BREAD-IGC Virtual PhD course on urban economics

‘The Connected city’: Urban road infrastructure and congestion

Gilles Duranton
University of Pennsylvania

On-line, October 16, 2024

Introduction

Cities are about *connections*: home to work, work to suppliers or customers, and home (or work) to shopping, schools, friends, government offices, leisure, places of worship, etc.

These connections are why we have cities in the first place. They allow for better sharing, matching, and learning of goods (consumption, intermediate, and public), factors, and ideas

⇒ Urban connections are the very source of agglomeration benefits.

To function properly and provide agglomeration benefits to its residents, a `good' transportation infrastructure is a necessary condition for easy access to places within the city.

We share this lecture with Gabriel Kreindler.
I will cover:

- Roadway infrastructure and determinants of travel time (traffic and travel effects)
- Traffic congestion

Gabriel will cover:

- Public transit
- Broader effects of urban mobility

Preliminary 1:

What are roads for?

Let us first step back.

The organization of the city is a land use problem: how should urban land be allocated across its different uses? How big should the city be given population? (Solow and Vickrey JET 1971, Pines et al. JUE 1985, Fujita 1989)

Urban residential and business locations need public spaces to allow access in and out.

- Public spaces "in-between" businesses and residences can be useful for transportation. However, they are also valuable for sellers and residents to transact, residents to hang out, children to play, animals to rest, etc.
  How much of the land useable for transportation should be used for the benefit of motorized vehicles?
- Land allocated to transportation must form a network, hopefully an efficient one (Fajgelbaum and Schaal E 2020)
- There are several modes of transportation (a continuum?), operating at different speeds, road space intensity (congestion), and other externalities (pollution, noise, accidents).
- Land allocated to transportation is also a durable investment (infrastructure) with effects on the efficiency and quality of transportation services.

Preliminary 2:

Developing vs. developed cities

Possibly apocryphal story:

Scott Fitzgerald: "The rich are different from you and me,"

Reply from Ernest Hemingway: "Yes, they have more money."

In our context: are cities in rich countries just a richer version of cities in poor countries?

Duranton (2024): in terms of levels of development, Latin American cities today look like 1920s European cities but operate with weak institutions in a 2024 world economy:

⇒ Emergence of large poor cities with no agricultural constraints (e.g., Kinshasa).

⇒ High and persistent levels of informality (labor, but also housing, transportation, and finance).

⇒ Informal housing is low quality and only supplies small quantities for its residents. Informal housing also generates relatively low population density per unit of residential land and leaves too little land for other uses, including transportation.

⇒ The division of public spaces into transportation and other uses is often inexistent or unenforced, leading to spaces dominated by either motor vehicles crowding out non-motorized users or by other users making transportation problematic.

So, this is not just a delay in development; poor cities appear to follow a different trajectory.

The transportation wedge: US vs. Bogota

In a typical US metropolitan area (US NHTS):

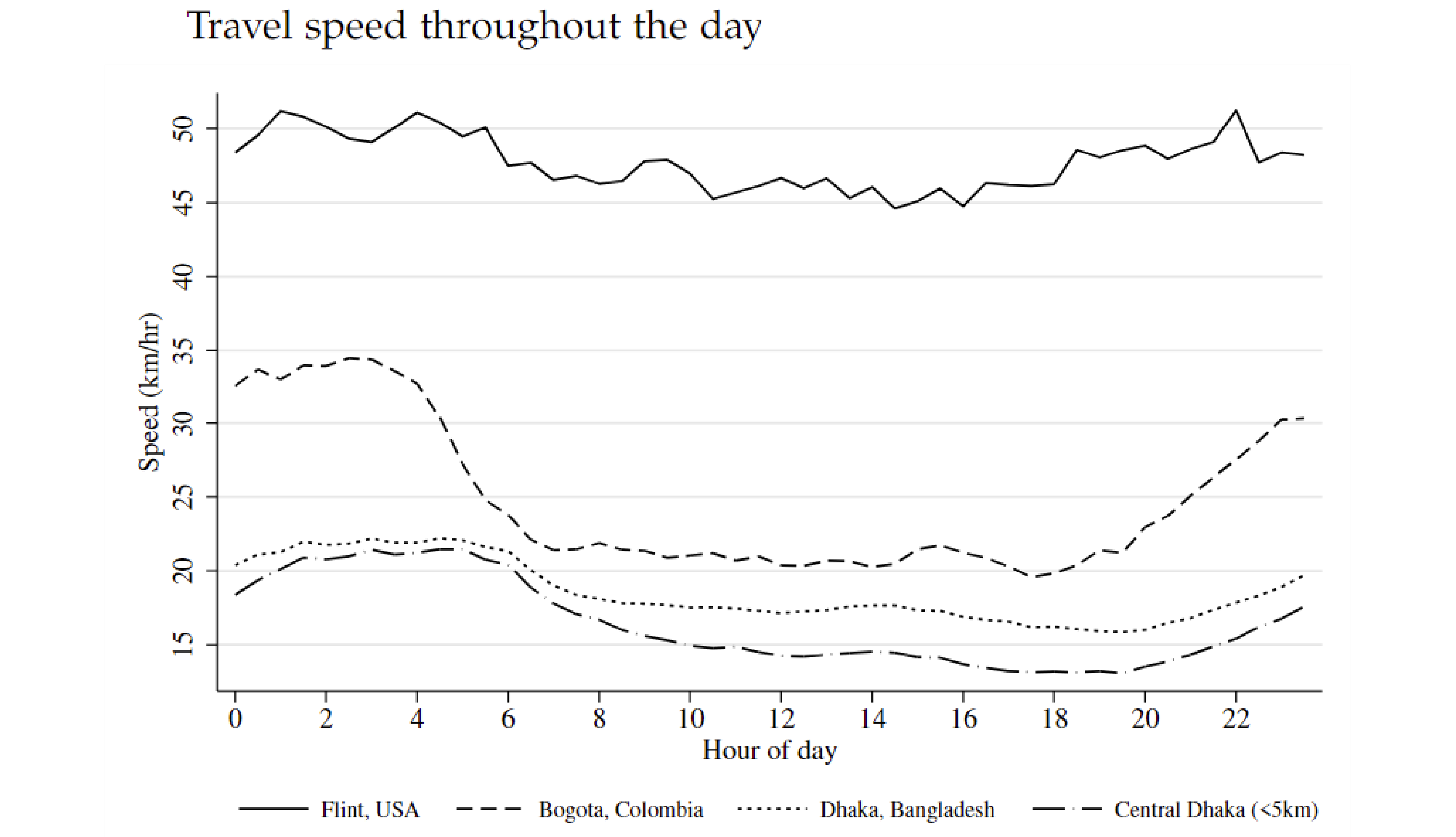
- A traveler takes **4.2** trips per day.
- Each trip is, on average, 12.8 kilometers long and takes **17.5** minutes.
- Reported travel speed is 38.5 km/h and overwhelmingly by car.

In Bogota, Colombia (Bogota transportation survey):

- A traveler takes **2.7** trips per day.
- Each trip is, on average, 10.9 kilometers long and takes **38.2** minutes.
- Reported travel speed is 17.1 km/h, and car and taxi are only about a quarter of all trips.

So, in poorer places, people take fewer trips that are shorter and slower. This is true much beyond this contrast between Bogota and US metropolitan areas.

How fast do people move around?



Akbar et al. (2023): Flint (MI, USA) is the fastest city in the world, Bogota is the most congested, and Dhaka is the slowest (among 1,215). Underlying data: simulated trips on Google Maps.

How fast do people move around?

Using the Google Maps travel-time data collected by Akbar et al. (2023), we can perform a series of city comparisons throughout the world using World Bank income groups of their country of location

Income Level City Details

WB income class	City Count	GDP per capita, mean	GDP per capita, SD	City population, mean	City population, SD
High income	436	47,373	13,755	1,168,586	2,456,771
Upper middle income	339	16,667	5,140	1,409,210	3,029,231
Lower middle income	377	6,326	1,759	1,677,752	3,435,010
Low income	63	2,111	1,365	1,202,884	1,269,827

GDP per capita is measured in 2017 US dollars.

WB income class	Speed mean	Speed SD
High income	0.111	0.161
Upper middle income	-0.009	0.130
Lower middle income	-0.127	0.141
Low income	-0.131	0.107

Speed is measured in trip distance (km)/trip duration (minutes). Speed is centered around world mean.

⇒ Travel speed is much higher in cities from richer countries (elasticity of speed wrt GDP =0.13)

How do people move around?

- In the US (NHTS 2017), private vehicles represent 83% of trips and 97% of the mileage. Transit is less than 2% of all trips.
- In the UK (NTS 2020), private vehicles represent 60% of trips and 80% of the mileage. Transit is 8% of all trips.
- In urban India (MoSPI) cars and rickshaw represent 10% of trips and 10% of the mileage. Transit is 13% of all trips but 30% of the mileage. Walking is 25% of trips but only 4% of the mileage. Motorbikes are 40% of trips and close to 50% of the mileage.

- Cars dominate travel in the rich world, and not only in the US.
- Motorbikes represent a high share of all motorized vehicles throughout the developing world – 50 to 80% of the mileage.
- People also walk a lot in developing cities.

The quantity of transportation infrastructure (OSM data)

Ratio of road length to city area

WB income class	Principal road to area, mean	Principal road to area, SD	Total road to area, mean	Total road to area, SD
High income	1.69	0.60	11.39	2.14
Upper middle income	1.60	0.73	12.60	2.90
Lower middle income	0.72	0.41	8.61	3.91
Low income	0.65	0.57	9.77	4.48

Principal road to area is the ratio of principal (motorway, trunk, primary, secondary) roadway length measured in kilometers and city area measured in square kilometers. Total road to area uses the same units but for all roads in the city.

Ratio of road length to city population

WB income class	Principal road to population, mean	Principal road to population, SD	Total road to population, mean	Total road to population, SD
High income	0.85	0.48	5.76	2.89
USA	1.37	0.36	9.32	2.29
Other	0.64	0.35	4.39	1.68
Upper middle income	0.45	0.30		1
Lower middle income	0.16	0.11	1.95	1.13
Low income	0.15	0.18	2.22	1.43

Principal road to area is the ratio of principal (motorway, trunk, primary, secondary) roadway length measured in meters to city population. Total road to population uses the same units but for all roads in the city.

The quantity of transportation infrastructure

- Cities in poor countries have much less principal roadway (per area, per person, and as a share of total roadway).
- They also have less total roadway per person due to their higher density.
- These differences between cities in rich and poor countries do necessarily reflect some inefficiencies as (principal) roads are (publicly) expensive to build, and there may not be a strong demand for fast travel that is (privately) expensive.
- Guerra et al. (2024) argue that US metropolitan areas are almost surely over-paved.
- The issue is more about allowing for urban expansion and infrastructure development as cities in poor countries develop.

The quality of transportation infrastructure

Acknowledging clear data imitations in the measurement of road quality by OpenStreetMap:

City level share of high quality roads

WB income class	Share roads w/ high quality paving, mean	Share roads w/ high quality paving, SD
High income	0.928	0.153
Upper middle income	0.837	0.198
Lower middle income	0.809	0.289
Low income	0.422	0.309

Road quality measured as the nonmissing share of total roads with quality pavement. We define quality roads as having a concrete, asphalt, or otherwise smooth paved surface.

Because better roads are more likely to be reported in poor countries, the table most likely understates the true differences in road quality between cities in rich and poor countries.

Networks

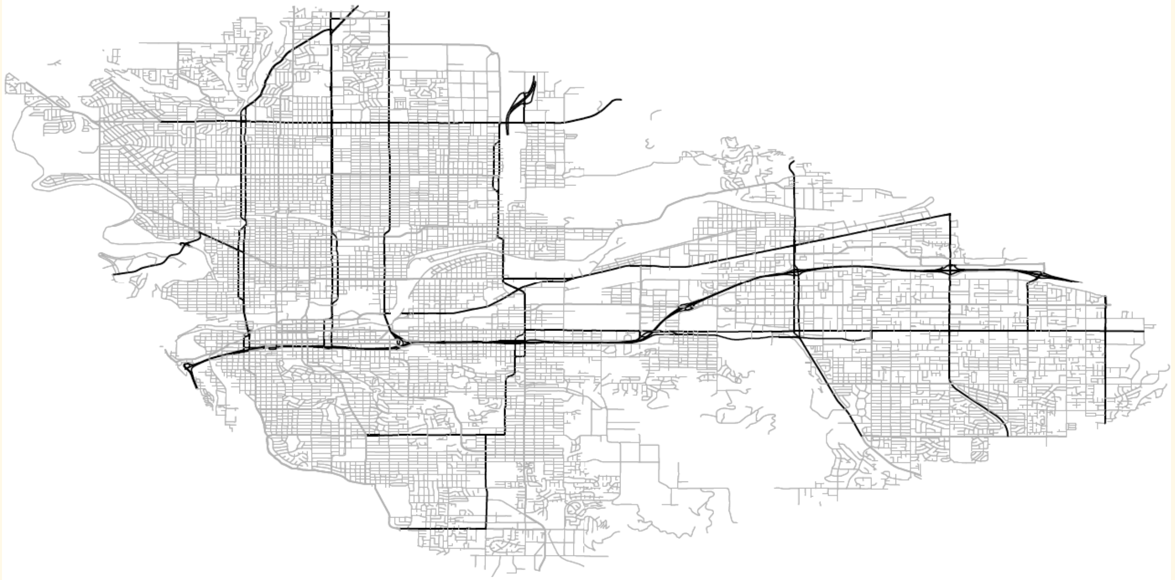
Road network measures

WB income class	Grid-like mean	Grid-like SD	Orientation entropy mean	Network circuity mean
High income	0.194	0.164	3.412	1.088
USA	0.364	0.194	3.184	1.100
Other	0.129	0.088	3.498	1.083
Upper middle income	0.154	0.118	3.436	1.055
Lower middle income	0.145	0.086	3.417	1.063
Low income	0.148	0.073	3.441	1.053

Grid-like is a measure of the share of road edges at intersections conforming to the main grid orientation. Orientation entropy is the Shannon entropy of the degree orientation of road edges at intersections. Network circuity is the ratio of road to effective distance.

- Road networks are more "grid-like" in the US but do not appear to systematically differ elsewhere.
- "Active" traffic management with more one-way streets and more (fast) principal roads may lead to more circuitous trips in richer cities.

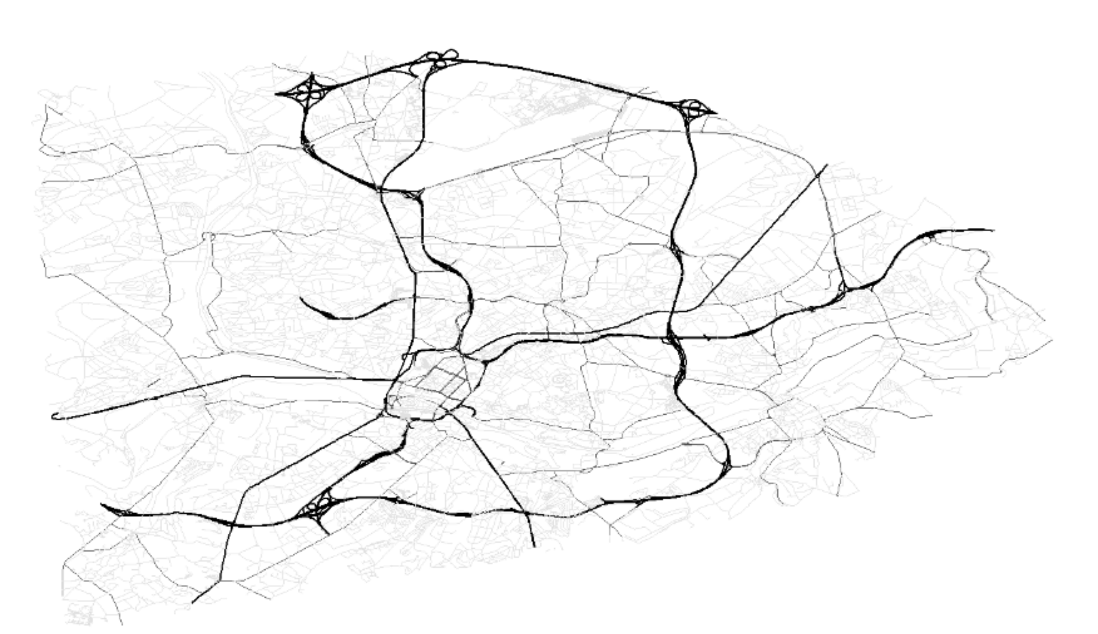
Networks



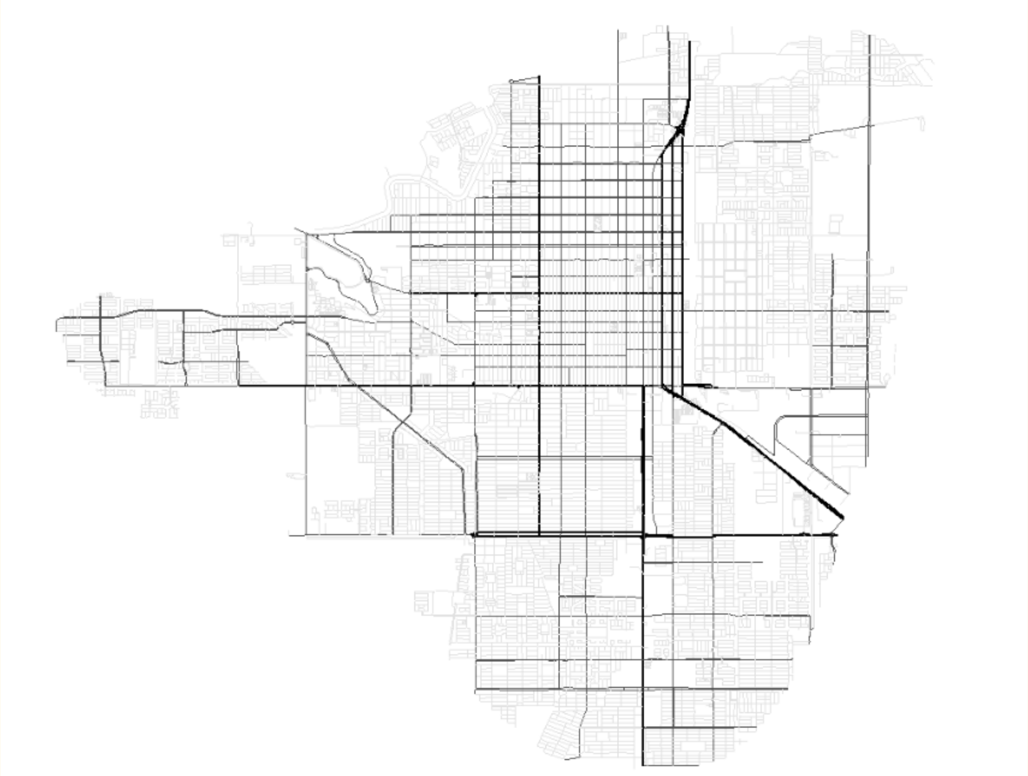
Spokane
(WA, USA)

Kigali
(Rwanda)

Networks



Charleroi (Belgium)



Ciudad Obregon (Mexico)

Bringing it together

The determinants of urban travel speed

		Speed index	
		(1) Base model	(2) Full regression
	log country GDP (pc)	0.13^a	0.042^a
		(0.023)	(0.011)
City size	log population		-0.17^a
			(0.018)
	log area		0.085^a
			(0.022)
Infrastructure	asinh major road length		0.080^a
			(0.012)
	Network griddiness (0-1)		0.18^a
			(0.057)
Topography	Elevation variance		-0.0024^b
			(0.00092)
	asinh water body length		-0.071^a
			(0.019)
	Observations	1,119	1,119
	R^2	0.46	0.70

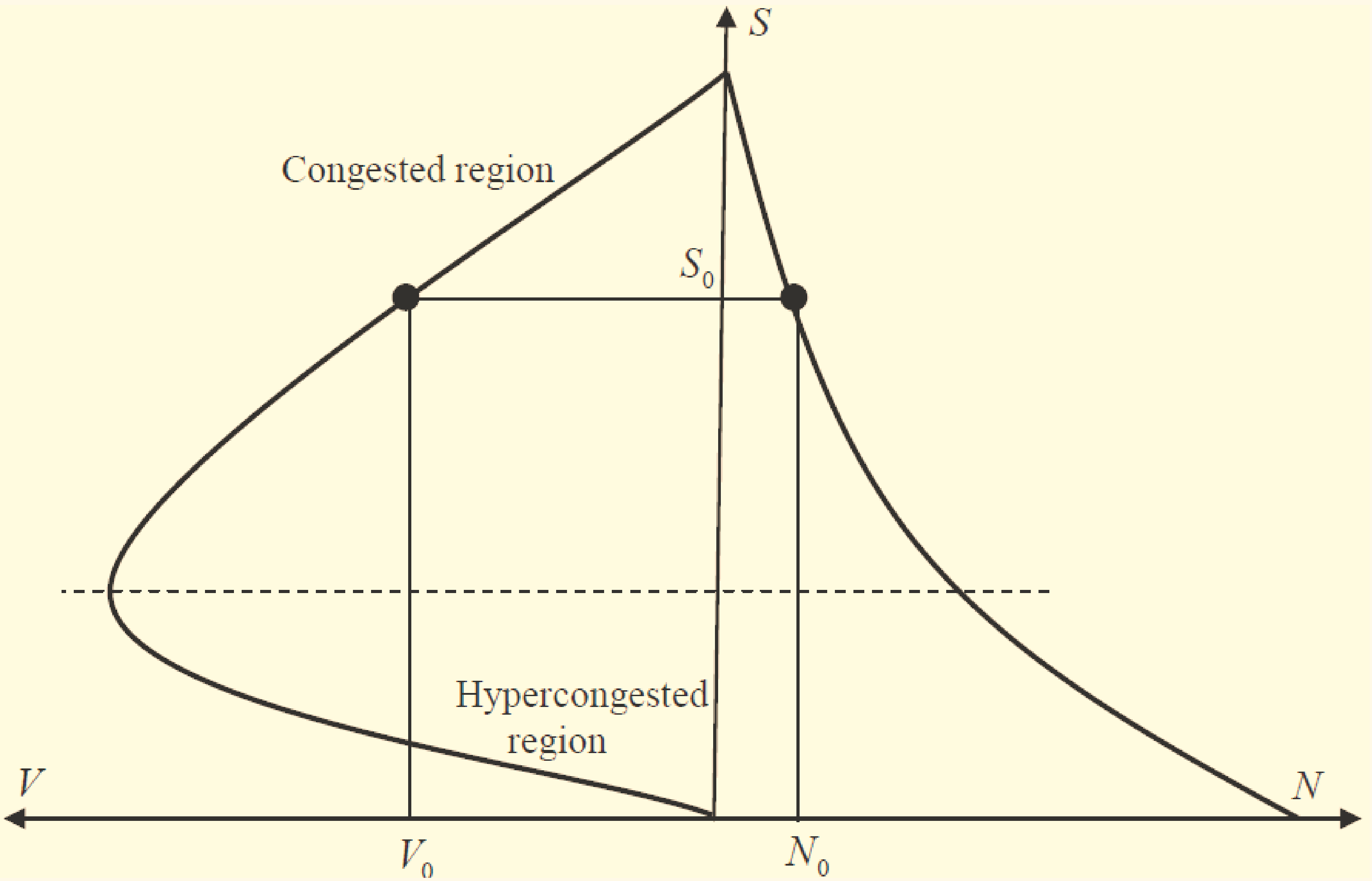
Preliminaries on congestion

As I will argue, congestion is a problem for urban transportation in developing cities, but it is not *the* problem

Before that, we need to understand congestion and its externalities better:

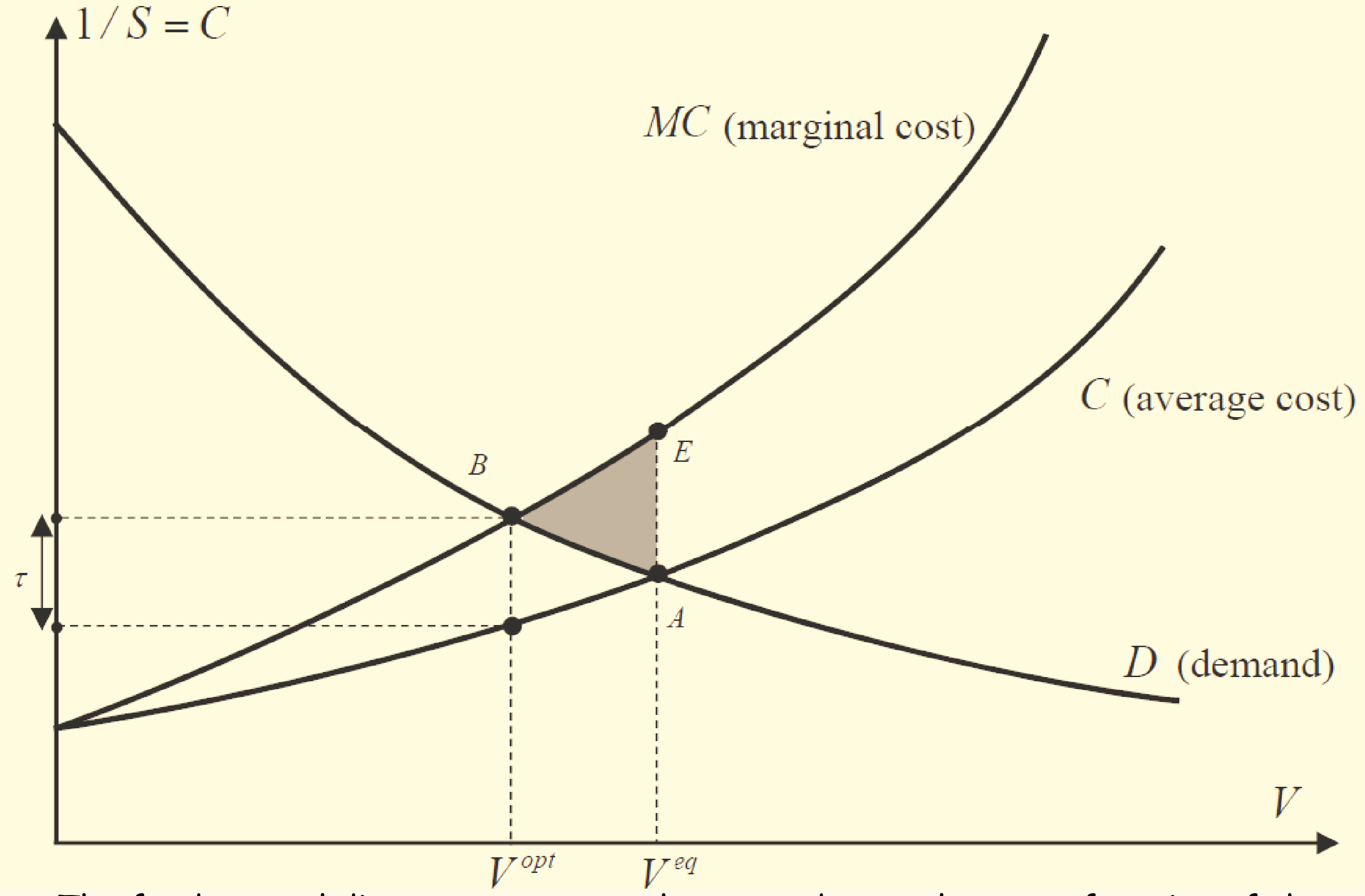
- The "fundamental diagram" of traffic congestion
- Equilibrium and welfare with congestion
- Welfare analysis: increasing supply vs. curing congestion

Preliminaries on congestion: The "fundamental diagram"



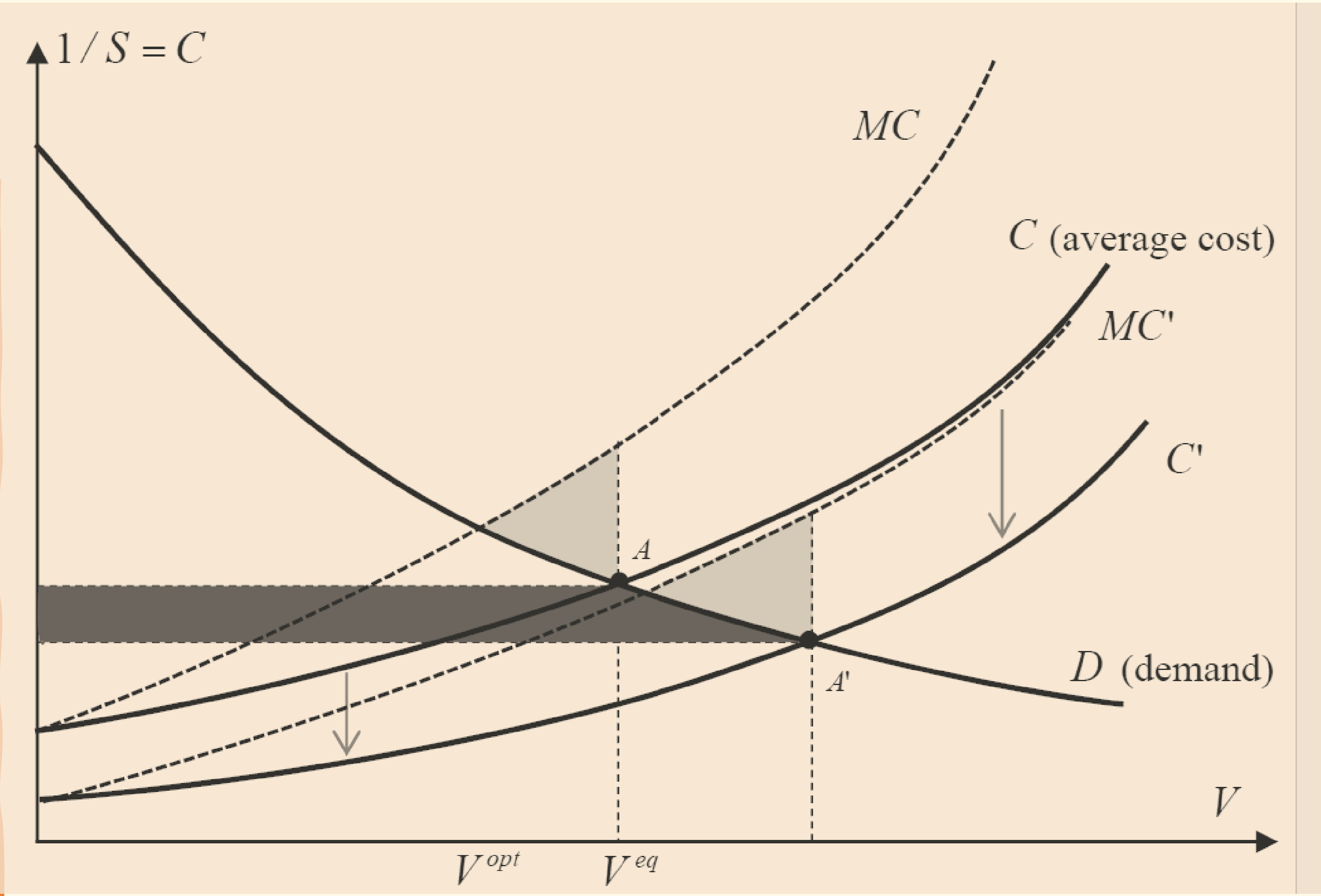
- N: number of vehicles.
- S: travel speed with $S = \bar{S}\ N^{-\sigma(N)}$, where $\bar{S}$ is free-flow speed, and $\sigma(N)$ is the congestion elasticity.
- $V = S \times N$: "flow" (total mileage per unit of time, also throughput when road length is 1).
- When $\sigma(N)$ >1, the flow declines with the number of vehicles. We call this situation hypercongestion.

Preliminaries on congestion: equilibrium and welfare loss



- The fundamental diagram represents how speed, S, evolves as a function of the mileage/flow, V, through the road technology. Since speed is the inverse of the time cost of travel (or its price), $C = \frac{1}{S}$, the speed-flow relationship is a supply curve in the (V, C) space (ignoring hypercongestion)
- Then in an urban setting, "speed" is, to a first approximation, the average speed faced by all (most) vehicles. From the average, it is easy to recover the marginal cost, ie the social marginal cost, and to compute the congestion externality.
- Welfare losses are much higher with hypercongestion (imagine a backward bending marginal cost curve).

Preliminaries on congestion: expanding supply vs. reducing congestion



- An optimal congestion charge increases welfare by the light grey triangle.
- Increasing supply increases welfare by the dark grey trapezoid. The latter looks much bigger because the reduction in the time cost (price) of travel affects all travel.
- Three key considerations: (i) increasing the supply of roads is costly, (ii) charging for congestion is unpopular and technically difficult to implement, (iii) the reduction in the time cost of travel may be small.

“Macro” evidence on urban road congestion

Akbar et al. (2023) define:

$$\text{Congestion} = \frac{\text{Travel time}}{\text{Uncongested travel time}}$$

Hence:

log congestion = log travel time – log uncongested travel time

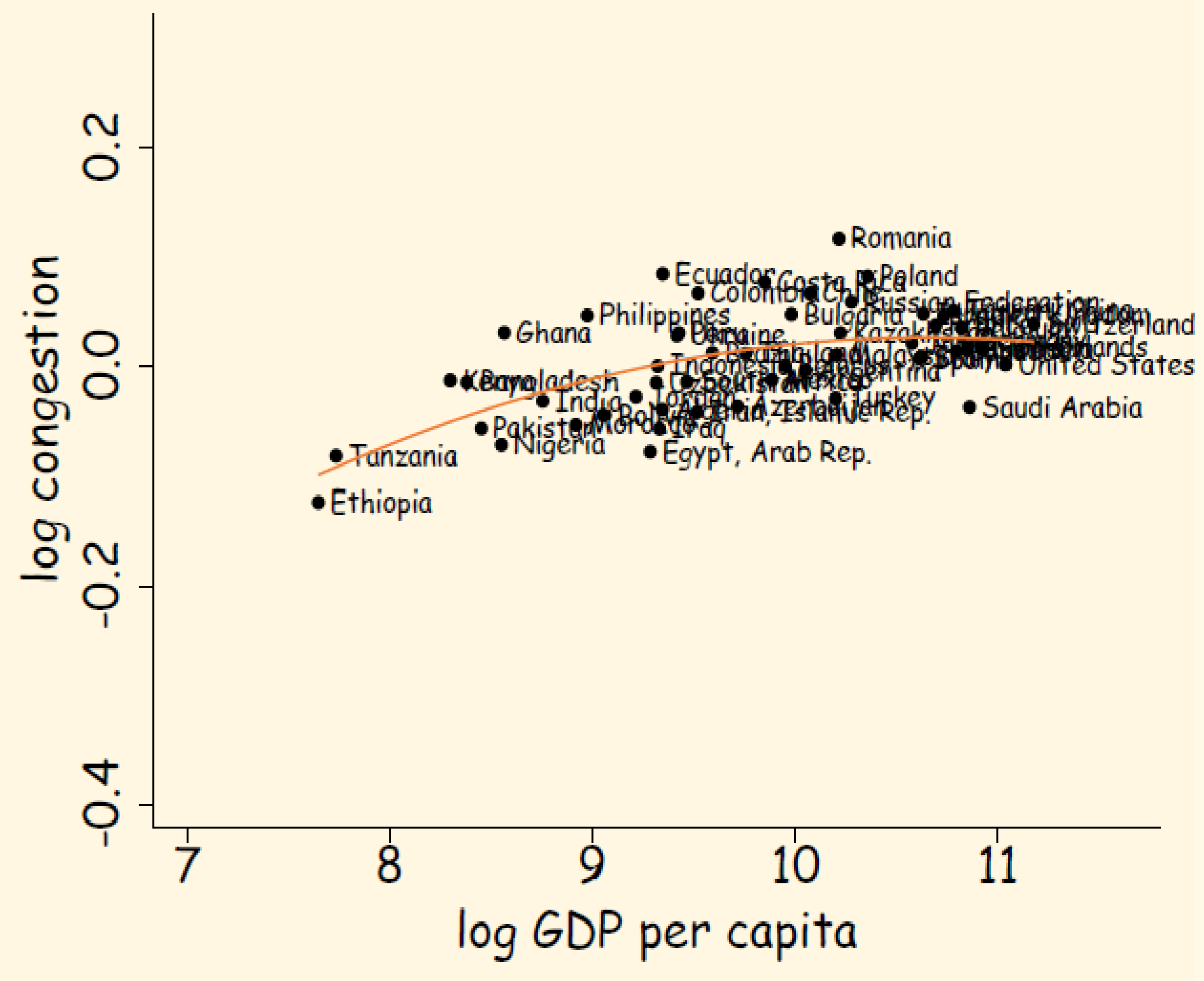
= log uncongested speed – log speed

The congestion index can be normalized to zero for the world average (on average, travel at peak hours is about 25% slower than free flow).

“Macro” evidence on urban road congestion

WB income class	Congestion mean	Congestion SD
High income	0.026	0.046
Upper middle income	0.001	0.054
Lower middle income	-0.031	0.064
Low income	-0.078	0.092

Congestion factor is measured as uncongested speed/speed. Congestion factor is centered around world mean.



“Macro” evidence on urban road congestion

Since, log congestion = log free-flow speed – log speed, we can perform a variance decomposition across cities:

	Cities	Uncongested Speed	Congestion factor	Covariance
All	1,119	0.934	0.117	0.025
Peak hour radial trips	1,119	0.910	0.174	0.042
Trips within 10 km of center	1,119	0.903	0.104	0.004
High income	434	0.653	0.083	-0.132
Upper middle income	330	0.638	0.175	-0.093
Lower middle income	345	0.631	0.205	-0.082
Low income	10	0.397	0.746	0.071

Speed and uncongested speed is measured in trip duration (km)/trip time (minutes). Congestion factor is measured as uncongested speed/speed.(in log.)

Differences in travel speed between rich and poor cities are mostly accounted for by free-flow speed (uncongested) speed.

⇒ Congestion is a problem for urban transportation in cities, including cities in developing countries, but it is not *the* problem

“Macro” evidence on urban road congestion

Determinants of urban travel speed

		Speed index		Uncongested speed		Congestion factor	
		(1) Base model	(2) Full regression	(3) Base model	(4) Full regression	(5) Base model	(6) Full regression
	log country GDP (pc)	0.13^a	0.042^a	0.15^a	0.074^a	0.019^a	0.032^a
		(0.023)	(0.011)	(0.018)	(0.0090)	(0.0059)	(0.0067)
City size	log population		-0.17^a		-0.13^a		0.040^a
			(0.018)		(0.015)		(0.0069)
	log area		0.085^a		0.064^a		-0.021^b
			(0.022)		(0.019)		(0.0087)
Infrastructure	asinh major road length		0.080^a		0.076^a		-0.0040
			(0.012)		(0.013)		(0.0080)
	Network griddiness (0-1)		0.18^a		0.12^a		-0.058^a
			(0.057)		(0.046)		(0.015)
Topography	Elevation variance		-0.0024^b		-0.0011		0.0013^a
			(0.00092)		(0.0010)		(0.00040)
	asinh water body length		-0.071^a		-0.054^a		0.017^b
			(0.019)		(0.019)		(0.0083)
	Observations	1,119	1,119	1,119	1,119	1,119	1,119
	R^2	0.46	0.70	0.63	0.77	0.11	0.43

Notes: The unit of analysis is the city. Columns 1–6 each report parameter estimates of separate OLS regressions on the sample of 1,119 cities with high-quality speed data and country GDP data. Standard errors clustered at the country level in parentheses. a, b, c: significant at 1%, 5%, 10%.

Micro evidence on urban road congestion: Three city studies

Bangalore (Kreindler E 2024)

- Pioneering RCT on congestion pricing.
- Uses an app and gives financial incentives to drivers to assess the value of scheduling and the value of travel time.

⇒ Moderate substitutability across departure times.

⇒ High values of travel time (dislike of driving, drivers overestimate time losses of alternative routes, risk aversion of new routes).

- Also combines GPS data with Google Maps to estimate the congestion elasticity.

⇒ Modest congestion externality and no evidence of convexity.

⇒ Overall modest gains from congestion pricing.

Micro evidence on urban road congestion: Three city studies

Bogota (Akbar and Duranton, partial draft 2017)

- Combine travel surveys with Google Maps to estimate the congestion elasticity for the city using a variety of identification strategies.

⇒ Congestion elasticity close to zero at night, moderate (c.0.2) during "transition hours", and modest at peak hours (c. 0.05).

- To explain these results, they propose a model where drivers arbitrage between fast, congestible (principal) roads and slow, non-congestible (local) roads.
- To verify this explanation, they use data about traffic measurements at major intersections.

⇒ Low throughput elasticity for the number of vehicles at major intersections (with fewer cars at peak hours!)

⇒ Higher throughput elasticity for the number of travelers (because of bigger and fuller buses).

- Policy implication: road pricing should take place at the road level, not the area level (cordon).

Micro evidence on urban road congestion: Three city studies

Manila (Duranton, Hall, and Yi, in progress)

- Provide a new methodology to distinguish between bottleneck and frictional congestion.
- Use simulated origin-destination Google Maps data. To detect bottlenecks:
  1. Estimate travel speed for each hour-1km pixel-direction of travel.
  2. Detect pixels of maximum congestion at early peak hours.
  3. Assess whether these high levels of congestion radiate to neighboring pixels at the next hour.

⇒ Congestion in Manila is broadly monocentric despite the city being fairly multicentric.

⇒ The procedure detects *many* "bottlenecks", making congestion in the city better described as "frictional".

⇒ Modest apparent congestion spillovers from bottlenecks.

- Policy implication: Fixing a few bottlenecks is unlikely to improve traffic conditions by much.

Policies: price instruments

Despite (Pigouvian) road pricing being the natural instrument to handle the congestion externality, price instruments are seldom used in rich cities and are essentially unused in poor cities. Why?

- A large literature emphasizes an adverse political economy.
- Road pricing is technically hard to implement and extremely costly, even for a simple cordon.
- The literature explored here also emphasizes the small size of the benefits and some intricacies in how these policies should be implemented.

What should we do? My two cents:

- More research: Explore how the costs of other externalities (pollution, noise, accidents) vary with the level and location of road traffic.
- Bundle congestion charging with other policies like transit (Eg, London, Almagro et al. WP 2024)
- Use substitutes for road pricing, like parking pricing.
- “Do no harm”: avoid damaging and wasteful policies.

Policies: quantity instruments

Instead of price instruments, many cities in the developing world have opted for quantity restrictions based on plate numbers, occupancy, type of vehicle, age of vehicle, etc.

- Examples include Jakarta, Dehli, Quito, Mexico City, several Colombian cities, etc. See Guerra et al. (TR 2022) for a review.
- The literature is heavily divided on the effectiveness of these policies. Hanna, Kreindler and Olken (S 2017) or Carrillo, Malik, and Yoo (CJE 2016) provide evidence of sizeable traffic reductions. Davis (JPE 2008) and Guerra and Millard-Ball (TRD 2017) are a lot more skeptical. Montero, Sepulveda, and Basso (wp 2024) are showing that road pricing dominates plate restrictions.
- My conclusion on this literature is that plate restrictions lead errands at preferred times and locations to be replaced with errands at less preferred times and locations. These restrictions may also lead households to buy more cars.

Promising avenues for future research

What we need to know about (IMHO):

- Other road traffic externalities, especially pollution (spatial reach, is it worth concentrating traffic on certain roads, etc.?)
- More traffic-focused evaluations of specific road infrastructure projects/cities (or assessment of many projects at once).
- Bring in more novel sources of data (e.g., cellphones) to traffic studies.
- Studies about specific new modes of urban transportation, e.g., motorbikes, bike lanes, micromobility.
- Better measures of roadway quality beyond the US (smoothness, but not only – traffic lights, etc.)
- Role of driving norms and behavior in congestion.

There is a rich(er) agenda to study the broader effects of transportation improvements, but the findings will not stand if they are inconsistent with changes in traffic patterns and travel times.