Journal of Urban Economics 81 (2014) 65–76

Contents lists available at ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

The fundamental law of highway congestion revisited: Evidence from national expressways in Japan ☆

Wen-Tai Hsu [a], Hongliang Zhang [b,*]

[a] School of Economics, Singapore Management University, Singapore
[b] Department of Economics, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 30 October 2012
Revised 5 February 2014
Available online 27 February 2014

Keywords:
Traffic congestion
Fundamental law of highway congestion
Returns to scale
Speed function
Coverage effect

ABSTRACT

The fundamental law of highway congestion states that when congested, the travel speed on an expanded expressway reverts to its previous level before the capacity expansion. In this paper, we propose a theory that generalizes this statement and finds that if there exists a coverage effect, that is, the effect of longer road length on traffic conditional on capacity, then the new equilibrium travel speed could be lower than its previous level. Given the fundamental law, the theory predicts that the elasticity of traffic to road capacity is at least 1. We estimate this elasticity for national expressways in Japan and test this prediction. Using the planned national expressway extension as an exogenous source of variation for capacity expansion, we obtain elasticity estimates ranging between 1.24 and 1.34, consistent with the prediction of our theory. We further investigate the sources of the larger-than-unity elasticity and find that the coverage effect plays a critical role, compared with the effect due to lane expansion.

"[E]xpansion of road capacity – no matter how large, within the limits of feasibility – cannot fully eliminate periods of crawling along on expressways at frustratingly low speeds."
– *Anthony Downs (2004, p. 85)*

1. Introduction

When urban residents complain about traffic congestion, or, in other words, they are not satisfied with the current travel speed on a certain portion of the road system, the most likely improvement option adopted is to expand the capacity of the congested roads. Whereas expanding road capacity seems intuitive, many economists have argued that this "building your way out of congestion" approach is likely to be fruitless. As Downs (1962) and many other authors[1] explain, this approach may fail because when there are alternatives to driving on congested routes, such as driving on less congested routes, using alternative transport modes, scheduling alternative travel times, or simply not traveling, latent travel demand exists. That is, potential traffic flows are not observed simply because the congestion itself deters them. When road capacity is expanded, however, the resulting increase in travel speed brings back the previously deterred potential traffic, thus leaving the congested routes as congested as they were before. This paradox is called the fundamental law of traffic/highway congestion, hereafter the fundamental law.

As the fundamental law is concerned with how traffic responds to road capacity expansion, it has implications for this elasticity. A large body of literature has estimated this elasticity to investigate whether, and how much, capacity expansion induces new traffic. The elasticity estimates obtained are almost always positive, confirming the existence of induced travel demand, but are typically significantly below 1, ranging from 0.2 to 0.8.[2] Recently, a strikingly different result was found by Duranton and Turner

☆ We have benefited from helpful comments from Jiahua Che, Gilles Duranton, Charles Y. Horioka, Tom Holmes, Nayoung Lee, Wallace Mok, Sung Yong Park, Liugang Sheng, Matt Turner, two anonymous referees, the Co-Editor, and seminar participants at Tsinghua University, the 2010 HKEA Sixth Biennial Conference, and the 2011 Annual Meeting of the Urban Economics Association. We would like to thank Kin Tai Cheung, Vincent Chow, Chuantao Cui, and Jia-Jin Wu for their research assistance. We are also grateful to Tomoya Mori for providing us the definition of urban employment areas. All errors are ours.

* Corresponding author.
E-mail addresses: wentaihsu@smu.edu.sg (W.-T. Hsu), hongliang@cuhk.edu.hk (H. Zhang).

[1] For example, Holden (1989), Arnott and Small (1994), and Small (1997).

[2] For example, Hansen and Huang (1997), Noland and Cowart (2000), and Cervero and Hansen (2002). For more thorough reviews, see Goodwin (1996), Cervero (2002), and Small and Verhoef (2007).

http://dx.doi.org/10.1016/j.jue.2014.02.002
0094-1190/

(2011, hereafter DT) who obtained estimates close to 1. Most importantly, they argue that this is evidence for the fundamental law. DT's work differs from previous studies in two important ways: (i) they estimate elasticity for the most congested type of roads in the US – the interstate highways in metropolitan statistical areas (MSAs); and (ii) they employ innovative and sensible instrumental variables (IV) to account for the possible endogeneity between traffic and road capacity.[3] Note that quantitative differences in the elasticity estimates have qualitatively different implications regarding whether the "building your way out of congestion" approach is likely to have some degree of success. As we will explain shortly, an elasticity smaller than 1 indicates that congestion could still be somewhat relieved by road expansion even though latent demand exists, whereas an elasticity of 1 or higher may suggest a complete failure of this approach.[4] Thus, DT's results convey a strong message, and their approach deserves further examination and/or application.

This paper contributes to both the theoretical and empirical literature on the fundamental law. We propose a simple and yet general theory of road capacity and traffic to guide our empirical analysis. We first clarify the conditions under which the fundamental law holds and then derive the equilibrium road elasticity of traffic. The theory postulates that road users care about not only the travel speed, which is a function of capacity and traffic, but also the coverage of the road system.[5] Under some weak conditions on the travel speed function, we show that this elasticity is at least 1 when the fundamental law holds. In particular, if there is no coverage effect, and if the travel speed function features constant returns to scale, that is, proportional increases in road capacity and traffic entail the same travel speed, then this elasticity equals unity. This is analogous to the demand-and-supply analysis for unit elasticity assuming that travel demand is perfectly elastic and total cost exhibits constant returns to scale (see Small and Verhoef, 2007; Duranton and Turner, 2009). In this case, testing unit elasticity is equivalent to testing the fundamental law. Nevertheless, the road elasticity of traffic could be larger than 1 either when there are increasing returns to scale, in the sense that an increase in capacity can accommodate a more than proportional increase in traffic while keeping the same travel speed, or when there is a significant coverage effect, or both.

On the empirical front, we estimate the road elasticity of traffic using another national-scale panel data set – road traffic data in Japan. Similar to DT, who focus on interstate highways in the US, we focus on national expressways, which are the highest-ranked roads in Japan. Following DT, we aggregate traffic (measured by vehicle kilometers traveled (VKT)) and road capacity (measured by lane kilometers) to the urban employment areas (UEAs), the Japanese version of MSAs, and conduct our analysis using the UEA-level aggregate data. In an earlier version of this paper, we conducted the empirical analysis at the prefecture level. The results are qualitatively similar to the UEA-level analysis. We begin with the ordinary least squares (OLS) estimations of the road elasticity of VKT by pooling the data from five traffic censuses. To address the potential correlation between road capacity and unobserved determinants of VKT at the UEA level, we also employ a UEA fixed-effect model to examine the relationship between growth in VKT and roadway capacity expansion conditioning out time-invariant UEA-level determinants. Our OLS and fixed-effect estimates of the road elasticity of VKT are always larger than 1, consistent with the necessary condition of the fundamental law – a road elasticity of at least 1 – predicted by our theory. Moreover, as their differences to 1 are rather small, our estimates are also generally consistent with DT's test of unit elasticity.

To further address the potential endogeneity of roadway capacity expansion in the fixed-effect model, we identify a new instrument for the growth of national expressways and carry out IV estimations using a fixed-effect model. Our instrument is based on Japan's 1987 National Expressway Network Plan. More specifically, we use the planned extension in the 1987 plan for a UEA smoothed by the national-level completion rate of the plan as the instrument for the growth of national expressways in this UEA over time. This identification strategy is similar to that of Baum-Snow (2007) who employs the 1947 US national highway plan to instrument for the growth of the number of highway rays in central cities in an MSA fixed-effect model to examine the effect of new highway rays on suburbanization. We obtain elasticity estimates ranging between 1.24 and 1.34 in the fixed-effect IV estimations, which are consistent with the prediction of our theory but can reject unit elasticity at the conventional significance levels. In light of these larger-than-unity estimates, we investigate the possible reasons for the elasticity to be larger than 1 by extending the fixed-effect specifications to further incorporate the roadway length and the capacity share of one-lane routes to account, respectively, for the coverage effect and the increasing returns to scale in the speed function, both of which can lead to the larger-than-unity elasticity as implied in our model. The empirical evidence suggests that the coverage effect may play a more important role in explaining the larger-than-unity elasticity, compared with the effect due to lane expansion.

In sum, whereas our general message is similar to that conveyed by DT, we differ from them in three important ways. First, we propose a theory that generalizes the statement of the fundamental law and the link between the law and road elasticity of traffic. Second, we find some evidence suggesting that the elasticity may be larger than 1 and further investigate the possible reasons for the larger-than-unity elasticity. Third, as no instrument was employed in DT's fixed-effect estimations, our fixed-effect IV exercise is also an innovation over DT's empirical analysis.

The remainder of this paper is organized as follows. Section 2 presents a model to guide our empirical work. Section 3 describes the data sets. Section 4 presents our empirical results. Section 5 concludes.

2. A model of road capacity and traffic

Building on the work of several authors, including Downs (1962), Holden (1989), Arnott and Small (1994), and Duranton and Turner (2009), we present in this section a theory of the fundamental law and the road elasticity of traffic to guide our empirical analysis. We then analyze whether increasing road capacity is welfare-improving in various situations.

2.1. Model setup

Suppose there are two different (sets of) routes, e and a, for drivers to choose from, where e annotates expressways and a annotates alternative routes. The two routes e vs. a can also be interpreted as two different modes, such as highways vs. subways, or as different travel schedules such as peak hours vs. off-peak hours, or as driving vs. staying at home. We adopt the standard assumption on how speed and traffic are related on roads. For $i = e, a$, let K_i and Q_i denote, respectively, the total capacity and traffic in route/system i, and assume that the average speed on route/system i is given by

[3] The three instruments used by DT are the routes of the major exploration expeditions between 1835 and 1850, the major rail routes in 1898, and the routes proposed in the 1947 interstate highway plan.

[4] Whether the "building your way out of congestion" approach does fail completely depends on the relative size of this elasticity to the returns to scale in the speed function, as we illustrate in further detail in Section 2.

[5] As we explain in Section 2, conditional on travel speed, a larger coverage of a given system may bring extra utility for drivers because it allows them to reach more places that would previously be reachable only by other systems (perhaps with a lower speed) or simply not reachable.

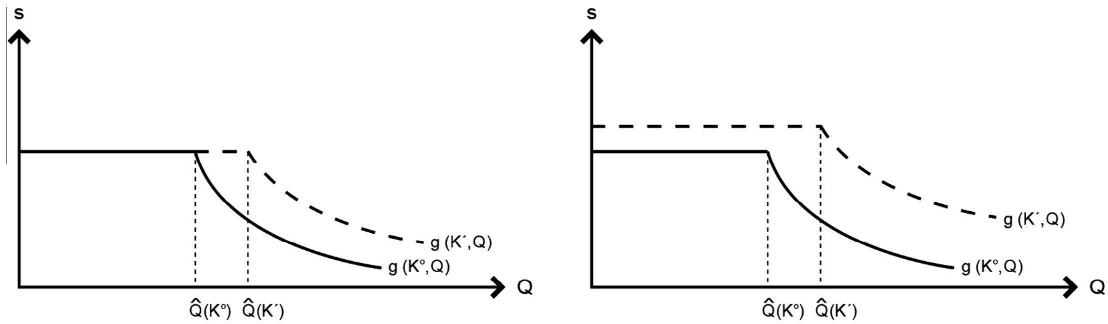

Fig. 1. Shifts in the speed function when capacity increases. The left panel illustrates shifts in the speed function with a capacity increase from K to K' when $b_i'(K) = 0$. The right panel illustrates shifts in the speed function with a capacity increase from K to K' when $b_i'(K) > 0$.

$$s_i = g_i(K_i, Q_i) = \begin{cases} b_i(K_i) & Q_i \leqslant \widehat{Q}_i(K_i) \\ f_i(K_i, Q_i) & Q_i > \widehat{Q}_i(K_i) \end{cases},$$

where $\widehat{Q}_i'(\cdot) > 0$, $b_i'(\cdot) \geqslant 0$, and $f_i(K_i, Q_i)$ is strictly increasing in K_i and strictly decreasing in Q_i. That is, for a given K_i, travel speed on i is constant (at its legal limit or natural limit) when traffic Q_i is no more than the threshold $\widehat{Q}_i(K_i)$, i.e., route i is not congested; and travel speed decreases in Q_i when Q_i is above the threshold, i.e., route i is congested. Moreover, the threshold $\widehat{Q}_i(K_i)$ strictly increases in the capacity of i. Given Q_i, whereas congested speed is strictly increasing in K_i, the uncongested speed b_i is weakly increasing in K_i. It is natural to think that if route i consists of roads, then capacity expansion will not lead to an increase in speed when the roads are uncongested (the case of $b_i'(.) = 0$). However, if such an expansion is an upgrade of the roads (e.g., from local roads to one level higher in the road system), one can think of $b_i'(\cdot) > 0$. In the case of route i being a mode, such as a subway system, it is natural to consider $b_i'(\cdot) > 0$ and that the threshold $\widehat{Q}_i$ is very large so that it is seldom reached. See Fig. 1 for an illustration of shifts in the speed function when capacity increases. Finally, we assume that $b_e > b_a$ to ensure a unique equilibrium.

Assume that the indirect utility a driver derives from using a particular route only depends on the average travel speed s_i and the coverage of that route/system L_i. This indirect utility is denoted as $v(s_i, L_i)$ with

$$\frac{\partial v}{\partial s_i} > 0, \quad \frac{\partial v}{\partial L_i} \geqslant 0.$$

Hence, a driver's utility depends on capacity K_i and traffic Q_i indirectly through speed function g_i. The case of $\frac{\partial v}{\partial L_i} = 0$ essentially reduces the arguments of the indirect utility function v to only the travel speed. However, $\frac{\partial v}{\partial L_i} > 0$ is also conceivable. In principle, capacity expansion of a road system may take either the form of lane expansion or increasing the coverage by extending the length of existing routes or by creating a new route. Pure lane expansion implies that K_i increases with L_i fixed. When the coverage L_i increases, K_i as the total capacity necessarily increases, say, from K_i to K_i'. In this case, initial traffic Q_i is shared among existing roads and new roads, inducing average travel speed $s_i(K_i', Q_i)$ to increase. However, the speed of the system does not necessarily capture the full benefits of a larger coverage. For example, a larger coverage allows drivers to reach more places that would previously be reachable only by the other system (perhaps with a lower speed) or simply not reachable.

2.2. Equilibrium and the fundamental law

Conditional on both routes $i = e, a$ being used, equilibrium traffic Q_e^* and Q_a^* must be such that

$$v(g_e(K_e, Q_e^*), L_e) = v(g_a(K_a, Q_a^*), L_a).$$

There are different cases of equilibrium, depending on the states of both routes. As our empirical analysis is centered around the fundamental law, we first analyze the case where the fundamental law arises. Consider capacity expansion of route e when route a is initially uncongested ($Q_a \leqslant \widehat{Q}_a(K_a)$). Because we would like to single out the effect of an increase in K_e, we fix K_a and L_a. Hence, travel speed on route a is constant at some b_a. We first consider the case when L_e is fixed. Then, at the initial and new level K_e and K_e', it must be true that

$$v(g_e(K_e, Q_e^*(K_e, L_e)), L_e) = v(b_a) = v(g_e(K_e', Q_e^*(K_e', L_e)), L_e). \quad (1)$$

The equilibrium adjustment goes as follows. At the initial equilibrium traffic level, speed increases as $g_e(K_e', Q_e^*(K_e, L_e)) > g_e(K_e, Q_e^*(K_e, L_e))$, which leads to a higher utility on route e and provides incentives for drivers to switch from route a to route e. If travel demand is endogenous, faster speed may also provide incentives for existing drivers on route e to drive more. Both effects imply that traffic will increase on route e and hence bring down the travel speed. This adjustment will not stop until drivers become indifferent between the two routes again or until all drivers have switched to route e. Suppose that the capacity expansion of route e is not so great as to attract all the drivers. As L_e is held constant, the new equilibrium speed must equal the original one, $g_e(K_e', Q_e^*(K_e', L_e)) = g_e(K_e, Q_e^*(K_e, L_e))$. In other words, the fundamental law holds in the sense that the travel speed on expressways reverts to its previous level before the capacity expansion. See Fig. 2a for this scenario in which the new equilibrium traffic is $Q_e^*(K_e', L_e)$.[6]

Now, consider a capacity expansion that is accompanied by a larger coverage. Suppose that the new total capacity level is still K_e' but the new coverage is L_e'. Suppose $\frac{\partial v}{\partial L_e} > 0$, then at the initial traffic volume, the utility of drivers using route e increases even more, and in equilibrium,

$$v(g_e(K_e, Q_e^*(K_e, L_e)), L_e) = v(b_a) = v(g_e(K_e', Q_e^*(K_e', L_e')), L_e').$$

What is more interesting in this case is that as the equilibrium utility remains at $v(b_a)$, the fact that the increase of coverage brings extra utility implies that in equilibrium $g_e(K_e', Q_e^*(K_e', L_e')) < g_e(K_e, Q_e^*(K_e, L_e))$. That is, congestion becomes worse in the sense that the travel speed is *lower* than the original speed. See Fig. 2a for this scenario in which the new equilibrium traffic is $Q_e^*(K_e', L_e')$.

[6] Note that the adjustment of traffic on either route might involve two margins, the number of drivers and the travel demand per driver. The way in which Figs. 2 and 3 are depicted implicitly assumes that the total traffic $\overline{Q} \equiv Q_e + Q_a$ is fixed. This can be interpreted as if there is only one margin in the number of drivers, but not the other margin in travel demand per driver. However, such a depiction is only for convenience, as it is easy to verify that the analysis here holds true when $Q_e + Q_a$ is not a constant.

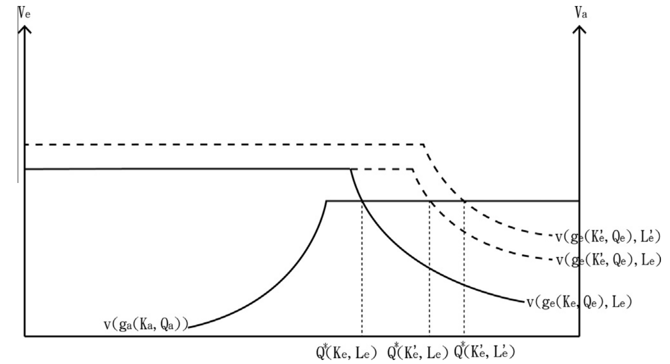

Fig. 2a. Capacity expansion of route e when route a is uncongested. Case (i): The capacity of route e expands from K_e to K'_e with coverage (L_e) held constant. The equilibrium traffic increases to $Q^*(K'_e, L_e)$ and the travel speed reverts to its original level. Case (ii): The capacity of route e increases from K_e to K'_e, accompanied with an increase in coverage from L_e to L'_e. With the increase in coverage generating additional utility, the equilibrium traffic further increases to $Q^*(K'_e, L'_e)$ while the travel speed, $g_e(K'_e, Q^*(K'_e, L'_e))$, is lower than its original level, $g_e(K_e, Q^*(K_e, L_e))$. In either case, the fundamental law holds and the capacity expansion of route e leads to welfare loss.

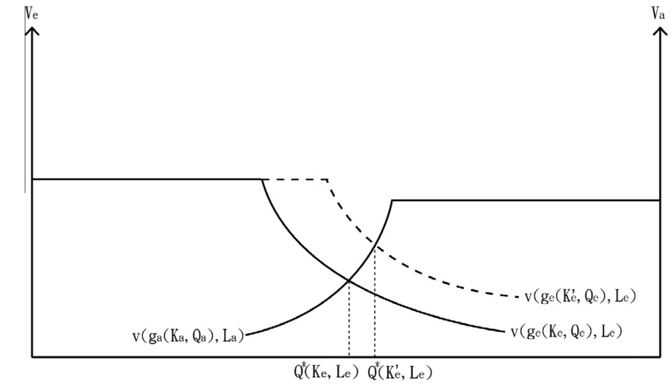

Fig. 2b. Capacity expansion of route e when route a is congested. The capacity expansion of route e increases the travel speed in both routes when route a is also congested. Welfare potentially improves.

Combining both cases (L_e increases or not), we reach a generalized statement of the fundamental law:

Proposition 1 (*Fundamental law of highway congestion*). *Suppose that the alternative routes are uncongested. When the capacity of the congested highways expands, the equilibrium travel speed on the expanded highways is* **no higher than** *its original level, provided that the expansion is not so great that it attracts all drivers.*

2.3. Elasticity of traffic to capacity

Next, we consider the implication of the fundamental law for the elasticity of traffic to road capacity. To build this link, we assume that the travel speed function g_i takes the following form on the congested portion:

$$f_i(K_i, Q_i) = \phi\left(\frac{K_i^\lambda}{Q_i}\right),$$

where ϕ is a strictly increasing function, and the parameter $\lambda > 0$ can be viewed as a measure of returns to scale. When $\lambda = 1$, there are constant returns to scale as an increase in capacity can accommodate a proportional increase in traffic while keeping the same travel speed. A $\lambda > 1 (< 1)$ corresponds to increasing (decreasing) returns to scale in the sense that an increase in capacity can accommodate more (less) than the proportional increase in traffic while keeping the same travel speed.

Continuing from the previous subsection, route a is assumed to be uncongested. Let the new capacity be $K'_e = (1+\mu)K_e$. In the case of the new equilibrium speed reverting to the original level after a capacity expansion (the case of $\frac{\partial v}{\partial L_i} = 0$ or K_e increasing without an increase in coverage L_e), we have $Q'_e = (1+\mu)^\lambda Q_e$. This implies that the elasticity of traffic to capacity, $\rho \equiv d \ln Q_e / d \ln K_e$, equals λ.[7] Unit elasticity follows simply from the constant-returns case. In the case that the new equilibrium speed is lower than the original one (the case of $\frac{\partial v}{\partial L_i} > 0$ and K_e increasing with an increase in coverage L_e), we have $Q'_e > (1+\mu)^\lambda Q_e$, and hence $\rho > \lambda$.

The available empirical evidence suggests that $\lambda \geqslant 1$. Constant returns have empirical support when the number of lanes is two or more in each direction (Keeler and Small, 1977, p. 3). Increasing returns may arise from lane expansion, particularly the expansion of a single-lane route to multiple lanes, which creates more opportunities for passing. Such "lane efficiency" is evidenced by Mohring (1976, pp. 140–45). Suppose the fundamental law holds. If $\frac{\partial v}{\partial L_i} = 0$ or K_e increases without an increase in coverage L_e, then the road elasticity of traffic $\rho = \lambda \geqslant 1$. If $\frac{\partial v}{\partial L_i} > 0$ and K_e increases with an increase in coverage L_e, then $\rho > \lambda \geqslant 1$. Therefore, if the fundamental law holds, then $\rho \geqslant 1$. If the hypothesis $\rho \geqslant 1$ is rejected, then so too is the fundamental law.

However, as our empirical evidence points to $\rho \geqslant 1$, it is worth noting that $\rho \geqslant 1$ may have several different causes. As we study aggregate data at the metro level, it is safe to assume K_e increases with an increase in coverage L_e. When the fundamental law holds, $\rho = 1$ if $\lambda = 1$ and $\frac{\partial v}{\partial L_i} = 0$, whereas $\rho > 1$ if $\lambda > 1$ or $\frac{\partial v}{\partial L_i} > 0$ or both. However, if the fundamental law does not hold, for example, the alternative routes are also congested (as illustrated in Fig. 2b), then $\rho \geqslant 1$ may still result from substantial increasing returns in the speed function or the coverage effect. In sum, $\rho < 1$ clearly rejects the fundamental law, and $\rho \geqslant 1$ is consistent with the fundamental law, although the relative relevance/importance of different causes needs careful examination.

If one assumes that $\lambda = 1$ and $\frac{\partial v}{\partial L_i} = 0$, then DT's test of $\rho = 1$ is equivalent to the test of the fundamental law. As we do not impose these assumptions, we are essentially testing $\rho \geqslant 1$, which is a necessary condition of the fundamental law. Although the resulting message is clearer with these assumptions, we believe it is important to examine the possibility of the violations of these assumptions, i.e., whether $\lambda > 1$ and/or $\frac{\partial v}{\partial L_i} > 0$, and if so, their magnitudes.

2.4. Welfare analysis

Suppose capacity costs are evenly shared by all individuals. Denote the common utility level in equilibrium for all individuals in the previous subsections as v^*. As v^* is the utility level before consideration of capacity costs, the welfare level of the society can therefore be measured by every individual's utility $v^* - h(K_e, L_e, K_a, L_a)$, for some capacity cost function h that is strictly increasing in all its arguments. A scenario in which v^* increases is said to be *potentially* welfare-improving, as we do not explicitly model h, and welfare does improve if capacity costs are sufficiently small. However, when v^* decreases or remains the same, welfare necessarily decreases as there are capacity costs.[8]

We assume route e is always congested and distinguish the following cases. The first case is concerned with the expansion of

[7] To see this, note that a μ percent increase in K_e induces a $(1+\mu)^\lambda - 1$ percent increase in Q_e. Hence, the point elasticity is $\lim_{\mu \to 0} \frac{[(1+\mu)^\lambda - 1]}{\mu} = \lambda$.

[8] Here, we only discuss the directions of welfare changes. The simple and general formulation here does not allow us to determine the optimal level of capacity or how a congestion tax may help.

route e, keeping route a's capacity unchanged. If route a is uncongested, then the fundamental law holds, and in this situation, increasing road capacity obviously reduces welfare, as v^* remains the same (see Fig. 2a). However, as shown in Fig. 2b, if route a is congested, the expansion of K_e is potentially welfare-improving. Although the illustration in Fig. 2b assumes $\frac{\partial v}{\partial L_i} = 0$ and K_e increases without an increase in L_e, this result holds regardless of whether these assumptions hold or not.

The second case is concerned with the expansion of route a, while keeping route e's capacity unchanged. If route a is congested, then similar to the logic illustrated in Fig. 2b, route a's expansion is potentially welfare-improving. However, when route a is not congested, there are a number of different scenarios. If $b_a'(K_a) = 0$, then welfare decreases when K_a expands. However, if $b_a'(K_a) > 0$ or if $\frac{\partial v}{\partial L_i} > 0$ and K_a increases with an increase in L_a, then the expansion of K_a becomes potentially welfare-improving. Fig. 3 illustrates the case that the expansion of route a is potentially welfare-improving when $b_a'(K_a) > 0$.[9]

In sum, welfare potentially improves in cases where both routes are congested or when the utility of using the alternative increases. The latter case may result from an increase in speed or the coverage effect or both. The case where all routes are congested may apply to many large cities in developing countries where road construction is not up to the rapid growth rate of the city population. As mentioned above, the case of alternative speed increases may be particularly relevant if the road system is upgraded or the subway system is expanded.

3. Data description

Our data on national expressways in Japan come from a panel data set of road traffic censuses collected non-periodically in 1990, 1994, 1997, 1999, and 2005 by Japan's Ministry of Land, Infrastructure, Transport, and Tourism. In each census, the national expressway network is divided into hundreds of road segments, and for each segment the census reports the road length (in km), number of lanes, and average weekday daytime traffic (AWDT), which is measured by the average number of vehicles passing through the segment observation point between 7 am and 7 pm on a weekday. For each road segment, we calculate the lane kilometers by multiplying the road length and number of lanes, and calculate the average weekday daytime VKT (hereafter, VKT) by multiplying the AWDT by the road length.

We take Japan's UEAs as our observation units and aggregate the road segment data to the UEA level by traffic census year. Here, UEAs are metropolitan areas that are constructed using commuter flow data and in a similar way to how core based statistical areas (CBSAs, which include both the metropolitan and micropolitan statistical areas) are constructed in the US.[10] The building blocks of UEAs are municipalities (*shi-ku-cho-son*), and the UEA definition we use is based on the 2001 municipality boundaries using 2005 commuter flow data.[11] Covariates at the municipality level are obtained from Toyo Keizai, Inc.

Japan is a densely populated urban country. Over 94% of the population of Japan reside in the country's 189 UEAs, which make up two-thirds of its territory. By the end of our study period (2005), the national expressway network extended to 117 UEAs. Nonetheless, we focus on the 93 UEAs with national expressways at the beginning of the study period (1990) to ensure a balanced panel. In 2005, the 93 UEAs in our sample represented over 83% of the country's population. Table 1 reports the UEA-level descriptive statistics by traffic census year. During the study period of 1990–2005, the average road length and lane kilometers of national expressways increased by 37% (from 44 to 61 km) and 35% (from 179 to 241 km), respectively, whereas the VKT rose 64%, from 0.82 million to 1.27 million km. Unlike the road length and lane kilometers, both of which grew steadily over time, most of the growth in VKT occurred during 1990–1997. The slowdown of VKT growth was largely due to the inverted U-shaped pattern of AWDT, which grew between 1990 and 1997 but declined thereafter, in part reflecting the deepening stagnation of Japan's economy after the late 1990s. One-lane routes account for 9.4–12.4% of the expressway capacity measured in lane kilometers during our study period. The changes over time in the capacity share of one-lane routes in a UEA reflect a combination of two dynamics with opposite effects: the expansion of a single-lane route to multiple lanes (which reduces this share) and the addition of a new single-lane route to the existing network (which increases this share). These two processes also have different implications for traffic volume changes, as is discussed in Section 4.4. Table 1 also reports the statistics for population size and per capita income, which are used as control variables in some of our empirical specifications. Both of these variables changed only slightly during the study period.

4. Empirical results

We now turn to our empirical analysis of the effect of roadway capacity on VKT using the UEA-level aggregate data on national expressways in Japan. In Section 4.1, we begin with the cross-sectional estimations of the road elasticity of VKT by pooling data from the five traffic censuses. In Section 4.2, we use a fixed-effect model to eliminate any potential correlation between the road stock and the UEA-level time-invariant determinants of VKT. In Section 4.3, we further employ an IV approach to address the potential endogeneity of road capacity expansion in the fixed-effect model. We then investigate evidence for the possible reasons for the elasticity estimates to be larger than 1 in Section 4.4. Finally, we perform some robustness checks in Section 4.5 using alternative samples and identification strategies.

4.1. Cross-sectional estimation

The focus of our empirical analysis is the relationship between road capacity (measured by lane kilometers) and traffic (measured by VKT) for national expressways at the UEA level in Japan. Our cross-sectional specification of this relationship takes the following form:

$$\ln Q_{it} = \rho \ln K_{it} + \mathbf{X}_{it}'\beta + \tau_t + \varepsilon_{it}, \quad (2)$$

where Q_{it} denotes the VKT in UEA i at time t, K_{it} is the lane kilometers in UEA i at time t, X_{it} is a vector of covariates, τ_t is a time fixed effect, and ε_{it} represents the error term. Coefficient ρ is our parameter of interest, that is, the road elasticity of VKT. The OLS estimate of ρ is consistent if the log lane kilometers variable is uncorrelated with the error term, conditional on the covariates and time fixed effects, i.e., $cov(\ln K, \varepsilon | X, \tau_t) = 0$.

We pool the five traffic censuses to estimate the elasticity of weekday daytime VKT to lane kilometers, ρ. Columns 1–3 of

[9] Another potential welfare improvement is at a corner equilibrium in which all drivers are attracted to route e and no one is using route a. If at this corner equilibrium, speed on route e, s_e, is still at the congested portion of the speed function, then capacity expansion is potentially welfare-improving.

[10] Unlike in the US where CBSAs are compiled by government statistical agencies, these UEAs are compiled by urban economists, e.g., Kanemoto and Tokuoka (2002) and Mori et al. (2008). The UEA definitions that we use are provided by Tomoya Mori, whom we gratefully acknowledge. For more details of UEA construction, see Mori et al. (2008).

[11] Boundaries of municipalities changed quite often during our study period. We use the information on boundary changes obtained from the Municipality Map Maker website (http://www.tkirimura.com/mmm/) to make necessary adjustments to ensure that the mapping from municipalities to UEAs is consistent over time.

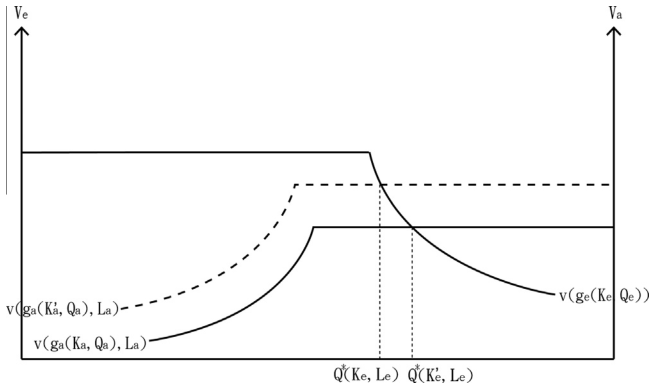

Fig. 3. Capacity expansion of route a accompanied with a higher speed. When the capacity expansion of route a (when uncongested) is accompanied with a higher speed, i.e., $b'_a(K_a) > 0$, the travel speed increases in both routes. Welfare potentially improves.

Table 2 present the OLS estimates of the elasticity of VKT to lane kilometers using the pooled panel data for the 93 UEAs in our analysis sample. Each column corresponds to one empirical specification. We include only our key variable of interest, log lane kilometers, and year dummies in column 1, and add log population and log per capita income sequentially in columns 2 and 3.[12] In column 1, the road elasticity of VKT is estimated to be 1.17 (with a standard error of 0.07) without controls and is significantly above 1 at the 5% level. The inclusion of log population in column 2 (and log per capita income in column 3) reduces the estimate of ρ to 1.02, consistent with both DT's unit elasticity test $\rho = 1$ and our test $\rho \geqslant 1$.

4.2. Fixed-effect estimation

For the OLS estimates to be consistent, we require the road capacity to be uncorrelated with the unobserved error term in Eq. (2). However, a UEA's stock of roads may be partly determined by VKT or be correlated with unobserved UEA-level characteristics that affect VKT. In either case, the road capacity will be endogenous in Eq. (2), violating the orthogonality assumption required for consistent estimation of the elasticity ρ. A number of extant studies have taken advantage of the availability of panel data to adopt a fixed-effect approach to mitigate the potential correlation between road capacity and the unobserved determinants of VKT (e.g., Hansen and Huang, 1997; Noland, 2001). Using the five traffic censuses in Japan, we also estimate the following fixed-effect model,

$$\ln Q_{it} = \rho \ln K_{it} + \mathbf{X}'_{it}\beta + \tau_t + \delta_i + \eta_{it}, \quad (3)$$

where the error term ε_{it} in Eq. (2) is decomposed into a time-invariant UEA-specific component (δ_i) and a time-varying UEA-specific component (η_{it}). Although both components are unobservable, the former (δ_i) can be controlled by adding a set of UEA dummies to the cross-sectional regressions. The potential correlation between the road stock and the unobserved determinants of VKT is mitigated in Eq. (3) because the inclusion of UEA fixed effects eliminates any correlation between the road stock and the UEA-level time-invariant determinants of VKT. The fixed-effect estimates of ρ are consistent if the log lane kilometers variable is uncorrelated with the time-varying error component (η_{it}) of Eq. (3), conditional on the covariates and UEA and time fixed effects, i.e., $cov(\ln K, \eta|\mathbf{X}, \delta_i, \tau_t) = 0$, or, put in another way, $cov(\Delta \ln K, \Delta\eta|\Delta X, \Delta\tau_t) = 0$.

Columns 4–6 of Table 2 replicate the same regressions as those in columns 1–3, but add UEA fixed effects. The fixed-effect estimates of ρ are insensitive to the inclusion of population size and income as additional controls and range narrowly from 1.13 to 1.15. Although all point estimates of ρ are greater than 1, they cannot reject the unit elasticity test at the conventional significance levels owing to the larger fixed-effect standard errors (0.10), which can be attributable to the reduction in the variation in log lane kilometers after the inclusion of UEA fixed effects. The coefficient on log per capita income is positive and significant, suggesting a positive correlation between income and VKT growth at the UEA level. The coefficient on log population has an unexpected negative sign but is small and insignificant in both columns 5 and 6. The imprecision in the estimate of the coefficient on log population is not very surprising given the lack of time variation at the UEA level during the study period.

Although the fixed-effect estimations mitigate the possible bias of the OLS estimations by eliminating the potential correlation between lane kilometers and the unobserved UEA-level permanent determinants of VKT, the fixed-effect estimates may still be biased if changes in lane kilometers are correlated with the unobserved time-varying UEA-level determinants of VKT. There are two potential sources of these correlations. First, changes in lane kilometers and changes in VKT may be simultaneously determined. Not only can changes in lane kilometers induce changes in VKT, but the causality can also run in the opposite direction. For instance, road planners who expect traffic to soar in certain UEAs may assign more roadways to them. Second, changes in lane kilometers in a UEA may be correlated with time-varying omitted variables that affect changes in the VKT through contemporary shocks. For instance, a government adopting an expansionary fiscal policy may build new roadways and boost economic activities (in both the public and private sectors) that generate new traffic. In either case, lane kilometers will be correlated with the error term η_{it} in the fixed-effect model. Therefore, roadway capacity can also be endogenous in the fixed-effect model, leading to biases in the fixed-effect estimates of ρ. A possible way to address this issue is to find an instrument for roadway expansion that is uncorrelated with the time-varying determinants of VKT growth at the UEA level. In the next subsection, we adopt this approach and conduct an IV estimation of the fixed-effect model.

4.3. Fixed-effect IV estimation

Our instrument for the expansion of the national expressway network is taken from Japan's fourth National Comprehensive Development Program (NCDP-4), which was approved in June 1987. The NCDP-4 is a fundamental plan for the use, development, and conservation of land in Japan and defines directions for the construction of infrastructure. An important component of the NCDP-4 is the national expressway network plan shown in Fig. 4, which aims to extend the country's national expressway network to 11,520 km, an almost threefold expansion of the existing national expressway network in 1987. As of March 2005, 7363 km of national expressways, approximately 64% of the planned total, had been completed. Our strategy is to employ *the planned national expressway extension* in the NCDP-4 to instrument for the actual growth of national expressways at the UEA level.

[12] In addition to log per capita income, DT also control for other socioeconomic characteristics, such as the poverty rate, the share of the population with a college education, and the share of manufacturing employment, and find that the estimates of ρ are insensitive to the inclusion of these additional socioeconomic variables in the US. We are unable to include these socioeconomic variables for Japan, as they are available only from the bidecennial population censuses, whereas the traffic censuses are collected non-periodically. More importantly, the socioeconomic variables of a UEA are likely to be affected by its road capacity, and hence their inclusion might even bias the estimates of ρ. Thus, DT also prefer the OLS specifications that do not control for additional socioeconomic variables, even when these variables are available for use.

Table 1
Summary statistics.

Year	1990	1994	1997	1999	2005
Road length (km)	44	51	54	55	61
	(53)	(57)	(62)	(64)	(67)
Lane km	179	204	214	221	241
	(240)	(274)	(297)	(308)	(321)
Average weekday daytime traffic	13,773	16,454	18,171	17,995	17,502
	(9792)	(10,200)	(11,027)	(10,803)	(10,944)
Average weekday daytime VKT ('000 km)	820	1067	1210	1215	1265
	(1982)	(2384)	(2716)	(2698)	(2703)
Capacity share of one-lane routes (in lane km)	0.094	0.121	0.124	0.094	0.104
	(0.281)	(0.283)	(0.283)	(0.229)	(0.216)
Population ('000)	1066	1095	1106	1113	1099
	(3533)	(3681)	(3719)	(3756)	(3955)
Real per capita income ('000 yen)	982	1212	1212	1264	1134
	(233)	(261)	(268)	(264)	(245)
Number of observations	93	93	93	93	93

Notes: The table reports the mean and standard deviation (in parentheses) of each variable by the traffic census year for the sample of 93 UEAs with national expressways since 1990.

Table 2
The elasticity of VKT to lane kilometers, OLS and fixed-effect estimations.

	OLS			FE		
	(1)	(2)	(3)	(4)	(5)	(6)
Log (lane km)	1.174***	1.022***	1.022***	1.146***	1.145***	1.133***
	(0.074)	(0.077)	(0.079)	(0.100)	(0.101)	(0.100)
Log (population)		0.218***	0.226***		−0.047	−0.127
		(0.049)	(0.050)		(0.076)	(0.083)
Log (per capita income)			0.295*			0.970**
			(0.154)			(0.383)
UEA fixed effects				Y	Y	Y
p-Value of the t-test of $\rho = 1$	0.021	0.778	0.781	0.149	0.154	0.186
R^2	0.797	0.826	0.831	0.985	0.985	0.986
Number of observations	465	465	465	465	465	465

Notes: The dependent variable is log VKT. All regressions include year fixed effects. Columns (2) and (4) control for log population, whereas columns (3) and (6) control for both log population and log per capita income. Columns (4)–(6) include UEA fixed effects. Robust standard errors clustered by UEA are reported in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

The working of this IV strategy can be illustrated as follows. Let γ_t^* denote the national-level *incremental completion rate* of the NCDP-4 from time $t-1$ to time t, defined as

$$\gamma_t^* = \frac{\ln L_t - \ln L_{t-1}}{\ln L_T - \ln L_0}, \tag{4}$$

i.e., the ratio of the increment of expressway length (in logarithm) from time $t-1$ to time $t(\ln L_t - \ln L_{t-1})$ and the planned expressway extension (in logarithm) in the NCDP-4 $(\ln L_T - \ln L_0)$. If we apply this national-level incremental completion rate to the UEA level, we will predict the growth of roadway capacity from time $t-1$ to time t for UEA i as

$$\widehat{\Delta \ln K_{it}} = \gamma_t^* R_i = \gamma_t^* \, (\ln L_{iT} - \ln L_{i0}), \tag{5}$$

where R_i is defined as the planned expressway extension (in logarithm) for UEA i in the NCDP-4. Concerns over the consistency of the fixed-effect estimates arise from the possible correlation between the observed roadway capacity expansion $(\Delta \ln K_{it})$ and the change in the time-varying UEA-specific determinants of VKT $(\Delta\eta_{it})$. As a local government may speed up roadway construction during both the upturn and downturn of the regional economy (e.g., to meet the rising travel demand in the former case and to create jobs in the latter case), the bias can go in either direction. However, the predicted growth of roadway capacity in a UEA in Eq. (5), $\widehat{\Delta \ln K_{it}}$, is not subject to this concern because it does not respond to the change in the unobserved time-varying local conditions affecting VKT $(\Delta\eta_{it})$. Rather, the predicted growth depends on the national-level completion rate (γ_t^*) and the planned local extension in the NCDP-4 (R_i). It is important to note that while R_i is an endogenous variable at the UEA level based on expectations about future travel demand, there is no *a priori* reason for R_i to be correlated with the change in the time-varying local determinants of VKT $(\Delta\eta_{it})$ at any time once the UEA fixed effects are accounted for.[13]

When applying this IV strategy to our data, we make two adjustments to the procedures illustrated above. First, because we lack information on the length of the existing national expressways in 1987, we measure the planned extension of national expressways for each UEA (R_i) by the difference (in logarithm) between its planned roadway length in the NCDP-4 and its actual

[13] A violation of the exclusion restriction (i.e., $E[(\gamma_t^* R_i)\Delta\eta_{it}] \neq 0$) would require not only that R_i and $\Delta\eta_{it}$ be correlated at times (i.e., $E_i[R_i\Delta\eta_{it}|t] \neq 0$ for some t's), but also that this correlation needs to vary systematically over time to exhibit a correlation with γ_t^*. For example, a positive (negative) correlation between $E_i[R_i\Delta\eta_{it}|t]$ and γ_t^* arises if $\Delta\eta_{it}$ and R_i are positively (negatively) correlated when γ_t^* is large and are negatively (positively) correlated when γ_t^* is small. Nevertheless, it is unlikely that R_i, which was predetermined at the beginning of our study period for each UEA, would be correlated with $\Delta\eta_{it}$ at any given point in time given that the UEA fixed effects are accounted for. Moreover, even if such correlation exists, there is also no *a priori* reason for it to exhibit a specific time pattern to be correlated with γ_t^*.

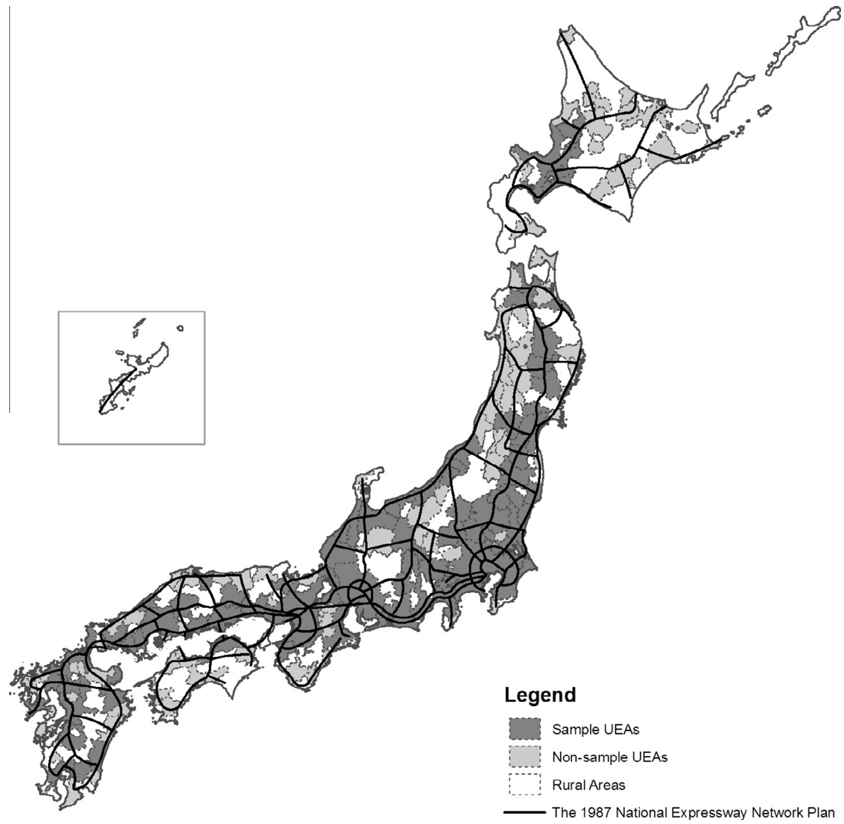

Fig. 4. The 1987 national expressway network plan for Japan.

roadway length in 1990, the earliest year for which data are available. Second, instead of calculating the national-level incremental completion rate for each period using Eq. (4), which is defined based on changes in roadway length ($\Delta \ln L_{it}$), we employ a first-stage regression to estimate it according to the cross-sectional relationship between changes in roadway lane kilometers ($\Delta \ln K_{it}$), which is our direct measure of roadway capacity change, and R_i in each period. Specifically, our first-stage regression takes the following form,

$$\ln(K_{it}) = \sum_{s=1}^{4} \gamma_s R_i \cdot \mathbf{1}(s \leqslant t) + \mathbf{X}'_{it}\phi + \kappa_t + \mu_i + v_{it}, \tag{6}$$

where $\gamma_1 - \gamma_4$ correspond to the incremental completion rates of the NCDP-4 for the 1990–1994, 1994–1997, 1997–1999, and 1999–2005 periods, respectively. An advantage of the first-stage specification in Eq. (6) is that it allows the speed of construction of the extension in the NCDP-4 to vary across periods, as observed in the data. However, to account for differences in the speed of construction across periods, the specification essentially uses four instruments (i.e., the interactions between R_i and the four time dummies), leading to concern over weak identification due to the use of multiple instruments. To address this concern, we assume a constant annual completion rate of the NCDP-4 during the study period and estimate an alternative first-stage equation using a single instrument as follows:

$$\ln K_{it} = \theta(Y_t R_i) + \mathbf{X}'_{it}\phi + \kappa_t + \mu_i + v_{it}, \tag{7}$$

where Y_t is the number of years since 1990 at time t, and the coefficient θ is the annualized completion rate of the NCDP-4 during the study period. Note that θR_i is the expected annual growth rate of national expressways in UEA i and $\theta Y_t R_i$ is the expected cumulative growth of national expressways in UEA i at time t relative to the initial level in 1990. Both rates vary across UEAs because of their variation in R_i.

Panel A of Table 3 presents our first-stage estimation results for the fixed-effect model. Columns 1–3 estimate the specification in Eq. (6), and columns 4–6 estimate that in Eq. (7). The inclusion of additional control variables has almost no effect on the coefficient(s) of the instrument(s) in the first-stage regressions. We thus discuss only the results in columns 1 and 4. Our estimates of γ_s in Eq. (6) suggest that the actual expansion of roadway capacity in a UEA is significantly correlated with its planned extension in the NCDP-4 for three out of four periods under study (the 1994–1997 period is the exception). Specifically, the incremental completion rate of the planned extension in the NCDP-4 is estimated to be 18.2% for the period 1990–1994, 7.6% for the period 1997–1999, and 10.9% for the period 1999–2005. While roadway capacity still increased by 4.9% at the national level during the period 1994–1997, changes in roadway capacity ($\ln K_{it}$) are almost uncorrelated with R_i at the UEA level in this period. With the use of the four instruments, the first-stage F-statistic is 4.33. Although this falls below the conventional critical value for the weak-instrument test using the 2SLS tabulated by Stock and Yogo (2005), it is still above the critical value for 15% maximal LIML size (3.87). Column 4 estimates the annualized completion rate of the NCDP-4 to be 2.4% during our study period, significant at the 1% level. The first-stage F-statistic is 11.83, above the rule-of-thumb threshold of 10 for the weak-instrument test, suggesting that at least the first-stage specification in Eq. (7) is not subject to the weak-instrument problem.

Panel B of Table 3 presents the 2SLS estimations of Eq. (3) corresponding to the first-stage regressions in Panel A. The estimates of the coefficient on log lane kilometers are insensitive to

Table 3
The elasticity of VKT to lane kilometers, fixed-effect IV estimations.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A (First-stage). Dependent variable: log (lane km)						
Planned extension in the NCDP-4 X number of years since 1990				0.024***	0.023***	0.024***
				(0.007)	(0.007)	(0.007)
Planned extension in the NCDP-4 X (year ⩾ 1994)	0.182**	0.182**	0.186**			
	(0.077)	(0.077)	(0.078)			
Planned extension in the NCDP-4 X (year ⩾ 1997)	0.007	0.007	0.010			
	(0.014)	(0.014)	(0.014)			
Planned extension in the NCDP-4 X (year ⩾ 1999)	0.076*	0.076*	0.074*			
	(0.039)	(0.039)	(0.040)			
Planned extension in the NCDP-4 X (year ⩾ 2005)	0.109*	0.107*	0.106*			
	(0.061)	(0.061)	(0.062)			
Log (population)		Y	Y		Y	Y
Log (personal income)			Y			Y
First-stage *F*-statistic	4.33	4.30	4.43	11.83	11.62	11.72
R^2	0.968	0.968	0.968	0.967	0.967	0.968
Panel B (2SLS). Dependent variable: log (VKT)						
Log (lane km)	1.282***	1.281***	1.299***	1.238***	1.235***	1.249***
	(0.130)	(0.131)	(0.129)	(0.125)	(0.126)	(0.126)
p-Value of the *t*-test of $\rho = 1$	0.030	0.032	0.021	0.058	0.063	0.048
R^2	0.985	0.985	0.985	0.985	0.985	0.985
Panel C (LIML). Dependent variable: log (VKT)						
Log (lane km)	1.308***	1.308***	1.335***	1.238***	1.235***	1.249***
	(0.149)	(0.151)	(0.151)	(0.125)	(0.116)	(0.126)
p-Value of the *t*-test of $\rho = 1$	0.039	0.041	0.026	0.058	0.063	0.048
R^2	0.985	0.985	0.985	0.985	0.985	0.985
Number of observations	465	465	465	465	465	465

Notes: Panels A, B and C report the first-stage, 2SLS, and LIML estimation results, respectively. Columns (1)–(3) use the interaction terms between the planned expressway extension in each UEA and time dummies as the instruments. Columns (4)–(6) employ the product of the planned extension in each UEA and the number of years since 1990 as the instrument. All regressions include UEA and year fixed effects. Columns (2) and (5) control for log population, and columns (3) and (6) control for both log population and log per capita income. Robust standard errors clustered by UEA are reported in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

the choice of either the first-stage specifications or the control variables, and range very narrowly from 1.24 to 1.30. As all of the point estimates are above 1, they are all consistent with our test of $\rho \geqslant 1$. However, unit elasticity ($\rho = 1$) can be rejected at the ten percent level by all these estimates and at the five percent level by four of them. To address the potential concern over weak instruments, especially for the estimations in columns 1–3, we also conduct the LIML estimations and report the results in Panel C. The LIML estimates are essentially the same as the 2SLS estimates in Panel B, suggesting evidence against the unit elasticity hypothesis.

4.4. Evidence for the sources of the larger-than-unity elasticity

Given that our point estimates of elasticity are consistently larger than 1 in all specifications and the unit elasticity hypothesis can be rejected in the fixed-effect IV estimations, it is worth investigating the possible reasons for the elasticity to be larger than 1. As our model suggests that a larger-than-unity elasticity can result from either the coverage effect or the increasing returns to scale in the speed function when the fundamental law holds, we empirically examine evidence for both possibilities in this subsection. To do so, we add two other features – the roadway length and the capacity share of the one-lane routes (in lane kilometers) – to characterize the capacity of a roadway system in addition to lane kilometers. If the fundamental law holds, then conditional on lane kilometers, the existence of a significant coverage effect would imply a positive relationship between VKT and roadway length, whereas increasing returns to scale in the speed function would suggest a negative relationship between VKT and the capacity share of one-lane routes. Because of the difficulty of locating appropriate additional instruments for different components of roadway expansion, we are unable to conduct IV estimations of the extended specifications that further incorporate roadway length and/or the capacity share of one-lane routes. However, this limitation is unlikely to be a real problem given that the concern over the endogeneity of roadway expansion did *not* materialize in our previous estimations, that is, the fixed-effect IV estimates and the fixed-effect estimates are not significantly different from each other in all specifications. Thus, we only use the fixed-effect estimations in the exercises in this subsection.

Table 4 reports the fixed-effect estimations of alternative extended specifications of Eq. (3) that incorporate roadway length or the capacity share of one-lane routes or both to characterize roadway capacity in addition to lane kilometers. All columns

Table 4
Fixed-effect estimations of the extended specifications.

	(1)	(2)	(3)
Log (lane km)	0.888***	1.146***	0.652*
	(0.092)	(0.041)	(0.383)
Log (road length)	0.264***		0.480
	(0.085)		(0.370)
Capacity share of one-lane routes (in lane km)		0.178**	−0.174
		(0.064)	(0.279)
R^2	0.986	0.986	0.986
Number of observations	465	465	465

Notes: The dependent variable is log VKT. All regressions include log population, log per capita income, and UEA and year fixed effects. Robust standard errors clustered by UEA are reported in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Table 5
Robustness analysis.

	Honshu Island			Rest of Japan		
	OLS (1)	FE (2)	FE-IV (3)	OLS (4)	FE (5)	FE-IV (6)
Panel A. Honshu Island vs. rest of Japan						
Log (lane km)	1.105***	1.202***	1.381***	0.898***	1.101***	1.030***
	(0.098)	(0.121)	(0.159)	(0.053)	(0.157)	(0.217)
p-Value of the t-test of $\rho = 1$	0.284	0.099	0.016	0.069	0.944	0.891
R^2	0.833	0.985	0.972	0.869	0.988	0.985
Number of observations	340	340	340	125	125	125
	Coastal UEAs			Inland UEAs		
Panel B. Coastal UEAs vs. Inland UEAs						
Log (lane km)	1.002***	1.214***	1.086***	0.977***	1.054***	1.255***
	(0.095)	(0.139)	(0.223)	(0.124)	(0.138)	(0.116)
p-Value of the t-test of $\rho = 1$	0.863	0.129	0.699	0.851	0.698	0.028
R^2	0.863	0.990	0.989	0.787	0.974	0.972
Number of observations	285	285	285	180	180	180

Notes: The dependent variable is log VKT. All regressions include log population, log per capita income, and year fixed effects. Columns (2)–(3) and (5)–(6) also include UEA fixed effects. Columns (3) and (6) employ the product of the planned extension in each UEA and the number of years since 1990 as the instrument. Robust standard errors clustered by UEA are reported in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

include log population and log per capita income as covariates and also control for UEA and time fixed effects. Therefore, the benchmark for comparison is column 6 of Table 2 with an estimated coefficient on log lane kilometers of 1.133. Column 1 of Table 4 estimates a specification that includes log lane kilometers ($\ln K_{it}$) and log road length ($\ln L_{it}$). In this specification, the coefficient on $\ln K_{it}$ corresponds to the elasticity of VKT to an increase in roadway capacity measured by the lane kilometers due to lane expansion only, i.e., an increase in K holding L constant, whereas the coefficient on $\ln L_{it}$ corresponds to the additional elasticity effect on VKT of an increase in road length net of that arising from the increase in K. Specifically, if a new road were built with exactly the same number of lanes as the average of the existing roadway system, i.e., K and L increased by the same proportion, the overall elasticity effect on VKT of such a roadway extension would be the sum of the two coefficients. The coefficient on log lane kilometers in column 1, 0.888, is significantly smaller than the benchmark estimate (1.133), but is still consistent with both the unit elasticity test $\rho = 1$ and our test $\rho \geqslant 1$. The estimated coefficient on log road length (0.264) is positive and significant at the one percent level, indicating that roadway extension generates additional traffic compared with the same increase in lane kilometers purely due to lane expansion, thereby supporting the existence of the coverage effect.

Column 2 of Table 4 adds the capacity share of one-lane routes (denoted by π) to the benchmark fixed-effect specification with log lane kilometers as the only measure of roadway capacity. If there exist increasing returns to scale in the speed function, one-lane routes should be less efficient in accommodating traffic while holding lane kilometers constant. However, the estimated coefficient on the capacity share of one-lane routes (0.178) has an unexpected positive sign and is significant at the five percent level. As mentioned in Section 3, changes in the capacity share of one-lane routes reflect a combination of two dynamics that imply opposite correlation with changes in VKT. First, the expansion of a single-lane route to multiple lanes reduces π but can increase VKT more than proportionally to K if there exist increasing returns to scale in the speed function; that is, $corr(\Delta\pi, \Delta \ln Q | \Delta \ln K) < 0$ when $\Delta\pi < 0$. Second, the addition of a new single-lane route to the existing roadway network increases π but can also increase VKT more than proportionally to K if there exists a coverage effect; that is, $corr(\Delta\pi, \Delta \ln Q | \Delta \ln K) > 0$ when $\Delta\pi > 0$. The positive coefficient on the capacity share of one-lane routes found here is likely to be due to the dominance of the second force over the first one.

In column 3 of Table 4, we further include log road length to separately account for the coverage effect so that the coefficient on the capacity share of one-lane routes is not driven by the coverage effect in the second force. The sign of the coefficient on the capacity share of one-lane routes is reversed to negative, consistent with the increasing returns to scale hypothesis, whereas the sign of the coefficient on log road length is positive, consistent with the coverage effect hypothesis. However, both these coefficients are insignificant and the coefficient on log lane kilometers becomes only marginally significant because of the high multicollinearity among these related measures of roadway capacity. Taken together, the results in columns 1 and 2 of Table 4 suggest that the coverage effect may play a key role in explaining the larger-than-unity elasticity, whereas those in column 3 are also consistent with the existence of increasing returns to scale in the speed function.

4.5. Robustness

In this subsection, we examine whether our elasticity estimates are robust to the choice of samples and identification strategies. Japan is a maritime nation and comprises four major islands (Hokkaido, Honshu, Shikoku, and Kyushu, from north to south), of which Honshu Island is the largest one and makes up 60% of the country's total land area. It is possible that geographic factors play a role in affecting the relationship between expressway capacity and traffic. To examine whether our results are robust to the influences of geographical factors, we conduct separate estimations for UEAs on Honshu Island and those in the rest of Japan as well as for coastal areas and inland areas. Panel A of Table 5 shows separate estimates for the 68 UEAs on Honshu Island (columns 1–3) and the remaining 25 UEAs not on Honshu Island (columns 4–6). Except for the OLS coefficient for the subsample of UEAs not on Honshu Island (0.898), which is smaller than 1 with a marginal significance level of 0.069, all other coefficients are larger than 1 and hence consistent with our test $\rho \geqslant 1$. Moreover, the coefficients for Honshu Island subsample are consistently larger than those for the other

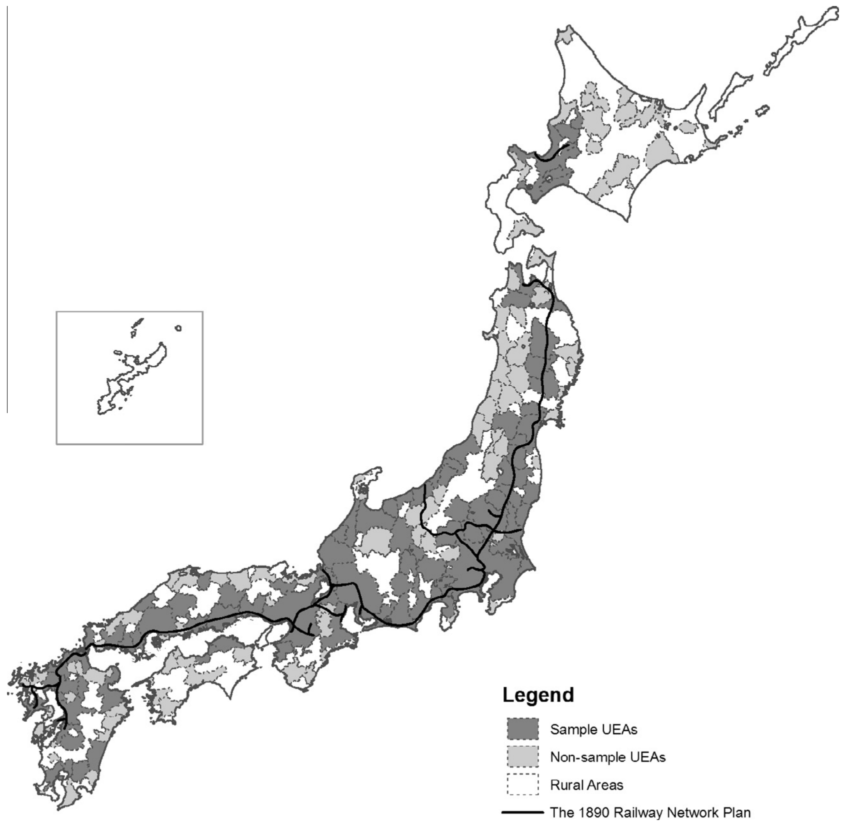

Fig. 5. The1890 railway network plan for Japan.

subsample and can reject unit elasticity in both the fixed-effect and fixed-effect IV specifications, suggesting that our larger-than-unity estimates of the elasticity for the entire sample stem mainly from the results for Honshu Island. Panel B of Table 5 shows separate estimates for the 57 coastal UEAs (columns 1–3) and the 36 inland UEAs (columns 4–6). All coefficients are consistent with our test $\rho \geqslant 1$. While the fixed-effect IV coefficient for the inland UEAs can reject unit elasticity at the five percent level, we cannot reject the equality in the coefficient between the coastal and inland UEAs for all three specifications.

Next, we consider an alternative empirical strategy that employs an IV estimation of the cross-sectional model, Eq. (2), to address the potential endogeneity of the stock of roadway capacity. Following Duranton and Turner (2011) and Duranton and Turner (2012) use of the historical rail routes as an instrument for the road capacity of interstate highways in the US, we employ the 1890 railway network plan in Japan as our instrument for the stock of national expressways. Our measures of the UEA-level railway length in the 1890 plan are based on a digital image of its map extracted from the *1890 Official Document Appendix No. 2011*. We convert this image into a digital map and use the Geographic Information System (GIS) to calculate the length of railway in each UEA. Fig. 5 shows a GIS image of the 1890 railway network plan: among the 93 UEAs in our sample, 44 UEAs had no railways planned within their jurisdictions, whereas the other 49 UEAs had the planned railway length ranging from 8.9 km to 252.2 km. A practical challenge we face for employing the historically planned railway length as an instrument for $\ln K_{it}$ in Eq. (2) is to deal with the log transformation of zero values. To address this issue, we generate a railway dummy taking the value 1 if a UEA was designated a positive railway length in the 1890 plan and 0 otherwise, and assign a value of 0 for the log planned railway length for UEAs with no planned railway in 1890, and include both variables in the first-stage regressions. It is important to note that our second-stage (i.e., IV) estimation results would remain the same if we replace the 0 value we assign to log planned railway length for UEAs without planned railways by any other constant value (such as the logarithm of the minimum positive railway length observed in our sample). This is because, regardless of the value assigned for the subsample of UEAs without planned railways, the "transformed" log railway length variable is always perfectly collinear to the railway dummy for this subsample.[14]

Panel A of Table 6 presents our first-stage regressions with the log lane kilometers of national expressways as the dependent variable. Estimates of the elasticity of national expressway lane kilometers to the planned railway length range between 0.94 and 1.11, suggesting that, among UEAs with planned railways in 1890, the capacity of national expressways increases approximately proportionally to the planned railway length. However, not all of the UEAs assigned a positive railway mileage are more advantageous in national expressway capacity compared with an average UEA not assigned any railway in 1890. This is reflected by the negative and significant coefficient estimates on the railway dummy in all three specifications. Taking column 3 in Panel A of Table 6 as an example, the coefficients on the railway dummy (−3.167) and log planned railway length (0.951) indicate that only UEAs with more than 28 km[15] of railway planned in 1890, or 70% of those assigned a positive railway mileage, are predicted to have

[14] That is, the subsample of UEAs without planned railway in 1890 will not contribute to the estimate of the coefficient on log planned railway length in the first-stage estimation, regardless of the value assigned to this variable for this subsample. Differences in the value assigned to log planned railway length for this subsample will only shift the coefficient on the railway dummy accordingly, resulting in the same predicted log lane kilometers for national expressways from the first-stage regressions and therefore the same second-stage estimation results.

[15] This number is calculated as $e^{3.167/0.951}$.

Table 6
The elasticity of VKT to lane kilometers, cross-sectional IV.

	(1)	(2)	(3)
Panel A (First-stage). Dependent variable: log (lane km)			
Log (railway length)	1.105**	0.941***	0.951***
	(0.173)	(0.222)	(0.195)
Railway dummy	−3.644***	−3.141***	−3.167***
	(0.704)	(0.833)	(0.850)
Log (population)		Y	Y
Log (per capita income)			Y
First-stage *F*-statistic	42.37	11.29	10.86
R^2	0.454	0.463	0.465
Panel B (2SLS). Dependent variable: log VKT			
Log (lane km)	1.246***	0.801***	0.830***
	(0.085)	(0.127)	(0.120)
p-Value of the *t*-test of $\rho = 1$	0.004	0.118	0.159
R^2	0.794	0.807	0.817
Number of observations	465	465	465

Notes: Panels A and B report, respectively, the first-stage and 2SLS estimation results using the railway dummy and log planned railway length as the instruments for log lane kilometers. All regressions include year fixed effects. Column (2) controls for log population and column (3) controls for both log population and log per capita income. Robust standard errors clustered by UEA are reported in parentheses.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

more national expressways than those with no railway assigned in 1890. Panel B of Table 6 presents our IV estimations of Eq. (2), corresponding to the first-stage regressions in Panel A. When estimated without any controls, the coefficient on log lane kilometers is 1.25 and significantly larger than 1. The elasticity estimate is reduced to 0.80 (in column 2) and 0.83 (in column 3) when control variables are included, neither of which can reject unit elasticity or our test $\rho \geqslant 1$. Nonetheless, it is advisable to exercise caution over the exogeneity of the historical railway plan as an instrument as it may be potentially correlated with the unobserved time-invariant determinants of VKT at the UEA level (such as historical or geographical factors or even the railway itself).

5. Conclusion

The fundamental law of highway congestion states that when congested, the travel speed of an expanded highway reverts to its previous level before the capacity expansion. This paper generalizes the traditional theory of this law by allowing an additional effect on drivers' utility other than travel speed or time. In our modeling, this additional effect is due to a larger coverage. Together with potential increasing returns in the speed function, this coverage effect implies that the road elasticity of traffic can be larger than 1. Our theory therefore predicts that if the fundamental law holds, this elasticity is at least 1. As our point estimates of this elasticity are generally larger than 1, the Japanese traffic data are consistent with the fundamental law. However, as these point estimates are systematically larger than those of DT's, which are much closer to 1, we then investigate the two possible sources of the larger-than-unity elasticity suggested in our model, namely, the coverage effect and the increasing returns to scale in the speed function. While the results are stronger for the coverage effect, our empirical evidence is consistent with both hypotheses.

From a broad perspective, our results are drastically different from those in the literature prior to DT in which the estimates obtained are always significantly lower than 1. This strengthens the message that "building your way out of congestion" is often fruitless, as the expansion of congested roads can induce more than proportionate increases in traffic. Nevertheless, it is important to note the "local" nature of the fundamental law. The law does not necessarily apply to all roads, since there may be many uncongested roads (especially in cities in more developed countries), and what distinguishes our work and DT's from the previous literature is mainly the focus on the most congested roads (in addition to better identification). Moreover, the existence of the fundamental law does not necessarily imply that building roads is a bad idea, as there are other scenarios in which building roads can be welfare-improving. As discussed in Section 2.4, these are cases where *all* roads are congested (which is likely to occur in rapidly growing cities in developing countries) and where the utility of using alternatives increases.

In this paper, we study the fundamental law based on its link with the road elasticity of traffic. However, the validity of the fundamental law may be inferred more firmly by examining the relationship between travel speed and road capacity. Doing so, however, would require direct measures of travel speed and is thus beyond the scope of the current paper. Research in this direction would be desirable to achieve a better understanding of the fundamental law and other issues related to traffic congestion.

References

Arnott, Richard, Small, Kenneth, 1994. The economics of traffic congestion. American Scientist 82, 446–455.
Baum-Snow, Nathaniel, 2007. Did highways cause suburbanization? Quarterly Journal of Economics 122, 775–805.
Cervero, Robert., 2002. Induced travel demand: research design, empirical evidence, and normative policies. Journal of Planning Literature 17, 2–20.
Cervero, Robert, Hansen, Mark, 2002. Induced travel demand and induced road investment: a simultaneous equation analysis. Journal of Transport Economics and Policy 36, 469–490.
Downs, Anthony, 1962. The law of peakhour expressway congestion. Traffic Quarterly 16, 393–409.
Downs, Anthony, 2004. Still Stuck in Traffic: Coping With Peak Hour Traffic Congestion. Brookings Institution Press, Washington, DC.
Duranton, Gilles, Turner, Matthew A., 2009. The Fundamental Law of Road Congestion: Evidence from US Cities. NBER working paper 15376.
Duranton, Gilles., Turner, Matthew A., 2011. The Fundamental law of road congestion: evidence from US cities. American Economic Review 101, 2616–2652.
Duranton, Gilles, Turner, Matthew A., 2012. Urban Growth and Transportation. Review of Economic Studies 79 (4), 1407–1440.
Goodwin, Phil B., 1996. Empirical evidence on induced traffic. Transportation 23, 35–54.
Hansen, Mark, Huang, Yuanlin, 1997. Road supply and traffic in california urban areas. Transportation Research A: Policy and Practice 31, 205–218.
Holden, David J., 1989. Wardrop's third principle: urban traffic congestion and traffic policy. Journal of Transportation Economics and Policy 23, 239–262.
Kanemoto, Yoshitsugu, Tokuoka, Kazuyuki, 2002. The proposal for the standard definition of the metropolitan area in Japan. Journal of Applied Regional Science 7, 1–15 (in Japanese).
Keeler, Theodore E., Small, Kenneth A., 1977. Optimal peak-load pricing, investment, and service levels on urban expressways. Journal of Political Economy 85, 1–25.
Mohring, Herbert, 1976. Transportation Economics. Ballinger, Cambridge, MA.
Mori, T., Nishikimi, K., Smith, T.E., 2008. The number-average size rule: a new empirical relationship between industrial location and city size. Journal of Regional Science 48, 165–211.
Noland, Robert B., 2001. Relationships between highway capacity and induced vehicle travel. Transportation Research Part A: Policy and Practice 35, 47–72.
Noland, Robert B., Cowart, William A., 2000. Analysis of metropolitan highway capacity and the growth in vehicle miles of travel. Transportation 27, 363–390.
Small, Kenneth A., 1997. Economics and urban transportation policy in the United States. Regional Science and Urban Economics 27, 671–691.
Small, Kenneth A., Verhoef, Erik T., 2007. The Economics of Urban Transportation. Routledge, New York.
Stock, James H., Yogo, Motohiro, 2005. Testing for weak instruments in linear IV regression. In: Andrews, D.W.K., Stock, J.H. (Eds.), Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, Cambridge, pp. 80–108.