American Economic Journal: Economic Policy 2020, 12(1): 418–453
https://doi.org/10.1257/pol.20170195

The Marginal Cost of Traffic Congestion and Road Pricing: Evidence from a Natural Experiment in Beijing†

By Jun Yang, Avralt-Od Purevjav, and Shanjun Li*

Severe traffic congestion is ubiquitous in large urban centers. This paper provides the first causal estimate of the relationship between traffic density and speed and optimal congestion charges using real-time fine-scale traffic data in Beijing. The identification relies on plausibly exogenous variation in traffic density induced by Beijing's driving restriction policy. Optimal congestion charges range from 5 to 39 cents per km depending on time and location. Road pricing would increase traffic speed by 11 percent within the city center and lead to an annual welfare gain of ¥1.5 billion from reduced congestion and revenue of ¥10.5 billion. (*JEL* H23, O18, P25, R41, R48)

Traffic congestion is ubiquitous in large cities around the world, especially in the middle-income countries and emerging markets where road infrastructure and regulations lag behind the rapid rise in vehicle ownership. Based on real-time traffic data in 390 cities from 48 countries in 2016, TomTom Traffic Index shows that among the top 20 most congested cities (population over 0.8 million), all but 1 are located in developing and emerging economies with 8 of them being from China. Drivers in Mexico City (number 1 on the list) spent 66 percent more commuting time on average than they would have under the free-flow condition, while drivers in Beijing (number 10 on the list) spent 46 percent extra time on the road, amounting to 227 and 179 hours of extra travel time per year in these two cities respectively.[1]

The first-best policy to address urban traffic congestion dates back at least to Vickrey (1959, 1963) who proposed road pricing by recognizing traffic congestion as a classic externality. Identifying the mispricing of a resource (road capacity) as

*Yang: Beijing Jiaotong University, No 9. South Liu Li Qiao Road, Fengtai District, Beijing, China 100073, and Beijing Transport Institute (email: yangjun218@sina.com); Purevjav: Dyson School of Applied Economics and Management, Cornell University, Warren Hall, Ithaca, NY 20854 (email: ap884@cornell.edu); Li: Dyson School of Applied Economics and Management, Cornell University, 405 Warren Hall, Ithaca, NY 20854, and NBER (email: SL2448@cornell.edu). Dan Silverman was coeditor for this article. Authors contributed equally and are listed in reverse alphabetical order. We thank Panle Jia Barwick, Jonathan Hughes, Matthew E. Kahn, Daniel McMillen, Kenneth Small, Matthew Turner, Clifford Winston, Shuang Zhang, three anonymous referees, and seminar participants at Cornell University, Jinan University, Renmin University, the Brookings-Tsinghua Conference, and the 2017 AEA Annual Meeting for helpful comments. Binglin Wang and Jingyuan Wang provided excellent research assistance. We acknowledge financial support from the International Initiative for Impact Evaluation (3ie) under project DPW1.1106, the National Natural Science Foundation of China under project 71628303, Cornell Atkinson Center for a Sustainable Future, and Cornell Institute for the Social Sciences.

†Go to https://doi.org/10.1257/pol.20170195 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] The full ranking based on the TomTom Traffic Index 2016 is available at https://www.tomtom.com/en_gb/trafficindex/list?citySize=LARGE&continent=ALL&country=ALL. Los Angeles is the only one from a developed country and ranked twelfth with a congestion index of 45 percent.

the root cause of traffic congestion, Vickrey (1963, 452) states: "… in no other major area are pricing practices so irrational, so out of date, and so conducive to waste as in urban transportation." This statement remains largely true more than half a century later. Despite being continuously advocated by economists, road pricing has not been widely adopted in practice largely due to technical feasibility and political acceptability. During the last 15 years, several European and US cities have implemented different designs of road pricing while policymakers elsewhere including those in China are showing increasing interests in the policy and paying attention to these real-world implementations.[2]

In China, traffic congestion is one of the most pressing challenges in major urban areas, largely as a result of the unprecedented economic and social transformation during the last three decades. Since the turn of the century, the dramatic increase in vehicle ownership and urbanization has overwhelmed the provision of road infrastructure and public transit, leading to serious traffic congestion that affects both housing and employment decisions, as well as the quality of life of urban residents (Zheng and Kahn 2013).[3] Between 2001 and 2015, Beijing experienced a 55 percent increase in population while its per capita GDP increased from about \$1,000 to over \$8,000, and vehicle stock increased from 1 million to nearly 6 million (see Figure 1). Beijing has routinely been ranked as one of the most congested cities in the world, with average traffic speed often less than 15 miles per hour during rush hours. To deal with traffic congestion, central and local governments in China have been employing various policies such as driving restrictions and vehicle purchase restrictions, but these policies have achieved little or no visible impact because they fundamentally failed to set the right price for road use. In December 2015, the Beijing municipal government announced a plan to introduce road pricing in the near future while soliciting feedback from experts and the general public.

One of the key components of road pricing is to estimate the marginal external cost of traffic congestion (MECC) and the optimal congestion charge (Walters 1961, Keeler and Small 1977, Newbery 1988b, Parry and Small 2009). To estimate MECC, researchers typically specify traffic speed as a deterministic function of traffic density (or flow) following the transportation engineering literature, which treats this relationship as one-way and mechanical. In reality, traffic speed and density affect each other, and both are equilibrium outcomes subject to idiosyncrasies. This gives rise to endogeneity due to simultaneity in empirical estimation, as recognized by Small and Chu (2003).

[2] Singapore was the first in adopting congestion pricing in 1975. In recent years, London (2003), Stockholm (2006), Milan (2008), and Gothenburg (2013) have adopted area-based congestion pricing. Several area-based schemes have been proposed for US cities over the years, but the only instances of congestion pricing per se in the United States are high-occupancy toll (HOT) lanes. Several dozen HOT lanes with either variable or dynamic tolls are operating or planned around the United States. A recent New York State budget in 2019 proposes congestion pricing on vehicles that enter Manhattan below Sixtieth Street. If adopted, New York City will become the first US city to use congestion pricing.

[3] While having virtually zero private vehicle ownership 20 years ago, China surpassed the United States in 2009 to become the world's largest automobile market, with over 25 million new passenger vehicles sold in 2016. At the same time, China has witnessed the largest ever migration of people from rural areas to cities, with the share of urban population increasing from 25 percent to over 50 percent during the last two decades.

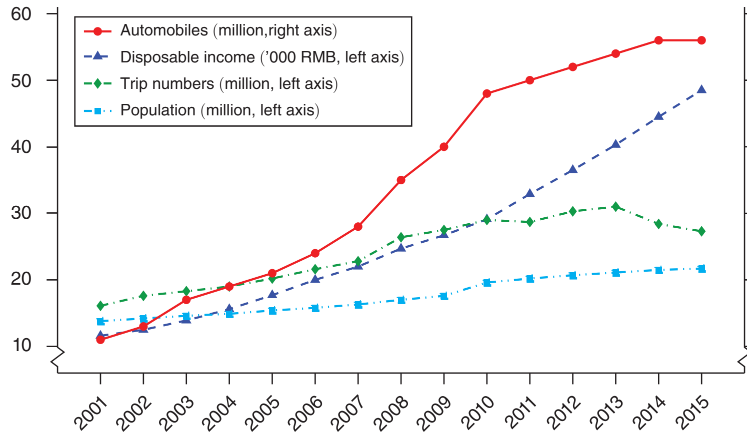

Figure 1. Population, Trips, Disposable Income, and Vehicle Stock in Beijing

Source: Beijing Transportation Annual Report

The objective of this study is to empirically estimate MECC and optimal congestion charges in Beijing while examining the consequences of road pricing on traffic speed and social welfare. We provide the first empirical estimate of the causal effect of traffic density on speed by relying on plausibly exogenous variations in traffic density introduced by Beijing's driving restriction policy. The policy restricts some vehicles from driving on a given workday depending on the last digit of the license plate number. The policy follows a preset rotation schedule in terms of which pair of numbers (one and six, two and seven, three and eight, four and nine, or five and zero) is restricted on a given day, and it is not adjusted based on traffic conditions. The policy affects the number of vehicles on the road due to the fact that the distribution of vehicles is not uniform with respect to the last digit of plate numbers. Vehicles with the license plate ending with number 4 only account for about 2 percent of all vehicles.[4] Therefore, when numbers four and nine are restricted, more vehicles are on the road, and congestion tends to be worse than other days.

We use a unique and fine-scale dataset from over 1,500 traffic-monitoring stations (microwave censors) throughout Beijing that recorded the real-time traffic speed and density at two-minute intervals during 2014. Our analysis based on the instrumental variable method shows that the average MECC during the rush hours is about ¥1.98 (about $0.30) per vehicle-km within the second ring road, over 60 percent larger than what OLS regressions would imply. The downward bias of the OLS estimate could be driven by unobservables such as big events, accidents, or road construction that reduce traffic density (for example due to avoidance behavior) and at the same time reduce traffic speed. The total value of lost time that drivers impose on other road users amounts to nearly ¥10 billion in Beijing for 2014, which is about

[4] The number four is considered an unlucky number in Chinese culture because it shares the same pronunciation as the word death in Mandarin.

half a percent of Beijing's GDP. The optimal congestion charges range from 5 to 39 cents per km depending on time and location. Following the time-varying and —location-specific congestion charges we estimated, road pricing would increase traffic speed by 11 percent within the city center (i.e., the third ring road) and lead to an annual welfare gain of ¥1.5 billion from reduced congestion. The welfare gain could reach over ¥4 billion if the benefit from reduced local air pollution, CO_2 emissions, and traffic accidents is taken into account. The tax revenue from road pricing would amount to over ¥10 billion per year, which is over 65 percent of the total annual operating subsidy for public transit (subways and buses) provided by the Beijing municipal government.

Our study makes the following two contributions to the literature. First, to our knowledge, this is the first attempt to empirically estimate the marginal cost of traffic congestion while addressing the endogeneity issue in the relationship between traffic speed and density (or flow) using real-time traffic data. Existing studies are mostly based on engineering estimates of the speed-flow relationship rather than empirical estimates of the relationship (Lindsey and Verhoef 2007, Parry and Small 2009). While the endogeneity issue in the speed-flow (or speed-density) relationship has been recognized (Small and Chu 2003), it has rarely been addressed in empirical settings perhaps due to the view that the relationship is a mechanical one and the challenge in finding a valid instrumental variable. One exception is Couture, Duranton, and Turner (2018), which investigates the determinants of driving speed in large US metropolitan statistical areas (MSAs) using household travel survey data. Their study addresses the endogeneity in the speed (or the inverse supply) function by instrumenting for trip distance using the mean distance of other trips for the same purpose in the same MSA or a set of trip type dummies. Our paper differs from their paper in terms of methodology, data, and identification strategy. Our empirical framework focuses on the relationship between traffic speed and density and utilizes real-time traffic data on speed and volume from traffic microwave censors for a large number of road segments in single city. We leverage the unique driving restriction policy as a natural experiment for identification. Consistent with their finding, our results show that OLS underestimates the slope of the (inverse) supply curve and hence the marginal external cost of congestion.

Second, our study provides the first empirical estimates of MECC and an empirical road map for designing road pricing schemes in China based on fine-scale traffic data rich in both temporal and spatial dimensions. This type of data is becoming increasingly available to researchers. Although there exist a number of studies that estimate the MECC in developed countries such as Canada and the United States (e.g., Walters 1961; Kraus, Mohring, and Pinfold 1976; Keeler and Small 1977; Dewees 1979; Parry and Small 2009), there is a lack of evidence from developing countries, where traffic congestion is becoming one of the most pressing urban challenges since it affects the quality of life in major urban areas. Rigorous empirical analysis to measure MECC is the first step toward designing market-based mechanisms to effectively address this challenge.

The next section presents the theoretical framework of road pricing. Section II provides the background and describes the data. Section III lays out the empirical strategy, and Section IV presents model estimation results. Section V estimates optimal congestion charges and investigates the consequences on traffic speed and welfare. Section VI concludes.

I. Theoretical Framework

The conventional approach to the economic analysis of traffic congestion is centered on the relationship between the cost of using a road and the traffic volume (flow), which is derived from the speed-volume relationship. This approach is widely employed to estimate MECC. The speed-volume relationship can be divided into two nonlinear parts with a turning point: *ordinary congestion* with a negative speed-volume relationship and *hyper-congestion* with a positive relationship. *Ordinary congestion* occurs when an increase in the number of vehicles on the road reduces average vehicle speed but increases traffic volume (throughput) because the effect of an additional vehicle is greater than the effect of average speed reduction on volume. When traffic volume exceeds the road capacity, additional vehicles entering the road lead to decreases in both vehicle speed and traffic volume. Beyond this capacity threshold, the road is said to be *hyper-congested*.

The empirical estimation of a speed-volume relation or the turning point from ordinary congestion to hyper-congestion is difficult as the relationship from volume to speed does not have a one-to-one correspondence and is nonlinear. Economic analyses typically focus on the ordinary congestion portion of the speed-volume relation (Newbery 1988b, Button 2010, Quinet and Vickerman 2004), and we follow this practice by focusing on ordinary congestion in the empirical analysis and welfare calculation. While taking hyper-congestion into account in road pricing could be potentially important as pointed out by Verhoef (2003), Fosgerau and Small (2013), and Fosgerau (2015), it would necessitate richer data and a different framework to estimate travel demand at a spatially finer level (e.g., road segments) given the dynamic, local, and transient nature of hyper-congestion.[5]

Following Newbery (1988a, b) and Quinet and Vickerman (2004), the marginal social cost (MSC) is defined as

$$MSC(V) = \frac{\partial TSC}{\partial V} = T(V) \cdot o \cdot VOT + MECC(V), \tag{1}$$

where travel time T, the reciprocal of vehicle speed S, is measured in hours per kilometer, and traffic volume V, the product of traffic density D and vehicle speed S, is measured in vehicles per hour. The variable o is vehicle occupancy, the average number of passengers per vehicle. The term VOT denotes the value of travel time for a

[5] The theoretical challenge and empirical treatment for hyper-congestion are still unsettled and under active research in the literature (Walters 1961; Verhoef 1999, 2001, 2003; Small and Chu 2003; Daganzo and Geroliminis 2008; Arnott 2013; Fosgerau 2015; Anderson and Davis 2018). As noted by Arnott (1990), hyper-congestion is dynamic in nature and occurs as a local transient response to demand fluctuations. While it may occur locally due to, for example, a nearby bottleneck, it could have a wider impact on the road network, for example by causing gridlock in a dense street network (Small and Chu 2003). In order to capture the full marginal external cost of hyper-congestion and its welfare effects, a dynamic model with network features might be needed to estimate the impact of additional driving on travel time and traffic volume over the congested period for a specific road segment as well as for the broader network.

representative driver, measured in dollars per passenger-hour. The function *MECC* in $/vehicle-km is the marginal external cost of *ordinary* congestion and is defined as

$$(2)\quad MECC(V) = V \cdot o \cdot VOT \cdot \frac{\partial T}{\partial V} = o \cdot VOT \cdot T(V) \cdot \frac{\epsilon}{1-\epsilon}, \quad 0 \leq \epsilon < 1,$$

where $\epsilon = -\partial S/\partial D \cdot D/S$ is the elasticity of vehicle speed S, measured in kilometers per hour, with respect to traffic density D, measured in vehicles per kilometer. This equation shows that once we recover the speed-density relationship (or the elasticity) and travel time as a function of traffic volume, we can calculate MECC in terms of volume, given certain assumptions on occupancy per vehicle and the value of travel time.

Speed and travel time (the inverse of speed) T are monotonic functions of density regardless of what kind of traffic congestion is occurring. Our empirical analysis specifies and estimates a linear relationship between traffic speed and density first proposed by Greenshields et al. (1935):

$$(3)\quad S = \alpha + \beta \cdot D,$$

with the elasticity of speed with respect to traffic density being defined as

$$\epsilon = -\beta \cdot D/S.$$

Travel time T under ordinary congestion can then be expressed as a function of traffic volume V as follows:

$$(4)\quad T = T(V) = \frac{2/\alpha}{1+\sqrt{1-V/V^m}}, \quad 0 \leq \epsilon < 1,$$

where $V^m = -\alpha^2/(4\beta)$ is the road capacity. Equation (4) shows that travel time is an increasing function of traffic volume under ordinary congestion.

The optimal congestion charges and resulting impacts on traffic congestion and social welfare are empirical questions. The goal of our analysis below is to estimate the two cost curves for the city of Beijing: average social cost curve *ASC* and the marginal social cost curve *MSC*. Based on these curves, we then calculate optimal congestion charges and examine the congestion and welfare impacts of road pricing.

II. Background and Data

In this section, we first discuss Beijing's transportation challenges and review various policies implemented to address the congestion issues. We then present our data.

A. *Traffic Congestion in Beijing*

Major urban cities in China, including Beijing, have been experiencing the world's worst traffic congestion due to the dramatic increase in vehicle ownership and travel demand over the past decade. As of 2014, Beijing's total number of vehicles hit 5.6 million; 4.5 million of those vehicles are privately owned. The share of travel done using private vehicles and taxis reached 63 percent while the share of public transportation was around 26 percent (Beijing Transportation Annual Report 2014). Beijing now is routinely ranked by numerous organizations as one of the most polluted and traffic-congested cities in the world (TomTom Traffic Index 2015). According to the Beijing Transportation Annual Reports, the average speed of vehicles on arterial roads during the peak hours on workdays decreased to 23 km per hour in 2014 from 60 km per hour in 2001. To address traffic congestion, central and local governments have adopted various policies including investment in public transportation (bus, road, and rail), driving restrictions, and restrictions on purchases of new vehicles. Table 1 lists major transportation policies in Beijing.

As a supply-side strategy to address traffic congestion, the Beijing municipal government has been expanding urban road and subway systems extensively. The number of subway lines increased from 2 before 2002 to 18 in 2016, totaling 554 km in length. However, the continuous increase in road supply and infrastructure development have not been adequate to accommodate the increasing travel demand. Previous experiences elsewhere have shown that the supply-side policy alone is unlikely to fully address traffic congestion in the long run because road expansion reduces the private cost of traveling and hence increases travel demand in the long run (Duranton and Turner 2011).

In addition to the road and subway expansion, Beijing imposed a parking certificate system in 1998 on newly purchased vehicles in eight urban districts to restrict the total number of vehicles on city roads. However, due to slow registration and administration of parking certificates, the high costs of verification, and the arbitrary collection of fees, the system was largely ineffective. In early 2002, the parking restriction system was replaced by a vehicle purchase tax of 10 percent, but Beijing continued to experience high rates of growth in motor vehicle purchases.

In 2005, the Beijing Transport Development Outlines (2004–2020) were issued, and two major strategic plans were highlighted including enhancing public transportation, adjusting the city's spatial structure, and gradually implementing travel demand management. In accordance with this outline, Beijing implemented a low-cost public transportation policy in 2007 and expanded bus-only lanes to increase the attractiveness of using buses and the subway system.

Due to the limited impact of these policy measures, the Beijing municipal government has decided to introduce more administrative means of policy measures aimed to combat traffic, one of which is the restriction of vehicle ownership and use.[6] However, road users responded to these policies by changing their travel behavior to circumvent the measures. For example, many motorists would enter the traffic zones prior to the start of the restriction period, and some would even purchase additional vehicles. Although driving restriction policies were effective in reducing vehicle usage in the short run (Viard and Fu 2015), the improvement in traffic condition

[6]During the Olympics in August 2008, Beijing implemented short-term driving restriction policies in which cars could be driven on the road only on certain days according to whether they had an even- or odd-numbered licensed plate.

TABLE 1—CONGESTION REDUCTION POLICIES IN BEIJING

Policies	Year	Actions
Public infrastructure	1986–2010 2007–2011	Road expansion Railway, subway expansion
Parking restrictions	1998–2002	Parking certificate Parking fee increase in selected areas
Public transportation	2007	Low-cost fare and subsidy for bus and subway Increase in bus-only lanes
Taxes and fees	2009–2011	Vehicle purchase tax 10 percent Taxi fuel surcharge for trips over 3 kms (¥1–2) Gasoline price adjustments
Air quality regulations	2008	Temporary shutdowns of factories and construction sites Revised commercial and light-duty vehicle emission standards
Driving restriction	2008	Restriction by odd-even license plate numbers one day per week for a certain period of time
Ownership restriction	2011	Private car license lottery

diminished over time. To curb the growth of automobile ownership, the Beijing municipal government started to implement a vehicle purchase restriction policy where lotteries are held to allocate vehicle licenses from 2011.[7]

B. *Road Pricing Policy*

The supply-side policies such as road expansion and the command-and-control policies such as driving restrictions may have had some impact in the short term but have ultimately done little to reduce traffic congestion in the long run because they fundamentally failed to set the right price for road use. Economists and transportation experts since Vickrey (1959, 1963) have continually advocated road pricing as the first-best policy to reduce traffic congestion. Singapore was the first to adopt road pricing in early 1975 with the help of the World Bank. In recent years, several cities and countries in Europe (Norway, London, Milan, Stockholm, and Gothenburg) have adopted the policy in various forms. Studies show that road pricing has been effective in reducing congestion in these cities, improving traffic speed by 10–30 percent (Small and Gomez-Ibañez 1998, Olszewski and Xie 2005, Leape 2006, Anas and Lindsey 2011).[8] Despite worsening traffic congestion in many middle-income countries and emerging markets, road pricing has not been introduced in these places. Equity and political concerns have been argued as the major reasons for not adopting road pricing (Rouwendal and Verhoef 2006, Eliasson and Mattsson 2006, de Grange and Troncoso 2011).

With increasing discontent from residents regarding traffic congestion and air pollution, the Beijing municipal government announced a plan to adopt road

[7] Since 2011, lotteries have been used to allocate about 20,000 licenses each month, with 88 percent to individuals, 10 percent going to companies, and the remaining 2 percent to operators of transportation service.

[8] See tables 1 and 2 in Anas and Lindsey (2011) for a detailed review on congestion pricing schemes and their impacts.

pricing in 2015. The scheme under consideration is to be implemented in the area encompassed by the third ring road, and all vehicles except buses entering this zone at any time will be charged ¥8 (about $1.25) each time of entry. According to Linn, Wang, and Xie (2016), the individuals likely to be affected by this road pricing plan will tend to be a relatively wealthier and small proportion of the population. Although it should reduce the number of road users within the third ring road, it is unlikely to reduce the travel distance for those who decide to enter the congestion-charge zone because the congestion charge does not vary with distance traveled. The proposed policy may actually increase the distance traveled for this group of consumers because of reduced congestion in the short run and the need to stay within the third ring road to avoid paying multiple entry fees.

The flat charge for entering the cordon area proposed above does not take into account the heterogeneity in the amount of congestion externality, which varies over time and especially across space as documented in our analysis below. Our analysis also shows that road pricing schemes that did not take into account this heterogeneity would be less effective in reducing traffic congestion and less efficient in improving social welfare. Existing road pricing schemes mainly take the form of zone-based charges where flat fees are levied for access to heavily congested urban centers. As GPS technology improves and the cost of implementation decreases, it is becoming practically feasible to approximate a first-best road pricing policy that is distance based and whose charge varies over time and across space to take into account the heterogeneity in congestion externality, especially hyper-congestion cases.[9] For example, Singapore's electronic road pricing system currently features about 80 entry points located around the city that record passing vehicles, but it is slated to upgrade in 2020 to a GPS-based system with the ability to charge for distance traveled.[10] In our analysis below, we consider several schemes of distance-based road pricing with varying degrees of approximation to the first-best road pricing policy.

C. *Data Description*

Our empirical analysis is based on two datasets. The first dataset (nearly 0.4 billion observations) contains real-time traffic volume and speed data in two-minute intervals from over 1,500 remote traffic microwave sensors covering all major roads (freeways and expressways, but not the secondary roads) throughout Beijing for 2014. The data is aggregated on an hourly basis at the road-segment level, and we have 1,528 road-level records representing over 12 million observations ($1{,}528 \times 365 \times 24$) of average vehicle speed (measured in kilometers per hour) and traffic volume (measured in number of vehicles per hour). The locations of the traffic-monitoring sensors are shown in Figure 2.

[9] This would allow researchers to build a dynamic model of road pricing capturing the "bottleneck" or "bathtub" issues and to estimate them empirically for a certain roadway or a system of road networks. These are interesting areas for future research, especially as more of data required to estimate these models are becoming available in some major cities.

[10] The estimated setup cost for the system is about $440 million.

FIGURE 2. LOCATIONS OF THE ROAD TRAFFIC MONITORING SENSORS

Notes: Real-time traffic data (in two minute intervals) in Beijing were monitored by a network of over 1,500 road traffic monitoring sensors in 2014. The dots in the graph show the locations of the 1,528 road traffic monitoring sensors operated by the Beijing Police Bureau.

Figure 3 plots key variables pairwise: vehicle speed versus traffic density (traffic density is determined as the ratio of volume to speed per lane), vehicle speed versus traffic volume, and traffic volume versus traffic density. The speed-density relationship depicted in the top left plot is largely linear, which informs our empirical specification. Different colors and shapes in the plot represent different ring roads. The observations within the second ring road (the city core) exhibit the lowest speed but highest density, while the observations outside the fifth ring road show the highest speed but the lowest density.

The second dataset includes hourly weather variables obtained from the ISD-Lite dataset published by the National Oceanic and Atmospheric Administration (NOAA). The dataset contains hourly records of important weather controls for the average vehicle speed: wind speed (km/hour), visibility (km), air temperature (℃), relative humidity (percent), wind direction—dummy variables for each of 16 cardinal directions—and sky coverage, dummies for clear, scattered, broken, overcast, and obscured. It is important to control for the weather conditions, as they could affect both the travel demand and traffic speed on the road. For example, though some people may choose to postpone their trips due to rain, traffic at any given density may move slower on

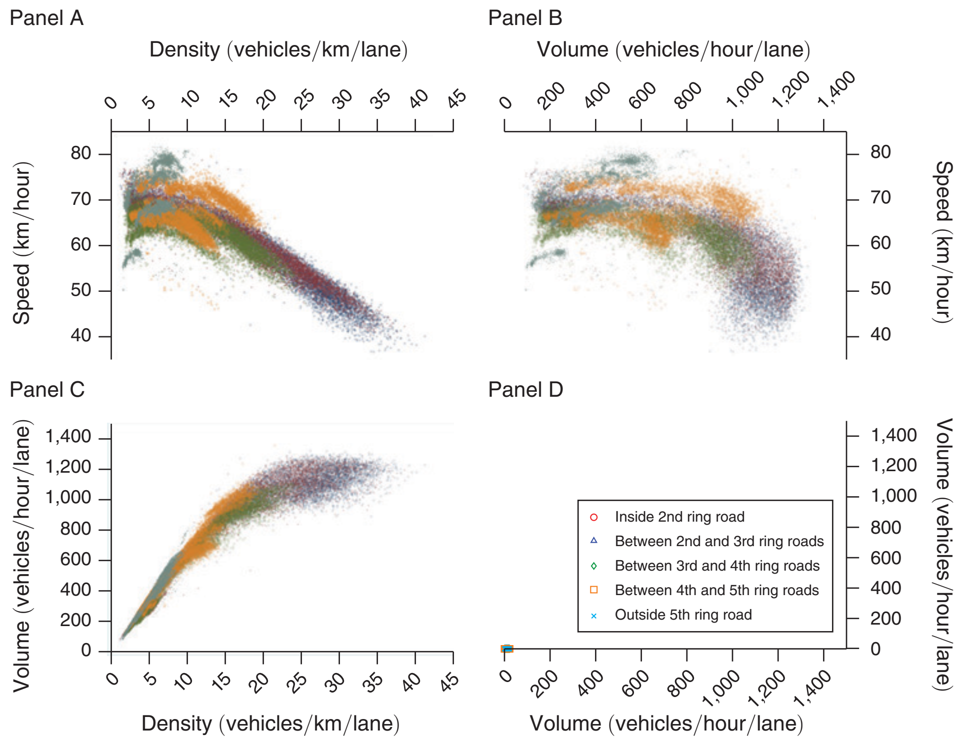

Figure 3. Speed-Density, Speed-Volume, Volume-Density Relationships

Note: The figure plots the observations of the variables pairwise: (panel A) average vehicle speed (km/hour) versus traffic density (vehicles/km/lane); (panel B) average vehicle speed (km/hour) versus traffic volume (vehicles/hour/lane); and (panel C) traffic volume (vehicles/hour/lane) versus traffic density (vehicles/km/lane).

a rainy day than on a non-rainy day. Table 2 presents summary statistics of the main variables from traffic monitoring sensors and weather variables.

III. Empirical Strategy

In this section, we first describe the empirical specification for the relationship between traffic speed and traffic density, which is the basis for estimating the MECC. We then discuss the empirical challenge in estimating the traffic speed-density relationship and our identification strategy.

A. *Speed-Density Relationship*

We specify the following relationship between traffic speed and density:

$$Speed_{it} = \alpha + \beta Density_{it} + Weather_t \gamma + Month_t + Hour_t + Week_t + Holiday_t + Road_i + Ring_i \times Hour_t + \varepsilon_{it}, \tag{5}$$

Table 2—Summary Statistics

Main variables	Mean	SD	Min	Max	Observations
Traffic variables (*road* × *hour*)					
Speed (*km*/*hour*)	67.34	14.45	0.41	165.70	12,088,092
Volume (*vehicles*/*hour*/*lane*)	564.98	451.66	0.25	2,000.00	12,088,092
Density (*vehicles*/*km*/*lane*)	9.55	10.66	0.06	206.44	12,088,092
Weather controls (*hourly*)					
Air temperature (°*C*)	13.80	11.51	−13.00	42.00	8,688
Wind speed (*m*/*s*)	2.60	1.93	0.00	20.00	8,688
Visibility (*km*)	10.81	8.65	0.00	30.08	8,760
Relative humidity (*percent*)	51.79	25.04	2.69	100.00	8,680
Wind direction dummies (*hourly*)					
Wind direction (*cat.*)	7.17	4.62	0.00	16.00	8,056
Calm = 0	0.02	0.12	0.00	1.00	8,760
North = 1	0.11	0.31	0.00	1.00	8,760
North-Northeast = 2	0.06	0.24	0.00	1.00	8,760
Northeast = 3	0.06	0.25	0.00	1.00	8,760
East-Northeast = 4	0.06	0.24	0.00	1.00	8,760
East = 5	0.06	0.24	0.00	1.00	8,760
East-Southeast = 6	0.05	0.23	0.00	1.00	8,760
Southeast = 7	0.06	0.23	0.00	1.00	8,760
South-Southeast = 8	0.06	0.24	0.00	1.00	8,760
South = 9	0.09	0.28	0.00	1.00	8,760
South-Southwest = 10	0.07	0.26	0.00	1.00	8,760
Southwest = 11	0.04	0.20	0.00	1.00	8,760
West-Southwest = 12	0.02	0.13	0.00	1.00	8,760
West = 13	0.02	0.13	0.00	1.00	8,760
West-Northwest = 14	0.03	0.16	0.00	1.00	8,760
Northwest = 15	0.04	0.20	0.00	1.00	8,760
North-Northwest = 16	0.06	0.23	0.00	1.00	8,760
Sky coverage dummies (*hourly*)					
Sky coverage (*cat.*)	2.07	1.62	0.00	4.00	8,745
Clear = 0	0.29	0.45	0.00	1.00	8,760
Scattered = 1	0.07	0.26	0.00	1.00	8,760
Broken = 2	0.24	0.43	0.00	1.00	8,760
Overcast = 3	0.05	0.23	0.00	1.00	8,760
Obscured = 4	0.34	0.47	0.00	1.00	8,760

Note: The unit of observation of the traffic variables is the road segment by hour.

where $Speed_{it}$ is the average speed (measured in km/hour) of a vehicle on road i at time t; $Density_{it}$ is the average traffic density (measured in vehicles/lane-km) on road i at time t; $Weather_t$ is a vector of hourly weather indicators including wind speed (km/hour), visibility (km), temperature (°C), relative humidity (%), wind direction dummies (16 cardinal directions), and sky coverage dummies—clear, scattered, broken, overcast, obscured; and ε_{it} is the unobserved, time-varying, and road-specific factors such as traffic accidents, road constructions, and big events.

We also include a full set of time fixed effects ($Month_t$, $Hour_t$, $Week_t$, and $Holiday_t$) and road-specific fixed effects ($Road_i$). The time fixed effects capture the unobserved, temporal, and common shocks associated with the traffic conditions during month of the year, hour of the day, day of the week, and holidays. Road-specific fixed effects control for road-related, time-invariant spatial factors such as road attributes that affect vehicle speed and traffic density.[11] The relationship could vary across types of vehicles (slow, wide vehicles may cause more congestion than fast, narrow vehicles) and across time (volume of vehicles varies at different times of day) and across different roads. To control for possible unobserved spatial and temporal differences in traffic conditions at different times of day, we interact the hour-of-the-day fixed effect with ring road dummies ($Ring_i$).

B. *Identification*

In the theoretical model presented above and in the transportation engineering literature, it is assumed that traffic density affects average vehicle speed and not vice versa, and that this relationship is deterministic. In practice, average vehicle speed can be affected by a variety of other (human and nonhuman) factors that researchers do not observe, as discussed above. In addition, both average vehicle speed and traffic density are realized simultaneously and affect each other. Road users decide whether they should take a trip or not based on the prevailing cost of the trip, which includes the level of traffic congestion on the road. If a road user observes or expects traffic congestion due to accidents, big events, or road construction, she might consider rescheduling the trip or changing route, which would in turn affect traffic density. This simultaneity would give rise to endogeneity in traffic density in equation (5).

To deal with the endogeneity, we leverage the once-a-week driving restriction policy for identification by arguing that this natural experiment generates exogenous variation in traffic density that in turn affects average vehicle speed. The policy satisfies the exogeneity assumption because it has a preset schedule and rotates exogenously regardless of the traffic congestion on a given day. On any given weekday, the policy restricts vehicles with license plates ending in one of two different digits (one and six, two and seven, three and eight, four and nine, and five and zero) from driving in areas inside the fifth ring road from 7:00 AM to 8:00 PM. The schedule rotates every 13 weeks.[12] Table 3 shows the policy schedule and the deterministic rotation schedule of the restricted digits on each weekday over time. Therefore, the driving restriction policy can be argued to be

[11] These road attributes include qualities (grade, surface), speed limits, engineering design (e.g., one-way, two way, number of lanes, restriction on vehicle types, bending, visibility), traffic lights, stops signs, sidewalks, bike lanes, number of intersections, and locational characteristics—street parking, pedestrian crossing, nearness to the school areas, business districts, parking areas, subway or bus stations.

[12] On October 11, 2008, the Beijing municipal government announced the implementation of a half-year trial of the driving restriction until April 10, 2009. During the trial period, the restricted day of the week for different numbers rotated every four weeks. The driving restriction was in force within (and including) the fifth ring road, from 6:00 AM to 9:00 PM. When this half-year trial ended, the government started a new round of the driving restriction lasting one year. This time, the restricted day of the week changed every 13 weeks, and the restriction area was narrowed to inside (and excluding) the fifth ring road and was in force from 7:00 AM to 8:0 0 PM. The third round of the driving restriction began immediately after the previous round on April 11, 2010. Since then, there have been no changes in the policy, and the restriction remains in force. Also, the penalty for violating the regulation has changed over time. Initially, drivers who violated the restriction were stopped and fined ¥100 (around $16.3) for the day. Since there would be no extra penalty if the violator was caught more than once per day, some people were willing to risk being caught and pay the daily fine. To improve enforcement, since 2011, the government has changed this daily penalty to a ¥100 penalty every three hours.

Table 3—Driving Restriction Policy Schedule

Policy period		Restricted pair of ending digits				
Starting date	Ending date	Mon	Tue	Wed	Thu	Fri
10/11/2008	11/10/2008	(1, 6)	(2, 7)	(3, 8)	(4, 9)	(5, 0)
11/11/2008	12/07/2008	(5, 0)	(1, 6)	(2, 7)	(3, 8)	(4, 9)
12/08/2008	01/04/2009	(4, 9)	(5, 0)	(1, 6)	(2, 7)	(3, 8)
01/05/2009	02/01/2009	(3, 8)	(4, 9)	(5, 0)	(1, 6)	(2, 7)
02/02/2009	03/01/2009	(2, 7)	(3, 8)	(4, 9)	(5, 0)	(1, 6)
03/02/2009	04/10/2009	(1, 6)	(2, 7)	(3, 8)	(4, 9)	(5, 0)
04/11/2009	07/10/2009	(5, 0)	(1, 6)	(2, 7)	(3, 8)	(4, 9)
⋮	⋮	⋮	⋮	⋮	⋮	⋮
04/08/2013	07/06/2013	(4, 9)	(5, 0)	(1, 6)	(2, 7)	(3, 8)
07/07/2013	10/05/2013	(3, 8)	(4, 9)	(5, 0)	(1, 6)	(2, 7)
10/06/2013	01/04/2014	(2, 7)	(3, 8)	(4, 9)	(5, 0)	(1, 6)
01/05/2014	04/11/2014	(1, 6)	(2, 7)	(3, 8)	(4, 9)	(5, 0)
04/14/2014	07/12/2014	(5, 0)	(1, 6)	(2, 7)	(3, 8)	(4, 9)
07/13/2014	10/11/2014	(4, 9)	(5, 0)	(1, 6)	(2, 7)	(3, 8)
10/12/2014	01/10/2015	(3, 8)	(4, 9)	(5, 0)	(1, 6)	(2, 7)
01/11/2015	04/10/2015	(2, 7)	(3, 8)	(4, 9)	(5, 0)	(1, 6)

Notes: Each column lists the pair of ending digits of license plates that were restricted on a certain weekday over different policy periods shown in column 1 (starting date) and column 2 (ending date). The policy applied to within (and including) the fifth ring road from 7:00 am–8:00 pm.

Source: Beijing Transportation Research Center

exogenous to the unobserved, time-varying, and road-specific shocks that affect traffic speed.

The policy is relevant and generates meaningful variation in traffic density. The distribution of the last digit of license plate number is determined by Chinese culture. Since the number four shares the same pronunciation with death, Chinese people avoid using the number in many aspects of the daily life, including door, mobile, floor, and license plate numbers. Only 2 percent of vehicles have license plates ending with 4, while the share is around 12 to 13 percent for vehicles with license plates ending with 6 or 8, which are traditionally considered lucky numbers. Table 4 shows the percentage of vehicles with license plates ending in each digit. Given that there are fewer vehicles with license plate numbers ending in the digit four, the driving restriction policy in Beijing unintentionally allows more vehicles on the road when four and nine are restricted.

Figure 4 illustrates the temporal and spatial effects of restricting vehicles with plate numbers ending in four or nine on the traffic density and average vehicle speed. Panel A of Figure 4 suggests that traffic density is higher when plate numbers ending in four or nine are restricted, and the magnitude of the effects varies over different times of day and across ring roads. The magnitude of the effect on traffic density of roads located inside the third or second ring roads increases

Table 4—Ending Digit Distribution of License Plate Numbers

Ending digit	2009	2010	2011	2012	2013	2014
1	10.0	9.9	9.9	9.9	9.9	9.8
2	10.2	9.9	9.9	9.9	9.9	9.8
3	9.9	9.7	9.6	9.6	9.6	9.7
4	2.8	2.3	2.2	1.9	1.7	1.5
5	10.4	10.4	10.5	10.6	10.7	10.7
6	11.7	12.1	12.3	12.3	12.3	12.4
7	10.1	10.1	10.2	10.3	10.4	10.4
8	12.7	13.0	12.9	12.9	12.8	12.9
9	11.6	12.1	12.2	12.3	12.2	12.3
0	10.6	10.5	10.5	10.5	10.5	10.5
(1, 6)	21.7	22.0	22.1	22.2	22.2	22.2
(2, 7)	20.3	20.1	20.0	20.1	20.3	20.2
(3, 8)	22.6	22.6	22.5	22.4	22.3	22.6
(4, 9)	14.4	14.4	14.4	14.1	13.9	13.9
(5, 0)	21.0	20.9	21.0	21.1	21.3	21.2

Notes: Each column shows the percent of vehicles with license plates ending with the digits 1, 2, 3, 4, 5, 6, 7, 8, 9, and 0 for each year. The ending digits of license plates are grouped into five pairs of (1, 6), (2, 7), (3, 8), (4, 9), and (5, 0). Each column sums to 100.

Source: Beijing Transportation Research Center

during the morning and evening peak hours while the effect is marginal during nonrestricted hours or when roads are located outside the fourth ring road. These results confirm that restricting vehicles with plate numbers ending in four or nine from driving inside the fifth ring road has a positive impact on the number of vehicles on the road and hence traffic density. On the other hand, during the days in which plate numbers ending in four or nine are restricted, traffic congestion is expected to be worse, as shown in panel B of Figure 4.

This policy hence induces plausibly exogenous variation in traffic density that is not correlated with the unobserved, time-varying, and road-specific shocks to average vehicle speed. A potential concern with the validity of the instrument is that owners of vehicles with license plates ending in four or nine could be different from other owners in their driving behavior, which could lead to a different relationship between traffic density and speed. If this were the case, one would need to include the policy variable in the speed-density relationship, hence invalidating the research design. We argue that this is unlikely to be the case: for the two groups to have differing speed-density relationships, one group would need to drive in a way that is systematically different from the other group (e.g., slower or faster) holding all else constant. Based on the 2010 Household Travel Survey in Beijing, we compare the observed demographics of these two types of vehicle owners in Table 5. There

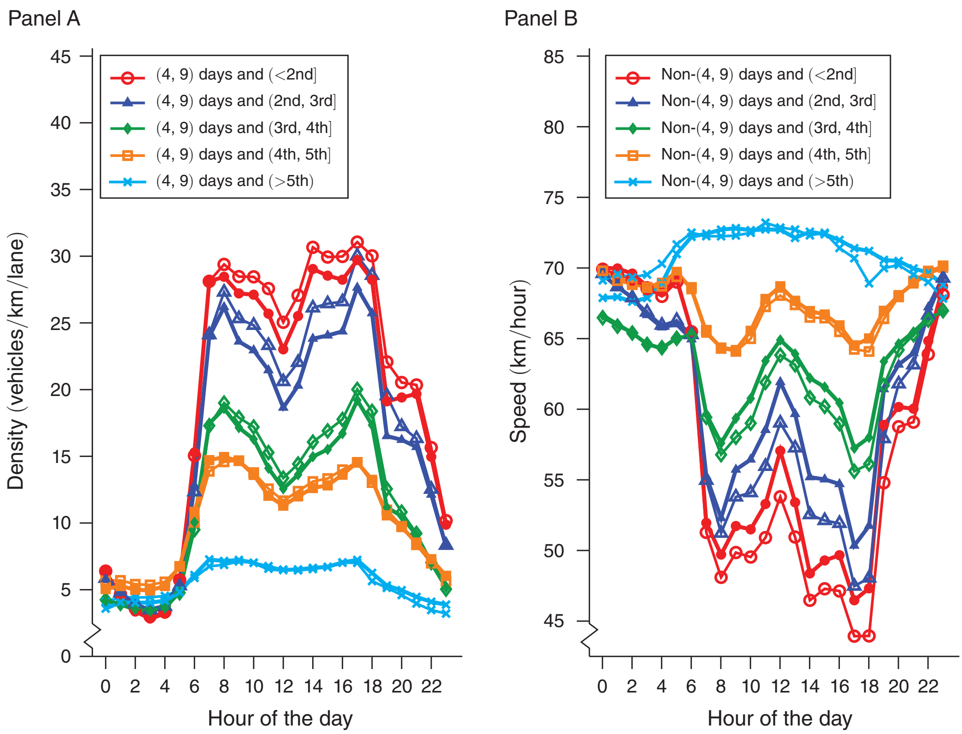

Figure 4. Average Traffic Density and Average Speed of Vehicles

Notes: The figure compares the hourly variation of average traffic density (left) and average vehicle speed (right) between the (4, 9) days and other weekdays for different locations (based on the ring roads). The data shows that the average traffic density is denser (or average speed is slower) during the days on which the plate numbers ending with four or nine are restricted and the difference varies over different times of day and across ring roads.

Table 5—Balance Test between (4, 9) Owners and Non-(4, 9) Owners

Variable	Ending digits				
	non-(4, 9)	(4, 9)	Difference	SE	p-value
Household income (seven categories)	1.97	1.96	0.01	0.06	0.86
House ownership (six types)	1.47	1.46	0.02	0.03	0.60
House size (m^2)	89.36	88.85	0.51	1.48	0.73
Annual travel distance (1,000 km)	15.35	15.66	−0.31	0.83	0.71

Notes: The data source is the 2010 Household Travel Survey in Beijing, a representative household survey on travel behavior among 46,899 households. Household income and house ownership are seven and six categorical variables respectively. Distance is annual travel distance measured in 1,000 kilometers. There are 1,975 vehicles that have four or nine as the last digit of their license plate, and 12,611 vehicles without four or nine as the last digit of the license plate.

is no statistically significant and economically meaningful difference between the two groups in household income, house ownership type, household size, and annual vehicle kilometers traveled.

The first step of the IV strategy specifies traffic density as a function of the policy variables (IVs) and other controls,

$$\begin{aligned} (6) \qquad Density_{it} \;=\; & Tail49_t + Tail49_t \times Ring_i + Tail49_t \times Hour_t \\ & + \; Weather_t \delta + Month_t + Hour_t \\ & + \; Week_t + Holiday_t + Road_i + Ring_i \times Hour_t + \xi_{it}, \end{aligned}$$

where the first three terms serve as exclusion restrictions. The term $Tail49_t$ is a dummy variable indicating if the digits four and nine are restricted on a given day. Although the driving restriction policy itself is applied uniformly inside the fifth ring road during hours from 7:00 AM to 8:00 PM, the policy does have heterogeneous effects across space and time of day as shown in Figure 4; the impact is more salient in the city center (e.g., within the second or third ring road). To capture the spatial and temporal variations in the policy impact, we interact the dummy $Tail49_t$ with hourly dummies ($Hour_t$) and ring road dummies ($Ring_i$).

IV. Estimation Results

We first present the OLS and IV results for equation (5) and the estimates of MECC.

A. *Regression Results*

Table 6 presents the results of equation (5) for five different specifications where we add more control variables successively. All model specifications are estimated on the full sample using the OLS estimation. Standard errors were clustered at the road-segment level. The coefficient estimate on the key regressor, traffic density, ranges from −0.67 to −0.75 in the five regressions with the last two regressions producing the smallest estimates. The coefficients of the other variables are all intuitively signed and consistent with the traffic congestion literature. We use the last regression as our preferred specification because the inclusion of road-segment fixed effects controls for time-invariant spatial factors such as road attributes and locational characteristics that could affect both average vehicle speed and traffic density.

Before discussing the IV results, we first present evidence on how the driving restriction policy affects traffic density, the relevance assumption for a valid IV. Figure 5 illustrates the effect of restricting vehicles with plate numbers ending in four or nine on traffic density. It shows that when four and nine are restricted, the traffic density (or number of vehicles per lane-km) on the roads located inside the third ring road is higher by around 1.5 vehicles, and the magnitude of the effects decreases as one moves further from the city center, such as between the third and fourth ring roads, between the fourth and the fifth ring roads, and outside the fifth ring road. Also, the effect is much stronger during the morning peak hours (6:00 AM to 9:00 AM) and evening peak hours (5:00 PM to 7:00 PM), while the effect is marginal

TABLE 6—TRAFFIC SPEED AND DENSITY FROM OLS REGRESSIONS

	Dependent variable: Speed (km/hour)				
	(1)	(2)	(3)	(4)	(5)
Density	−0.685	−0.692	−0.752	−0.668	−0.671
	(0.013)	(0.013)	(0.014)	(0.018)	(0.018)
Temperature		−0.091	−0.059	−0.056	−0.064
		(0.006)	(0.003)	(0.003)	(0.003)
Wind speed		−0.028	−0.011	−0.008	−0.008
		(0.003)	(0.002)	(0.002)	(0.002)
Visibility		−0.062	0.007	0.008	0.007
		(0.002)	(0.001)	(0.001)	(0.000)
Humidity		−0.047	−0.003	−0.003	−0.004
		(0.002)	(0.000)	(0.000)	(0.000)
Constant	73.882	78.275	76.045	78.053	75.703
	(0.221)	(0.264)	(0.242)	(0.906)	(0.202)
Wind directions	No	Yes	Yes	Yes	Yes
Sky coverage	No	Yes	Yes	Yes	Yes
Month fixed effects	No	No	Yes	Yes	Yes
Week fixed effects	No	No	Yes	Yes	Yes
Hour fixed effects	No	No	Yes	Yes	Yes
Ring road fixed effects	No	No	No	Yes	No
Ring × Hour fixed effects	No	No	No	Yes	Yes
Road specific fixed effects	No	No	No	No	Yes
Observations	12,088,092	11,977,139	11,977,139	11,977,139	11,977,139
R^2	0.26	0.27	0.32	0.34	0.32

Notes: Each column reports results from an OLS regression where the dependent variable is average vehicle speed (km/hour) and the key explanatory variable is traffic density (vehicles/km/lane). The unit of observation is road-hour. The weather control include hourly variables: air temperature (°C), wind speed (km/hour), visibility (km), relative humidity (percent), and two sets of dummies for wind direction and sky coverage. The time fixed effects include day of week, month of year, hour of day, and holiday dummies. Ring road fixed effects include dummies for ring roads. Road-segment fixed effects include segments (or monitoring sensors), and the interactions terms include the interactions of ring road dummies with hour of day fixed effects (Ring × Hour). Parentheses contain standard errors clustered by road segment.

(negative in some cases) during the nonrestricted hours or when roads are located outside the fifth ring road.

Table 7 presents the first-stage results for the five specifications listed in Table 6 where different sets of fixed effects are included. The IVs include the policy variable and its interactions with the ring-road-specific and hour of day fixed effects. The policy variable is a dummy variable indicating if four and nine are restricted on a given day. The variable itself has a positive and statistically significant effect on traffic density, consistent with the fact that vehicles having the last digit four or nine account for a smaller share of vehicles relative to other combinations. Therefore, there are more vehicles on the road when vehicles with license plates ending in four or nine are restricted from driving. The coefficient estimates on the interactions terms tend to be statistically significant, suggesting that the effects of the driving restriction policy on traffic density vary across different ring roads and over time. For all specifications, the joint F-statistics on the excluded instruments are above 40. The driving restriction policy therefore provides exogenous variations in traffic density that we can leverage to identify a causal effect of traffic density on average vehicle speed.

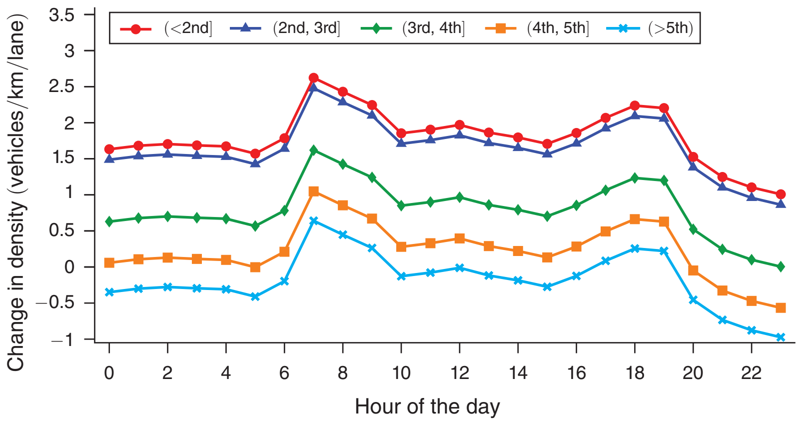

FIGURE 5. EFFECT OF RESTRICTING (4, 9) ON TRAFFIC DENSITY

Notes: The figure shows the effect of restricting vehicles with plate numbers ending with four or nine on the traffic density. When the plate numbers ending with four or nine are restricted, the traffic density on the road located inside the third ring road is higher by around 1.5 vehicles, and the magnitude of the effects decreases as one moves outward from the city center, such as between the third and fourth ring roads, fourth and the fifth ring roads, and outside the fifth ring road.

TABLE 7—FIRST-STAGE RESULTS ON TRAFFIC DENSITY

	Dependent variable: Density (vehicles/lane-km)				
	(1)	(2)	(3)	(4)	(5)
(4,9) days	5.331	5.508	10.185	1.569	1.241
(4,9) × (second, third]	−2.418	−2.464	−2.474	−0.117	−0.034
(4,9) × (third, fourth]	−8.044	−8.075	−8.091	−0.969	−0.715
(4,9) × (fourth, fifth]	−9.538	−9.518	−9.526	−1.575	−1.224
(4,9) × (>fifth)	−14.249	−14.270	−14.299	−1.974	−1.478
(4,9) days × Hour	Yes	Yes	Yes	Yes	Yes
Weather controls	No	Yes	Yes	Yes	Yes
Time fixed effects	No	No	Yes	Yes	Yes
Ring road fixed effects	No	No	No	Yes	No
Ring × Hour fixed effects	No	No	No	Yes	Yes
Road-specific fixed effects	No	No	No	No	Yes
Observations	12,088,092	11,977,139	11,977,139	11,977,139	11,977,139
R^2	0.04	0.05	0.15	0.35	0.30
F-statistic on IVs	126.60	87.57	70.07	58.93	53.09

Notes: The dependent variable is traffic density (vehicles/km/lane). Each column reports OLS coefficient estimates on the IVs—the interactions of the indicator for (4, 9) with ring-road-specific and hour of day fixed effects. The unit of observation is road-hour. The weather controls include hourly variables: air temperature (℃), wind speed (km/hour), visibility (km), relative humidity (percent), and two sets of dummies for wind direction and sky coverage. The time fixed effects include day of week, month of year, hour of day, and holiday dummies. Ring road fixed effects include dummies for ring roads. Road-segment fixed effects include segments (or monitoring sensors). Standard errors are clustered by road segment.

Table 8 presents the estimates from the IV regressions for equation (5) and the five specifications corresponding to the OLS specifications in Table 6. Compared to the OLS results in Table 6, the coefficient estimates in Table 8 are fairly similar except for those on traffic density. The coefficient estimates on traffic density from the IV regressions are negative and statistically significant, ranging from −0.814

Table 8—Traffic Speed and Density from IV Regressions

	Dependent variable: Speed (km/hour)				
	(1)	(2)	(3)	(4)	(5)
Density	−0.814	−0.822	−1.024	−1.136	−1.098
	(0.024)	(0.025)	(0.038)	(0.043)	(0.044)
Temperature		−0.093	−0.066	−0.070	−0.077
		(0.006)	(0.003)	(0.003)	(0.003)
Wind speed		−0.040	−0.019	−0.022	−0.021
		(0.004)	(0.002)	(0.002)	(0.002)
Visibility		−0.070	0.005	0.004	0.003
		(0.002)	(0.001)	(0.001)	(0.001)
Humidity		−0.048	−0.003	−0.003	−0.005
		(0.002)	(0.000)	(0.000)	(0.000)
Constant	75.110	79.670	77.433	81.397	77.903
	(0.265)	(0.333)	(0.282)	(0.974)	(0.329)
Wind directions	No	Yes	Yes	Yes	Yes
Sky coverage	No	Yes	Yes	Yes	Yes
Month fixed effects	No	No	Yes	Yes	Yes
Week fixed effects	No	No	Yes	Yes	Yes
Hour fixed effects	No	No	Yes	Yes	Yes
Ring road fixed effects	No	No	No	Yes	No
Ring × Hour fixed effects	No	No	No	Yes	Yes
Road specific fixed effects	No	No	No	No	Yes
Observations	12,088,092	11,977,139	11,977,139	11,977,139	11,977,139
R^2	0.25	0.26	0.29	0.27	0.25

Notes: Each column reports results from a 2SLS regression where the dependent variable is average vehicle speed (km/hour) and the key explanatory variable is traffic density (vehicles/km/lane). The IVs are the policy indicator for (4, 9) and its interactions with ring-road-specific and hour of day fixed effects. The unit of observation is road-hour. The weather control include hourly variables: air temperature (ºC), wind speed (km/hour), visibility (km), relative humidity (percent), and two sets of dummies for wind direction and sky coverage. The time fixed effects include day of week, month of year, hour of day, and holiday dummies. Ring road fixed effects include dummies for ring roads. Road-segment fixed effects include segments (or monitoring sensors), and the interactions terms include the interactions of ring road dummies with hour of day fixed effects (Ring × Hour). Parentheses contain standard errors clustered by road segment.

to −1.136. This is much larger (in magnitude) than the estimates from the OLS regressions.

Specification (5) in Table 8, the preferred specification, shows that a one unit increase in traffic density (i.e., an additional vehicle per lane-km) would result in approximately a one unit (i.e., 1.1 km/hour) decrease in the average vehicle speed. This is 1.6 times larger (in magnitude) than the effect from the OLS regression. The comparison suggests that the OLS results are biased toward zero. The bias could be due to unobservables such as big events, accidents, or road construction that reduce traffic density (for example due to avoidance behavior) and at the same time reduce traffic speed. These unobservables would attenuate the impacts of traffic density on vehicle speed, biasing the coefficient estimate toward zero.[13]

[13] The relationship between travel time and traffic volume is commonly described in the traffic engineering literature by the US Bureau of Public Roads (BPR) formula, $T - T^f = \alpha V^\beta$, where T^f refers to the travel time under free-flow speed. The literature commonly uses $\alpha = 0.15$ and $\beta = 4.0$ (or between 2.5 to 5) as in Parry et al. (2014) and Parry (2009). Based on the estimated relationship between traffic speed and density, we can recover the relationship between travel time and volume under ordinary congestion as in equation (4). Taking the estimated

B. *Marginal External Cost of Ordinary Congestion*

As defined in equation (2), the MECC in terms of traffic volume, $MECC_{it}(V)$ (measured in Yuan per vehicle-km), is

$$(7) \qquad MECC_{it}(V) = o \cdot VOT \cdot Traveltime_{it} \cdot \frac{\epsilon_{it}}{1 - \epsilon_{it}}, \quad 0 \leq \epsilon_{it} < 1,$$

where ϵ_{it} is elasticity of average vehicle speed with respect to traffic density on road i at time t:

$$\epsilon_{it} = -\frac{\partial Speed}{\partial Density} \cdot \frac{Density_{it}}{Speed_{it}} = -\hat{\beta} \cdot \frac{Density_{it}}{Speed_{it}}.$$

The function $MECC_{it}(V)$ captures the external cost from a one-unit increase in traffic volume (or one unit increase in vehicle-kilometers traveled). To calculate $MECC_{it}$, we use the parameter estimate on traffic density, $\hat{\beta} = -1.098$, from the 2SLS regression. The average number of passengers per vehicle (o) is assumed to be 1.34 (persons/vehicle) based on the 2010 Beijing Household Travel Survey. The value of time (VOT) is assumed to be 50 percent of the market wage in Beijing, which is ¥62.98 per hour ($9.5 per hour) based on the monthly market wage of ¥10,076.8 per month.[14]

Figure 6 depicts the MECC estimates based on equation (7) or the MECC in terms of traffic volume. The top figure shows that the MECC curve increases nonlinearly as traffic volume increases under ordinary congestion. As we can see from the bottom of Figure 6, when traffic volume is around 1,000 vehicles per lane-hour, MECC is estimated to be ¥0.30 per vehicle-km during ordinary congestion (when the average speed is around 60 km/hour). Since traffic conditions vary over time and space, the MECC varies accordingly. The bottom panel of Figure 6 illustrates the temporal and spatial pattern of the MECC in Beijing. Based on the estimate from the 2SLS regression, the average MECC per extra vehicle-kilometers traveled is ¥0.46 ($0.07). It is about ¥1.98 ($0.30) when the average traffic volume is around 1,200 vehicles per lane-hour (or when average traffic density is around 29 vehicles per km-lane) inside the second ring road during the evening peak hours. This is over 60 percent higher than what the estimate from the OLS regression would imply (¥1.2 or $0.19). The MECC estimates based on 2SLS for locations between the second and fifth ring roads during the peak hours are often more than twice as large as those based on OLS estimates.

Our estimates are based on the data sample containing freeways and expressways but not secondary roads. According to the 2015 Beijing Transportation Annual Report, the average speed of vehicles on arterial roads during the workday peak

relationship between traffic speed and density as a true data-generating process allows us to generate a sample of traffic volume and corresponding travel time. Our estimates based on the simulated sample suggest that the value for β should be about 7.6 instead of 4 (or between 2.5 to 5.0) under ordinary traffic congestion. The estimate of 7.6 implies that the calibrated BPR relationship tends to underestimate the ordinary congestion.

[14] The literature often suggests a rule of thumb for the value of time: half of the market wage for automobile travel in Canada, France, the United Kingdom, and the United States (Small and Verhoef 2007, Parry and Small 2009).

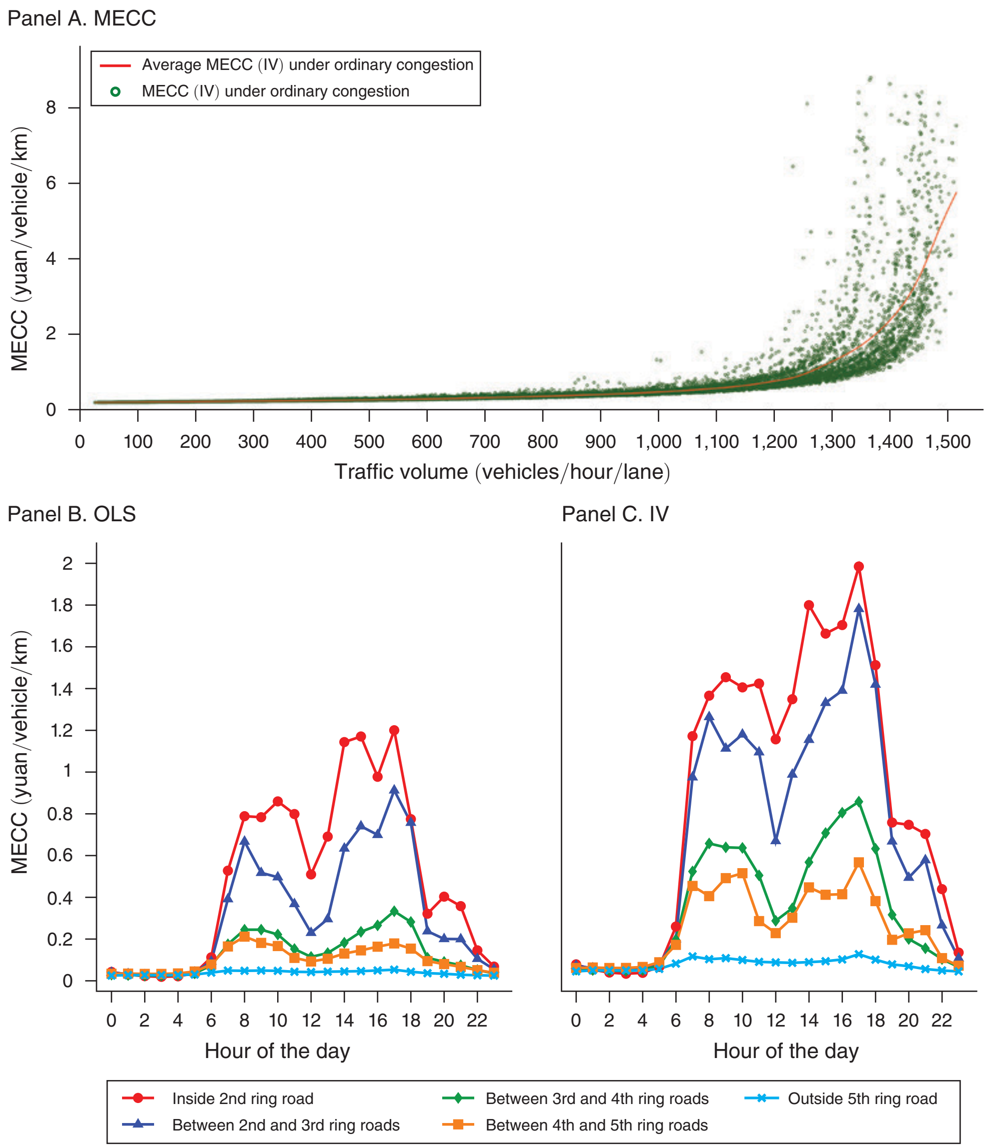

Figure 6. Marginal External Cost of Congestion

Notes: The figures depict the MECC based on the empirical speed-density relationship. The top figure shows the MECC curves from the sample averages for all observations over the levels of traffic volumes under ordinary congestion. The bottom panels show the MECC estimates over time and space with the left plot showing the OLS estimates and the right plot showing the 2SLS estimates.

hours was 23 km/hour in 2014, which implies that the average traffic density is around 53 vehicles per km-lane. MECC during the peak hours would be about ¥7.4 ($1.16) per vehicle-km under this condition.[15]

[15] Our data shows a higher average speed than that of the 2015 Beijing Transportation Annual Report for several reasons. First, the average speed from the report is based on the GPS data from a large number of taxis. Taxis

While our empirical analysis focuses on Beijing, the estimated speed-density relationship as well as the estimation strategy could also be applied to other major cities in China. The relationship between travel time and traffic volume is governed by two parameters, $\hat{\alpha}$ and $\hat{\beta}$. The value of $\hat{\alpha}$ can be adjusted to match the free-flow speed in other cities, and with these estimates one can approximate the relationship between travel time and traffic density in other cities. Information on average travel time or average vehicle speed is readily available (e.g., Google Maps), and the information would allow us to calculate MECC based on equation (7).[16]

V. Road Pricing and Impacts

In this section, we use the MECC estimates to inform the design of road pricing schemes and quantify the impacts of different policy designs on traffic speed, social welfare, and government revenue.

A. *Optimal Congestion Charges and Travel Demand*

The market failure under ordinary congestion and conventional diagram of optimal pricing are demonstrated in panel A of Figure 7. The inverse demand curve represents travelers' willingness to pay curve or the marginal private benefit $P(V_{it})$ from a trip using a road.[17] The travel decision is dictated by the $P(V_{it})$ and the average social cost $ASC(V_{it})$ from the travel. The difference between the $MSC(V_{it})$ and $ASC(V_{it})$ is the MECC $MECC(V_{it})$. The unregulated equilibrium occurs at the intersection of the $ASC(V_{it})$ and the MPB, resulting in an equilibrium price of p^0 and a traffic volume of V_{it}^0. At this point, there is an excess burden of congestion, defined as the net social cost of unregulated use of the road network. Under the road pricing, the socially optimal level of traffic volume V_{it}^* can be achieved with the optimal congestion charge τ_{it}^* as a Pigouvian tax to internalize the externality. The gain in social surplus from road pricing is shown by the shaded area with traffic volume being reduced from V_{it}^0 to V_{it}^*.

To conduct welfare analysis, we need ASC and MSC as functions of traffic volume, as well as the inverse demand curve for travel $P(V_{it})$. Our empirical analysis of the speed-density relationship allows us to estimate ASC and MSC. However, estimating travel demand would necessitate a different framework, as well as data on travel behavior and gasoline consumption. In particular, one would need a variable that captures the travel cost (such as fuel cost) as well as an IV to generate exogenous variation in the travel cost. While there is a large literature on travel demand in the United States using either household-level or aggregate data (Small and Van Dender 2007; Bento et al. 2009; Li, Linn, and Muehlegger 2014), we do not

stop more frequently to pick up and drop off passengers. Second, the traffic-monitoring sensors tend to be on road segments that are less congested. The real speed on the road is likely between what is reported from the Beijing Transportation Annual Report and our estimates.

[16] One should be cautious in adopting our speed-density estimates for other cities as the relationship may vary due to, for example, differences in road conditions and urban structure. Nevertheless, the framework can be adopted in other cities based on the increasingly available data as in this paper.

[17] We assume that the inverse demand curve only reflects private benefits from travel and there are no external social benefits.

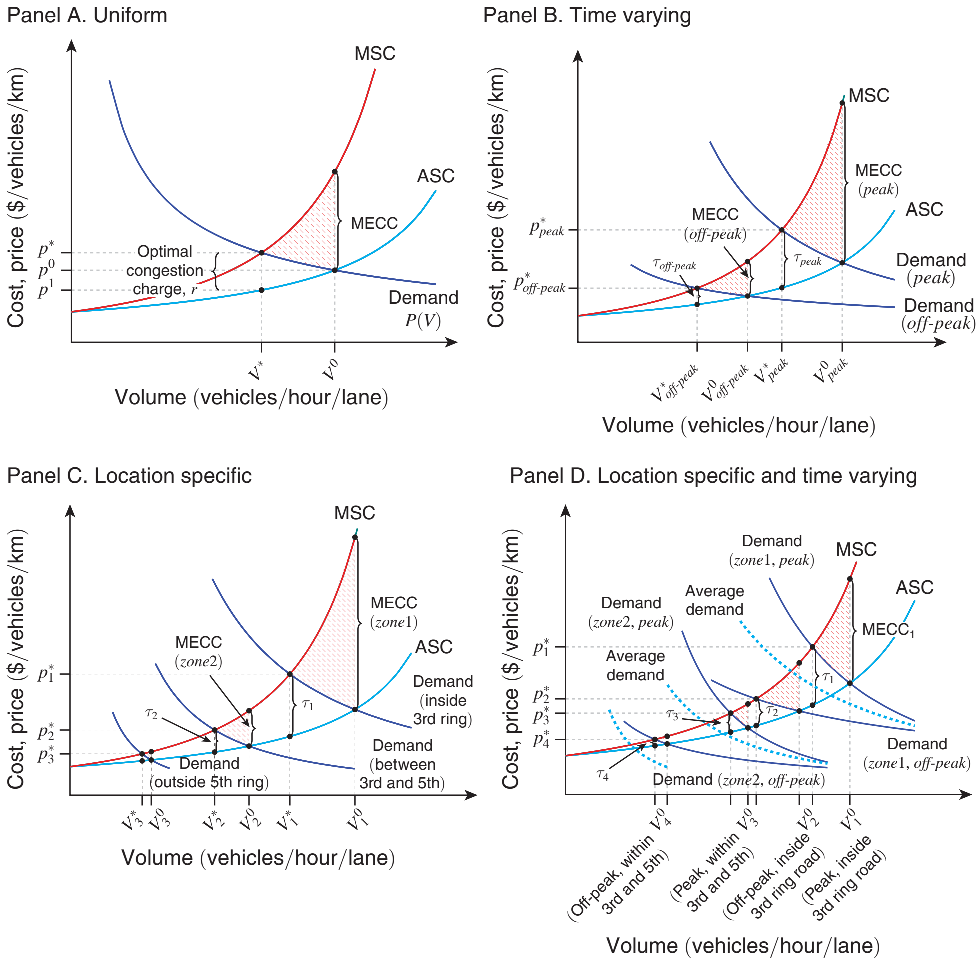

Figure 7. Road Pricing Schemes

Notes: The figure shows four different schemes of road pricing between the hours of 6:00 AM and 11:00 PM on workdays. Panel A illustrates a basic road pricing design with a uniform congestion charge, which is constant across different areas and over different hours. Panel B illustrates a time-varying congestion charge with two time periods (peak hours and off-peak), but that is constant across locations. Panel C depicts a location-specific congestion charge that has three zones (1 to 3) but is not time-specific. Panel D shows a time-varying and location-specific charging scheme with two time periods and two zones.

have data that allow us to credibly estimate the travel demand function for Beijing. Instead, we calibrate the travel demand function following this parsimonious form:

$$V_{it} = A_{it} P_t^{-\eta}, \tag{8}$$

where V_{it} is average traffic volume on road i at time t (or vehicle-kilometers traveled, VKT, measured in vehicles-km/hour); P_t is travel cost; A_{it} is a demand shifter; and $-\eta$ is a long-run elasticity of traffic volume with respect to travel cost. There is a large empirical literature on the overall responsiveness of traffic volume to travel cost, usually measured by fuel cost.

Previous studies have estimated the elasticity of VKT with respect to fuel costs or gasoline prices to be between −0.1 and −0.3 (Parry 2009; Goodwin 1992; Goodwin, Dargay, and Hanly 2004; Small and Van Dender 2007; Bento et al. 2009; Knittel and Sandler 2011). We use −0.1 as the long-run elasticity of VKT with respect to travel costs during the peak period and −0.3 during the off-peak period.[18] Travel should be less elastic during the peak hours because they are usually related to work trips rather than discretionary trips such as shopping or dining during off-peak hours. The average of these two estimates of −0.2 is used as the long-run VKT elasticity with respect to fuel cost for the daily average. Following Anderson (2014), we convert these elasticities to the long-run traffic volume elasticities with respect to travel costs (i.e., fuel cost plus time costs) of −0.44 during the peak hours, −1.32 during the off-peak hours, and −0.88 during the day.[19] We perform robustness checks with respect to the elasticities of travel demand in the next section.

Given the average density on road i at time t, we calculate the expected traffic volume (V_{it}^0 in panel A of Figure 7, the equilibrium traffic volume without road pricing) based on the empirical speed-density relationship. The demand shifter, denoted by A_{it}, is calculated by finding the value of A_{it} that satisfies the equilibrium without the congestion charge at which $P\left(V_{it}^0\right) = ASC\left(V_{it}^0\right)$. We then solve for the long-run equilibrium traffic volume, V_{it}^*, at which $P\left(V_{it}^*\right) = MSC\left(V_{it}^*\right)$, using the demand equation in (8) with $\eta = 0.44$ during the peak hours, $\eta = 1.32$ during the off-peak hours, and $\eta = 0.88$ during the day.

B. *Road Pricing Schemes and Impacts*

Although the estimation of the density-speed relationship is based on the data at the road segment-hour level, our analysis on MECC and road pricing is conducted at a more aggregated level in both temporal and spatial dimensions, e.g., ring roads by peak/off-peak periods rather than individual road segments by hours of the day. Once the data are averaged over road segments within certain a region (e.g., second to third ring roads) and time periods (e.g., off-peak), Figures 3, 4, and 6 show that we do not observe hyper-congestion.

We examine four different schemes of road pricing between the hours from 6:00 AM to 11:00 PM on workdays while assuming no congestion charges outside of this window as illustrated in Figure 7. These four schemes offer varying degrees of approximation to the first-best road pricing policy. Panel A illustrates the first scheme with a uniform congestion charge. The second congestion charge scheme is time-varying but constant across locations as shown in panel B. For simplicity, we consider only two time periods: peak hours (the morning peak hours from 7:00 AM to 9:00 AM and the evening peak hours from 5:00 PM to 7:00 PM) and off-peak hours. The third congestion charge scheme is location-specific but not time-varying in panel C. In this case, the congestion charge varies across three zones: zone one

[18] When faced with increasing traffic congestion, individuals may adopt strategies in the long run that are not feasible in the short to medium run as suggested by the "fundamental law of road congestion." Potential adaptation strategies that may not be available in the short run include increasing telecommuting, ride sharing, moving closer to work or school, work schedule changes, and changing the job location entirely.

[19] Delay-penalized time costs are approximately 240 percent higher than fuel costs.

Table 9—MECC, Total Value of Lost Time, and Congestion Charges

	Uniform	Time varying		
	All day	Peak	Off-peak	Total
Panel A. $MECC(V^0)$ without road pricing (yuan/vehicle-km)				
Uniform				
Beijing	0.15	0.18	0.14	
Location specific				
within third ring	0.53	0.82	0.47	
between third and fifth	0.18	0.24	0.16	
outside fifth ring	0.06	0.07	0.06	
Panel B. Total value of lost time (annualized, billion yuan)				
Uniform				
Beijing	7.5	3.3	4.5	7.8
Location specific				
within third ring	4.9	2.2	3.0	5.2
between third and fifth	3.8	1.8	2.2	4.0
outside fifth ring	1.0	0.4	0.6	1.1
Total	9.8	4.4	5.8	10.2
Panel C. Road pricing $\tau^* = MECC(V^*)$ (yuan/vehicle-km)				
Uniform				
Beijing	0.12	0.15	0.10	
Location specific				
within third ring	0.26	0.39	0.22	
between third and fifth	0.13	0.19	0.12	
outside fifth ring	0.06	0.07	0.05	

Notes: Panel A reports the average MECC (yuan/vehicle-km) under the four different road pricing schemes: (i) uniform across time and space, (ii) uniform across space and time varying, (iii) location specific and uniform across time, and (iv) time varying and location specific. Panel B reports the total value of lost time (TVLT) without road pricing. Panel C reports the optimal congestion charges.

(central Beijing or inside the third ring road), zone two (between the third and the fifth ring roads), and zone three (outside the fifth ring road). The fourth road pricing scheme illustrated in panel D uses time-varying and location-specific charges with two time periods (peak and off peak) and three zones (zones one to three).[20]

Panel A in Table 9 reports the average MECC in terms of traffic volume without road pricing. The average MECC from 6:00 AM to 11:00 PM throughout Beijing is estimated to be ¥0.15/vehicle-km. There is a large heterogeneity across time and especially space with a high of ¥0.82/vehicle-km within the third ring road during the peak hours and a low of ¥0.06/vehicle-km outside the fifth ring road during the off-peak hours. Panel B in Table 9 presents the total value of lost time (*TVLT*) that drivers impose on other road users (relative to the free-flow speed scenario).[21] It amounts to nearly ¥10 billion in Beijing in 2014, or nearly half percent of Beijing's

[20] For simplicity, panel D of Figure 7 illustrates a time-varying and location-specific charging scheme with two time periods and only two zones. However, the fourth road pricing scheme used in the welfare analyses has two time periods and three zones.

[21] Total value of lost time (*TVLT*) in the absence of road pricing under ordinary congestion is defined as

$$TVLT = \int_{V^f}^{V^0} \left[MSC(\nu) - p^f\right] d\nu = \int_{V^f}^{V^0} MSC(\nu)\, d\nu - p^f \cdot \left(V^0 - V^f\right),$$

where V^f is the free-flow traffic volume and p^f is the free-flow *ASC* under ordinary congestion.

GDP. These costs, which are not internalized when drivers make their trip decisions, are a substantial loss to society, confirming Vickrey's (1963) statement that the lack of an adequate road pricing policy in urban transportation is conducive to waste.

Panel C in Table 9 reports the optimal congestion charges under the four road pricing schemes. The uniform pricing scheme suggests a congestion charge of 12 cents per kilometer while the time-varying scheme calls for a congestion charge of 15 cents for peak hours and 10 cents for non-peak hours. The larger heterogeneity emerges under the location-specific scheme, which suggests a charge of 26, 13, and 6 cents in zones one, two, and three, respectively. This is driven by the fact that the heterogeneity in MECC estimates is larger in the spatial dimension than the temporal dimension. The time-varying and spatial-specific scheme suggests congestion charges from 5 to 39 cents, with the highest being within the third ring road during peak hours and the lowest being outside the fifth ring road during non-peak hours.

Table 10 presents the impacts of road pricing on traffic speed in panel A, social welfare in panel B, and revenue in panel C. Under the first two schemes, the speed improves by less than 3 percent because neither scheme accounts for the large heterogeneity in congestion across space. The third scheme of location-specific road pricing has a more significant effect on speed improvement: the average speed in central Beijing (within the third ring road) is predicted to rise by 11.4 percent. The fourth scheme achieves roughly the same results as the third scheme: it would increase the average speed by 0.5–11.4 percent with the largest improvement occurring within the third ring road. The predicted impacts of road pricing on congestion are at the lower end of the estimates from the road pricing policies in several cities (Singapore, London, Milan, and Stockholm) where it has been shown that road pricing improves speed by 10–30 percent (Anas and Lindsey 2011). Given that the traffic speed in our data is higher than those reported by the 2015 Beijing Transportation Annual Report as we discussed above, our estimates of both optimal congestion charges and their resulting speed improvements could serve as a lower bound of the optimal congestion charges and their potential speed improvements from road pricing.

Panel B suggests that the total welfare gain ranges from ¥0.7 billion under the uniform scheme to ¥1.5 billion under the location-specific and time-varying scheme.[22] It is interesting to note that the third and fourth schemes, which take into account heterogeneity across locations, have much larger welfare gains than the first two schemes. The increase in welfare gain from adding the time-varying component (e.g., from the first to the second scheme and from the third to the fourth scheme) is much smaller than that from adding the location-specific component (e.g., from the first to the third scheme and from the second to the fourth scheme). These comparisons are consistent with the fact that the heterogeneity in traffic congestion is

[22] The welfare gain is the deadweight loss resulting from the deviation from the socially optimal outcome in the absence of the corrective policies for ordinary congestion. It is defined as

$$\Delta W = \Delta TSC - \Delta CS = \int_{V^*}^{V^0} \left[MSC(\nu) - P(\nu) \right] d\nu,$$

where V^f is the free-flow traffic volume and p^f is the free-flow ASC under ordinary congestion.

Table 10—Congestion Reduction, Welfare, and Revenue

	Uniform	Time varying		
	All day	Peak	Off-peak	Total
Panel A. Increase in average speed (percent)				
Uniform				
Beijing	2.7	2.2	3.2	
Location specific				
within third ring	10.9	11.4	11.4	
between third and fifth	3.4	3.1	3.9	
outside fifth ring	0.8	0.5	1.0	
Panel B. Welfare gain (annualized, billion yuan)				
Uniform				
Beijing	0.7	0.2	0.5	0.7
Location specific				
within third ring	1.0	0.4	0.6	1.1
between third and fifth	0.4	0.1	0.3	0.4
outside fifth ring	0.0	0.0	0.0	0.1
Total	1.4	0.6	1.0	1.5
Panel C. Tax revenue (annualized, billion yuan)				
Uniform				
Beijing	7.6	4.0	4.1	8.1
Location specific				
within third ring	4.8	2.8	2.5	5.3
between third and fifth	3.8	2.2	2.0	4.2
outside fifth ring	1.0	0.5	0.6	1.1
Total	9.7	5.4	5.1	10.5

Notes: Panel A shows the predicted improvement in average travel speed under the different road pricing schemes. Panel B reports the estimated annual welfare gain after imposing the four different road pricing schemes in panel B of Table 8: (i) uniform across time and space, (ii) uniform across space and time varying, (iii) location specific and uniform across time, and (iv) time varying and location specific. Panel C reports the estimated annual tax revenue from the four schemes.

small in the temporal dimension but large in the spatial dimension. A scheme using a flat cordon charge for entering the city center (e.g., the fifth ring road) that ignores the spatial heterogeneity is likely to be less effective in improving traffic speed and social welfare than the first scheme because it does not incentivize drivers to drive less once they have entered the cordons.

In addition to the large gain in social welfare, road pricing would also generate substantial revenue for the Beijing municipal government. Panel C of Table 10 shows that while there are large differences in social welfare impacts across different road pricing schemes, all the schemes would generate a similar revenue of about ¥10 billion. This revenue could be used to cover the fixed and operating costs of the road pricing scheme and/or to improve the public transit system.[23] The improvement in public transit would provide better alternatives for road users and help to address the equity concerns that often come with road pricing.

[23] In 2014, the total operating subsidy on the public transit system was ¥15.3 billion in Beijing, accounting for over 4 percent of the total government spending. The total capital investment on transportation in Beijing was ¥88.6 billion, including ¥43 billion on subway construction.

TABLE 11—ROAD PRICING WITH MORE ELASTIC DEMAND

	Uniform	Time varying		
	All day	Peak	Off-peak	Total
Panel A. Road pricing (yuan/vehicle-km)				
Uniform				
Beijing	0.10	0.14	0.09	
Location specific				
within third ring	0.21	0.31	0.18	
between third and fifth	0.12	0.16	0.10	
outside fifth ring	0.05	0.06	0.05	
Panel B. Increase in average speed (percent)				
Uniform				
Beijing	4.1	3.6	4.6	
Location specific				
within third ring	13.8	14.9	14.1	
between third and fifth	5.0	4.8	5.4	
outside fifth ring	1.3	0.9	1.6	
Panel C. Welfare gain (annualized, billion yuan)				
Uniform				
Beijing	1.0	0.3	0.7	1.1
Location specific				
within third ring	1.2	0.5	0.8	1.3
between third and fifth	0.6	0.2	0.4	0.6
outside fifth ring	0.1	0.0	0.1	0.1
Total	1.9	0.8	1.2	2.0
Panel D. Tax revenue (annualized, billion yuan)				
Uniform				
Beijing	6.2	3.3	3.3	6.6
Location specific				
within third ring	3.7	2.1	1.9	4.0
between third and fifth	3.1	1.8	1.6	3.4
outside fifth ring	0.9	0.4	0.5	0.9
Total	7.7	4.3	4.0	8.3

Notes: The demand elasticities are assumed to be twice as large (in magnitude) as the baseline case in Tables 9 and 10. Panel A reports the estimates of the optimal congestion charges under the four different road pricing schemes: (i) uniform across time and space, (ii) uniform across space and time varying, (iii) location specific and uniform across time, and (iv) time varying and location specific. Panel B shows the predicted improvement in average travel speed under different road pricing schemes. Panel C reports the estimated annual welfare gain, and Panel D reports the estimated annual tax revenue.

C. *Discussion and Robustness Checks*

To check the robustness of our findings with respect to the travel demand elasticity from the literature, we examine three alternative scenarios: the first one with a demand elasticity twice as large (in magnitude) as the baseline model, the second one with demand elasticity half as elastic as the baseline, and the third one with a very large demand elasticity of -16 as in Couture, Duranton, and Turner (2018). Table 11 presents the results for the first scenario of more elastic demand. The optimal congestion charge decreases slightly from 39 to 31 cents per vehicle-km, while the traffic speed improves by 15 percent instead of 11 percent during peak hours within the third ring road. The comparison is intuitive because with a more elastic

Table 12—Road Pricing with Less Elastic Demand

	Uniform	Time varying		
	All day	Peak	Off-peak	Total
Panel A. Road pricing (yuan/vehicle-km)				
Uniform				
Beijing	0.13	0.16	0.11	
Location specific				
within third ring	0.31	0.48	0.26	
between third and fifth	0.15	0.20	0.13	
outside fifth ring	0.06	0.07	0.06	
Panel B. Increase in average speed (percent)				
Uniform				
Beijing	1.7	1.3	2.0	
Location specific				
within third ring	7.9	8.0	8.6	
between third and fifth	2.1	1.9	2.5	
outside fifth ring	0.4	0.3	0.6	
Panel C. Welfare gain (annualized, billion yuan)				
Uniform				
Beijing	0.4	0.1	0.3	0.4
Location specific				
within third ring	0.7	0.3	0.5	0.8
between third and fifth	0.2	0.1	0.2	0.3
outside fifth ring	0.0	0.0	0.0	0.0
Total	1.0	0.4	0.7	1.1
Panel D. Tax revenue (annualized, billion yuan)				
Uniform				
Beijing	8.9	4.4	4.9	9.3
Location specific				
within third ring	6.2	3.6	3.2	6.8
between third and fifth	4.5	2.5	2.4	4.9
outside fifth ring	1.1	0.5	0.7	1.2
Total	11.9	6.6	6.3	12.9

Notes: The demand elasticities are assumed to be half as large as the baseline case in Tables 9 and 10. Panel A reports the estimates of the optimal congestion charges under the four different road pricing schemes: (i) uniform across time and space, (ii) uniform across space and time varying, (iii) location specific and uniform across time, and (iv) time varying and location specific. Panel B shows the predicted improvement in average travel speed under different road pricing schemes. Panel C reports the estimated annual welfare gain, and Panel D reports the estimated annual tax revenue.

demand, a lower congestion charge is needed in order to induce the same change in travel behavior. The welfare gain increases from ¥1.5 billion to ¥2.0 billion under the fourth pricing scheme. The total government revenue decreases slightly to ¥8.3 billion from ¥10.5 billion due to lower congestion charges.

The results for the less elastic travel demand are presented in Table 12. Under a less elastic demand curve, the road pricing policy would call for higher congestion charges but produce a smaller speed improvement and lower welfare gain relative to the baseline model. The comparisons suggest that the more elastic the demand is, the larger the welfare gain from congestion pricing. An important implication of this finding is that the effectiveness of the road pricing policy can be enhanced by improving the attractiveness of alternative travel modes (e.g., by decreasing transit

Table 13—Road Pricing with Very Elastic Demand

	Uniform	Time varying		
	All day	Peak	Off-peak	Total
Panel A. Road pricing (yuan/vehicle-km)				
Uniform				
Beijing	0.07	0.09	0.06	
Location specific				
within third ring	0.14	0.19	0.13	
between third and fifth	0.08	0.10	0.07	
outside fifth ring	0.04	0.05	0.03	
Panel B. Increase in average speed (percent)				
Uniform				
Beijing	7.7	8.1	7.8	
Location specific				
within third ring	19.2	22.5	18.4	
between third and fifth	9.0	10.1	8.9	
outside fifth ring	3.4	3.1	3.6	
Panel C. Welfare gain (annualized, billion yuan)				
Uniform				
Beijing	1.9	0.8	1.2	2.0
Location specific				
within third ring	1.6	0.8	1.0	1.7
between third and fifth	1.0	0.4	0.6	1.1
outside fifth ring	0.2	0.1	0.1	0.2
Total	2.8	1.3	1.7	3.0
Panel D. Tax revenue (annualized, billion yuan)				
Uniform				
Beijing	3.2	1.7	1.8	3.5
Location specific				
within third ring	2.1	1.1	1.2	2.3
between third and fifth	1.6	0.9	0.9	1.8
outside fifth ring	0.5	0.2	0.3	0.5
Total	4.2	2.2	2.3	4.6

Notes: The demand elasticities are assumed to be -16 as in Couture, Duranton, and Turner (2018). Panel A reports the estimates of the optimal congestion charges under the four different road pricing schemes: (i) uniform across time and space, (ii) uniform across space and time varying, (iii) location specific and uniform across time, and (iv) time varying and location specific. Panel B shows the predicted improvement in average travel speed under different road pricing schemes. Panel C reports the estimated annual welfare gain, and Panel D reports the estimated annual tax revenue.

fare or investing in public transit) in order to increase the elasticity of demand for driving.

Table 13 shows the results for the third scenario of very elastic demand as in Couture, Duranton, and Turner (2018). By imposing an optimal congestion charge of just 19 cents, the traffic speed during the peak hours within the third ring road could improve by 22.5 percent and the welfare gain could be doubled relative to the baseline (Table 10). Overall, the comparison between the baseline results in Table 9 and those under the three alternative assumptions in Tables 11, 12, and 13 suggests that the key results are robust to a range of plausible travel demand elasticities from the literature.

Table 14—Welfare Gain from Road Pricing (*in billion yuan*)

	Uniform	Time varying		
	All day	Peak	Off-peak	Total
Panel A. The baseline scenario (*Table 10*)				
Uniform				
Beijing	2.8	0.8	2.2	3.0
Location specific				
within third ring	2.1	0.8	1.5	2.3
between third and fifth	1.5	0.4	1.1	1.5
outside fifth ring	0.4	0.1	0.3	0.4
Total	4.0	1.3	2.9	4.2
Panel B. More elastic demand (*Table 11*)				
Uniform				
Beijing	4.3	1.3	3.2	4.5
Location specific				
within third ring	2.7	1.0	1.8	2.9
between third and fifth	2.1	0.7	1.5	2.2
outside fifth ring	0.6	0.2	0.5	0.7
Total	5.5	1.9	3.9	5.8
Panel C. Less elastic demand (*Table 12*)				
Uniform				
Beijing	1.7	0.5	1.4	1.8
Location specific				
within third ring	1.5	0.5	1.1	1.6
between third and fifth	0.9	0.3	0.7	1.0
outside fifth ring	0.2	0.0	0.2	0.2
Total	2.6	0.8	2.0	2.8
Panel D. Very elastic demand (*Table 13*)				
Uniform				
Beijing	8.4	3.0	5.5	8.5
Location specific				
within third ring	3.8	1.5	2.4	4.0
between third and fifth	3.9	1.5	2.5	4.0
outside fifth ring	1.7	0.5	1.2	1.7
Total	9.5	3.5	6.1	9.7

Notes: The four panels correspond to different demand elasticities: the baseline scenario, more elastic demand (twice as elastic as the baseline), less elastic demand (half as elastic as the baseline), and very elastic demand (-16). The welfare gain is on an annual basis and takes into account reductions not only in congestion but also in local air pollution, CO_2 emissions, and traffic accidents.

Vehicle usage generates multiple externalities including congestion, local air pollution, CO_2 emissions, and traffic accidents (Parry, Walls, and Harrington 2007). The welfare gain discussed above only reflects the benefit from reduced travel time but not other types of externalities. Table 14 presents the welfare gain under different travel demand elasticities by adding the auxiliary benefit of reducing other types of externalities as a result of reduced travel and congestion. Parry et al. (2014) estimates the total external cost from gasoline consumption for over 100 countries, and the external cost in China is estimated as \$0.55 (¥3.7) per liter of gasoline, 64 percent of which is from local pollution, CO_2 emissions, and accidents. Based on the average fuel economy of 9.55 liters/100km in Beijing (Li 2018), the total external cost from local pollution, CO_2 emissions, and accidents is ¥223 per 1,000 vehicle-km.

As shown in panel A in Table 14, the total welfare gain from the road pricing under the baseline scenario is estimated at ¥4.2 billion per year under the location-specific and time-varying scheme. The estimate is nearly three times as large as the welfare gain from congestion reduction alone in panel C of Table 10. The difference is consistent with the finding from Parry et al. (2014) that congestion externality accounts for about 36 percent of total externalities from automobile usage in China. The welfare gain from reduced congestion relative to that from other types of externalities is similar to Milan but smaller compared to London and Stockholm (see Table 3 in Anas and Lindsey 2011). The total welfare gain is larger with a more elastic demand in panels B and especially D.

Previous practices on road pricing show that well-designed road pricing systems in heavily congested cities can generate significant economic benefit. Our welfare analysis suggests that the total benefits from a road pricing in Beijing could be much higher than those found in London, Stockholm, and Milan. The net annual benefits of these road pricing programs are positive, and the net benefit of the London scheme was the highest (Anas and Lindsey 2011). According to the cost-benefit analysis of the congestion charging in London (Leape 2006), the total estimated annual costs (including the initial setup costs, scheme operation costs, supervisory costs, traffic management costs, and compliance costs) of the congestion charging scheme are £163 million (or ¥1.4 billion), while the total annual benefits (including time savings and reliability benefits, reduced accidents, reduced CO_2 emissions, and other resource savings) are £230 million (or ¥1.9 billion).

It is important to note that our MECC estimates are based on the assumption that all travelers have the same value of time and are driving the same type of vehicle with average occupancy of 1.34. However, MECC varies not only across space and time but also across vehicle types such as bus, tax, truck, and bicycle. Optimal congestion charges can vary based on the values of the passenger car equivalence (PCE) and vehicle occupancy rate. MECC should thus be scaled up for vehicles with a higher PCE and scaled down for the vehicles with a higher occupancy rate. For instance, a representative vehicle can have a significantly greater impact on increasing travel times for other road users when a greater portion of vehicles on the road are carrying a large number of passengers. Incorporating this heterogeneity in welfare analysis would necessitate information on the shares of the different vehicle types on a specific road segment for a given time of the day (Parry et al. 2014). Our dataset does not have this critical information to make such an adjustment.

VI. Conclusion

This study applies a fundamental economic principle to big data to investigate the first-best policy to address urban traffic congestion. Leveraging a natural experiment, it provides the first empirical estimates of the MECC and optimal congestion charges, as well as their welfare consequences in Beijing. The analysis presents two important departures from the literature. First, we focus on the traffic speed and density relationship in the empirical analysis instead of the commonly-used speed and flow relationship. Second, ours is the first empirical analysis to address the endogeneity issue due to simultaneity when quantifying MECC using real-time traffic data.

The driving restriction in Beijing provides plausibly exogenous variation in traffic density for the causal identification.

Our analysis shows that addressing endogeneity in the speed and density relationship leads to a large increase in the estimate of MECC. In addition, there is a large heterogeneity in MECC over time and especially across space: it ranges from ¥1.98 ($0.30)/vehicle-km within the second ring road during peak hours to ¥0.06/vehicle-km outside the fifth ring road during off-peak hours. Traffic congestion imposes a very large cost to society: the total value of lost time amounted to nearly half percent of Beijing's GDP in 2014. Under time-varying and location-specific road pricing, the optimal charges should be ¥0.39/km within the third ring road during peak hours and ¥0.05/km for outside the fifth ring road during off-peak hours. This scheme of road pricing would lead to at least an 11 percent increase in traffic speed within the third ring road during peak hours and an increase of social welfare of ¥1.5 billion per year. The total charges would amount to over ¥10 billion annually that could be used to cover the capital and operating costs of the system, improve public transit in order to address the equity concerns of a road pricing policy, and increase the effectiveness of the policy.

We conclude with some caveats and directions for future research. First, our analysis represents an initial key step toward using road pricing to address traffic congestion in large urban centers such as Beijing. Effectively addressing traffic congestion would require a comprehensive approach that includes demand- and supply-side strategies as well as better urban planning. Road pricing should be an essential element, but its effectiveness can be enhanced by improving road infrastructure as well as access to and quality of public transit.

Second, our analysis is based on the road conditions in 2014, and the optimal congestion charges should be adjusted based on how the policy affects the congestion level and hence the MECC across time and space. In addition, the travel demand function, one essential component for estimating the congestion charges, is calibrated based on previous studies rather than real-world data in Beijing. The estimation of this function presents its own identification challenges and warrants future research.

Third, we employ a partial equilibrium framework to estimate the welfare gain from different road pricing schemes. It is well understood that regulations such as road pricing could interact with the tax system, and this interaction could lead to a divergence between the partial and general equilibrium welfare impacts (Goulder et al. 1999). In the context of road pricing, Parry and Bento (2001) shows that the form of revenue recycling could crucially affect the magnitude and even direction of the welfare impacts from road pricing in the presence of preexisting tax distortions such as the labor tax. Future research should help understand the distributional and general equilibrium impacts of road pricing and the impacts from different revenue recycling mechanisms.

Lastly, hyper-congestion when existent could represent a large source of congestion cost and welfare loss (Verhoef 2003, Fosgerau and Small 2013). Our welfare analysis shows that road pricing schemes that do not take into account the heterogeneity in congestion externality across space and over time would be less effective in reducing congestion and less efficient in improving social welfare.

Therefore, ignoring hyper-congestion in road pricing schemes, particularly in a dense street network as in our context, could reduce the effectiveness and efficiency of the policy. Future research could empirically investigate the frequency and severity of hyper-congestion as well as its welfare consequences.

REFERENCES

Anas, Alex, and Robin Lindsey. 2011. "Reducing Urban Road Transportation Externalities: Road Pricing in Theory and in Practice." *Review of Environmental Economics and Policy* 5 (1): 66–88.

Anderson, Michael L. 2014. "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion." *American Economic Review* 104 (9): 2763–96.

Anderson, Michael L., and Lucas W. Davis. 2018. "Does Hyper-congestion Exist? New Evidence Suggests Not." https://pdfs.semanticscholar.org/75b5/5b24661ef57365d32d2298f1b1d5380b7930.pdf.

Arnott, Richard. 1990. "Signalized Intersection Queuing Theory and Central Business District Auto Congestion." *Economics Letter* 33 (2): 197–201.

Arnott, Richard. 2013. "A Bathtub Model of Downtown Traffic Congestion." *Journal of Urban Economics* 76: 110–21.

Beijing Transportation Research Center. 2014. Beijing Transportation Annual Report 2014. http://www.bjtrc.org.cn/List/index/cid17.html.

Bento, Antonio M., Lawrence H. Goulder, Mark R. Jacobsen, and Roger H. von Haefen. 2009. "Distributional and Efficiency Impacts of Increased US Gasoline Taxes." *American Economic Review* 99 (3): 667–99.

Button, Kenneth. 2010. *Transport Economics*. Cheltenham, UK: Edward Elgar Publishing.

Couture, Victor, Gilles Duranton, and Matthew A. Turner. 2018. "Speed." *Review of Economics and Statistics* 100 (4): 725–39.

Daganzo, Carlos F., and Nikolas Geroliminis. 2008. "An Analytical Approximation for the Macroscopic Fundamental Diagram of Urban Traffic." *Transportation Research Part B: Methodological* 42 (9): 771–81.

de Grange, Louis, and Rodrigo Troncoso. 2011. "Impacts of Vehicle Restrictions on Urban Transport Flows: The Case of Santiago, Chile." *Transport Policy* 18 (6): 862–69.

Dewees, Donald N. 1979. "Estimating the Time Costs of Highway Congestion." *Econometrica: Journal of the Econometric Society* 47 (6): 1499–1512.

Duranton, Gilles, and Matthew A. Turner. 2011. "The Fundamental Law of Road Congestion: Evidence from US Cities." *American Economic Review* 101: 2616–52.

Eliasson, Jonas, and Lars-Göran Mattsson. 2006. "Equity Effects of Congestion Pricing: Quantitative Methodology and a Case Study for Stockholm." *Transportation Research Part A: Policy and Practice* 40 (7): 602–20.

Fosgerau, Mogens. 2015. "Congestion in the Bathtub." *Economics of Transportation* 4 (4): 241–55.

Fosgerau, Mogens, and Kenneth A. Small. 2013. "Hyper-congestion in Downtown Metropolis." *Journal of Urban Economics* 76: 122–34.

Goodwin, P.B. 1992. "A Review of New Demand Elasticities with Special Reference to Short and Long Run Effects of Price Changes." *Journal of Transport Economics and Policy* 26 (2): 155–69.

Goodwin, Phil, Joyce Dargay, and Mark Hanly. 2004. "Elasticities of Road Traffic and Fuel Consumption with Respect to Price and Income: A Review." *Transport Reviews* 24 (3): 275–92.

Goulder, Lawrence W., Ian W.H. Parry, Roberton C. Williams III, and Dallas Burtraw. 1999. "The Cost Effectiveness of Alternative Instruments for Environmental Protection in a Second-Best Setting." *Journal of Public Economics* 72 (3): 329–60.

Greenshields, B.D., J.R. Bibbins, W.S. Channing, and H.H. Miller. 1935. "A Study of Traffic Capacity." In *Proceedings of the Fourteenth Annual Meeting of the Highway Research Board*, Vol. 14, 448–77. Washington, DC: Highway Research Board.

Keeler, Theodore E., and Kenneth A. Small. 1977. "Optimal Peak-Load Pricing, Investment, and Service Levels on Urban Expressways." *Journal of Political Economy* 85 (1): 1–25.

Knittel, Christopher R., and Ryan Sandler. 2011. "Cleaning the Bathwater with the Baby: The Health Co-benefits of Carbon Pricing in Transportation." NBER Working Paper 17390.

Kraus, Marvin, Herbert Mohring, and Thomas Pinfold. 1976. "The Welfare Costs of Nonoptimum Pricing and Investment Policies for Freeway Transportation." *American Economic Review* 66 (4): 532–47.

Leape, Jonathan. 2006. "The London Congestion Charge." Journal of Economic Perspectives 20 (4): 157–76.

Lindsey, Robin, and Erik Verhoef. 2007. "Congestion Modelling." In *Handbook of Transport Modelling*, Vol. 1, 2nd ed., edited by David A. Hensher and Kenneth J. Button, 417–41. Bingley, UK: Emerald Group Publishing Limited.

Linn, Joshua, Zhongmin Wang, and Lunyu Xie. 2016. "Who Will Be Affected by a Congestion Pricing Scheme in Beijing?" *Transport Policy* 47: 34–40.

Li, Shanjun. 2018. "Better Lucky than Rich? Welfare Analysis of Automobile License Allocations in Beijing and Shanghai." *Review of Economic Studies* 85 (4): 2389–2428.

Li, Shanjun, Joshua Linn, and Erich Muehlegger. 2014. "Gasoline Taxes and Consumer Behavior." *American Economic Journal: Economic Policy* 6 (4): 302–42.

Newbery, David M. 1988a. "Road Damage Externalities and Road User Charges." *Econometrica* 56 (2): 295–316.

Newbery, David M. 1988b. "Road User Charges in Britain." *Economic Journal* 98 (390): 161–76.

Olszewski, Piotr, and Litian Xie. 2005. "Modelling the Effects of Road Pricing on Traffic in Singapore." *Transportation Research Part A: Policy and Practice* 39 (7–9): 755–72.

Parry, Ian W.H. 2009. "Pricing Urban Congestion." *Annual Review of Resource Economics* 1: 461–84.

Parry, Ian W.H., and Antonio Bento. 2001. "Revenue Recycling and the Welfare Effects of Road Pricing." *Scandinavian Journal of Economics* 103 (4): 645–71.

Parry, Ian W.H., Dirk Heine, Eliza Lis, and Shanjun Li. 2014. *Getting Energy Prices Right: From Principle to Practice*. Washington, DC: International Monetary Fund.

Parry, Ian W.H., and Kenneth A. Small. 2009. "Should Urban Transit Subsidies Be Reduced?" *American Economic Review* 99 (3): 700–724.

Parry, Ian W.H., Margaret Walls, and Winston Harrington. 2007. "Automobile Externalities and Policies." *Journal of Economic Literature* 45 (2): 373–99.

Quinet, Emile, and Roger W. Vickerman. 2004. *Principles of Transport Economics*. Cheltenham, UK: Edward Elgar Publishing.

Rouwendal, Jan, and Erik T. Verhoef. 2006. "Basic Economic Principles of Road Pricing: From Theory to Applications." *Transport Policy* 13 (2): 106–14.

Small, Kenneth A., and Xuehao Chu. 2003. "Hyper-congestion." *Journal of Transport Economics and Policy* 37 (3): 319–52.

Small, Kenneth A., and José A. Gomez-Ibañez. 1998. "Road Pricing for Congestion Management: The Transition from Theory to Policy." In *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*, edited by Kenneth J. Button and Erik T. Verhoef, 213–46. Cheltenham, UK: Edward Elgar Publishing.

Small, Kenneth A., and Kurt Van Dender. 2007. "Fuel Efficiency and Motor Vehicle Travel: The Declining Rebound Effect." *Energy Journal*: 25–51.

Small, Kenneth A., and Eric Verhoef. 2007. *The Economics of Urban Transportation*. Abington, UK. Routledge.

Verhoef, Erik T. 1999. "Time, Speeds, Flows and Densities in Static Models of Road Traffic Congestion and Congestion Pricing." *Regional Science and Urban Economics* 29 (3): 341–69.

Verhoef, Erik T. 2001. "An Integrated Dynamic Model of Road Traffic Congestion Based on Simple Car-Following Theory: Exploring Hyper-congestion." *Journal of Urban Economics* 49 (3): 505–42.

Verhoef, Erik T. 2003. "Inside the Queue: Hyper-congestion and Road Pricing in a Continuous Time–Continuous Place Model of Traffic Congestion." *Journal of Urban Economics* 54 (3): 531–65.

Viard, V. Brian, and Shihe Fu. 2015. "The Effect of Beijing's Driving Restrictions on Pollution and Economic Activity." *Journal of Public Economics* 125: 98–115.

Vickrey, William. 1959. "Statement on the Pricing of Urban Street Use." *Hearing of US Congress Joint Committee on Metropolitan Washington DC Problems* 11: 466–77.

Vickrey, William. 1963. "Pricing in Urban and Suburban Transport." *American Economic Review* 53 (2): 452–65.

Walters, A.A. 1961. "The Theory and Measurement of Private and Social Cost of Highway Congestion." *Econometrica* 29 (4): 676–99.

Zheng, Siqi, and Matthew E. Kahn. 2013. "Understanding China's Urban Pollution Dynamics." *Journal of Economic Literature* 51 (3): 731–72.