Market Design for Road Use

Peter Cramton, R. Richard Geddes, and Axel Ockenfels

11 January 2018

Abstract

Traffic congestion is a pervasive worldwide problem. We explain how to harness existing technologies together with new methods in time-and-location markets to eradicate traffic congestion along with its attendant social harms. Our market design for road use builds on congestion pricing and models of efficient pricing in the electricity sector. The market maximizes the value of a transport network through efficient scheduling, routing and pricing of road use. Privacy and equity concerns are addressed. The market makes the best use of the existing road network. Transparent price information provides essential information for efficient long-term investment in transport.

JEL: D47, L90, R48

Traffic congestion is a pervasive and growing worldwide problem. Global congestion costs were estimated at about $1 trillion in 2013 (Economist 2014). According to a recent study, an average driver in the United States spent 42 hours in congestion during peak hours in 2016, an average driver in Los Angeles spent 104 hours, and an average driver on the 4.7 mile stretch of the Cross Bronx Expressway in New York City spent 86 hours (Cookson and Pishue 2017; Pishue 2017). Moreover, congestion’s social costs are growing over time. The United States congestion “invoice” for added costs in terms of fuel and time grew from $42 billion in 1982 to about $160 billion in 2014 (in 2014 dollars)—almost a three-fold increase—in the 471 urban areas studied by the Texas Transportation Institute (Schrank et al. 2015). Similar increases are occurring worldwide as car ownership rises with development. Many roads are failing to perform their basic task of safely facilitating vehicle movement, and the problem is getting worse.

The instant reaction is often to call for more roads. However, numerous studies have shown that an increase in road capacity does not relieve congestion. “The Fundamental Law of Road Congestion” (Duranton and Turner 2011) states that, if new unpriced capacity is added, the road will become as congested as it was before the capacity addition. Indeed, even multibillion investments in new roads in cities like Los Angeles and Houston did little to reduce commuting times. There is also little support for the widespread belief that an increase in ride-hailing services like Uber and Lyft, or in self-driving cars, will cure traffic congestion. Ride-hailing services appear to increase traffic (Schaller Consulting 2017), which probably contributes to Uber’s support of road pricing as “the most effective way to manage vehicles on

---

1 Peter Cramton is Professor of Economics at the University of Maryland and University of Cologne; Rick Geddes is Professor in the Department of Policy Analysis and Management and Director of the Program in Infrastructure Policy at Cornell University; Axel Ockenfels is Professor of Economics at the University of Cologne and Speaker of Cologne Excellence Center for Social and Economic Behavior. Ockenfels thanks the German Research Foundation (DFG) for financial support through the Research Unit “Design & Behavior” (FOR 1371). The authors also thank Arpent for financial support. The views expressed are our own.
the road” (Salzberg 2017), and why Lyft suggests that “congestion pricing . . . has not caught on in a big enough way” (Zimmer and Green 2017). Although self-driving cars use roads more efficiently, it is not obvious that they will relieve traffic congestion. The associated decrease in costs and increase in the demand for mobility may swamp the efficiency gain (Fulton et al. 2017, Henzelmann 2017, SBD 2016). The one approach known to solve the gridlock, and that advocated by most economists, is congestion pricing.

Vickrey (1963, 1969) was the first to recommend congestion pricing. He observed that as roads get congested, additional drivers lower the speed of the drivers behind them. Indeed, as the FHA (2009) explains, the “number of vehicles that get through per hour can drop by as much as 50 percent when severe congestion sets in. At high-traffic levels, the freeway is kept in this condition of ‘collapse’ for several hours after the rush of commuters has stopped.” That is, as a road gets close to its physical limits, even a small increase in the number of vehicles can lead to a large drop in average speed and throughput. Dynamic congestion pricing can internalize the costs that are imposed by individual drivers on the system, and thus maximize throughput at any time. Available transportation capacity is instead currently allocated by queuing. Queuing is especially wasteful in transport because the queue degrades throughput, is blind towards drivers' preferences, and leads to much less predictable and longer travel times.

Efficient road-use pricing generates many additional benefits. It improves environmental quality since vehicles operate more efficiently, and spend less time idling. Safety improves due to more consistent, predictable traffic flow. Moreover, dynamic road prices provide essential information to direct scarce investment resources toward projects where those dollars are most highly valued by road users, while generating the funds that underlie that investment with non-distortionary taxes (Geddes 2011). Additional benefits include: (i) enabling road charges to reflect social cost, rather than tying road charges to fuel taxes; (ii) adopting the basic horizontal-fairness principle that the motorists using a road should pay in proportion to use, which enhances social equity; (iii) allowing scarce road space to be allocated to motorists who value it most highly at that specified time of day; (iv) incentivizing technological innovations that reduce the cars’ demand on scarce road capacity; and (v) encouraging commuters through current road-use prices to explore the travel alternatives of their choice during peak times.

This list suggests that the social benefits created by dynamic road-use pricing will be substantial. Yet, when Vickrey proposed his solution, the time was not right for actual implementation. As Harstad (2008) described it,

Vickrey, though, was appalled at the notion of adding to traffic congestion to collect tolls, and railed against tollbooths, urging the development of a system where small radio transmitters would transmit vehicle or driver IDs over a distance of a few feet, and a computerized system connected to roadbed receivers would calculate liabilities and bill drivers periodically. A few years afterward, Vickrey was challenged that the system he proposed was infeasible. He responded in typical fashion: in the mid-1960s, he first built a rudimentary computer in his home and connected it to a radio receiver, then limited himself to a $3 budget for parts with which he built a small radio transmitter placed under the hood of his car. He could then show anyone who asked a printout of the times his own car went up or down his driveway.
Now, however, the time is right. Advances in mobile communications make it possible to identify and communicate the location of a vehicle to within one cubic meter—allowing precise measurement of road use. User preferences can be communicated both in advance to determine scheduled transport and in real time to optimize routes based on current information. Computer advances facilitate efficient scheduling and pricing of road use. Consumer apps help road users translate detailed price information into preferred transport plans. Computers also allow an independent system operator to better model demand and adjust prices to eliminate congestion and maximize the total value of road infrastructure.

So far, however, the use of road-use fees is limited. Although Oregon debuted a system-wide user-fee program (Morris 2015), road pricing is most often limited to new lanes, such as high-occupancy toll (HOT) lanes or to conversions from high-occupancy vehicle (HOV) to HOT lanes. This has left existing transportation facilities—often older roadways in need of fresh investment—out of user-fee-generated funding streams. Moreover, only a few small sections of the U.S. highway system are dynamically priced. Looking abroad, Singapore, London, Oslo, Stockholm, and Trondheim, among other cities, have adopted some form of congestion pricing. However, its application is often limited to a cordon around the city, or to a small set of roads within, and prices are typically not responsive to real-time changes in demand.

We explain how to harness existing technological developments which, when combined with advanced markets in road use, can fully eradicate traffic congestion along with its attendant social harms. Our proposal represents a significant leap forward to the most complete and efficient end-state for road pricing. We propose a comprehensive system of direct, variable road-use charges. This creates a market for transport that maximizes the value of a transport network through efficient scheduling, routing and pricing of road use.

Congestion pricing for road use

Although universal pricing of road use has only recently become feasible, efficient pricing of network capacity is not new. In restructured electricity markets in the United States, energy is priced at each time and at every location in the system. Electricity supply and demand is scheduled to maximize gains from trade subject to all physical constraints. The prices that support this efficient outcome are (approximate) competitive equilibrium prices—known in the industry as locational marginal prices. The electricity is produced at least cost and consumed by those valuing it most. Our proposal for road-use markets builds on this model and on the advances made in the electricity sector.

In the simplest approach, the market for road use is conducted by an independent system operator (ISO). The ISO’s mission is to maximize the value of scarce existing road resources. The key instrument available to the ISO is the ability to set efficient congestion prices. The ISO models road-use demand and establishes prices at each congested segment and at each time to maximize the value of the road network subject to free-flow capacity constraints. Usage is monitored and charged to each user based on the marginal social

---

2 System uncertainties mean that the ISO cannot operate the road at 100 percent of capacity. Some capacity is reserved to handle momentary surges in demand or drops in supply, as is done in electricity markets. Supply uncertainty arises from road construction, accidents, and weather; demand uncertainty arises from weather and
cost of use. The user price depends on the vehicle-specific demand for road capacity (which depends on the availability of optimized driving algorithms and enhanced capabilities, such as platooning) and pollution. Computer apps armed with this price information then guide consumers in making transport choices consistent with their preferences. For autonomous vehicles, real-time road price data can guide the vehicle’s route decisions without driver intervention. All of this is possible with existing technology. Indeed, adding road-use prices to those apps is straightforward (Cramton et al. 2017).

A wholesale market for transport

A more sophisticated market design is based on a wholesale market model, as in electricity markets (Cramton 2017). In the wholesale market, the ISO has the same mission—to maximize the value of road use. However, it does so by having service providers (e.g., Google, Uber, and Lyft) compete for road use in forward markets as well as in real time. Forward trading mitigates risks. To do this, service providers will likely develop user apps that enable users to easily and effectively express demand. The service providers also guide users, both in scheduling future demand as well as routing during real time. Of course, schedules can change in response to unexpected events and new information. The wholesale market for transport can handle such changes efficiently.

The market model is portrayed in Figure 1. Users provide the underlying demand for road use. Service providers compete for users in the retail market. Providers that offer more attractive plans are likely to be more successful. Some large companies, such as UPS and FedEx, will participate directly in the wholesale market.

The product being traded in the wholesale market is the right to use a well-defined road section for a specific time slot. Time is broken into discrete intervals, such as ten minutes, to keep the number of products manageable. The transport market’s underpinning is the real-time market, which prices road use in real time. The real-time market is a physical market, based on actual (i.e., physical) road use. The system

Figure 1: Wholesale market model

The market model is portrayed in Figure 1. Users provide the underlying demand for road use. Service providers compete for users in the retail market. Providers that offer more attractive plans are likely to be more successful. Some large companies, such as UPS and FedEx, will participate directly in the wholesale market.

The product being traded in the wholesale market is the right to use a well-defined road section for a specific time slot. Time is broken into discrete intervals, such as ten minutes, to keep the number of products manageable. The transport market’s underpinning is the real-time market, which prices road use in real time. The real-time market is a physical market, based on actual (i.e., physical) road use. The system

special events, such as the end of a football game, as well as other shocks. For related issues surrounding real-time pricing in the electricity sector see Joskow and Wolfram (2012), and in the airline industry see McAfee and te Velde (2007). Real-time markets can create complications due to strategic timing of transactions, which, however, can be dealt with (Budish et al. 2015, Roth and Ockenfels 2002).
operator adjusts the price to clear the market at each time and location. Effectively, there is an auction to balance supply and demand.

Identifying equilibrium prices and quantities for many interrelated products is complicated, but it can be done with modern optimization techniques (Milgrom 2009, 2017). Indeed, the number of congested road segments is limited, the congested segments are highly predictable, and integer constraints do not play an important role. The number of vehicles that can travel on a road segment in a ten-minute interval is at least several hundred and over one thousand for major highways with three or more lanes. This makes the product a divisible good; it can be treated as such in forward markets. Further, with demands expressed as piecewise linear curves, the objective of maximizing gains from trade is quadratic and positive definite. With linear constraints, the problem is readily solved with standard techniques from convex optimization (Bazaraa et al. 2013). Feasibility is assured since zero trade satisfies each trader’s constraints. Even for large instances with thousands of products and constraints, the unique competitive equilibrium prices and quantities that maximize gains from trade (i.e., value of the network) can be computed in seconds using today’s computers and algorithms.

Since real-time congestion prices tend to be volatile, participants wish to manage price risks. To mitigate risk and promote efficient scheduling, we propose four forward (primary) markets: yearly, monthly, weekly, and daily. Forward markets are financial, not physical. Service providers take positions in forward markets, while subsequent markets allow adjustment of positions as uncertainty is resolved. The ISO determines supply offered in each forward market consistent with customers’ and service providers’ interest in taking forward positions. The ISO offers more supply at higher prices, creating a rising supply curve. Service providers bid in forward and real-time markets to maximize net value to users and manage risk. This typically involves purchasing some fraction of user demand in each of the markets and to adjust positions as demand uncertainty is resolved.

Having multiple opportunities to trade reduces risk to the service provider and facilitates planning. Forward markets also facilitate price discovery through price transparency. Finally, forward trading mitigates market power in the real-time market by putting wholesale customers and service providers in a more balanced position entering the real-time market.\(^3\)

Under the wholesale-market model, service providers compete for retail customers. Offering a service plan that creates the most value for the consumer is one form of competition. The most sensible and efficient plans will let the consumer express a realistic estimate of usage, informed by past driving behavior, then let the consumer purchase this expected usage on a forward basis to reduce exposure from real-time price volatility. The monthly settlement of such a plan would eliminate deviations from expected

\(^3\) See, e.g., Wolak (2000) for the argument that forward trading mitigates market power. In electricity markets, real-time prices are significantly more volatile than day-ahead prices, because, as we get closer to real time, participants have fewer options to address shocks to supply or demand. The supply and demand curves that determine price are steeper—quantity varies less in response to price. While most energy is traded in advance of real time, the real-time market plays an essential role in efficiently pricing deviations from day-ahead schedules. Such markets are known to function robustly, even under extreme stress. For instance, the Texas electricity system withstood the huge shocks to supply and demand that Hurricane Harvey produced in August 2017: market pricing and system stability were maintained throughout the multiday event (Cramton et al. 2017).
usage using real-time prices. Thus, the consumer is exposed to the real-time price on the margin—as required for efficiency—but most users’ expenditure would be at forward prices.

Service providers would also compete by providing useful apps that help customers easily express their preferences. An example retail app is shown in Figure 2. The app lets the user purchase a trip on a forward basis, which may be either a one-time trip or a recurring trip at a specified time. In this case, the user is buying a daily roundtrip from home to work for all the weekdays in November. The user enters his preferred departure times and the app calculates a cost per trip. The app also shows how the cost per trip would change if the user departed earlier or later.

![Figure 2: Example of a retail app](image)

Promoting acceptance

Before communities shift to efficient congestion pricing, important questions must be addressed to overcome opposition in the public and in politics: How does pricing alter consumer choice? Can privacy concerns be addressed? Can equity concerns be addressed?

Trade-offs in transport

Road-use pricing will be successful if it delivers tangible benefits. The elimination of congestion delays will be the most salient benefit driving consumer acceptance. Traffic congestion comes at huge individual and social costs, and makes millions of people’s lives miserable. Moreover, because road pricing assures free-flow throughput, the capacity on congested road segments is up to twice as large at the most popular times.

Another key benefit of road pricing is simplified decision making. Absent road pricing, drivers face uncertainty about trip time. Those uncertainties depend on departure time. Even adding substantial extra
time to account for typical delays may still result in a late arrive, frustrating not only the driver but others who are counting on the vehicle’s on-time arrival.

Although uncertainty about travel time is eliminated by efficient congestion pricing, it will create price uncertainty. Price volatility is however lower due to the reduced volatility of travel times on previously congested roads. Moreover, the harm from price risk is less. What matters for the monthly budget is the volatility of the monthly expense, which is smaller due to the law of large numbers. Price risk can be further mitigated via forward purchases. In contrast, delay uncertainty requires drivers who need to arrive on time to take costly actions that exceed the expected cost of delay.

**Equity and fairness**

Road pricing frequently raises equity concerns in that pricing may benefit the rich at the expense of the poor. This is a critical question. However, road pricing benefits the entire population; surveys show that resistance in road pricing projects in operation is particularly low among low-income users.⁴ One reason for acceptance of pricing is that pricing maintains free-flow throughput, and the road-use cost paid by drivers generates revenue, which can be used to build and maintain road infrastructure and public transportation. Moreover, various options to enhance affordability are available and have been partly implemented, such as a fixed travel allowance for commuters (Small 1992), allocating monthly budgets to spend on congestion tolls (Kalmanje and Kockelmann 2004), and reduction of distortionary taxation (Parry and Bento 2001). Although it is true that poorer users are more apt to shift to less expensive travel times, this implies that low-income households benefit the most from replacing existing transport taxes with an efficient pricing scheme (Martin and Thornton 2017; West and Williams 2004; Schweitzer and Taylor 2008).⁵ Other options include ‘grandfathering’ vouchers to poor or needy people, which they could then either use themselves or sell on the market.⁶

Viewing road-use pricing as unfair implies that the status quo is fairer. But if we had started with a market for road use, there would be fairness arguments against switching to free road use; e.g., paying the same price regardless of whether the driver contributes to a traffic gridlock or causes more emissions in densely

---

⁴ For instance, 70 percent of the lowest income users strongly support San Diego’s HOT lanes (FHA 2008). Hall (2017) gives a simple illustration that road pricing does not need to create losers; a carefully designed, time-varying toll on a portion of the lanes of a highway that can make everybody better off, even before the toll revenue is spent and even with realistic driver heterogeneity. On the other hand, Arnott et al. (1994) and Parry and Bento (2001) show that, under certain conditions, many drivers may not want to support congestion pricing solely based on resulting travel-time savings in the short run.

⁵ For a similar reason, the poor are likely better able to take advantage of profit opportunities that arise in markets for road use near real time because of supply or demand shocks. For example, a poor commuter may buy a monthly pass to commute at her preferred time, thereby locking in an attractive price per trip of $10. On occasions when there is a lane closure the real-time price may surge to $30. The poor commuter may opt to take public transit on such days and earn $30 per event (her forward purchase would automatically be sold in real time for $30).

⁶ Extreme surge prices in exceptional circumstances seem unacceptable to many (Kahenman et al. 1986, Irwin 2017). To further promote acceptability, the market design could include special rules for well-defined exceptional circumstances, as we also see them in other markets.
populated areas is unfair; that rich people drive more and so benefit more from eliminating road charges is unfair; that poor people suffer more from increased traffic externalities is unfair. Indeed, once congestion pricing can be established, it often gains strong support. Eliasson (2014) describes how the attitudes changed “from fiercely hostile to overwhelmingly positive” after Stockholm introduced congestion charges.7

The norm of a market-based economy is for consumers to buy goods and services at competitive market prices. This is true even for essential services such water, gas, electricity, and communications. The non-pricing of roads stems from the fact that roads were originally uncongested, and that pricing, collecting, and enforcing payments was too costly. In today’s urban areas, those arguments to avoid pricing are no longer valid.

**Privacy**

Road pricing also raises privacy concerns. Monitoring and enforcement require that the system operator know the location of each vehicle during use. Technically this is easy. Each vehicle would have a device for this purpose,8 and there would be stiff fines for vehicle operation with a non-functioning device. But users may be concerned that the location information would be inappropriately used—an invasion of privacy. However, nearly everyone with a smart phone uses at least one map app and nearly everyone agrees to give the map app this location information. This consent is given with the belief that the information will be used to help, not hurt, the user. The same could be done for measuring road use. The system operator would have strict rules about how the information is used. The information would be used only to meter road use for purposes of scheduling, routing, and pricing. No individual data would be shared with others. Controls would be in place to assure that the information policy is followed, and the data remained secure. Indeed, modern cryptography makes it possible for the system operator to run the market without any human having access to the individual data and still prove that the market rules are faithfully followed (Parkes et al. 2008, 2009, 2015; Thorpe and Parkes 2007, 2009).

**Conclusion**

Traffic congestion is one of the biggest challenges of modern societies. Worldwide congestion costs are estimated at about $1 trillion per year, and traffic is expected to only become worse over the next years and decades. The “Fundamental Law of Road Congestion” implies that building or widening new roads, or other supply-based policies, will not solve the problem. Thus, under existing policies, traffic is getting worse and will inevitably end in a crisis.

Fortunately, technological advances allow us to largely eliminate congestion through efficient congestion pricing, which eliminates the inefficiency, frustration and unfairness that comes with the current system.

---

7 As Eliasson (2014) describes in detail “congestion charges overcame fierce initial hostility, survived a heated and complicated political and legal process, including a referendum initially forced through by opponents to the charges, and has eventually gained support by more than 2/3 of the population..... The Stockholm charges went from ‘the most expensive way ever devised to commit political suicide’ ... to something that the initially hostile media eventually declared to be a ‘success story’.”

8 Singapore has already procured such devices for every vehicle in Singapore, beginning in 2020 (LTA 2016).
Congestion pricing for road use is apt to be introduced gradually as technology evolves and opportunities for implementation are identified. In the simplest approach, an independent system operator models demand and computes real-time prices for road use to maximize the value of road use. Prices on congested road segments are set to induce demand to eliminate the congestion and maximize throughput. User-friendly computer apps armed with this price information then guide consumers in making transport choices consistent with their preferences. Equity and privacy concerns can be effectively dealt with.

In a sense, responding to prices in transport is nothing new. Today one’s “price” is paid in throughput reduction and delay cost. Users do respond to this price, but the price is waste and is set incorrectly—the delay cost does not reflect the negative externality one user imposes on others. In the market model, congestion is eliminated: the real-time price is set at the marginal value of demand at the point of supply and demand balance.

Congestion pricing is not a novel concept. Similar variable charges have been successfully utilized in transport and many other industries. For example, airfares, cell phone rates, electricity rates, room rates at hotels and resorts, train fares, and some local transit systems use variable pricing. Empirical studies strongly confirm the effectiveness and efficiency of pricing mechanisms to address traffic congestion (Cramton et al. 2017). In transport, demand response takes one of four forms: (i) time shifters, who shift transport to a less congested time; (ii) route shifters, who shift to a less congested route; (iii) mode shifters, who decide to take a train, bus, or bike, rather than drive; and (iv) curtailers, who decide to ride-share, work at home or otherwise reduce demand. Experimental evidence such as Lopez (2017), controlled field studies such as Martin and Thornton (2017) as well as naturally occurring data such as (Eliasson 2014 and Foreman 2016) provide evidence that drivers strongly respond to congestion prices, and thus suggest that a market for road use, as we envision it, would work well.

A more sophisticated market design, built on congestion pricing, is based on a wholesale market model, as we see in electricity markets. The advantage of a wholesale market is that it allows relatively easy entry as a service provider. The ensuing competition among service providers then promotes innovation. That innovation helps service providers to better understand user demands, translate user demands into bids in the wholesale market, and develop forward trading strategies to mitigate risk. Although this market involves complex optimization, the problem is made easier by the limited number of bottlenecks and the large number of vehicles. The latter reduces the importance of integer constraints in the optimization problem making it more likely that efficient congestion prices exist. Another helpful factor is that aggregate traffic patterns tend to follow a predictable cycle—this greatly facilitates the system operator’s modeling of demand.

Modern mobile communications allow road use to be monitored and charged based on real-time scarcity. Doing so gets the most out of our existing transportation infrastructure and simultaneously provides essential funding of the roads as well as valuable price information to evaluate road enhancements. This is the inevitable future of roads.
References


