Effective and Equitable Congestion Pricing: New York City and Beyond

Michael Ostrovsky
ostrovsky@stanford.edu

Frank Yang*
franky@uchicago.edu

August 2024
Preliminary and incomplete

Abstract

In this paper, we argue that the New York City congestion pricing plan whose implementation was paused in the summer of 2024 had a major shortcoming: as designed, it would have had a much more severe impact on the drivers of personal vehicles than on the passengers of taxis and ride-hailing vehicles, as well as on the clients of delivery services. In addition to being inequitable, this shortcoming would have likely made the congestion pricing scheme ineffective at solving the traffic congestion problem, due to the fact that the drivers of personal vehicles constitute a minority of traffic in the congestion pricing zone. We propose a simple modification to the scheme that addresses this shortcoming and discuss its properties.

1 Introduction

On June 30, 2024, the City of New York was planning to launch its congestion pricing system, following a series of earlier failed initiatives, decades of political and legal debates, and hundreds of millions of dollars of investment in supporting infrastructure. Just weeks before that anticipated historic event, on June 5, 2024, New York State governor Kathy Hochul announced an "indefinite pause" to the plan[1], to the dismay of its supporters who viewed this as a major setback not just for New York City itself, but also for numerous other regions suffering from traffic congestion. In the words of a former executive of New York's Metropolitan Transportation Authority (MTA), "Kathy Hochul didn't just kill congestion pricing in New York, she killed it across the country for a long time."[2]

*Ostrovsky: Stanford University and NBER. Yang: University of Chicago. We thank the National Science Foundation (grant #1824317) and the Business, Government & Society Initiative at the Stanford Graduate School of Business for financial support. We are grateful to seminar participants at Stanford, NYU, and the Bay Area IOFest for helpful comments and suggestions.

[1] https://www.governor.ny.gov/news/what-they-are-saying-governor-hochul-announces-pause-congestion-pricing-address-rising-cost

[2] https://www.nytimes.com/2024/06/16/nyregion/congestion-pricing-delay-effects.html

In this paper, we argue that this pause to the implementation of congestion pricing in New York City may have been a blessing in disguise, because the delay makes it possible to address a major shortcoming of the plan's design. Moreover, to the extent that the congestion pricing plan implemented in New York City will serve as an example and a blueprint for future congestion pricing plans in other cities in the United States and the rest of the world, it is especially important to "get it right."

The core parts of New York's paused plan were as follows. Cars entering the Central Business District (CBD, the part of Manhattan south of the 60th street) during peak hours (5am–9pm on weekdays and 9am–9pm on weekends) would be charged $15 for crossing into the congestion pricing zone, once per day. Larger vehicles (excluding transit or commuter services) would be charged $24–$36, depending on their size. Toll rates during overnight off-peak hours would be reduced by 75%. Passengers of taxis and app-based for-hire vehicles (FHVs) would be exempted from the $15 charge, and would instead pay an additional surcharge of $1.25 for taxis and $2.50 for FHVs, 24 hours a day, on any trip to, from, or within the CBD. As explained in the recommendation by the Traffic Mobility Review Board (2023), on which New York's congestion pricing plan was based, "The recommended tolls of $1.25 for yellow taxis, green cabs and black cars, and $2.50 for app-based FHVs, were determined by dividing the daily base auto toll rate ($15) by the average number of trips that taxis and FHVs make in the CBD today, which are 12 and 6, respectively" (p. 23).

As we argue in this paper, the key shortcoming of this plan is that as designed, it charges different types of travelers dramatically different amounts per "unit" of congestion that they generate. Commuters who drive into the CBD in the morning, work during the day, and drive home in the evening in effect pay $7.50 per trip ($15 daily fee spread over two trips). Taxi passengers pay only $1.25 per trip—one sixth of that amount—even if one ignores the fraction of the time that taxis create congestion when driving empty. And an Amazon, FreshDirect, UPS, or Fedex delivery vehicle that might enter the CBD before 5am to avoid paying the full peak-hour toll would pay the off-peak rate of $6 (25% of the $24 regular rate for package delivery trucks) that would cover a full day of driving around CBD delivering packages, inducing vastly more congestion than one commuter or taxi trip.[3]

These dramatic differences between "per-unit-of-congestion" prices for different types of travelers are undesirable for two separate reasons. First, they are patently inequitable. It is not clear why a nurse from Queens driving to work in Manhattan should pay six times as much as a business traveler taking a taxi there from La Guardia, or one or two orders of magnitude more than a customer having a truck drive to his or her house to have an Amazon package delivered. The basic economic logic of Pigouvian taxation says that these prices should be the same (at least if they cover similar distances over similarly congested areas in Manhattan), and while in principle there may be other reasons for price discrimination or redistribution based on various group character-

[3]A typical delivery driver makes 100–200 stops per day (https://www.aboutamazon.com/news/transportation/photos-day-in-the-life-amazon-delivery-driver, https://www.businessinsider.com/im-a-ups-delivery-driver-what-my-job-is-like-2022-2, https://www.quora.com/How-many-stops-does-a-UPS-DHL-FedEx-delivery-truck-make-daily-in-NYC-With-how-many-packages-avg).

istics (e.g., income levels), it is not at all clear that they would align directionally with the price differences above, and even with such considerations, the magnitudes of suitable "per-unit" price differences are unlikely to be so dramatic.

Second, and perhaps more critically, such a toll scheme is unlikely to be effective at solving the traffic congestion problem. The key observation behind this claim is that the private vehicles that would be substantially affected by the scheme constitute only around one third of all vehicles traveling in the CBD. Traffic Mobility Review Board (2023, p. 11) reports that private autos and motorcycles constitute are responsible for only 35% of traffic, while taxis and FHVs are responsible for 52%, with the rest being split approximately equally among buses, trucks, and commercial vans. Similarly, a recent study conducted by the former Commissioner of NYC's Department of Transportation Lucius Riccio estimates that of the 2,000 vehicles recorded in randomly selected locations in Midtown New York City, only 32.7% were personal cars, while 50.5% were taxis and FHVs.[4] So even if the scheme is effective at reducing the number of private vehicles in Manhattan, it will barely affect the remaining two thirds of the traffic, and is thus unlikely to meaningfully relieve traffic congestion. This observation is consistent with the experience of London's congestion pricing scheme, which exempts taxis from the toll and charges FHVs only once per day, thus making a "per unit of congestion" toll for such vehicles minimal and ineffective.[5] Despite raising the daily fee on private vehicles from £5 at the scheme's introduction in 2003 to its current rate of £15, the city still suffers from heavy traffic congestion. E.g., a recent study by the navigation company TomTom has awarded London the dubious honor of being the most congested city in the world (out of 387 for which TomTom has collected traffic data).[6] Of course, it is possible that without London's congestion pricing scheme, traffic there would have been even worse, but it is nevertheless safe to say that the scheme has failed to solve London's traffic congestion woes.

So is there a solution? To a properly trained economist, the answer appears immediate and long-established: Pigouvian taxation, going back to Pigou (1920). The regulator should simply charge each vehicle for the externality it imposes on others, and the problem is solved. However, a moment's reflection makes it clear that such a scheme, taken literally, is too complex and impractical. First, the externality varies by the specific location in the CBD, the specific time and day of the week, weather conditions, traffic conditions on the neighboring roads, traffic accidents and special events, vehicle size, and so on. This leads to a massive, highly multidimensional space of prices, that is both impossible to compute and hard to communicate to the drivers. Second, imposing such a pricing scheme would require tracking all cars in real time, which raises potential privacy issues and imposes additional technological costs. While not insurmountable on its own, this issue adds to the political costs of congestion pricing that are already very large.

In this paper, we propose a congestion pricing scheme that is equitable (different vehicle types pay tolls commensurate with the externalities that they impose) and practical (technologically,

[4] https://www.gothamgazette.com/city/11828-congestion-pricing-for-hire-vehicles-midtown-traffic

[5] https://tfl.gov.uk/modes/driving/congestion-charge/discounts-and-exemptions

[6] https://www.tomtom.com/newsroom/explainers-and-insights/tomtom-traffic-index-2024-london-is-slowest/

it is only a minor deviation from New York's "paused" congestion pricing plan, and it can be launched quickly). The scheme can be viewed as an approximation of the idealized Pigouvian pricing discussed in the previous paragraph, and is based on two observations. The first observation, which we already mentioned, is that personal cars are responsible for only around a third of traffic in the CBD, while taxis, FHVs, delivery trucks, and other commercial vehicles are responsible for the remaining two thirds. The second observation is that driving patterns are very different between typical personal and commercial vehicles. A typical commuter in a personal vehicle drives to (and through) the CBD in the morning, spends the day at his or her workplace, and then drives back home in the evening: "All happy families are alike." By contrast, different commercial vehicles have dramatically different driving patterns inside the CBD, with some circling throughout the CBD for much of the day (e.g., delivery vehicles) and others driving through it a few times per day (e.g., Uber drivers) on trips of varying lengths: "Each unhappy family is unhappy in its own way."

These observations lead us to the following proposed congesting pricing scheme. For personal vehicles, the scheme is the same as the "paused" plan: such vehicles will pay a certain daily amount for entering the CBD (as discussed below, given the other modifications in the plan, this amount can be lower than in the "paused" plan while achieving similar revenue and traffic reduction goals). Given their minority share in the overall traffic volume and the relative similarity of their driving patterns, this is a reasonable approximation of the "idealized" Pigouvian taxation, and has the additional practical benefits of being very simple to communicate and not requiring tracking the vehicles throughout the CBD. For commercial vehicles, however, the scheme is different. Given the dramatic differences in their driving patterns, the scheme charges these vehicles on the per-mile basis, during peak times. In Section 2, we discuss specific calculations for the per-day tolls for personal vehicles and the per-mile tolls for commercial vehicles, based on the principle that on average, the different types of vehicles should pay comparable amounts per unit of congestion that they create. Note that introducing per-mile tolls is much easier for commercial vehicles than for personal ones. FHVs and taxis already track mileage on passenger trips, so no additional infrastructure is required. (We discuss the issue of "empty miles," when these vehicles travel without passengers, in Section 3.) For delivery vehicles (Amazon, Fedex, UPS, etc.), their movements are likewise tracked in real time by the companies (e.g., Amazon, Fedex, and UPS have customer apps that display real-time locations of their deliveries), and thus the information can be easily collected. The remainder of commercial vehicles is only a small fraction of the overall traffic, and can be required to either install a tracking device and pay per mile or be folded under the "pay-per-day" case without substantially affecting the overall congestion levels.

The "paused" NYC congestion pricing scheme does impose fees on taxis and FHVs, but they are very different from the ones in the scheme that we propose. First, as already discussed above, they are much smaller in magnitude than what we propose (roughly speaking, following the principle of "same congestion fee for the same amount of congestion created," an average taxi or FHV trip should pay a half of the daily fee paid by a commuter, i.e., \$7.50 per trip rather than \$1.25 or \$2.50, although as we discuss below, the number should actually be higher due to "empty miles" that

taxis and FHVs drive). Second, and more subtly, under our proposal the fees are per mile rather than per trip: a congestion fee for a short trip of a few blocks is substantially lower than a long trip across the CBD. Third, unlike the paused proposal, ours would charge these congestion fees only during the peak travel times. There is no need to charge a congestion fee to a passenger taking an Uber from a bar at 1am. By contrast, the paused proposal would charge the same congestion fees to FHVs and taxis throughout the day.[7]

The last part of our proposal does not directly affect economic incentives "on the margin," but is instead designed to alleviate the impact of the new congestion fee on the values of the medallions, which would otherwise be negatively affected. This reduction in medallion values is an important concern, and a commonly voiced suggestion to address it is to exempt taxis from the congestion fee. Of course, a major problem with such an exemption is that since taxis are a substantial part of traffic in the CBD, it would severely lower the effectiveness of the overall congestion scheme (in particular, the number of taxi trips in the CBD would not just remain the same, but would likely rise, as passengers would substitute in higher numbers from FHVs to taxis; see, e.g., Leccese (2024) for evidence of this effect after such an "asymmetric" exemption in Chicago). Our proposal is to instead compensate medallion holders in a lump-sum fashion (e.g., a fixed per-medallion payment per year for a certain number of years), with the amount commensurate with the estimated negative impact on the value of the medallion. Crucially, this approach allows regulators to make medallion holders "whole" without negatively impacting the overall effectiveness of the congestion pricing scheme. And to maintain the overall financial soundness of the congestion pricing plan, this compensation can be taken from just the overall congestion pricing fees collected from the taxi rides. We provide further details in Section 2.

2 Quantifying the Alternatives

In this section, we discuss quantitative estimates of the effects of our alternative plan and compare them to those of the baseline "paused" plan. Our model is deliberately streamlined, to the very basics, to make our calculations as transparent as possible. In particular, we focus on just two parts of traffic, personal vehicles and taxis/FHVs, leaving aside commercial vans and trucks of various sizes (incorporating such vehicles would only strengthen our conclusions). For our quantitative inputs we closely follow the assumptions made in the Balanced Transportation Analyzer,[8] which was the basis behind the official traffic and revenue estimates for the paused plan. While the numbers we compute do of course depend on the specific assumptions and parameter estimates, the analysis is so transparent and the magnitudes of the effects are so large that it is clear that the qualitative conclusions are robust to changes in these assumptions.

[7] Note that taxis and FHVs already pay certain per-trip congestion fees for trips south of the 96th street. Such fees may also be in principle replaced by a per-mile, time-varying fee. Our current analysis focuses only on the modifications to the most recent, paused proposal, but would be straightforward to extend to the other congestion fees.

[8] `https://nurturenature.org/pages/balanced-transportation-analyzer`.

2.1 Paused Plan

We start out by calculating the estimates for the paused plan.

There are approximately 400,000 personal cars entering the CBD on a typical day (slightly more during the workweek, slightly less on during weekends and holidays, though we will ignore these differences). The elasticity of demand for driving into the city is estimated at -0.425 (as a weighted average of elasticities from various subgroups, though again, for our purposes, we will ignore these distinctions and will focus just on the first-order averages). The average base price of driving into the city (fuel, wear and tear of the car, parking) is \$30. Finally, because of various credits and reductions (e.g., discount for using a tolled bridge or tunnel; exemptions for some groups; lower toll during off-peak hours), the effective average toll is assumed to be two thirds of the nominal one (i.e., \$10 when the nominal toll is \$15).

Assuming constant elasticity of demand, with the toll, we get the average predicted daily traffic from personal vehicles equal to

$$400,000e^{-0.425\frac{10}{30}} = 347,164,$$

with the corresponding annual toll revenue from personal drivers equal to

$$347,164 \times 365 \times \$10 = \$1,267,149,347.$$

Next, there are 100,000 taxi trips per day that involve the CBD. The elasticity of demand for taxi trips is assumed to be -0.3, and the average price of a taxi trip is \$25. With the per-trip toll of \$1.25 and a similar calculation to the one above, we get the predicted post-toll daily number of taxi trips as

$$100,000e^{-0.3\frac{1.25}{25}} = 98,511,$$

with the corresponding annual toll revenue from taxi passengers equal to

$$98,511 \times 365 \times \$1.25 = \$44,945,732.$$

Finally, there are 300,000 daily FHV trips that involve the CBD. The elasticity of demand for such trips is also assumed to be -0.3, and the average price is \$30. With the per-trip toll of \$2.50, the predicted post-toll daily number of FHV trips is then

$$300,000e^{-0.3\frac{2.50}{30}} = 292,593,$$

with the corresponding annual toll revenue from FHV passengers equal to

$$292,593 \times 365 \times \$2.50 = \$266,991,088.$$[9]

[9]While our model is much more stylized, our predictions are close to those of the BTA model, which estimates the expected annual toll revenues from personal drivers, taxi passengers, and FHV passengers to be \$1,207 billion, \$50 million, and \$220 million, respectively.

2.2 Alternative Plans

As discussed in the Introduction, the paused plan disproportionately affects personal cars relative to the passengers of taxis and FHVs. The calculation in Section 2.1 makes this asymmetry clear: despite comparable aggregate numbers of trips, the drivers of personal cars would pay $1.27 billion in tolls per year, while the passengers of taxis and FHVs would only pay $312 million. The reduction in traffic is also disproportionate: the number of personal car commutes would drop by more than 50 thousand, while the number of taxi and FHV trips would drop by less than 10 thousand per day. So how can the plan be made more equitable? As we discussed in the Introduction, our proposal for modifying the toll structure on taxis and FHVs involves three changes: the higher absolute level, the time-varying structure (high during peak times, zero during off-peak times), and the per-mile rather than per-trip computation of the fee. In this section, we focus just on the first of these three changes, for simplicity, transparency, and immediate comparability with the paused plan. We discuss the other two changes in subsequent sections.

So what should the toll on taxis and FHVs be to have a comparable payment "per unit of congestion generated" to that of personal cars? Note that $15 cordon fee pays for two trips: one trip to CBD in the morning, and one trip from CBD in the evening. So the per-trip fee is $7.50. However, to be comparable on the "per unit of congestion" basis, the fee on taxis and FHVs needs to be higher than that, because in addition to the trips for which they are paid (and on which the congestion fee is imposed), taxis and FHVs also drive a fraction of the time empty—and that needs to be taken into account. Using the estimate of "empty miles" of 0.35 (i.e., a typical taxi in the CBD is empty 35% of the time), we find that to be commensurate with the personal car cordon fee, the per-paid-trip congestion fee should be equal to $\frac{\$7.50}{.65} \approx \$11.50$. We call this congestion pricing scheme "Plan 1."

With this toll level, the predicted daily number of taxi trips drops to

$$100,000e^{-0.3\frac{11.50}{25}} = 87,110,$$

while the corresponding annual toll revenue from taxi passengers increases to

$$87,110 \times 365 \times \$11.50 = \$365,643,676.$$

The number of FHV trips drops to

$$300,000e^{-0.3\frac{11.50}{30}} = 267,410,$$

while the corresponding annual toll revenue from FHV passengers increases to

$$267,410 \times 365 \times \$11.50 = \$1,122,452,817.$$

With the per-trip toll of $11.50, both the overall toll revenue and the overall trip reduction

for taxis and FHVs are commensurate with those of personal cars: both types of vehicles pay comparable tolls "per unit of congestion." However, this is only a part of the story. The second observation is that, unsurprisingly, the overall toll revenue grows dramatically, by more than $1 billion per year. This means that if tolls and taxis and FHVs are set at the level commensurate with that of personal cars, the latter *can be substantially reduced* while the overall congestion pricing plan's revenue goals will continue to be met. Our second alternative plan (which we call "Plan 2") illustrates this possibility.

Consider setting the cordon toll for personal cars at $9 (instead of $15 – a 40% reduction). It translates to a $4.50 toll per trip, and, using the same logic as in the previous plan, the commensurate toll on FHVs and taxis would be equal to $\frac{\$4.50}{.65} \approx \$7$. Repeating the above calculations with the cordon fee of $9 and the FHV/taxi per-trip fee of $7, we get

Number of personal cars	$400,000e^{-0.425\frac{6}{30}}$	=	367,405,
Revenue from personal cars	$367,405 \times 365 \times \$6$	=	$804,616,761,
Number of taxi trips	$100,000e^{-0.3\frac{7}{25}}$	=	91,943,
Revenue from taxi passengers	$91,943 \times 365 \times \$7$	=	$234,914,686,
Number of FHV trips	$300,000e^{-0.3\frac{7}{30}}$	=	279,718,
Revenue from FHV trips	$279,718 \times 365 \times \$7$	=	$714,679,863.

Note that compared to the paused plan, Plan 2 reduces the overall number of trips by a similar amount; charges a lot less to the commuters ($9 vs. $15); and raises substantially higher toll revenue ($1.754 billion vs. $1.579 billion). This makes it possible to meet the same congestion pricing revenue goals as what the paused plan would achieve, *and* compensate the owners of taxicab medallions for the reduction in their value due to the introduction of congestion pricing (with 13,500 medallions, the extra revenue would be sufficient to compensate each one up to $\frac{\$1.754B - \$1.579B}{13500} \approx \$13,000$ per year, which, annualized, would be much higher than the total current value of a medallion, let alone the reduction in its value. The regulators could also lower other registration fees for taxicub drivers. Crucially, these would be fixed-type fees that would not affect the taxi drivers' marginal incentives to drive in CBD during congested times.

3 Additional Details

Plans 1 and 2 discussed in Section 2 are most directly comparable to the paused plan, and the comparison presented in that section is very transparent.

However, there are additional improvements and modifications that can be made to these plans.

First, following the design of the paused plan, Plans 1 and 2 charge the congestion pricing fee to taxis and FHVs 24 hours per day. When the level of those fees is low, like in the paused plan, this does not introduce a big distortion during off-peak hours. However, once the fees are set at the

level of $7 or $11.50, the distortion (relative to the ideal Pigouvian taxes, which are close to zero when there is no congestion) becomes substantial. Moreover, it violates the "similar payment for similar amount of congestion principle" when compared with personal car drivers, who pay 75% less when they drive into the CBD during off-peak hours. Thus, one improvement to Plans 1 and 2 is to only charge the high rates during peak hours, and then during off-peak hours, either charge zero, or apply the same 75% reduction factor as the drivers of personal cars get.

Second, when the level of fees is low, they do not introduce a big distortion to demand for short vs. long trips. However, once the level of fees becomes substantial, the relative effects become very different. An $11.50 surcharge on a short trip that would otherwise cost $15.00 has a very different effect from the same surcharge applied to a long, $45.00 trip. So this pricing structure not only violates the "similar payment for similar amount of congestion" principle, but also skews the overall passenger trip composition toward those who take long trips—and create more congestion! So another improvement to the plans is to replace the per-trip fee on taxis and FHVs with the commensurate (on average) per-mile fee, that much more accurately reflects the amount of congestion caused by such trips.

4 Conclusion

Congestion pricing has the potential to improve traffic and generate revenue in New York City and many other congested cities worldwide. However, to be equitable and effective, it needs to follow the "similar payment for similar amount of congestion" principle. We discuss how to modified New York's paused congestion pricing plan to closer follow this principle, and show that such a change would make it possible to substantially lower the cordon fee paid by the commuters while achieving the paused plan's revenue and traffic reduction goals, and generating additional revenue that can be used to alleviate the effects of congestion pricing on taxicab medallion owners.

References

Leccese, M. (2024). Asymmetric taxation, pass-through and market competition: Evidence from ride-sharing and taxis. University of Maryland Working Paper, available at `https://ssrn.com/abstract=3824453`.

Pigou, A. C. (1920). *The Economics of Welfare*. Macmillan.

Traffic Mobility Review Board (2023). Congestion pricing in New York: A toll structure recommendation from the Traffic Mobility Review Board. Available at `https://new.mta.info/document/127761`.