International Transport Forum

High-occupancy Toll Lanes
Their Distributional Impact and Effect on Congestion
Discussion Paper

183 Roundtable

Jonathan Hall
University of Toronto

High-occupancy Toll Lanes
Their Distributional Impact and Effect on Congestion
Discussion Paper

183

Roundtable

Jonathan Hall
University of Toronto

The International Transport Forum

The International Transport Forum is an intergovernmental organisation with 62 member countries. It acts as a think tank for transport policy and organises the Annual Summit of transport ministers. ITF is the only global body that covers all transport modes. The ITF is politically autonomous and administratively integrated with the OECD.

The ITF works for transport policies that improve peoples' lives. Our mission is to foster a deeper understanding of the role of transport in economic growth, environmental sustainability and social inclusion and to raise the public profile of transport policy.

The ITF organises global dialogue for better transport. We act as a platform for discussion and pre-negotiation of policy issues across all transport modes. We analyse trends, share knowledge and promote exchange among transport decision-makers and civil society. The ITF's Annual Summit is the world's largest gathering of transport ministers and the leading global platform for dialogue on transport policy.

The Members of the Forum are: Albania, Armenia, Argentina, Australia, Austria, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Canada, Chile, China (People's Republic of), Croatia, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, India, Ireland, Israel, Italy, Japan, Kazakhstan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Mexico, Republic of Moldova, Mongolia, Montenegro, Morocco, the Netherlands, New Zealand, North Macedonia, Norway, Poland, Portugal, Romania, Russian Federation, Serbia, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Tunisia, Turkey, Ukraine, the United Arab Emirates, the United Kingdom, the United States and Uzbekistan.

International Transport Forum
2 rue André Pascal
F-75775 Paris Cedex 16
contact@itf-oecd.org
www.itf-oecd.org

ITF Discussion Papers

Cite this work as: Hall, J. (2020), "High-occupancy Toll Lanes: Their Distributional Impact and Effect on Congestion", *International Transport Forum Discussion Papers*, No. 2020/07, OECD Publishing, Paris.

Acknowledgements

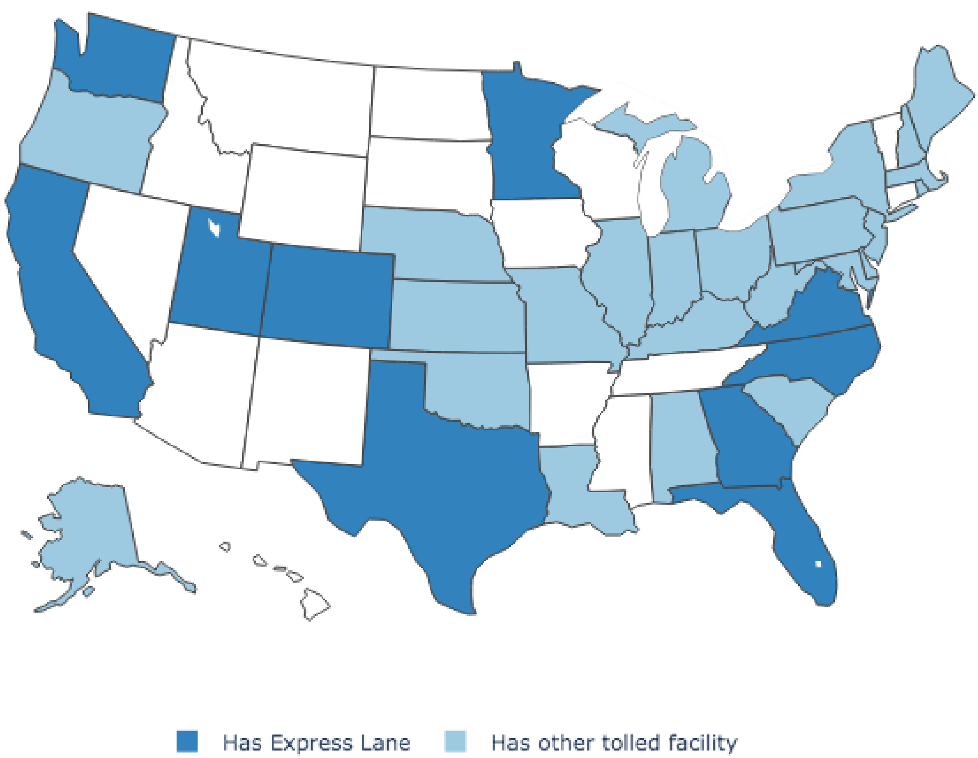
The author thanks Arsalan Esmaeili, Jonathan Blumenthal and Catherine Luongo for help gathering data and constructing tables and graphs.

Table of contents

Introduction

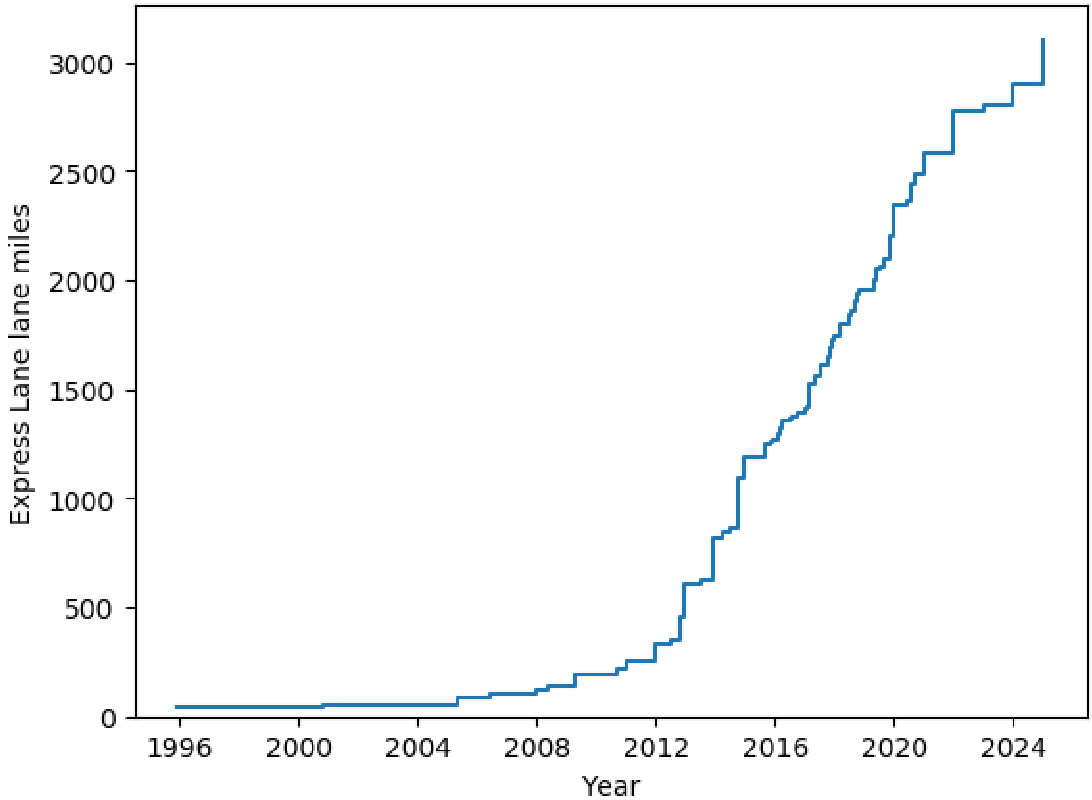
Traffic congestion is a serious problem in major cities worldwide and there is growing interest in using tolling to address it. Well-designed tolls can help reduce congestion to socially optimal levels because they help drivers internalise the cost they impose on others. However, historically, tolls in the United States were designed to recoup the costs of building and maintaining the tolled road, bridge, or tunnel, rather than to manage congestion. This started to change in 1995 when the California State Route 91 (SR-91) Express Lanes opened. While the primary motivation for tolling the lanes was that the state did not have the money to expand SR-91, they also implemented time-varying tolls designed to keep the lanes free from congestion. Since that time the practice of tolling a portion of the lanes of a highway, while leaving other lanes unpriced has grown. This practice is often called value pricing, and the lanes are called Express Lanes or, when carpools can use them free or at a discount, high-occupancy toll (HOT) lanes. These new implementations are expressly designed to improve congestion and provide travellers with a congestion-free option. Currently, ten states have Express Lanes, as shown in Figure 1, and the extent of these lanes continues to increase, with over 2 500 lane miles of Express Lanes currently operational and over 500 additional lane miles scheduled to open in the next few years, as shown in Figure 2.

Figure 1. States with Express Lanes and tolled facilities



Furthermore, the technological implementation of Express Lanes continues to improve. While the SR-91 Express Lanes update their toll schedule every six months, newer Express Lanes update their tolls every six minutes, or less, to ensure the lanes maintain high travel speeds.

Figure 2. Growth in Express Lanes over time



The innovation of Express Lanes is important because while a major barrier to introducing tolls is the widespread belief that it hurts many, if not most, travellers, carefully designed Express Lanes can help most, if not all, travellers. This paper reports the theoretical and empirical evidence on the distributional effects of Express Lanes. It also reports the evidence regarding a key determinant of their distributional impact: how they affect congestion, both in the Express Lanes themselves and in the parallel general-purpose lanes. Finally, to help put Express Lanes in context, this paper discusses the merits of alternative approaches to managing congestion such as ramp metering, permits, and vehicle kilometres travelled charges.

The theoretical argument for why Express Lanes can help most, if not all, travellers has four components. First, only pricing a portion of the lanes preserves the ability of travellers to continue to choose to not pay a toll. This greatly reduces the harm done relative to pricing all the lanes. Second, if the toll is designed to maximise throughput (vehicles per hour), it is possible for tolling to increase the throughput in the Express Lanes, and thereby improve congestion in the parallel general-purpose lanes. If so, then tolling has helped all travellers, even before using the toll revenue. Achieving this requires the toll be time-varying. Third, the toll revenue can be used in ways to help improve the distributional impacts. Most directly, it can be used to pay for the Express Lanes. If the Express Lanes are adding new capacity, then they will improve travel times in the general-purpose lanes, and thereby help everyone. The revenue can also be used to improve public transportation or used to reduce regressive taxes, both of which help improve the distributional impacts. Fourth, Express Lanes provide a reliable and fast option for those times travellers who need it. While most people do not use them daily, it is very valuable to be able to use them on those days when it is essential to be on time.

Empirically, users of Express Lanes come from all income classes and demographic groups, though those with higher incomes use them more often. Furthermore, surveys typically show that most people in all income classes support Express Lanes. Both facts imply the distributional impact of Express Lanes, as typically implemented, is significantly less regressive than typically assumed.

Theory of distributional impacts of tolling

This section provides a theoretical background for understanding the distributional impacts of tolling. The argument for tolling starts with recognising that adding an additional vehicle to the road slows down all the other vehicles on the road. When travellers are deciding whether to drive, they do not account for how driving will slow down others; and as a result, travellers perceive the cost of driving as being lower than it truly is. Because of this, too many travellers decide to drive rather than either not taking the trip, travelling at a different time, or travelling via a different mode. A properly designed toll can charge drivers for the cost they impose on others, and in so doing, lead travellers to make the socially efficient choice regarding whether to drive.

As adding a toll raises the financial costs of travelling but also improves congestion, whether this helps or hurts a given traveller depends on how the value of the time savings compares to the cost of the toll. If the average value of time is USD 14.10 per hour (US Department of Transportation, 2016), then a USD 5 toll helps the average traveller if it saves her 22 minutes or more, and hurts if it saves less. As travellers differ in their value of time, the same toll can help some travellers and hurt others. Thus, a toll is more likely to hurt travellers with low values of time and to help drivers with high values of time.

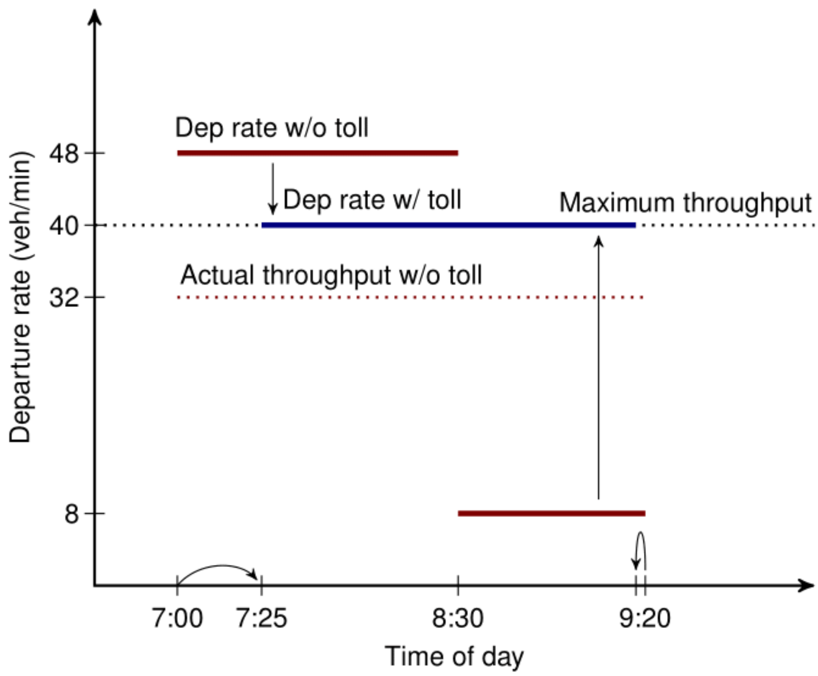
However, the distributional impacts also depend on how the toll is designed, how tolling affects throughput, and whether travellers have alternatives to paying the toll. Additionally, the distributional impacts are better than typically assumed because travellers value having an option that is reliable and because, in the case of pricing a portion of the lanes, travellers value having the ability to pay to save a few minutes when it really matters. Finally, the revenue can be used to improve the distributional impacts of tolling.

Distributional impacts depend on how tolling affects throughput

The distributional impacts of tolling depend crucially on how tolling affects throughput. For the sake of showing this, this subsection assumes all travellers have the same value of time, though they can differ in how they value taking a trip by automobile. Also note that, holding fixed vehicle occupancy, throughput is proportional to the number of people travelling. The demand for automobile trips is a decreasing function of the total cost of an automobile trip, which itself is the sum of the toll and the time cost. Assuming all travellers have the same value of time means that tolling changes all travellers costs by the same amount. Since the demand for automobile trips is a decreasing function of total cost, if tolling reduces the equilibrium number of travellers, then it must have increased the cost of travelling. Likewise, if tolling increases the equilibrium number of travellers, then it must have decreased the cost of travelling.

This naturally raises the question of how tolling affects throughput. It is long been assumed that tolling should reduce throughput (e.g., Knight, 1924); however, scholars began to appreciate the dynamic nature of traffic congestion and that tolling could reduce congestion by rearranging when people travel, potentially leaving throughput unchanged (Arnott, de Palma and Lindsey, 1993; e.g., Vickrey, 1969). Recently, Hall (2018, forthcoming) has argued that it is theoretically possible for tolling to increase throughput.

Figure 3. Tolls can smooth the departure rate, preventing queuing, and increasing throughput



The argument that it is possible for tolling to increase throughput builds on research by transportation engineers showing that not only do additional vehicles slow others down but, in heavy enough traffic, they cause frictions that reduce throughput by 10-25%.[1] There are two causal mechanisms by which additional vehicles can create frictions that reduce throughput. The first causal mechanism is that queues behind a bottleneck can grow long enough that they block other traffic. For example, when a queue forms at a highway exit, it can spill onto the mainline of the highway and block through traffic, or a queue on the highway can block other exits. This is the reason ring roads, such as the Boulevard Périphérique, which encircles Paris, are so prone to terrible congestion, as on such roads it is possible for all drivers to simultaneously be in each other's way (Daganzo, 1996; Vickrey, 1969). This same problem can occur in city centres, where roads can become jammed as drivers are in each other's way, reducing throughput (Geroliminis and Daganzo, 2008). "Blocking the box" is the most direct way this happens in cities.

The second causal mechanism by which an additional vehicle can reduce throughput is from destructive lane changes. When a highway lane is ending, or at the typical highway on-ramp, vehicles in the lane that is ending need to merge into the other lanes. When traffic is heavy, doing so is challenging, and at some point there will be a vehicle that forces its way over rather than waiting for a gap. This slows down the traffic, reducing throughput. There is a large transportation engineering literature showing that throughput falls at bottlenecks once a queue forms, though recent evidence questions whether this is causal (Anderson and Davis, 2020).

If either of these two causal mechanisms exists, then tolling can increase throughput by smoothing the rate of departures, decreasing it at the start of the peak period and increasing it at the end of the peak period, so that the average departure rate is higher than it was before.[2] Figure 3 shows a numerical example of how this can happen. When there is no toll, the first departure occurs at 7:00 a.m., and the departure rate is 48 vehicles per minute. However, if the highway's maximum capacity is only 40 vehicles per minute, a queue will form. Because of this queue, the two frictions described above reduce throughput to 32 vehicles per minute. Between 7:00 and 8:30, the departure rate is greater than throughput, and so the queue grows longer, and so travel times likewise grow longer. Starting at 8:30, the departure rate falls to eight vehicles per minute, and the queue shortens and travel times fall. By 9:20, everyone has departed and the peak period draws to a close. This can be an equilibrium as drivers choose between departing early (or late) in exchange for shorter travel times, or departing to arrive on-time but facing long travel times.

An optimal time-varying toll will reduce the departure rate at the start of the peak period while increasing it at the end. By smoothing the departure rate, the toll prevents a queue from forming and so prevents the frictions from reducing throughput. Since throughput is higher, the peak period is shorter.

This simple illustrative example shows how considering the dynamics of traffic matters, how smoothing drivers' departure rate can increase the average departure rate, and gives an alternative method of showing how the impact of tolling on throughput affects the distributional impact of tolling. In this example, adding a toll allowed the first driver to delay her departure by 25 minutes. As the first driver to depart never faces any congestion nor pays a toll, this means that tolling made her better off. If, on the other hand, tolling had reduced throughput, then the peak period would become longer, requiring this driver to leave earlier and making her worse off. Were all travellers identical, the effect would be the same for all travellers, either helping or hurting them all.

While travellers are not identical, this subsection has shown that how tolling affects travellers depends on how tolling affects throughput. If tolling increases throughput, then it has more beneficial impacts than if it reduces throughput.

Distributional impacts depend on whether travellers have an alternative to paying the toll

The distributional impacts of tolling also depend on whether travellers have an alternative to paying the toll. Since travellers differ in their value of time, some travellers prefer to pay a toll in exchange for faster travel times while other travellers prefer not to do so.[3] Travellers can reduce or eliminate the need to pay a toll by changing to an alternative mode (such as public transportation), time of departure, or destination; indeed, the social welfare gains from tolling come from travellers making socially optimal choices of all these decisions; however, there is likely to be at least some travellers with a low value of time who do not have the option of substituting to a different mode, time, or destination. These travellers will be significantly hurt if they must pay a toll to continue making the same choice as before the toll was implemented. By providing travellers with a good alternative to paying a toll, it is possible to improve the distributional impacts of tolling.

In the case of tolling highways, it is possible to provide an alternative to paying the toll by leaving some of the lanes untolled. When doing so, the tolled lanes are often called "Express Lanes" or "HOT Lanes" and the untolled lanes are called "General Purpose Lanes". Doing so preserves travellers' ability to choose not to pay a toll in exchange for faster travel times.

That said, if adding tolls to some of the lanes makes congestion worse in the remaining lanes, this will still hurt many travellers. However, as Hall (2018) explains, if tolls can increase throughput, then it is possible for tolling some lanes to increase speeds in the adjacent free lanes. If this happens, then the tolls help all travellers, even before using the toll revenue. The intuition for why is as follows. Imagine a two-lane highway. Before any toll is introduced, each lane carries roughly half of the vehicles, and both lanes are congested and have long travel times. Due to the congestion, the frictions described above reduce capacity in both lanes. If a time-varying toll is added to one lane, and set at a level that smooths the rate at which travellers get in the lane so that the frictions are eliminated and throughput increased, then the toll lane is able to carry more than half the traffic. Because the toll lane is carrying more than half the traffic, the remaining free lane now moves fewer vehicles than it did before. For all the traditional reasons, this results in faster travel times in the free lane. Because those in the free lanes have faster travel times, and still do not pay a toll, they must be better off than prior to tolling. Because those in the toll lane have the option of using the free lane, but are choosing to use the toll lane, they too must be better off. All travellers are better off, even before the revenue is used.

Hall (2018) shows that this intuition does not universally hold. It is possible that traveller preferences are such that tolling a portion of the lanes still hurts some travellers. Hall (forthcoming) estimates the distribution of traveller preferences and shows that, depending on assumptions, tolling 25—50% of the lanes can help all travellers.

While it is not always possible for tolling to help all travellers, the result that pricing a portion of the lanes greatly improves the distributional impacts is robust. Hall (forthcoming) finds that pricing all the lanes can hurt some travellers significantly. The worst-off traveller is harmed by as much as USD 3 420 per year, while pricing half of the lanes reduces this harm by 75-100%.

Importantly, tolling a portion of the lanes captures a more-than-proportionate share of the available social welfare gains from tolling. This is because, while the travel time saving from pricing a portion of the lanes is essentially proportional to the share of lanes priced, since those with a high value of time choose to use the priced lanes, the *value* of the travel time savings is more than proportional.

Value of reliability improves distributional impacts

Many analyses of the distributional impacts of road tolls are too pessimistic because they ignore the fact that tolling can provide a more reliable travel time option. This includes the analyses in Hall (2018, forthcoming) discussed in the previous subsection. Accounting for reliability is important because the magnitude of the uncertainty in travel times is large. More than half of the time lost to traffic congestion is unpredictable, being due to crashes, bad weather, and other shocks (Dowling et al., 2004; Kwon, Mauch and Varaiya, 2006). This lack of reliability accounts for 30—90% of the total cost of congestion (Bento, Roth and Waxman, 2020; Small, Winston and Yan, 2005).

Hall and Savage (2019) use a dynamic model with endogenous congestion and endogenous reliability to show that accounting for the value of reliability yields more positive estimates of the distribution impacts of tolling. They find it is possible for adding time-varying tolls to help all travellers, despite reducing throughput.

Bento, Roth, and Waxman (2020) use data from the I-10 Express Lanes in Los Angeles, California, to show that travellers' value of reliability accounts for 87% of their willingness-to-pay to use the toll lanes. They show this by documenting that when the time savings from taking the tolled lanes are less than five minutes, then the implied values of time are implausibly large and are over USD 60 per hour. To rationalise these high implied values of time they argue it is important to consider how saving five minutes is exceptionally valuable when those five minutes are the difference between being on-time or late. They call this the "value of urgency." To support this argument, they document how usage of the toll lanes has local maximums shortly before standard start times and has local minimums at the start times, suggesting that people use the lanes to arrive on-time.

It is important to note that what Bento, Roth, and Waxman (2020) measure is broader than the traditional sense of the value of reliability. It is not simply that the ExpressLanes are more reliable than the parallel general-purpose lanes, but that taking the ExpressLanes allows travellers to save the crucial few minutes that allows them to be on-time.

If the reason travellers need to save those few minutes is because of unpredictable traffic conditions prior to the ExpressLanes, then this is just a broader definition of the value of reliability. However, inasmuch as the reason the travellers need to save those few minutes is because of starting their trip later than they should have, then the value of reliability is a distinct concept and suggests that pricing a portion of the lanes may be preferred to pricing all the lanes. If all lanes are tolled, then travellers are unable to pay to make up lost time; but when part of the lanes are tolled, they can plan their trip intending to take the untolled lanes, and then if something goes wrong, switch to the priced lanes.

In addition to simply being another benefit from tolling, there are two reasons accounting for the increased reliability of a tolled route improves the distributional impacts of tolling. First, Hall (2018) shows that those most hurt by adding tolls are those with inflexible schedules and low values of time. Given that travellers with inflexible schedules value reliability the most, this means that accounting for the value of reliability improves the estimated distributional impacts for those travellers most hurt by tolling. Second, travellers' schedule flexibility is not constant, rather, everyone has days when they need to be on time and when something goes wrong, and so having the ability to pay a toll to make up the crucial minutes is valuable to all travellers, not just those who typically have a high value of time.

Toll revenue can be used to improve distributional impacts

The discussion so far has ignored the use of the toll revenue, but the revenue can be used to improve the distributional impacts. In practice, the revenue is used to pay for the construction and operation of the Express Lanes, used to improve public transit along the same corridor, used to provide reduced tolls for low-income drivers, and otherwise targeted to improve the tolled corridor.

The distributional impacts are different depending on whether tolling pays for the construction of new lanes or an existing lane is converted. When the tolls pay for new construction, tolling still helps some travellers more than others, but it will not hurt any travellers.

Empirical evidence on tolling's effect on congestion and throughput

Since the distributional impacts of tolling are strongly affected by how tolling affects congestion and throughput, this section reviews the existing evidence on these questions.

Tolling's effect on congestion in tolled lanes

Appropriately set tolls improve congestion in the tolled lanes. All HOT lanes and Express Lanes have been able to design tolls that vary across time and that update either every few months or every few minutes to keep the traffic moving at high speeds (Goel and Burris, 2012; Liu et al., 2011; Zimmerman et al., 2015).

On the other hand, many older tolled roads and bridges have tolls that are constant and set with the goal of covering the cost of building or reconstructing the facility. These tolls are set at much lower levels and the roads remain congested.

Tolling's effect on throughput in tolled lanes

For determining the distributional impacts of tolling, it is important to know how tolling affects throughput in the tolled lane. Unfortunately, it is not yet known if it is possible for a carefully designed toll to increase throughput. Most HOT lanes and Express Lanes are conversions of pre-existing HOV lanes, and thus see an increase in throughput since those lanes had capacity to spare. Furthermore, in practice, the tolls schedules are often explicitly set to reduce throughput, with the ultimate goal of maintaining high speeds, rather than with the goal of maximising throughput.[4]

Tolling's effect on congestion in parallel general-purpose lanes

In practice, as Table 1 shows, most Express Lanes lead to faster travel times in the parallel general-purpose lanes, as they either add new lanes or allow single-occupancy vehicles to access underutilised HOV lanes. This does imply the existing implementations helped all travellers on those corridors, however, it does not imply that converting a general-purpose lane to a tolled lane would do the same.

Tolling's long-run effect on congestion

Tolling's long-run effect on congestion in the parallel general-purpose lanes is likely to be less than its short-run effect. Over time, travellers will change where they live and work such that congestion levels return approximately to their original levels (Downs, 1962; Duranton and Turner, 2011). Inasmuch as travel times return to their original levels, adding tolls to some of the lanes, even if doing so reduces throughput, does not hurt travellers who continue to use the untolled lanes. Since those in the tolled lanes could have chosen to use the untolled lanes, they must not be hurt either. This implies that, in the long run, tolling will not hurt any travellers regardless of how tolling affects throughput. While there is some debate on the magnitude of the long-run adjustment, there is agreement that the long-run effects are smaller than the short-run effects.

Table 1. Tolling's effect on throughput and speed

Facility	Change in throughput in tolled lane(s)	Change in speed in tolled lane(s)	Change in throughput in GP lanes	Change in speed in GP lanes	Change in total throughput
Atlanta – I-85		2.8%		-4.8%	-4.4%
Los Angeles – I-110		-13.2		0.1%	7.6%
Los Angeles – I-10		5.8%		-8.9%	21.5%
Miami – I-95		220%		240%	26%
Minneapolis – I-35W		-2.2%		9.1%	20.9%
Seattle – SR-167	1—3%	7—8%	3—4%	10%	
California – I-15	48%		-4%		6%
Minneapolis – I-394	9-33%	0%>		2-15%	5%
California – I-680		4%		13%	11—38%

Note: Miami both converted an existing HOV lane to be a HOT lane and added an additional HOT lane.

Sources: Zimmerman et al. (2015), Washington State Department of Transportation (2009), Supernack et al. (2001), Bhatt et al. (2008), Alameda County Transportation Commission (2013).

Empirical evidence on tolling's distributional impact

There is a large literature seeking to understand the distributional impacts of Express Lanes, and here I focus on two points. First, in contrast to the typical theory, low-income drivers regularly use Express Lanes, though at a lower frequency than high-income drivers. Second, low-income drivers typically, though by no means always, support Express Lanes. Table 2 summarises a large number of studies showing this. Both facts imply the distributional impact of Express Lanes, as typically implemented, is significantly better than typically assumed.

Table 2. Evidence on distributional effects of Express Lanes

Facility	Do low-income drivers use the lane?	Do low-income drivers support the lane?
I-110 Express Lanes, California	Yes, 61% of users have an annual household income less than USD 35 000.	Probably not, 54% of survey respondents agreed that the tolls on I-110 are unfair to people on limited incomes.
I-10 Express Lanes, California	Yes, 58% of users have an annual household income less than USD 35 000.	Probably not, 55% of survey respondents agreed that the tolls on I-10 are unfair to people on limited incomes.
I-15 Value Pricing Project, California	Yes, 22% of users have an annual household income less than USD 50 000 .	Yes, over 80% of low-income users (those with a household income less than USD 40 000/year) support the I-15 HOT lane.
I-25 HOV/Tolled Express Lanes, Colorado	Yes, however, fewer than 10% of users have been found to earn less than USD 50 000.	Yes, 45% of drivers earning less than USD 35 000 per year support the lanes, with 33% undecided and 22% opposed.
I-95 Express Lanes, Florida	Yes, however, only 4% of users have an annual household income below USD 25 000, while 87% earn more than USD 76 000 per year.	Probably not, before the implementation 40% of low-income residents actively opposed the project, and 38% indicated little to no support.
I-85 Express Lanes, Georgia	Yes, 26% of HOT drivers have low to very low income.	No, only 8% of those with a household income USD 50 000 per year say the Express Lanes have improved their travel.
I-10 Katy Managed Lanes, Texas	Yes, however only 13% of users earn less than USD 50 000 per year.	
US-290 Northwest Freeway, Texas	Yes, 7% of users have a household income below USD 50 000 per year.	Yes, focus groups held during project planning did not find concerns about social equity among either corridor users or the public at large and no equity concerns have been raised during operations.
SR 167 HOT Lanes Pilot Project in Seattle, Washington	Yes, 16% of users earn less than USD 50 000 per year.	Yes, evaluators found through outreach efforts that low-income drivers are as supportive of the HOT lanes as are drivers from other income groups.
I-394 MnPass in Minneapolis, Minnesota	Yes, 55% of low-income survey respondents reported using the HOT lanes.	Yes, support for the lanes was found to be high across all income levels, including by 64% of low-income respondents.
SR 91 Express Lanes, California	Yes, 19% of users earn less than USD 40 000 per year.	Yes, over half of commuters with household incomes less than USD 25 000 a year approved of providing toll lanes.

Sources: US Department of Transportation (2020, 2017, 2014, 2008), Norman (2002), Toor and Salisbury (2014), Ungemah, Swisher, and Tighe (2005), Barton (2014), Cleland and Winters (2000), Khoeini and Guensler (2014), Petrella et al. (2014), Burris and Hannay (2003), and Burris and Appiah (2004).

Alternative approaches to managing congestion

While adding tolls directly targets the externality travellers impose on others when they choose to drive, there are alternative approaches to managing congestion. This section briefly summarises the benefits and costs of these different approaches.

Before going through several alternative approaches, it is helpful to consider the six different choices that an ideal toll affects: whether to travel, where to travel, when to travel, how to travel, what route to take, and land use.

Changes to parking supply and price

Since most trips end with the vehicle parked at the destination, changes to the supply or price of parking can have, in theory, a large effect on congestion. Furthermore, since these charges can vary with time and location, they can affect nearly all the traveller decisions that an optimal toll would affect. Their only weakness is that they do not impact route choice and are unlikely to be able to be as time-varying as the ideal toll. That said, changing parking supply and price can be easier and cheaper to implement.

San Francisco implemented significant parking reforms that helped drivers find parking spots faster and reduced cruising for parking, but did not find that this improved congestion (Zimmerman et al., 2015).

Ramp metering

Ramp metering is the practice of putting a traffic light on highway on-ramps and letting one or two vehicles enter the highway every few seconds. Ramp metering can increase highway throughput and speed by preventing the highway from getting overcrowded, though travel times (including waiting at the ramp) may not be much better (Levinson and Zhang, 2004; Zhang and Levinson, 2010).

Ramp metering has three weaknesses. First, it still allocates access to the highway using waiting time, which is a social waste, rather than money, which is just a transfer and so not wasted. Second, two vehicles entering at the same on-ramp at the same time pay the same waiting cost, even if one is travelling further than the other, and so imposing a greater cost on other travellers. Third, on-ramps nearer to downtown have longer waits then ramps further out, even though, holding the destination fixed, those further out impose a greater cost on other travellers.

Ramp metering and tolling are complements. Ramp metering is better at smoothing entrance to the highway over a small interval of a few minutes, while tolling is better at smoothing entrance to the highway over larger intervals.

Vehicle permits

Several cities, including Beijing, restrict the number of vehicles in the city by requiring residents to have a permit to own a vehicle, and then limiting the number of permits. Some cities allocate this via a lottery while others use auctions. In either case, reducing the number of vehicles will reduce congestion, though it is a blunt instrument and the social welfare gains may be limited or even negative.

VKM based pricing

Another approach to improving congestion is to charge drivers for the total number of vehicle kilometres travelled in a year, which is similar to raising gas taxes. Both raise the cost of driving and reduce the amount of driving. This is a great approach to having drivers pay for their share of the roads, and gas taxes are a great approach to charging drivers for the external cost of pollution, but neither works well for addressing congestion.

The problem is that the amount a driver slows down others varies wildly across time and space, and so the optimal VKM charge in the downtown area is different than in the suburbs, which is different from in the surrounding countryside. Even in a congested city, the average externality can be quite small (Akbar and Duranton, 2020), and so the optimal VKM charge would be quite small, and not lead to a large improvement in congestion.

Cordon charges

Several cities, including London and Stockholm, charge a toll to enter their downtown core. These cordon charges address downtown congestion while Express Lanes address congestion on highways, and so they serve different purposes. That said, cordon charges have been successful at addressing congestion in the downtown core by up to a third in both Stockholm and London (Eliasson et al., 2009; Leape, 2006) and have had benefits outside the core (Green, Haywood and Navarro, 2016).

GPS-based systems

A GPS-based system would be ideal. Such a system is very flexible and can charge a time-varying toll that differs for every segment of roadway. This would allow for charging drivers perfectly for the marginal external cost they impose on others.

However, such an ideal system of toll may be too complex for travellers to accept, and so something simpler would likely be used.

A GPS-based system has a higher fixed cost than a gantry-based system since every vehicle needs to get the device and pay for the cost of a data plan that allows the device to communicate with the tolling authority, while a gantry-based system has a lower cost per vehicle but a higher cost per kilometre tolled. Thus, a GPS-based system is better when the intent is to toll many roads. This potential problem can be overcome if it is possible to use mobile phones, rather than dedicated devices.

The other negative of a GPS-based system is that it raises privacy concerns, as a government agency will have access to real-time data on the location of every vehicle. There are technological ways of addressing these concerns, but it may still be a barrier for public acceptance.

Implementing such a system will become easier as autonomous vehicles are introduced. Such vehicles will already have the GPS and data transmission capacities needed, and will also be able to help travellers manage the information to make wise choices. For example, the autonomous vehicle may present route options that include travel times and tolls, without requiring the traveller to know what the toll is on each segment; or an app may allow travellers to plan when to take the trip, presenting information on travel time and toll.

Conclusion

This paper discussed the theoretical and empirical evidence on the distributional effects of Express Lanes, with a focus on how these lanes affect congestion. It also briefly discussed the merits of alternative approaches to managing congestion.

While a major source of opposition to tolling is that it hurts many, if not most, travellers, there are reasons to believe this is not the case with Express Lanes. First, only pricing a portion of the lanes preserves the ability of travellers to continue to choose to not pay a toll. This greatly reduces the harm done relative to pricing all the lanes. Second, if the toll is designed to maximise throughput (vehicles per hour), it is possible for tolling to increase the throughput in the Express Lanes, and thereby improve congestion in the parallel general-purpose lanes. If so, then tolling has helped all travellers, even before using the toll revenue. Third, the toll revenue can be used in ways to help improve the distributional impacts, such as paying for the Express Lanes or improved public transportation. Fourth, Express Lanes provide a reliable and fast option for those times travellers need it, which further helps improve the distributional impacts. Finally, people from all income classes use Express Lanes and report that they like them.

Notes

1 For example, see Persaud, Yagar, and Brownlee (1998), Muñoz and Daganzo (2002), and Srivastava and Geroliminis (2013).

2 In the specific model being used, departures start high and then decrease. Empirically, the departure rate climbs and then falls, so that the optimal toll will shift departures from the peak to the shoulders, causing some travellers to leave earlier and others to leave later.

3 For sufficiently compressed distributions of value of time, it is possible that all travellers have the same preference.

4 For example, the tolls on California SR-91 are set to maintain a target throughput of 1 360-1 600 vehicles per hour per lane.

References

Akbar, P., Duranton, G., 2020. Measuring the cost of congestion in a highly congested city: Bogotá. Work. Pap.

Alameda County Transportation Commission, 2013. Southbound I-680 Express Lane Performance Evaluation: An After Study. Alameda County Transportation Commission.

Anderson, M.L., Davis, L.W., 2020. An empirical test of hypercongestion in highway bottlenecks. J. Public Econ. 187, 104197. https://doi.org/10.1016/j.jpubeco.2020.104197

Arnott, R., de Palma, A., Lindsey, R., 1993. A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand. Am. Econ. Rev. 83, 161-179.

Barton, E., 2014. Toll Lanes Booming Across Florida. The Ledger.

Bento, A., Roth, K., Waxman, A., 2020. Avoiding Traffic Congestion Externalities? The Value of Urgency. Work. Pap.

Bhatt, K. et al., 2008. Value Pricing Pilot Program: Lessons Learned: Final Report (No. FHWA-HOP-08-023). U.S. Department of Transportation, Washington, D.C.

Burris, M.W., Appiah, J., 2004. Examination of Houston’s QuickRide Participants by Frequency of QuickRide Usage. Transp. Res. Rec. 1864, 22-30.

Burris, M.W., Hannay, R.L., 2003. Equity Analysis of the Houston, Texas,QuickRide Project. Transp. Res. Rec. 1859, 87-92. https://doi.org/10.3141/1859-11

Cleland, F., Winters, P.L., 2000. Value Pricing - HOT Lanes in South Florida: Results of Survey and Conclusions. Center for Urban Transportation Research, University of South Florida, Florida.

Daganzo, C., 1996. The Nature of Freeway Gridlock and How to Prevent It, in: Traffic and Transportation Theory. Presented at the 13th International Symposium on Transportation and Traffic Theory, Pergamon, Lyon, France, pp. 629-646.

Dowling, R. et al., 2004. Methodology for Measuring Recurrent and Nonrecurrent Traffic Congestion. Transp. Res. Rec. J. Transp. Res. Board 1867, 60-68. https://doi.org/10.3141/1867-08

Downs, A., 1962. The Law of Peak-Hour Expressway Congestion. Traffic Q. 16.

Duranton, G., Turner, M.A., 2011. The Fundamental Law of Road Congestion: Evidence from US Cities. Am. Econ. Rev. 101, 2616-2652. https://doi.org/10.1257/aer.101.6.2616

Eliasson, J., et al., 2009. The Stockholm congestion – charging trial 2006: Overview of effects. Transp. Res. Part Policy Pract., Stockholm Congestion Charging Trial 43, 240-250. https://doi.org/10.1016/j.tra.2008.09.007

Geroliminis, N., Daganzo, C.F., 2008. Existence of Urban-Scale Macroscopic Fundamental Diagrams: Some Experimental Findings. Transp. Res. Part B Methodol. 42, 759-770. https://doi.org/10.1016/j.trb.2008.02.002

Goel, R., Burris, M.W., 2012. Hot Lane Policies and Their Implications. Transportation 39, 1019-1033. https://doi.org/10.1007/s11116-011-9382-5

Green, C.P., Heywood, J.S., Navarro, M., 2016. Traffic Accidents and the London Congestion Charge. J. Public Econ. 133, 11-22. https://doi.org/10.1016/j.jpubeco.2015.10.005

Hall, J.D., 2018. Pareto Improvements from Lexus Lanes: The Effects of Pricing a Portion of the Lanes on Congested Highways. J. Public Econ. 158, 113-125. https://doi.org/10.1016/j.jpubeco.2018.01.003

Hall, J.D., forthcoming. Can Tolling Help Everyone? Estimating the Aggregate and Distributional Consequences of Congestion Pricing. J. Eur. Econ. Assoc. https://doi.org/10.1093/jeea/jvz082

Hall, J.D., Savage, I., 2019. Tolling Roads to Improve Reliability. J. Urban Econ. 113. https://doi.org/10.1016/j.jue.2019.103187

Khoeini, S., Guensler, R., 2014. Socioeconomic Assessment: Conversion of I-85 High-Occupancy Vehicle to High-Occupancy Toll in Atlanta, Georgia. J. Transp. Res. Board 2450, 52-61. https://doi.org/10.3141/2450-07

Knight, Frank.H., 1924. Some Fallacies in the Interpretation of Social Cost. Q. J. Econ. 38, 582-606. https://doi.org/10.2307/1884592

Kwon, J., Mauch, M., Varaiya, P., 2006. Components of Congestion: Delay from Incidents, Special Events, Lane Closures, Weather, Potential Ramp Metering Gain, and Excess Demand. Transp. Res. Rec. 1959, 84-91. https://doi.org/10.3141/1959-10

Leape, J., 2006. The London Congestion Charge. J. Econ. Perspect. 20, 157-176. https://doi.org/10.1257/jep.20.4.157

Levinson, D., Zhang, L., 2004. Evaluating Effectiveness of Ramp Meters: Evidence from the Twin Cities Ramp Meter Shut-off, in: Gillen, D., Levinson, D. (Eds.), Assessing the Benefits and Costs of ITS, Transportation Research, Economics and Policy. Springer US, Boston, MA, pp. 145-166. https://doi.org/10.1007/1-4020-7874-9_9

Liu, X., et al., 2011. Measuring the Quality of Service for High Occupancy Toll Lanes Operations. Procedia - Soc. Behav. Sci. 16, 15-25. https://doi.org/10.1016/j.sbspro.2011.04.425

Muñoz, J.C., Daganzo, C.F., 2002. The Bottleneck Mechanism of a Freeway Diverge. Transp. Res. Part Policy Pract. 36, 483-505. https://doi.org/10.1016/S0965-8564(01)00017-9

Norman, J., 2002. I-15 Managed Lanes Value Pricing Project Planning Study, Concept Plan. San Diego Association of Governments, San Diego.

Persaud, B., Yagar, S., Brownlee, R., 1998. Exploration of the Breakdown Phenomenon in Freeway Traffic. Transp. Res. Rec. J. Transp. Res. Board 1634, 64-69. https://doi.org/10.3141/1634-08

Petrella, M. et al., 2014. Effects of an HOV-2 to HOT-3 Conversion on Traveler Behavior : Evidence from a Panel Study of the I-85 Corridor in Atlanta (Technical Report). John A. Volpe National Transportation Systems Center, Massachusetts.

Small, K.A., Winston, C., Yan, J., 2005. Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. Econometrica 73, 1367-1382. https://doi.org/10.1111/j.1468-0262.2005.00619.x

Srivastava, A., Geroliminis, N., 2013. Empirical Observations of Capacity Drop in Freeway Merges with Ramp Control and Integration in a First-Order Model. Transp. Res. Part C Emerg. Technol. 30, 161-177. https://doi.org/10.1016/j.trc.2013.02.006

Supernak, J. et al., 2001. Phase II Year Three Overall Report (Task No. 13), I-15 Congestion Pricing Project. San Diego State University, San Diego, CA.

Toor, W., Salisbury, M., 2014. Managed Highway Lanes in Colorado: Everyone Benefits from Including Carpools and Public Transit. Southwest Energy Efficiency Project, Colorado.

Ungemah, D., Swisher, M., Tighe, D.C., 2005. Discussing High-Occupancy Toll Lanes with the Denver, Colorado, Public. J. Transp. Res. Board 1932, 129-136.

US Department of Transportation, 2020. Urban Partnership Agreement Low-Income Equity Concerns of US Road Pricing Initiatives [WWW Document]. URL https://ops.fhwa.dot.gov/congestionpricing/resources/lwincequityrpi/ (accessed 7.8.20).

US Department of Transportation, 2017. Getting Started with Congestion Pricing: A Web-Based Workshop for Local Partners [WWW Document]. URL https://ops.fhwa.dot.gov/publications/fhwahop11035/wkshoppresentations/part2/index.htm (accessed 7.9.20).

US Department of Transportation, 2016. The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations Revision 2. U.S. Department of Transportation, Washington, D.C.

US Department of Transportation, 2014. Los Angeles Congestion Reduction Demonstration (LACRD) ExpressLanes Program National Evaluation: Technical Memorandum on Congestion, Tolling, Transit, and Equity Results. U.S. Department of Transportation, Washington, D.C.

US Department of Transportation, 2008. Income-Based Equity Impacts of Congestion Pricing (No. FHWA-HOP-08-040). U.S. Department of Transportation, Washington, D.C.

Vickrey, W.S., 1969. Congestion Theory and Transport Investment. Am. Econ. Rev. 59, 251-260.

Washington State Department of Transportation, 2009. SR 167 HOT lanes Pilot Project: First Annual Performance Summary. Seattle, Washington.

Zhang, L., Levinson, D., 2010. Ramp metering and freeway bottleneck capacity. Transp. Res. Part Policy Pract. 44, 218-235. https://doi.org/16/j.tra.2010.01.004

Zimmerman, C. et al., 2015. Contemporary Approaches in Congestion Pricing: Lessons Learned from the National Evaluation of Congestion Pricing Strategies at Six Sites (No. FHWA-JPO-2015-217). U.S. Department of Transportation, Washington, D.C.

High-occupancy Toll Lanes

This paper reports the theoretical and empirical evidence on the distributional effects of Express Lanes. It also provides evidence of how they affect congestion, both in the Express Lanes themselves and in the parallel general-purpose lanes. The paper also helps put Express Lanes in context by discussing the merits of alternative approaches to managing congestion such as ramp metering, permits, and vehicle-kilometres travelled charges.

All resources from the Roundtable on Congestion Control Experiences and Recommendations are available at: https://www.itf-oecd.org/congestion-control-experience-recommendations-roundtable.

International Transport Forum
2 rue André Pascal
F-75775 Paris Cedex 16
+33 (0)1 73 31 25 00
contact@itf-oecd.org
www.itf-oecd.org

Published: 09/2020 | Photo credit: JW_PNW/Shutterstock