategies: reducing wait time and providing fare discount. Reducing wait time (or increasing service frequency) is one of the most common transit improvement strategies ([Chen & Nozick, 2016](#)). Wait time reduction can also represent improvements in service reliability, as average wait time is a function of headway variance ([Osuna & Newell, 1972](#)). Transit fare discount, on the other hand, directly alleviates the monetary burden on travelers, which is adopted as a short-term strategy to enhance affordability and promote public transit usage, especially when pricing policies increase out-of-pocket travel costs ([Paulley et al., 2006](#)). We operationalize these strategies by applying reductions in average waiting times and population segment-specific fare discounts to corresponding trips.

Moreover, compensatory transit strategies are evaluated under two objectives: achieving Kaldor–Hicks efficiency ([Kaldor, 1939](#)) and ensuring Pareto improvement ([Varian, 1992](#)). The former emphasizes whether the total welfare losses are compensated, while the latter requires sufficient compensation for the origin county and population group with the largest loss. Under the Kaldor–Hicks efficiency scenario, we evaluate the two strategies independently and calculate the amounts of wait time reduction or fare discount required to offset the aggregate welfare loss, as shown in [Eqs. \(21\) – \(22\)](#).

$$WTC^{KH} = f_{wt}(\sum_{t \in T_c} CV_t) \quad (21)$$

$$FDC^{KH} = f_{fd}(\sum_{t \in T_c} CV_t) \quad (22)$$

where $\sum_{t \in T_c} CV_t$ represents the total welfare change aggregated across the set of markets to be compensated (T_c). $f_{wt}(\cdot)$ and $f_{cost}(\cdot)$ are functions that take changes in CS as inputs and return the corresponding amounts of wait time and fare discount compensation needed to offset those changes. Since CS changes nonlinearly with respect to alternative attributes, these functions do not have closed-form solutions. Instead, we use a numerical root-finding approach to identify the values of WTC^{KH} or FDC^{KH} that reproduce the welfare change induced by the congestion toll in each market. This is implemented as an optimization problem where WTC^{KH} or FDC^{KH} is defined as the decision variable, and the objective is to minimize the squared difference between the given and reproduced $\sum_{t \in T_c} CV_t$, solved using the SLSQP algorithm in SciPy.

Under the Pareto improvement scenario, we consider a set of combined strategies, since traveler groups differ in their value of time and may not be fully compensated through a single measure. Specifically, we treat trips made by each population and originating in each county as a group. Group-level welfare losses are first compensated through incremental reductions in wait time, with levels set at half-minute intervals. For each level of time reduction, we then calculate the remaining uncompensated welfare for each group, which is subsequently addressed through population segment-specific transit fare reductions. Finally, our evaluation quantifies the decreases in wait time (min) and the corresponding subsidies

(\\$) needed to achieve full welfare compensation. Given a one-minute reduction in wait time, the fare discount for population g to ensure Pareto improvement, FDC_g^{Pareto} , is defined in Eqs. (23) – (25).

$$FDC_g^{Pareto} = \max_{t \in T_{c,g}} FDC_t^{Pareto}, \quad \forall g \in G \quad (23)$$

$$FDC_t^{Pareto} = f_{wt}(CV_t^{remain}), \quad \forall t \in T_c \quad (24)$$

$$CV_t^{remain} = \min(CV_t + CV_t^{1min}, 0), \quad \forall t \in T_c \quad (25)$$

where CV_t represents the CV brought by congestion pricing; CV_t^{1min} represents the CV brought by one-minute reduction in wait time; and CV_t^{remain} represents the remaining CV (non-negative) to compensate the loss in market t ; $T_{c,g}$ is the set of compensated markets belonging to population g ; the maximum operator ensures that the compensation level for each population segment is sufficient to cover the largest welfare loss among its corresponding markets.

4. Results

This section presents the results of choice models, welfare analysis, and transit policy evaluations. The experiments were conducted on a local machine equipped with an Intel Core i7-10875H CPU and 32GB of RAM. The AER package in R was used for IPDL estimations, while the remaining analyses were performed in Python.

4.1 Estimated choice models

4.1.1 Basic statistics

[Table 2](#) summarizes the parameter estimates of four selected pre-implementation choice models, each corresponding to a different population group. The reported values include mean estimates, standard errors, and significance levels. For brevity, only the destination-specific constant for the CRZ is reported. Model parameters for other trip segments, as well as a comparison among the MNL, NL, and IPDL, are provided in [Appendix Tables A4–A8](#).

In general, the adjusted R^2 (based on inverse market share) for the four models is around 0.90, while the McFadden R^2 (based on loglikelihood value) is around 0.45. Most parameters are significant, and the estimates align with existing studies on mode and destination choice ([He et al., 2021](#)).

Transit wait time and transfer penalties are strongly significant for commuters but become less influential for seniors and students, suggesting that non-commute trips made by these populations may exhibit greater tolerance toward delayed schedules or additional transfers.

The significant nesting parameters confirm the appropriateness of the IPDL framework: travelers exhibit substitution patterns both across modes serving the same destination and across destinations accessible by the same mode.

Mode- and destination-specific constants provide insight into baseline preferences. Driving constants are positive and significant across groups, underscoring the relative attractiveness of private auto use, while biking and walking constants are strongly negative,

particularly for seniors and students, reflecting limited substitution toward active modes in these groups. The CRZ-specific constants are uniformly negative and large in magnitude, indicating substantial disutility associated with traveling to the area even before the implementation of congestion pricing. This is likely due to traffic congestion, limited availability of parking, and transit service unreliability.

NYC-specific interaction terms highlight spatial heterogeneity in preferences. For example, auto and transit travel time in New York City are more negatively perceived compared to other regions, suggesting that travelers in NYC often operate under more rigid time schedules. Conversely, cost interactions for trips destined to NYC show positive values, potentially indicating the higher baseline willingness to pay among travelers accessing NYC.

Table 2

Parameter estimates of selected pre-implementation choice models (each entry represents the average value, and the number in the parenthesis is the standard error).

	NotLowIncome, Commute, Peak	LowIncome, Commute, Peak	Senior, Non- commute, Peak	Student, Non- commute, Peak
Travel time and cost				
Auto travel time (θ_{autoTT})	-0.033*** (0.002)	-0.024*** (0.002)	-0.033*** (0.002)	-0.027*** (0.002)
Transit access time (θ_{AT})	-0.099*** (0.011)	-0.105*** (0.011)	-0.034 (0.052)	-0.107* (0.043)
Transit egress time (θ_{ET})	-0.100*** (0.017)	-0.063*** (0.018)	-0.269*** (0.062)	-0.142** (0.046)
Transit wait time (θ_{WT})	-0.104*** (0.005)	-0.066*** (0.005)	0.012 (0.022)	-0.023 (0.019)
Transit in-vehicle time (θ_{IVT})	-0.043*** (0.003)	-0.027*** (0.002)	-0.006 (0.009)	-0.022** (0.007)
Number of transfers (θ_{trans})	-0.680*** (0.184)	-0.313* (0.149)	-0.065 (0.316)	0.083 (0.294)
Non-vehicle travel time ($\theta_{nonautoTT}$)	-0.044*** (0.009)	-0.044*** (0.007)	-0.057*** (0.008)	-0.039*** (0.008)
Trip cost (θ_{cost})	-0.147*** (0.021)	-0.515*** (0.040)	-0.276*** (0.022)	-0.356*** (0.028)
NYC-specific interaction terms				
Auto travel time (θ_{autoTT}^{NYC})	-0.017*** (0.002)	-0.014*** (0.002)	-0.016*** (0.002)	-0.015*** (0.002)
Transit access time (θ_{AT}^{NYC})	-0.135*** (0.012)	-0.118*** (0.008)	0.029 (0.021)	-0.029 (0.016)
Transit egress time (θ_{ET}^{NYC})	-0.146*** (0.028)	-0.088*** (0.023)	0.224*** (0.069)	0.091 (0.049)
Transit wait time (θ_{WT}^{NYC})	-0.058*** (0.007)	-0.040*** (0.005)	0.017 (0.023)	-0.003 (0.021)
Transit in-vehicle time (θ_{IVT}^{NYC})	-0.034*** (0.004)	-0.016*** (0.003)	0.006 (0.012)	-0.005 (0.009)
Non-vehicle travel time ($\theta_{nonautoTT}^{NYC}$)	-0.020* (0.008)	-0.041*** (0.006)	-0.046*** (0.006)	-0.041*** (0.007)
Trip cost (θ_{cost}^{NYC})	0.012** (0.005)	0.054*** (0.010)	0.040*** (0.005)	0.050*** (0.006)
Mode and destination constant				

Driving constant ($\theta_{asc}^{driving}$)	0.459*** (0.048)	0.555*** (0.049)	0.491*** (0.045)	0.552*** (0.049)
Transit constant ($\theta_{asc}^{transit}$)	-0.763** (0.253)	0.081 (0.180)	0.607* (0.258)	0.427* (0.204)
FHV constant (θ_{asc}^{fhv})	-0.098 (0.274)	0.819* (0.370)	0.055 (0.152)	0.295 (0.194)
Biking constant (θ_{asc}^{biking})	-1.647*** (0.240)	-1.553*** (0.198)	-1.244*** (0.183)	-1.429*** (0.193)
Walking constant ($\theta_{asc}^{walking}$)	-0.622*** (0.103)	-0.126 (0.097)	-0.127 (0.089)	0.023 (0.090)
CRZ-specific constant (θ_{asc}^{CRZ})	-3.088*** (0.390)	-1.910*** (0.402)	-2.178*** (0.289)	-1.937*** (0.331)
Nesting parameter				
$\ln\left(\frac{S_{jt}}{\sum_{q \in J_{mode}} S_{qt}}\right)$	0.642*** (0.025)	0.710*** (0.026)	0.647*** (0.020)	0.736*** (0.022)
$\ln\left(\frac{S_{jt}}{\sum_{q \in J_{destination}} S_{qt}}\right)$	0.543*** (0.031)	0.409*** (0.033)	0.521*** (0.031)	0.490*** (0.033)
Meta information				
Instrumental variables	Yes	Yes	Yes	Yes
# Observations	3,859	2,887	3,291	2,643
# Trips per day	14,050,299	3,045,699	3,556,543	2,920,304
Adj. R ²	0.912	0.890	0.892	0.908
McFadden R ²	0.458	0.434	0.445	0.457
Estimation time	8 s	6 s	7 s	6 s

Note: ***p-value<0.001, **p-value<0.01, *p-value<0.05. Given the table length, only the destination constant of CRZ is reported in the table.

4.1.2 Model prediction and validation

The toll-related parameters in the post-implementation model are calibrated as follows: $\theta_{asc-toll}^{driving} = -0.287$, $\theta_{asc-toll}^{fhv} = -0.224$, $\theta_{asc-toll}^{carpool} = -0.214$, and $\theta_{asc-toll}^{CRZ} = -0.182$. Table 3 presents the comparison between model predictions and observed data from the MTA. For the calibration data, the model accurately reproduces the percentage changes in trips entering the CRZ from New York and New Jersey, with prediction errors of less than 1%.

Table 3

Comparison of model predictions with MTA observed transit ridership

	Model prediction	MTA observation	% error
Data used for parameter calibration			
% change in auto trips from New York	-12.95%	-13.06%	0.84%
% change in auto trips from New Jersey	-10.37%	-10.29%	-0.77%
Data used for model validation			
Transit ridership in 2023 Q2 (trips/day)	4,361,722	4,478,608	-2.61%
Transit ridership in 2025 Q2 (trips/day)	4,668,461	4,872,669	-4.19%
% change in transit ridership	7.03%	8.79%	-20.02%

Note: “auto” includes three trip modes: driving, FHV, and carpool.

For the validation data, the predicted transit ridership levels in Q2 of both 2023 and 2025 are slightly lower than MTA's observed ridership, with percentage errors of -2.61% and -4.19% , respectively. Our models predicts a 7.03% increase in transit ridership from 2023 to 2025, which is slightly lower than the observed increase of 8.79% . This underestimation may be because the models only consider the impacts of congestion pricing, while other transit-promoting initiatives in the study area may also have contributed to the observed ridership growth. Nevertheless, the underestimation is acceptable as we focus mainly on congestion pricing.

Fig. 3 illustrates the spatial and modal distribution of predicted trips. While the total trip volume remains nearly unchanged with only a marginal decline of 0.01% , trips destined for the CRZ and upper Manhattan are reduced more noticeably, by 1.62% and 1.15% , respectively. In contrast, trips to other areas of New York City (outside Manhattan) and New York State remain relatively stable, with only minor reductions of 0.09% . From a modal perspective, driving trips from Manhattan, NYC, and New Jersey show notable declines after the implementation of congestion pricing, accompanied by corresponding increases in transit usage, indicating a clear modal shift away from auto travel. These results indicate that congestion pricing primarily deters auto travel into the CRZ and reduce car usage throughout the NYC, while leaving overall trip-making behavior largely unchanged, suggesting potential modal and spatial substitution effects.

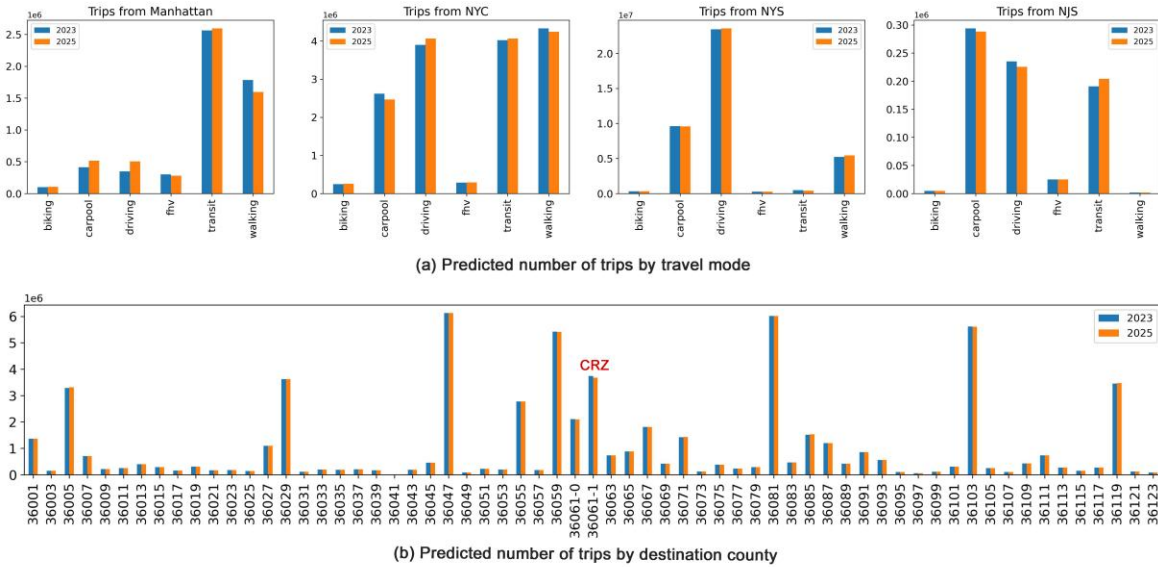


Fig. 3. Predicted number of trips by mode and destination. In panel (a), NYC refers to areas of New York City outside Manhattan, NYS refers to areas of New York State outside New York City, and NJS refers to the five selected counties in New Jersey State. In panel (b), “36061-1” refers to the CRZ and “36061-1” refers to the upper Manhattan.

Table 4 reports the estimated toll revenue by population group, mode, and time period, where driving and carpool trips are categorized as using passenger cars and FHV trips are categorized as using for-hired vehicles. The total toll revenue is estimated to be approximately \$1.077 billion per year. Not-low-income travelers account for the largest share, contributing nearly \$750 million annually, primarily from passenger car trips during peak periods (\$621 million). In comparison, other population groups contribute smaller

shares, ranging from \$80 –130 million each. This is due to the substantially larger number of trips into the CRZ made by not-low-income population relative to other groups.

It is important to note that our estimation (\$1.077 billion/year in total) is considerably higher than the \$500 million in annual net revenue reported by [MTA \(2025c\)](#), as our calculation does not account for infrastructure and administrative costs associated with implementing congestion pricing. Accordingly, our estimate represents the gross revenue, which is slightly higher than the gross revenue (\$889 million/year) estimated using MTA vehicle entries (see [Appendix Table A9](#)). Several reasons may account for the overestimation. First, the calculations are based solely on weekday trips, when commuting demand into the CRZ is higher, and do not consider variations in weekend travel. Second, we apply a uniform toll rate of \$1.50 for all for-hire vehicle (FHV) trips, as the available data does not distinguish between taxis (\$0.75/trip) and app-hail vehicles (\$1.50/trip), which may overestimate the revenue from this category. Third, seasonal variation is not considered, as the synthetic trip data reflect only the second quarter of the year.

Nevertheless, the estimates remain valuable as they provide a benchmark for understanding the scale of expected revenues and their distribution across population groups. Such insights are essential for evaluating the fiscal sustainability of the program and for informing decisions about the amount of toll revenues to be reinvested.

Table 4

Gross toll revenue estimated by our models

	Predicted tolled trips (trips/day)	Toll rate (\$)	Annual revenue (million \$)
NotLowIncome Population			
Passenger cars (Peak)	189,032	9.00	620.97
For-hired vehicles (Peak)	181,510	1.50	99.38
Passenger cars (Overnight)	23,435	2.25	19.25
For-hired vehicles (Overnight)	20,016	1.50	10.96
LowIncome Population			
Passenger cars (Peak)	24,943	9.00	81.49
For-hired vehicles (Peak)	32,821	1.50	17.97
Passenger cars (Overnight)	2,786	2.25	2.29
For-hired vehicles (Overnight)	3,736	1.50	2.05
Senior Population			
Passenger cars (Peak)	34,485	9.00	113.28
For-hired vehicles (Peak)	32,483	1.50	17.78
Passenger cars (Overnight)	4,528	2.25	3.72
For-hired vehicles (Overnight)	3,831	1.50	2.10
Student Population			
Passenger cars (Peak)	20,440	9.00	67.15
For-hired vehicles (Peak)	27,246	1.50	14.92
Passenger cars (Overnight)	2,365	2.25	1.94
For-hired vehicles (Overnight)	2,367	1.50	1.30

Note: Trips by driving and carpool modes are categorized as using passenger cars.

4.2 Distributional welfare impacts on accessibility

The overall compensating variation (CV) shows that congestion pricing without redistributing the revenue leads to a net consumer surplus (CS) loss of \$657,573 per day, or approximately \$240 million per year. In a report from [MTA \(2025c\)](#), the net toll revenue is projected to be \$500 million per year. The report also noted that \$48.66 million was generated from the first month, 9% of which comes from trucks and 1% from tourist buses and motorcycles. Since our models do not consider these vehicle categories, we reduce the reported net revenue estimate by 10% to ensure comparability.

Taken together, the total CS loss (\$240 million per year) is significantly smaller than the gross toll revenue estimated by our models (\$1,077 million per year) and the adjusted net toll revenue projected by the MTA (\$450 million per year), demonstrating that the policy satisfies Kaldor–Hicks efficiency ([Kaldor, 1939](#)): although tolls impose costs on drivers who continue to enter the charged zone, the aggregate benefits from reduced congestion, improved traffic speeds, and toll revenues for reinvestment outweigh these losses.

However, this does not ensure the policy is Pareto improving, where no one is made worse off by the implementation of a policy ([Varian, 1992](#)), as distributional impacts on accessibility persist due to heterogeneities in tolls burdens, travel time changes, and how travelers perceive them.

4.2.1 Value of time (VOT) by segment

Toll burdens and travel time savings are perceived differently across traveler groups, which can be captured using the estimated value of time (VOT). [Table 5](#) lists the VOT for 16 trip segments. By incorporating NYC-specific interaction terms into travel time and cost, we are able to differentiate VOTs between trips starting from NYC and those from other regions.

Table 5

VOT (\$/hour) by trip segment and region

	VOT (autoTT)		VOT (transitVT)		VOT (transitWT)		VOT (nonautoTT)	
	NYC	Other	NYC	Other	NYC	Other	NYC	Other
NotLowIncome Population								
Commute, Peak	22.24	13.52	34.39	17.54	72.67	42.58	19.48	17.84
Commute, Overnight	17.82	6.47	--	--	--	--	21.17	10.30
Non-commute, Peak	13.06	6.50	17.28	6.88	28.39	13.59	29.87	12.97
Non-commute, Overnight	14.55	5.29	--	--	6.72	5.79	29.88	12.44
LowIncome Population								
Commute, Peak	5.00	2.79	5.58	3.15	13.83	7.68	11.02	5.12
Commute, Overnight	6.05	2.60	--	--	--	--	14.61	6.62
Non-commute, Peak	5.91	3.04	9.81	4.07	22.11	10.18	14.75	6.99
Non-commute, Overnight	7.10	2.49	--	--	--	--	18.49	8.63
Senior Population								
Commute, Peak	13.46	7.18	3.61	3.13	11.82	10.24	31.46	15.22
Commute, Overnight	9.42	3.94	--	--	--	--	30.76	12.94
Non-commute, Peak	12.30	7.11	--	--	--	--	26.31	12.49

Non-commute, Overnight	10.31	4.82	--	--	--	--	31.44	13.81
Student Population								
Commute, Peak	8.07	5.79	2.38	2.07	3.49	3.03	8.62	7.49
Commute, Overnight	9.07	4.03	--	--	--	--	4.69	4.18
Non-commute, Peak	8.27	4.51	4.37	3.77	--	--	7.69	6.62
Non-commute, Overnight	3.43	3.43	1.45	1.45	2.93	2.93	6.68	6.68

Note: “autoTT” refers to auto travel time; “transitIVT” refers to transit in-vehicle time; “transitWT” refers to transit wait time; “nonautoTT” refers to nonauto travel time. “--” indicates that the relevant parameters are insignificant. VOT values for the same segment in NYC and other regions are the same when the interaction parameters are insignificant.

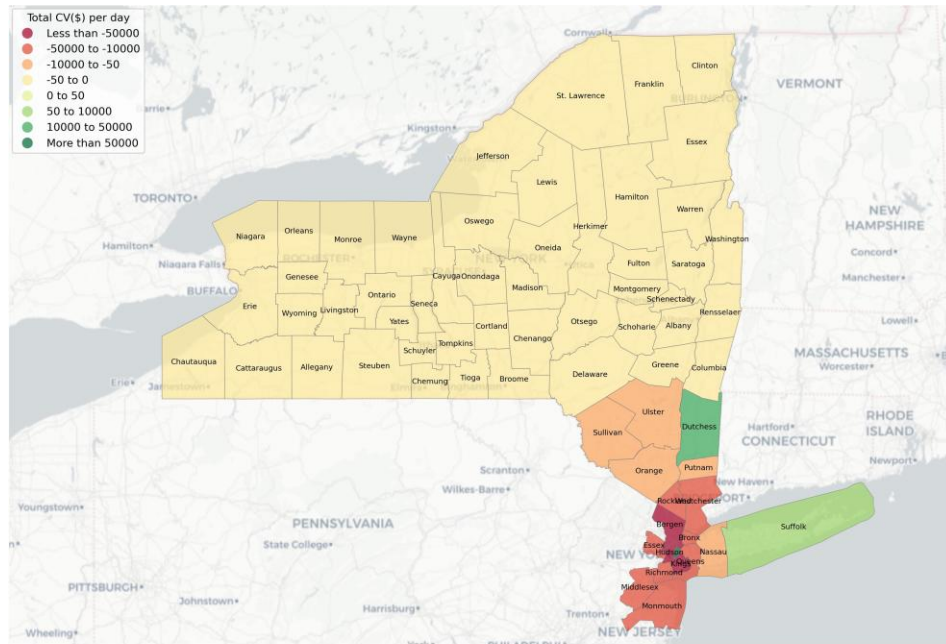
The results reveal substantial heterogeneity in VOT across segments, reflecting variations in both socio-demographic characteristics and travel contexts. Non-low-income commuters during peak hours exhibit the highest VOT, particularly for transit wait time, which exceeds \$70 per hour in NYC, underscoring the heightened disutility of delays in dense urban environments. By contrast, low-income travelers consistently exhibit lower VOTs across all time and purpose components, indicating that they place greater weight on monetary cost relative to travel time savings. Seniors demonstrate moderate sensitivity to auto and non-auto travel times but lower valuation of in-vehicle and wait times, suggesting greater flexibility in scheduling. Students report the lowest VOTs overall, especially for non-commute trips, consistent with reduced time constraints and budget limitations in this group. The NYC-specific interactions further highlight spatial variation, with consistently higher VOTs in the city compared to other regions, emphasizing the premium that travelers place on time savings in areas with more rigid activity schedules and severe congestion pressures.

4.2.2 Welfare impacts across regions and segments

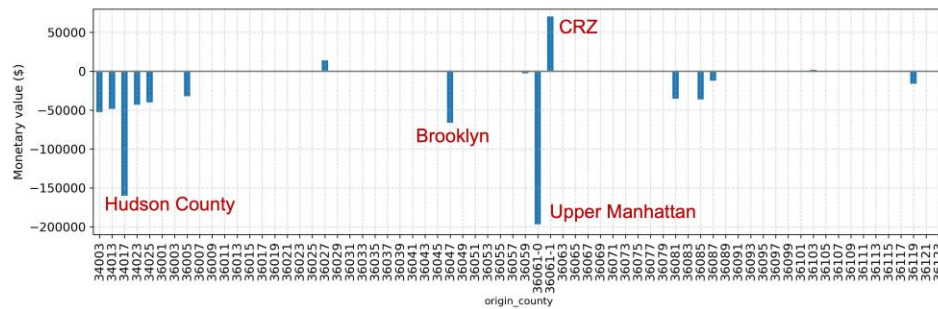
Fig. 4 highlights the uneven spatial distribution of welfare impacts under congestion pricing. The CRZ shows a welfare gain of around \$60,000 per day, while substantial daily welfare losses are concentrated in Upper Manhattan (around -\$200,000/day), Hudson County, NJ (around -\$160,000/day), and Brooklyn (around -\$60,000/day), reflecting their higher trip volumes into the CRZ and greater reliance on auto travel. Most outer New York State counties incur very small losses, which are consistent with their lower exposure to the toll. Some peripheral counties even exhibit slight positive compensating variation, likely attributable to travel time savings from reduced congestion. These results emphasize that welfare losses from congestion pricing are highly concentrated in a few dense and nearby counties.

Fig. 5 presents the average CV (\$) for a single trip across different origins, trip purposes, and population groups, where higher values reflect greater welfare impacts at the individual trip level. The results show that the welfare impacts also vary markedly across segments. Commuters from New Jersey are the most adversely affected, with average per-trip losses around -\$0.40/trip, reflecting their limited flexibility in choosing alternative destinations and strong reliance on auto travel into the CRZ. By contrast, the impacts on low-income and student populations are smaller than expected, as these groups are more likely to be substituted by public transit or redirected trips to other destinations. In addition, non-

commuter trips are generally less affected by the congestion toll compared to commute trips, probably because these trips have more flexible schedules.



(a) Spatial distribution of CV (\$) brought by the congestion toll



(b) Total CV (\$) aggregated by trip origin county

Fig. 4. Welfare impacts across trip origin counties. In panel (b), “36061-1” refers to the CRZ and “36061-1” refers to the upper Manhattan.

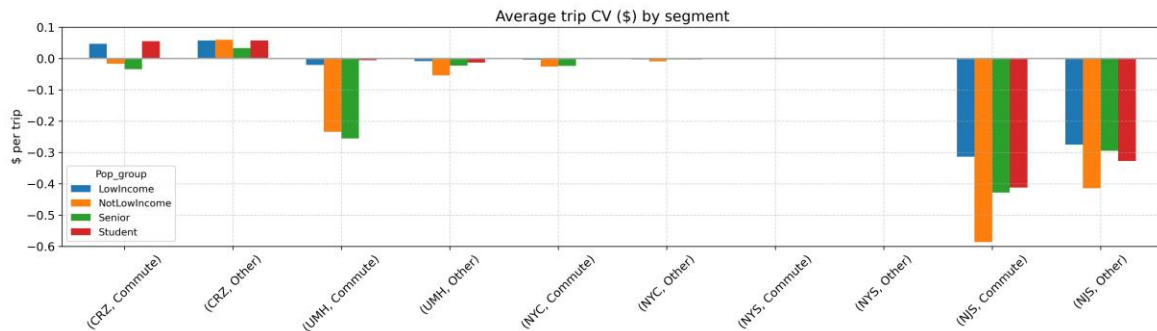


Fig. 5. CV (\$) per trip aggregated by segments. “UMH” refers to the upper Manhattan outside CRZ; “NYC” refers to the areas outside Manhattan; “NYS” refers to the areas outside NYC; “NJS” refers to the five selected counties in New Jersey; “Other” refers to non-commute trips.

4.3 Compensatory transit strategies

As the results in [Section 4.2](#) show, welfare losses are concentrated in NYC and New Jersey. Accordingly, we split the set of markets for compensation (T_c) and consider two subsets: (1) T_{NYC} , which includes all markets with trips originating in NYC, where reductions in wait time and fares are applied to all transit services operating within NYC; (2) T_{NJ} , which includes all markets with trips originating in five selected counties in New Jersey, where reductions in wait time and fares are applied to cross-state transit services between New Jersey and New York.

4.3.1 Evaluations under the scenario of Kaldor-Hicks efficiency

Under the Kaldor–Hicks efficiency scenario, we evaluate two compensatory strategies independently and estimate the extent of transit wait time reduction or fare discount needed to offset the aggregate accessibility-related welfare loss. As shown in [Table 6](#), achieving full welfare compensation requires a reduction of about 0.48 minutes in NYC and 5.32 minutes in New Jersey. Given the current average wait times (4.77 minutes in NYC and 7.98 minutes in New Jersey), these represent 10% and 67% reductions, respectively. Although the cost of reducing average wait times is not directly estimated in this study, empirical evidence suggests that achieving a 67% improvement in service frequency would require substantial capital investment, implying that compensating welfare losses through service enhancement in New Jersey would be considerably more challenging.

Alternatively, welfare compensation can be achieved through segment-specific fare discounts, requiring an annual subsidy of approximately \$135.59 million for NYC and \$108.53 million for New Jersey. The non-low-income population requires the largest fare discount per trip, as their sensitivity to monetary costs is relatively lower than time savings. In contrast, low-income and student travelers, who are more cost-sensitive, can be compensated with smaller reductions in fares. This imbalance raises an equity concern—compensating solely through fare discounts may inadvertently allocate a larger share of financial benefits to higher-income travelers, who are less burdened by monetary costs. This underscores the importance of designing integrated strategies that combine both time savings and fare reductions to achieve a more balanced and equitable outcome.

Table 6

Compensatory transit strategies under Kaldor-Hicks efficiency

	Wait time reduction (min)	Avg. wait time (min)	Segment-level fare discount (\$/trip)	Segment-level avg. transit travel cost (\$/trip)	Subsidy for discount (M \$/year)
NYC part	0.48	4.77	NotLowIncome:0.39	NotLowIncome:3.06	135.59
			LowIncome:0.07	LowIncome:3.08	
			Senior:0.45	Senior:1.55	
			Student:0.10	Student:2.99	
New Jersey part	5.32	7.98	NotLowIncome:2.23	NotLowIncome:4.48	108.53
			LowIncome:0.94	LowIncome:5.29	
			Senior:1.10	Senior:2.49	
			Student:1.37	Student:5.00	

4.3.2 Evaluations under the scenario of Pareto improvement

Under the Pareto improvement scenario, we require that all population groups and counties achieve non-negative welfare changes, implying that no group experiences a net loss after implementing compensatory strategies. Compared with the Kaldor–Hicks efficiency, which prioritizes aggregate welfare, achieving Pareto improvement demands substantially greater resources, as compensation must be sufficient for each group individually.

Fig. 6 presents combinations of wait time reduction and fare subsidy required to achieve Pareto improvement. When no reduction in transit waiting time is implemented, the required fare subsidy is approximately \$587.94 million/year for NYC and \$222.27 million/year for New Jersey, nearly several times of the amount estimated under the Kaldor–Hicks scenario. Moreover, even with a 10-minute reduction in transit wait time, a small fare subsidy is still required to fully offset the remaining welfare loss. These findings highlight the inefficiency of using a single compensatory strategy—either through wait times reductions or fare discounts—to achieve equitable welfare restoration across all traveler groups.

For New York City residents, the fare subsidy exhibits a steep decline between 1–2 min of wait time reduction, indicating that Pareto improvement can be achieved through a combined strategy of modest waiting time reduction and manageable fare subsidy. In contrast, New Jersey residents are more inelastic to transit improvements. For example, even with a 5–6 minute reduction in average wait time, an annual fare subsidy of approximately \$100 million is still required, which is nearly equivalent to the total fare subsidy needed to achieve Kaldor–Hicks efficiency (\$108.53 million/year). Given the inefficiency of transit improvement and considerable cost for Pareto improvement, pursuing Kaldor-Hicks efficiency while compensate mainly by fare discounts represents a more practical strategy for the New Jersey side.

Table 7 summarizes the required segment-level fare discounts and corresponding annual subsidies under various levels of wait time reduction from 0–5 min for both NYC and New Jersey residents. Evaluations under the 5.5–10 min time reduction are summarized in Appendix Table A10. With the decrease of wait time, not-low-income population can be fully compensated, as their higher value of time makes them more responsive to service frequency improvements. In contrast, senior and student populations exhibit lower time sensitivity, requiring substantial fare discounts even under notable wait time reductions. Moreover, the fare discounts required to achieve Pareto improvement are considerably higher than those needed for Kaldor–Hicks efficiency, especially when the wait time reduction is less than 5 minutes. This indicates that applying uniform discounts leads to overcompensation for some groups, thereby inflating the total subsidy. These results highlight the inefficiency of uniform fare discounts and underscore the importance of differentiated, region-specific adjustments to improve welfare without imposing excessive fiscal burden.

Table 7

Compensatory transit strategies under Pareto improvement (0–5 min wait time reduction)

Wait time reduction (min)	NYC part		New Jersey part	
	Segment-level fare discount (\$/trip)	Subsidy for discount (M \$/year)	Segment-level fare discount (\$/trip)	Subsidy for discount (M \$/year)
0	NotLowIncome:1.65 LowIncome:0.49	587.94	NotLowIncome:4.09 LowIncome:4.76	222.27

	Senior:1.51 Student:0.89		Senior:2.04 Student:4.16	
0.5	NotLowIncome:1.17 LowIncome:0.39 Senior:1.40 Student:0.88	439.23	NotLowIncome:3.82 LowIncome:4.66 Senior:2.02 Student:4.15	209.99
1.0	NotLowIncome:0.69 LowIncome:0.29 Senior:1.29 Student:0.86	290.52	NotLowIncome:3.55 LowIncome:4.56 Senior:2.01 Student:4.14	197.71
1.5	NotLowIncome:0.20 LowIncome:0.19 Senior:1.19 Student:0.84	141.81	NotLowIncome:3.28 LowIncome:4.46 Senior:1.99 Student:4.13	185.44
2.0	NotLowIncome:0 LowIncome:0 Senior:1.08 Student:0.82	72.80	NotLowIncome:3.01 LowIncome:4.36 Senior:1.97 Student:4.13	173.16
2.5	NotLowIncome:0 LowIncome:0 Senior:0.97 Student:0.81	61.46	NotLowIncome:2.75 LowIncome:4.26 Senior:1.96 Student:4.12	160.89
3.0	NotLowIncome:0 LowIncome:0 Senior:0.87 Student:0.79	56.81	NotLowIncome:2.48 LowIncome:4.16 Senior:1.94 Student:4.11	148.61
3.5	NotLowIncome:0 LowIncome:0 Senior:0.76 Student:0.77	52.17	NotLowIncome:2.21 LowIncome:4.06 Senior:1.92 Student:4.10	136.33
4.0	NotLowIncome:0 LowIncome:0 Senior:0.65 Student:0.75	47.53	NotLowIncome:1.94 LowIncome:3.96 Senior:1.91 Student:4.09	124.06
4.5	NotLowIncome:0 LowIncome:0 Senior:0.55 Student:0.74	42.89	NotLowIncome:1.67 LowIncome:3.86 Senior:1.89 Student:4.08	111.78
5.0	NotLowIncome:0 LowIncome:0 Senior:0.44 Student:0.72	38.25	NotLowIncome:1.40 LowIncome:3.76 Senior:1.87 Student:4.07	99.50

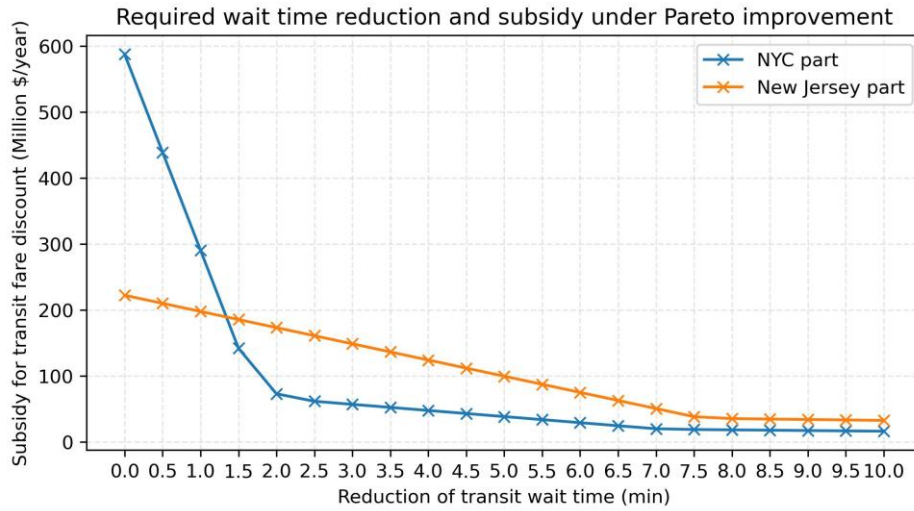


Fig. 6. Compensation strategy evaluation under Pareto improvement. The reduction of transit wait time differs between regions: for the NYC part, it is applied to all transit services operating within NYC; for the New Jersey part, it is applied to cross-state transit services operating between New Jersey and New York.

5. Discussion

The findings of this study highlight several important considerations for the long-term design and governance of congestion pricing. This section discusses three key policy implications: (1) the need to account for travelers' differing sensitivities to time and cost when designing compensations, (2) the trade-off between aggregate efficiency and equity across population groups, and (3) the necessity of broadening compensatory frameworks beyond transit to include multiple dimensions.

5.1 Compensatory levers: time saving vs. fare discount

Our results underscore that traveler responses to congestion pricing are mediated by heterogeneous sensitivities to time and money. Segment-specific VOT differ widely—non-low-income peak commuters in NYC place a very high premium on transit wait time, while low-income and student segments value monetary cost more heavily. These differences explain why a single lever can hardly restore welfare across all segments. In [Section 4.3](#) we showed that relying only on frequency improvements (i.e., reducing wait time) leaves cost-sensitive segments under-compensated, and relying only on fare discounts is expensive for high-VOT segments. The efficient frontier therefore combines both levers: modest, widely distributed reductions in scheduled headways (which benefit everyone and especially high-VOT users) paired with targeted fare relief for groups whose primary barrier is cost.

From an implementation standpoint, this argues for (1) service packages that prioritize frequent, all-day transit in high-demand corridors and (2) finely targeted, means-tested or trip-purpose-based fare programs (e.g., commuter-hour credits for low-income workers) rather than uniform, system-wide discounts. Such a portfolio minimizes deadweight compensation while maximizing perceived fairness and ridership response.

5.2 Balancing aggregate efficiency with equity goals

A second policy choice is the compensation criterion itself. Compensating the aggregate loss in consumer surplus (Kaldor–Hicks efficiency) requires substantially fewer resources than compensating every segment and county (Pareto improvement), but the former can mask significant distributional shortfalls. In our setting, Kaldor–Hicks compensation for NYC residents is attainable with modest wait-time reductions (1–2 minutes) plus limited fare subsidies; pushing to Pareto quickly becomes costly and, for some segments, infeasible through service improvements alone.

For New Jersey residents, we observed that even with a 5–6 minute reduction in average wait time, an annual fare subsidy of approximately \$100 million is still required, revealing that pursuing Kaldor–Hicks efficiency while compensate mainly by fare discounts is more efficient for the New Jersey side. Two implications follow. First, when the policy target is aggregate efficiency (e.g., to ensure the program’s fiscal viability), agencies should still track a small set of “equity sentinel” segments, such as auto-reliant New Jersey residents and specific NYC neighborhoods with limited alternatives. Second, where a jurisdiction aims for stronger equity guarantees, Pareto-style goals should be operationalized through more granular instruments (e.g., origin-specific discounts or commuter-product bundles) rather than uniform fare reductions, to avoid over-spending on groups already fully compensated. In practice, agencies can stage compensation: first meet a Kaldor–Hicks threshold system-wide, then add focused, data-driven transfers until identified gaps close.

5.3 Transit improvements and the broader compensation portfolio

While transit is the common reinvestment target, congestion pricing affects more than transit-eligible travelers. Trucks and commercial vehicles pay higher tolls, and their operators cannot simply substitute to bus and subway. This highlights the need for a broader compensation portfolio that complements transit with freight-oriented measures. Examples include delivery consolidation support, off-peak delivery incentives (paired with curb management and enforcement), and grants for zero-emission freight vehicles that both reduce operating costs and amplify air-quality benefits near the cordon. Similarly, some auto-dependent neighborhoods may require “first/last-mile” connectors (e-bike share, microtransit) to make transit frequency gains usable. Finally, environmental and health co-benefits—documented in other cities and emerging in New York—should be captured explicitly in the reinvestment calculus: bus priority and signal priority in the CRZ can lock in speed gains; targeted station accessibility upgrades ensure benefits accrue to seniors and people with disabilities; and sidewalk/bike-network investments expand viable non-auto substitutes. These complementary measures do not replace frequency and fare tools; they make them effective for populations whose constraints lie outside the transit farebox.

6. Conclusion

NYC’s congestion pricing program offers a rare, real-world testbed for how congestion pricing reshapes travel behavior, welfare, and policy priorities in a dense metropolitan region. Motivated by persistent concerns over fairness and practicality, this study measures distributional welfare impacts across New York and New Jersey and evaluates transit

reinvestment strategies that can credibly compensate losses. To that end, we leverage a regional, joint mode and destination framework that connects pre- and post-implementation conditions, allowing us to speak directly to questions that matter for program legitimacy: who bears losses, where, and by how much—plus what mix of transit frequency improvements and fare relief most effectively restores welfare.

Methodologically, the study advances welfare analysis with joint mode and destination choice models estimated using large-scale synthetic trips. Parameters reflecting toll-related preference changes are calibrated using MTA traffic counts and validated against ridership trends. Substantively, results show small changes in overall trip-making but clear modal and spatial substitution away from driving into the cordon. Net welfare is positive once toll revenue is included (approximately +\$210M/year on \$450M in annual revenue), yet losses are unequally distributed—concentrated in Upper Manhattan, Brooklyn, and Hudson County, NJ. Value-of-time heterogeneity is pronounced: high-VOT commuters are especially sensitive to waiting, while cost-sensitive groups respond more to fares. Consequently, single-lever compensation performs poorly. A mixed strategy—modest, broadly applied wait-time reductions paired with targeted fare discounts—can achieve Pareto improvement at manageable fiscal cost. Pareto-style guarantees become much more expensive and often infeasible via service improvements alone.

Several limitations point to a forward agenda. Aggregation to county markets and reliance on synthetic trips may mute within-market heterogeneity; post-implementation calibration draws on limited marginal counts; freight and commercial vehicles are outside the choice model; and temporal adaptations (e.g., departure-time shifts) are treated implicitly. Future work should exploit richer passenger-level panels and continuous post-rollout data, incorporate departure-time and destination attributes explicitly, co-optimize reinvestment with service design (headways, priority, first/last-mile), and extend compensation beyond transit (freight off-peak incentives, curb management, accessibility upgrades). Embedding equity constraints and uncertainty quantification in the revenue-allocation problem will further support durable, transparent reinvestment policies.

Acknowledgments

Funding support from C2SMARTER is appreciated. Data shared by [Replica Inc.](#) are gratefully acknowledged.

References

- American Association of State Highway and Transportation Officials. (2025). *2017-2021 Census Transportation Planning Products (CTPP) Data*. <https://aashtojournal.transportation.org/aashto-unveils-new-census-data-set-for-transportation/>
- American Trucking Associations. (2025). *Trucking Renews Push to End NYC Congestion Pricing Ahead of March 21 Deadline*. <https://www.trucking.org/news-insights/trucking-renews-push-end-nyc-congestion-pricing-ahead-march-21-deadline>
- Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, 15(4), 69–85. <https://doi.org/10.1257/jep.15.4.69>

- Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M., & Puchinger, J. (2019). Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 101, 254–275.
- Baghestani, A., Tayarani, M., Allahviranloo, M., Nadafianshahamabadi, R., Kucheva, Y., Reza Mamdoohi, A., & Oliver Gao, H. (2022). New York City cordon pricing and its impacts on disparity, transit accessibility, air quality, and health. *Case Studies on Transport Policy*, 10(1), 485–499. <https://doi.org/10.1016/j.cstp.2022.01.009>
- Basso, L. J., & Jara-Díaz, S. R. (2012). Integrating congestion pricing, transit subsidies and mode choice. *Transportation Research Part A: Policy and Practice*, 46(6), 890–900.
- Beevers, S. D., & Carslaw, D. C. (2005). The impact of congestion charging on vehicle emissions in London. *Atmospheric Environment*, 39(1), 1–5.
- Beria, P. (2016). Effectiveness and monetary impact of Milan's road charge, one year after implementation. *International Journal of Sustainable Transportation*, 10(7), 657–669.
- Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.
- Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics*, 25(2), 242. <https://doi.org/10.2307/2555829>
- Bills, T. S., Twumasi-Boakye, R., Broadus, A., & Fishelson, J. (2022). Towards transit equity in Detroit: An assessment of microtransit and its impact on employment accessibility. *Transportation Research Part D: Transport and Environment*, 109, 103341.
- Bloomberg, M. (2007). *PlaNYC: A Greener, Greater New York*. https://www.nyc.gov/html/planyc/downloads/pdf/publications/full_report_2007.pdf
- Börjesson, M., & Kristoffersson, I. (2015). The Gothenburg congestion charge. Effects, design and politics. *Transportation Research Part A: Policy and Practice*, 75, 134–146.
- Bowman, J. L., & Ben-Akiva, M. E. (2001). Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice*, 35(1), 1–28.
- Chen, C. (2025). *ROOTS OF RESISTANCE: UNDERSTANDING THE OPPOSITION TO CONGESTION PRICING IN NEW YORK CITY*.
- Chen, R., & Nozick, L. (2016). Integrating congestion pricing and transit investment planning. *Transportation Research Part A: Policy and Practice*, 89, 124–139.
- Cook, C., Kreidieh, A., Vasserman, S., Allcott, H., Arora, N., van Sambeek, F., Tomkins, A., & Turkel, E. (2025). The Short-Run Effects of Congestion Pricing in New York City. *NBER Working Paper No. 33584*. <https://www.nber.org/papers/w33584>
- De Palma, A., & Lindsey, R. (2011). Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies*, 19(6), 1377–1399.
- Downs, A. (2005). *Still stuck in traffic: Coping with peak-hour traffic congestion*. Rowman & Littlefield.
- Eliasson, J. (2009). A cost–benefit analysis of the Stockholm congestion charging system. *Transportation Research Part A: Policy and Practice*, 43(4), 468–480. <https://doi.org/10.1016/j.tra.2008.11.014>
- Eliasson, J. (2016). Is congestion pricing fair? Consumer and citizen perspectives on equity effects. *Transport Policy*, 52, 1–15.
- Fosgerau, M., De Palma, A., & Monardo, J. (2024). The Inverse Production Differentiation Logit Model. Available at SSRN: <https://ssrn.com/abstract=3141041>. <https://doi.org/10.2139/ssrn.3141041>
- Fosgerau, M., Monardo, J., & De Palma, A. (2024). The inverse product differentiation logit model. *American Economic Journal: Microeconomics*, 16(4), 329–370.
- Freeman Iii, A. M., Herriges, J. A., & Kling, C. L. (2014). *The measurement of environmental and resource values: Theory and methods*. Routledge.
- Harris, W., & Ley, A. (2024). How the Resurrected Congestion Pricing Plan Could Die in the Courts. *The New York Times*.

- He, B. Y., Zhou, J., Ma, Z., Wang, D., Sha, D., Lee, M., Chow, J. Y., & Ozbay, K. (2021). A validated multi-agent simulation test bed to evaluate congestion pricing policies on population segments by time-of-day in New York City. *Transport Policy*, 101, 145–161.
- Hensher, D. A., & Ho, C. Q. (2016). Experience conditioning in commuter modal choice modelling – Does it make a difference? *Transportation Research Part E: Logistics and Transportation Review*, 95, 164–176. <https://doi.org/10.1016/j.tre.2016.09.010>
- Hörl, S., & Balac, M. (2021). Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transportation Research Part C: Emerging Technologies*, 130, 103291.
- Huo, J., Dua, R., & Bansal, P. (2024). Inverse product differentiation logit model: Holy grail or not? *Energy Economics*, 131, 107379. <https://doi.org/10.1016/j.eneco.2024.107379>
- Isaksen, E. T., & Johansen, B. G. (2025). Congestion pricing with electric vehicle exemptions: Car-ownership effects and other behavioral adjustments. *Journal of Environmental Economics and Management*, 131, 103154. <https://doi.org/10.1016/j.jeem.2025.103154>
- Ji, Y. (2025). Distributional Impacts of Congestion Pricing in New York City. 2025 AAEA & WAEA Joint Annual Meeting. <https://doi.org/10.22004/ag.econ.360946>
- Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The Economic Journal*, 49(195), 549–552.
- Karlström, A., Waddell, P., & Fox, D. (2009). Scaling up the microeconomic dynamic discrete choice model of activity-based scheduling. *Proc. Eur. Transport Conf.*
- Kockelman, K. M., & Kalmanje, S. (2005). Credit-based congestion pricing: A policy proposal and the public's response. *Transportation Research Part A: Policy and Practice*, 39(7–9), 671–690. <https://doi.org/10.1016/j.tra.2005.02.014>
- Kraft, D. (1988). A software package for sequential quadratic programming. *Forschungsbericht- Deutsche Forschungs- Und Versuchsanstalt Fur Luft- Und Raumfahrt*.
- Krueger, R., Bierlaire, M., & Bansal, P. (2023). A data fusion approach for ride-sourcing demand estimation: A discrete choice model with sampling and endogeneity corrections. *Transportation Research Part C: Emerging Technologies*, 152, 104180. <https://doi.org/10.1016/j.trc.2023.104180>
- Leape, J. (2006). The London congestion charge. *Journal of Economic Perspectives*, 20(4), 157–176.
- Li, W., Kockelman, K. M., & Huang, Y. (2021). Traffic and Welfare Impacts of Credit-Based Congestion Pricing Applications: An Austin Case Study. *Transportation Research Record: Journal of the Transportation Research Board*, 2675(1), 10–24. <https://doi.org/10.1177/0361198120960139>
- Li, X., Shaw, J. W., Liu, D., & Yuan, Y. (2019). Acceptability of Beijing congestion charging from a business perspective. *Transportation*, 46(3), 753–776.
- Marazi, N. F., Majumdar, B. B., & Sahu, P. K. (2024). Examining Congestion Pricing Scheme Effectiveness Using the Travel Time Congestion Index. *Transportation Research Record: Journal of the Transportation Research Board*, 2678(11), 474–488. <https://doi.org/10.1177/03611981241242061>
- McFadden, D. (1977). *Modelling the choice of residential location*.
- McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447–470.
- Meier, A. (2024, November 15). New NYC congestion pricing plan: Details for drivers. FOX 5 NY. <https://www.fox5ny.com/news/nyc-congestion-pricing-plan-new-tolls-details-drivers-hochul>
- MobilityData. (2025). *Global Transit Data* [Data set]. <https://mobilitydatabase.org/>
- MTA. (2025a). *Central Business District Tolling Program*. <https://www.mta.info/project/CBDTP>
- MTA. (2025b). *Vehicle entries to the Congestion Relief Zone (CRZ) as detected by the tolling system* [Data set]. <https://metrics.mta.info/?cbdtp/vehicleentries>

- MTA. (2025c, February 24). *MTA Releases Revenue From Congestion Relief Zone Tolling Showing Program in Line With Projections*. <https://www.mta.info/press-release/mta-releases-revenue-congestion-relief-zone-tolling-showing-program-line-projections>
- National Bureau of Economic Research. (2025). Impact of New York City's Congestion Pricing Program. *NBER Digest (June 2025)*. <https://www.nber.org/digest/202506/impact-new-york-citys-congestion-pricing-program>
- National Environment Agency. (2010). *Singapore's second national communication under the UNFCCC: Addressing climate change for Singapore*. Singapore: National Environment Agency.
- Newman, J. P., & Bernardin Jr, V. L. (2010). Hierarchical ordering of nests in a joint mode and destination choice model. *Transportation*, 37(4), 677–688.
- Niemeier, D. A. (1997). Accessibility: An evaluation using consumer welfare. *Transportation*, 24(4), 377–396.
- Nogueira, M. (2025). The Impact of New York's 2025 Congestion Pricing on Traffic Volume. Available at SSRN 5383652.
- Olszewski, P., & Xie, L. (2005). Modelling the effects of road pricing on traffic in Singapore. *Transportation Research Part A: Policy and Practice*, 39(7–9), 755–772.
- Osuna, E. E., & Newell, G. F. (1972). Control strategies for an idealized public transportation system. *Transportation Science*, 6(1), 52–72.
- Parr, S., Wolshon, B., Renne, J., Murray-Tuite, P., & Kim, K. (2020). Traffic impacts of the COVID-19 pandemic: Statewide analysis of social separation and activity restriction. *Natural Hazards Review*, 21(3), 4020025.
- Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J., & White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, 13(4), 295–306. <https://doi.org/10.1016/j.tranpol.2005.12.004>
- Phang, S.-Y., & Toh, R. S. (2004). Road congestion pricing in Singapore: 1975 to 2003. *Transportation Journal*, 16–25.
- Regional Planning Association. (2025). *Congestion Pricing: Faster All Around*. <https://rpa.org/news/lab/congestion-pricing-getting-around-faster-all-around>
- Ren, X., & Chow, J. Y. J. (2022). A random-utility-consistent machine learning method to estimate agents' joint activity scheduling choice from a ubiquitous data set. *Transportation Research Part B: Methodological*, 166, 396–418.
- Ren, X., Chow, J. Y. J., & Bansal, P. (2025). Nonparametric mixed logit model with market-level parameters estimated from market share data. *Transportation Research Part B: Methodological*, 196, 103220. <https://doi.org/10.1016/j.trb.2025.103220>
- Ren, X., Chow, J. Y. J., & Guan, C. (2024). Mobility service design with equity-aware choice-based decision-support tool: New York case study. *Transportation Research Part D: Transport and Environment*, 132, 104255. <https://doi.org/10.1016/j.trd.2024.104255>
- Replica. (2024). *Annual Average Daily Traffic (AADT) validation*. <https://www.replicahq.com/traffic-datasets>
- Santos, G., & Bhakar, J. (2006). The impact of the London congestion charging scheme on the generalised cost of car commuters to the city of London from a value of travel time savings perspective. *Transport Policy*, 13(1), 22–33. <https://doi.org/10.1016/j.tranpol.2005.07.002>
- Schaller, B. (2010). New York City's congestion pricing experience and implications for road pricing acceptance in the United States. *Transport Policy*, 17(4), 266–273. <https://doi.org/10.1016/j.tranpol.2010.01.013>
- Schwartz, S., Soffian, G., Kim, J. M., & Weinstock, A. (2008). A Comprehensive Transportation Policy for the 21st Century: A Case Study of Congestion Pricing in New York City. *NYU Env'tl. LJ*, 17, 580.
- Shalma, M. (2025). Congestion pricing threatens small businesses, events and communities. *Op-Ed*. <https://www.amny.com/oped/op-ed-congestion-pricing-threatens-small-business/>
- Siena College Research Institute. (2024). *New York State poll: New Yorkers overwhelmingly oppose congestion pricing plan*. <https://sri.siena.edu>

- Simeonova, E., Currie, J., Nilsson, P., & Walker, R. (2021). Congestion pricing, air pollution, and children's health. *Journal of Human Resources*, 56(4), 971–996.
- Small, K. A. (2012). Valuation of travel time. *Economics of Transportation*, 1(1–2), 2–14.
- Small, K. A., & Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, 105–130.
- Standen, C., Greaves, S., Collins, A. T., Crane, M., & Rissel, C. (2019). The value of slow travel: Economic appraisal of cycling projects using the logsum measure of consumer surplus. *Transportation Research Part A: Policy and Practice*, 123, 255–268. <https://doi.org/10.1016/j.tra.2018.10.015>
- Tarduno, M. (2022). For whom the bridge tolls: Congestion, air pollution, and second-best road pricing. *Unpublished Manuscript*.
- Tonne, C., Beevers, S., Armstrong, B., Kelly, F., & Wilkinson, P. (2008). Air pollution and mortality benefits of the London Congestion Charge: Spatial and socioeconomic inequalities. *Occupational and Environmental Medicine*, 65(9), 620–627.
- Varian, H. R., & Varian, H. R. (1992). *Microeconomic analysis* (Vol. 3). Norton New York.
- Vij, A., & Walker, J. L. (2016). *Latent Class Choice Models with Feedback through Consumer Surplus*. Citeseer.
- West, J., & Börjesson, M. (2020). The Gothenburg congestion charges: Cost–benefit analysis and distribution effects. *Transportation*, 47(1), 145–174.
- Wongpiromsarn, T., Xiao, N., You, K., Sim, K., Xie, L., Frazzoli, E., & Rus, D. (2012). Road pricing for spreading peak travel: Modeling and design. *arXiv Preprint arXiv:1208.4589*.
- Young, M. (2018). OpenTripPlanner-creating and querying your own multi-modal route planner. URL: <https://Github.Com/Marcusyoung/Otp-Tutorial/Raw/Master/Introotp.Pdf>.