Remote work, COVID, and urban transportation in the United States

Prottoy Akbar, Victor Couture, Gilles Duranton, Braydon Neiszner, Adam Storeygard

May 2024

Introduction: What we do

1. We want to learn about the implications of COVID cases and work from home (WFH) for urban transportation

2. Use plausibly exogenous variations of travel demand from COVID and WFH to estimate the elasticity of urban congestion

Introduction: Why we should care

We provide some answers to three important questions

1. What is happening to cities post-COVID?
2. How big is the congestion elasticity?
3. What is the effect of WFH on urban traffic?

Introduction: Key contributions

1. Bring together unique travel data in large US cities for a period of close to 5 years
2. Provide novel description of what happened to urban traffic during this period
3. Develop a novel approach to estimate the road congestion elasticity...
4. ... and the effects WFH on urban mobility

Data

Core data: Simulated travel data over 2019 - to date

- For 2019, we use simulated trip data queried from Google Maps as in Akbar et al. (2023a,b): 45,194,664 instances of 2,174,984 trips in 139 US cities
- We resumed queries at the onset of COVID in March 2020 until now
- For 2020-2023, we use 123,236,041 instances of the same trips in the same 139 cities
- For each trip instance, we know real-time travel speed, uncongested speed, and travel distance

These data essentially inform us about the *price* of travel in cities over time

Data: city sample and validation

- City sample: We considered all US cities with population above 300,000 according to UN WUP and delineated these cities using two layers of the GHSL and separated "joined" cities (eg SF and Oakland) $\Rightarrow$ 139 cities
- Trip sample: We designed four types of trips meant for some to resemble existing trips (radial, circumferential, 'amenity', and gravity)

- Validation: We also mapped trips from the 2017 NHTS and collected instances of the same trips in the same manner (not used today)
- Validation: Akbar et al. (2023,ab) provide extensive internal and external validations of the data for 2019

Data issues

Google Maps and AWS periodically make minor changes to their systems leading to data collection issues affecting some or all cities

- Attrition. We have two periods with prolonged dropoffs in the number of queries: April-June 2022 and February-May 2021
- Failure of randomization affecting some cities during some periods of time

$\Rightarrow$ We eliminate 14,615,108 observations

Additional data

- Travel *quantities*:
  - Google COVID-19 Community Mobility Reports (GCCMR) from February 15th, 2020 to October 15th, 2022
  - Report number of weekly 'visits' (trips) for six categories (workplaces, retail & recreation, grocery & pharmacy, parks, train stations, and residential) by county *relative to pre-COVID baseline*
- Employment data by industry:
  - 2019 1-year ACS estimates
  - Industry employment aggregates from Statistics of US Businesses
- Predictors of work from home:
  - Dingel and Neiman (2020)'s classification by occupation interacted with employment data
  - Dingel and Neiman (2020)'s WFH prediction by geography
- *New York Times* data about Covid cases at the county level
- Future: changes in roadway from different layers of OSM, high frequency population data, etc

Estimate city speed and congestion indices over time

We estimate 'city speed indices' with two-way fixed effects for any instance at t of trip i

$$speed_{it} = \alpha_{c(i)k(t)p(t)} + \beta_i + \gamma_{h(t)} + \delta_{d(t)} + X_{it}\zeta + \epsilon_{it} \tag{1}$$

- Our main object of interest: $\alpha_{c(i)k(t)p(t)}$ is a city fixed effect for city c, time of day k (peak/off peak/night) and calendar period p (week by weekdays)
- β is a trip fixed effect; γ is an hour of day fixed effect; δ is a day of week fixed effect; and X_{it} are trip time-varying characteristics (weather - to be added)
- We use this regression to carefully construct predicted speed for each city-week-period of day: $speed_{ckp}$ (more formally $\widehat{speed}_{c(i)k(t)p(t)}$)
- Note: Alternative dependent variables are possible, including speed with no traffic, congestion factor (=speed with no traffic - speed), travel time, and trip distance
- Note 2: Alternative specifications are also possible with fixed effects α by distance bands/quantiles to the center, type of trips, etc.

City and year data characteristics

Table: Top and bottom three cities by number of observations

City	Observations
New York	4,403,372
Los Angeles	3,777,196
Chicago	2,972,123
Bakersfield (median)	869,553
Rockford	548,049
Shreveport	563,303
Laredo	548,049

Table: Number of observations by year

Year	Observations
2019	45,194,664
2020	21,437,336
2021	24,599,587
2022	21,375,798
2023	41,208,212
Total	153,815,597

Counts by and counts of city-weeks

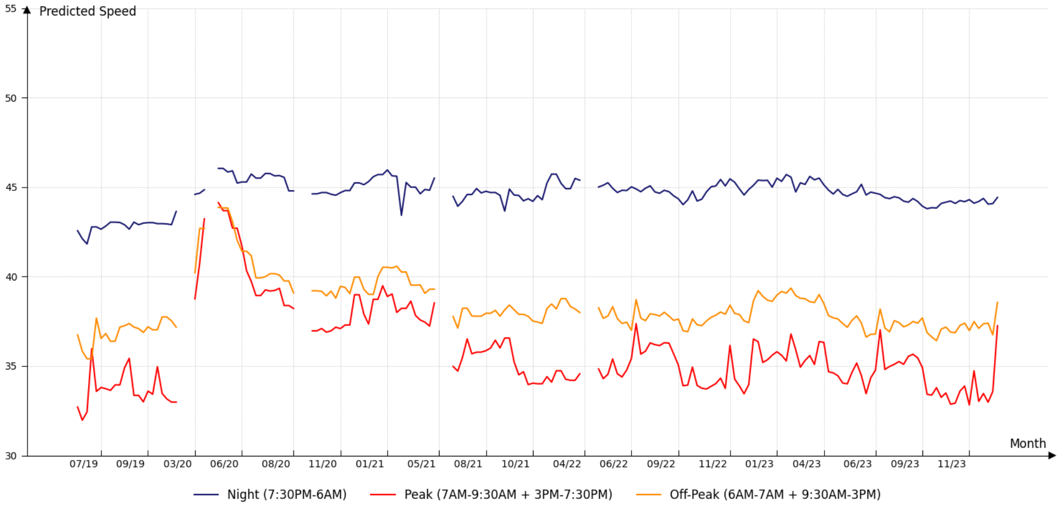
Total city-week pairs	25,331

Percentile	Observations by city-week
5th	2,353
25th	3,077
50th	4,179
75th	6,436
95th	14,579

Fastest and slowest cities (km/h)

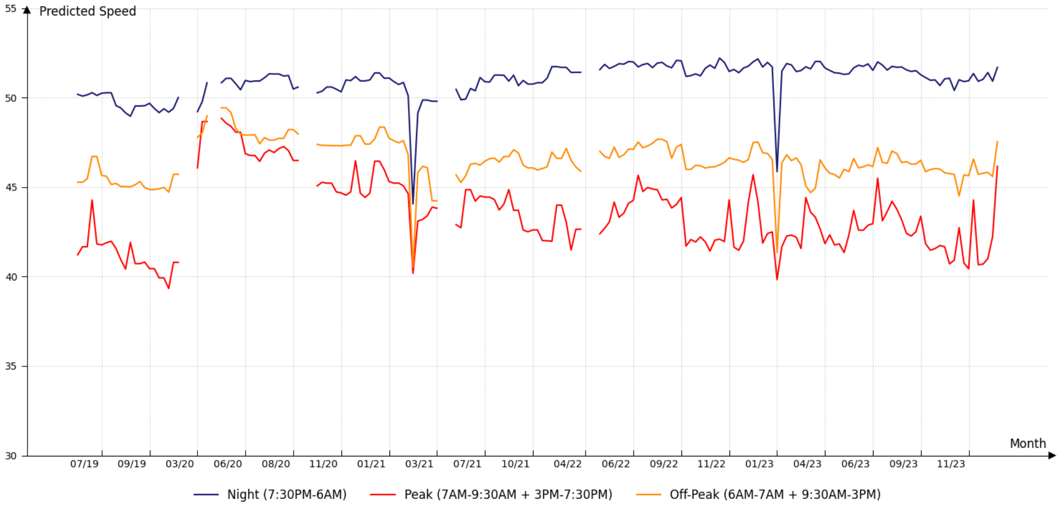
Percentile (2019)	City	2019	2020	2021	2022	2023
Fastest	Tulsa, OK	44.31	46.66	45.44	45.73	45.30
75th	Columbus, OH	39.88	44.35	43.36	43.00	42.83
50th	Cleveland, OH	38.26	41.73	40.77	40.93	40.99
25th	Birmingham, AL	35.69	40.41	39.84	40.12	40.03
Slowest	Boston, MA	29.36	37.09	33.81	31.99	31.53

Speed in NYC (km/h)

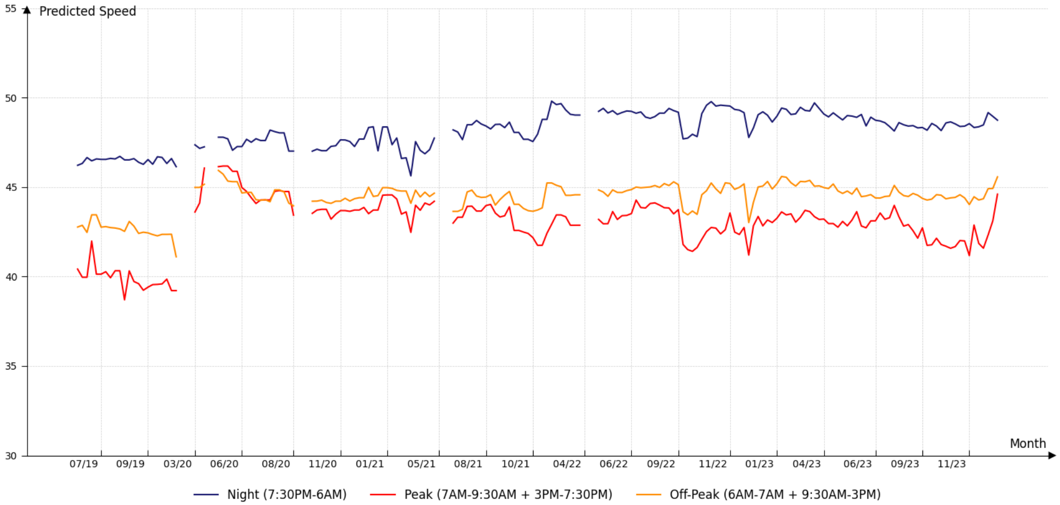


Speed in San Francisco

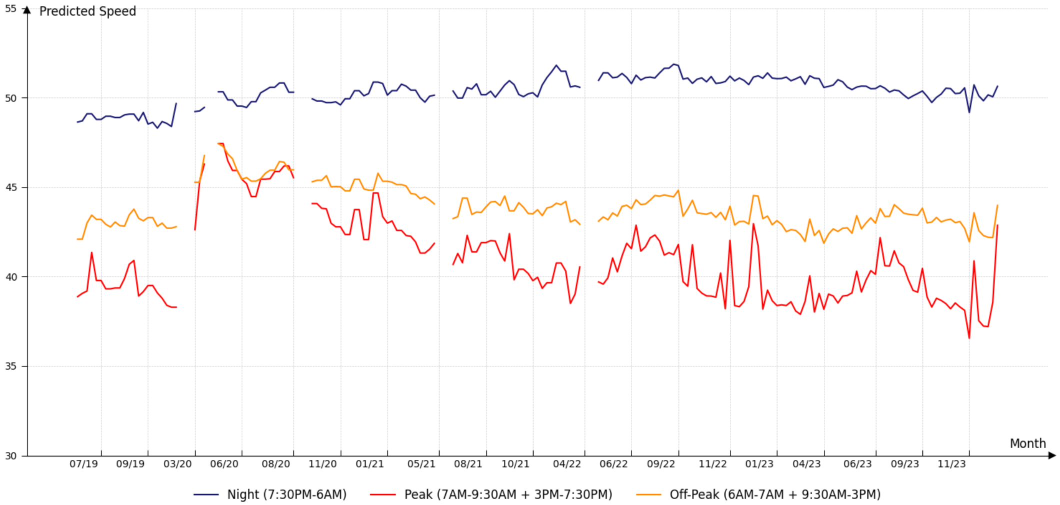
Speed in Dallas



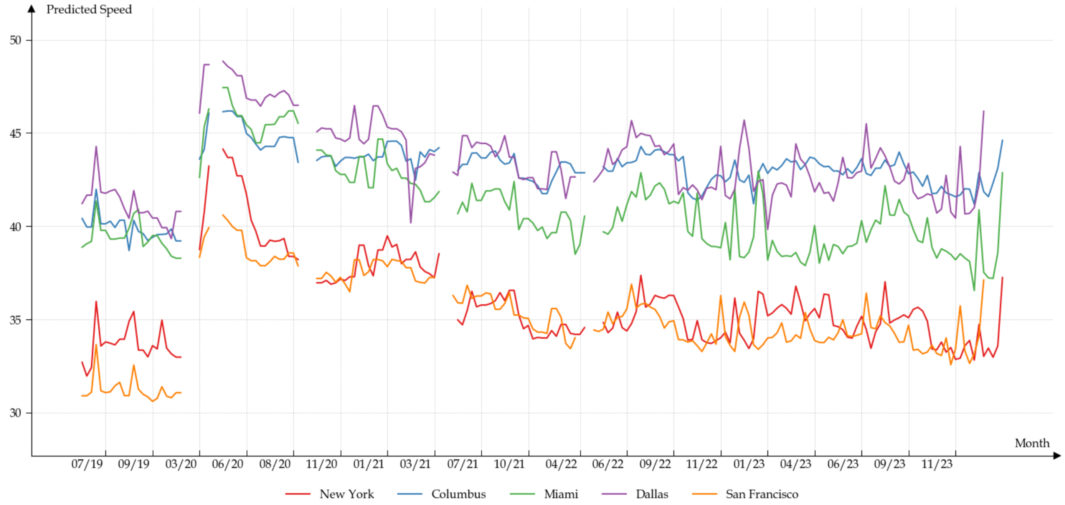
Speed in Columbus



Speed in Miami

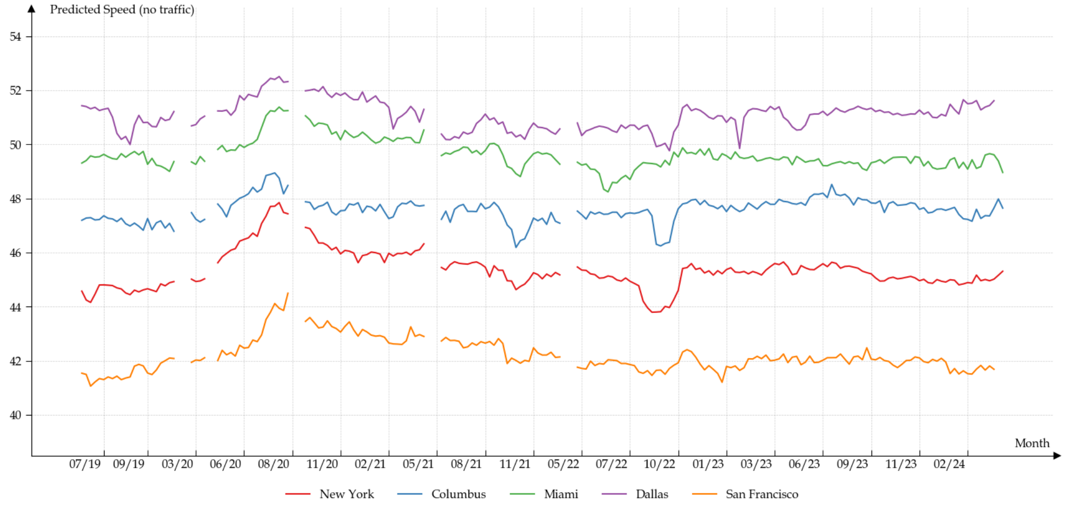


Peak hour speed in five cities

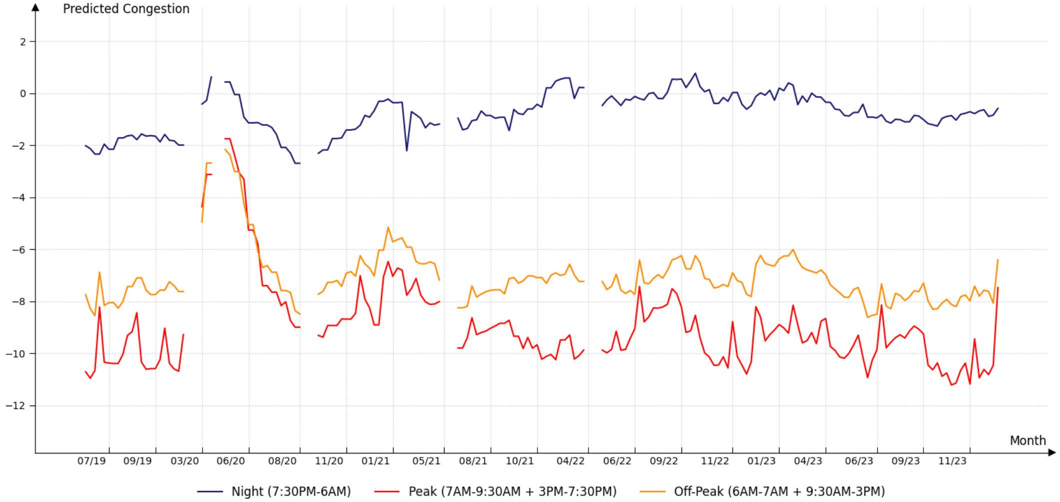


Speed in absence of traffic in New York

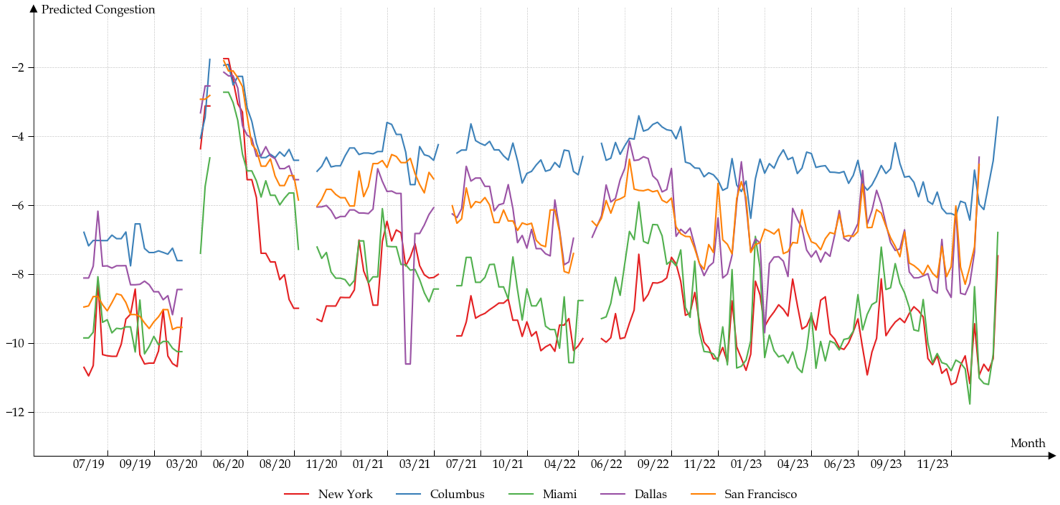
Night speed in absence of traffic in five cities



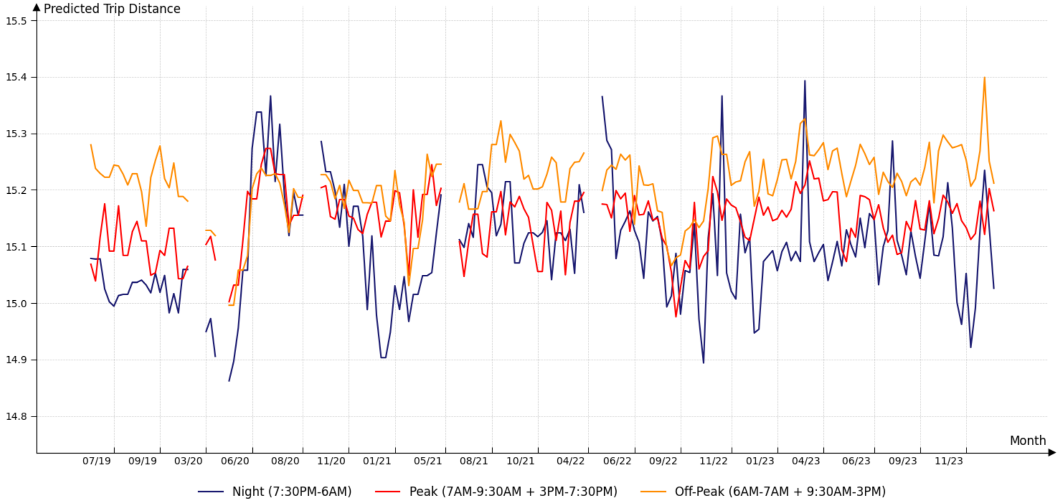
Congestion in NYC



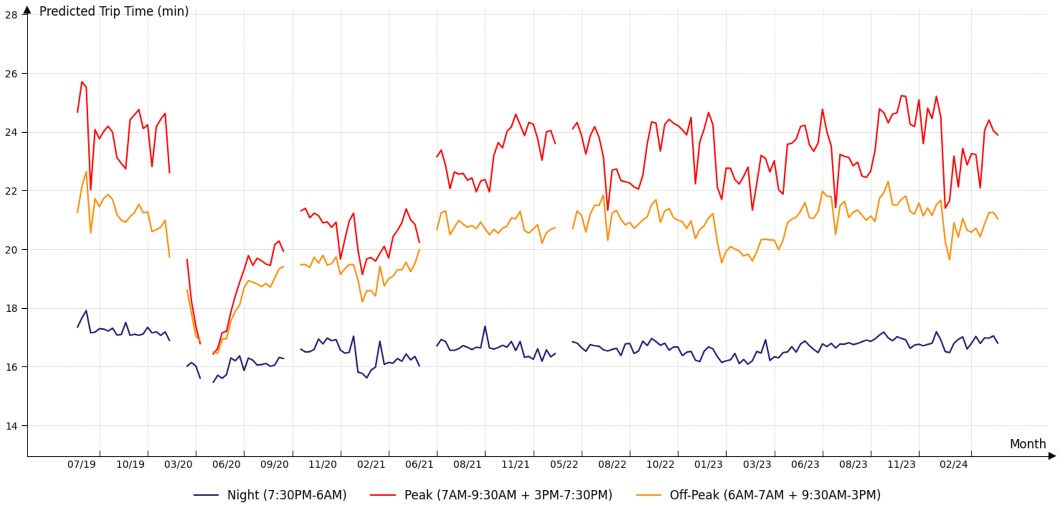
Peak hour congestion in five cities



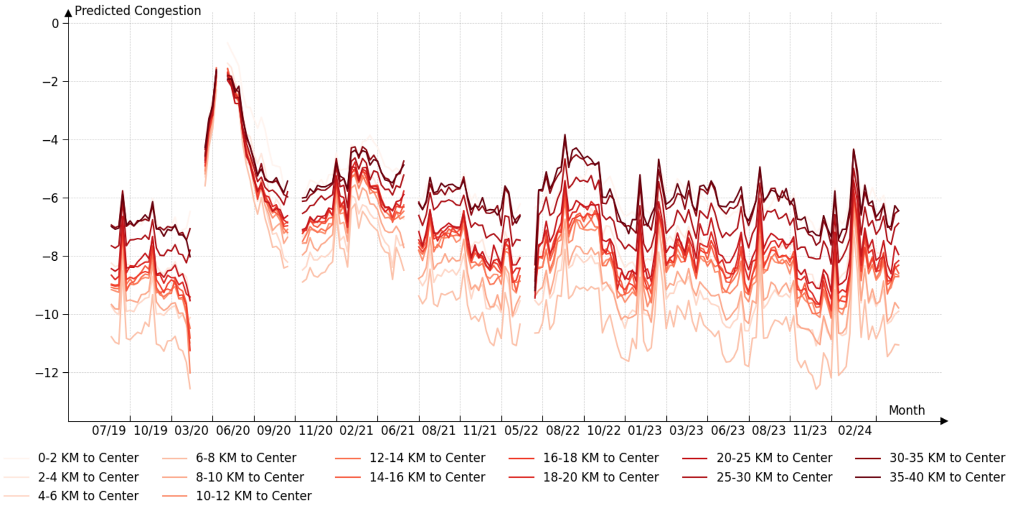
Trip distance (kilometers)



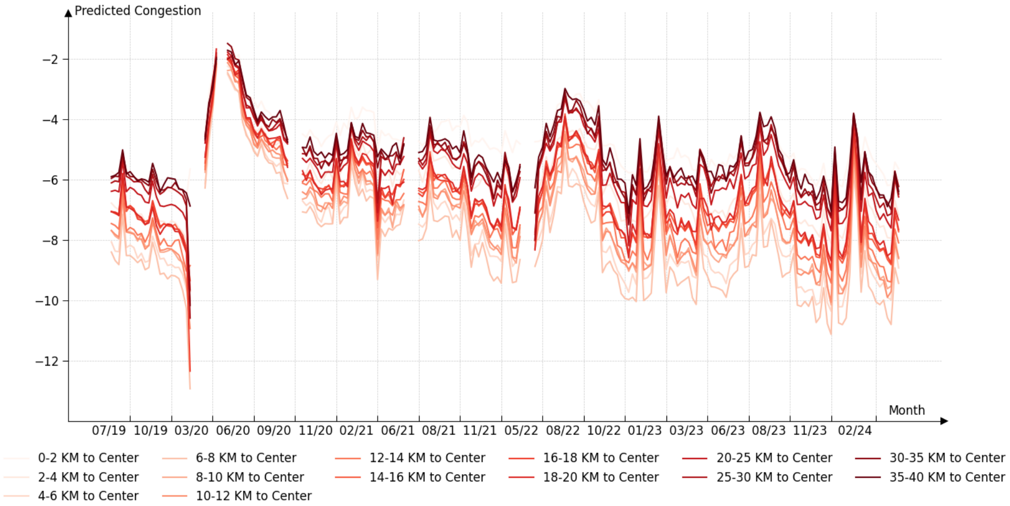
Travel time (minutes):



Peak hour congestion by distance to the center: Major northern cities



Peak hour congestion by distance to the center: Major southern cities



Taking stock

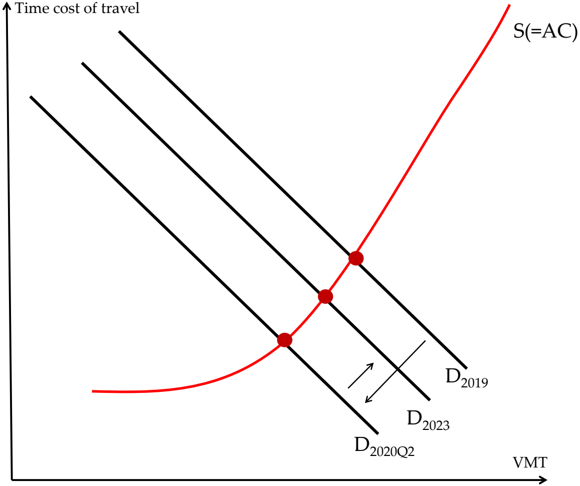
So far:

1. Collected a large volume of mobility data in major US cities
2. Estimated speed indices for (city, week, period of day) triplets
3. Interlude: some interesting graphs about what happened to traffic in the last five years

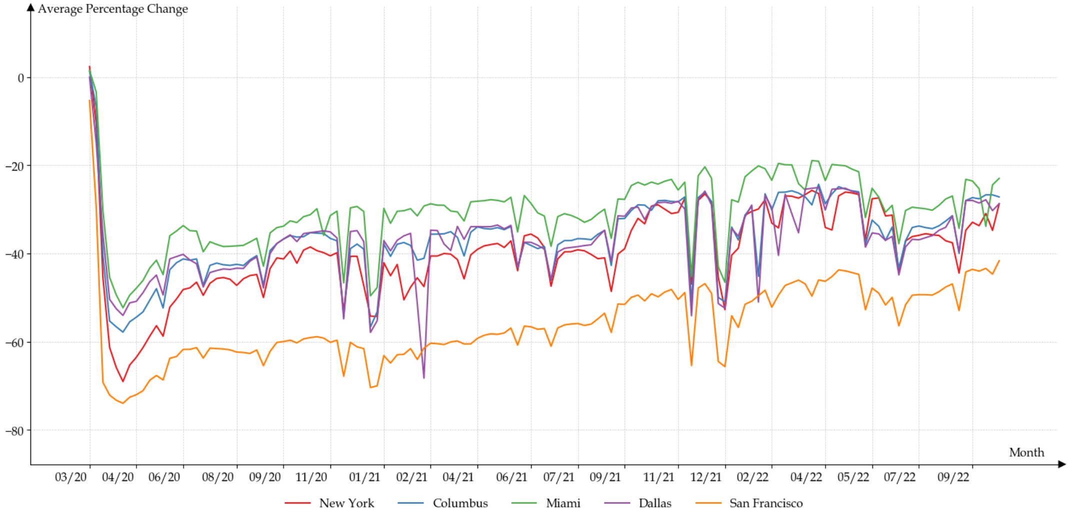
Now:

4. Basic conceptualization of the city traffic equilibrium
5. Some further suggestive data
6. More conceptualization of the city traffic equilibrium
7. Estimate the congestion 'elasticity'
8. Estimate the causal effect of WFH on mobility and congestion

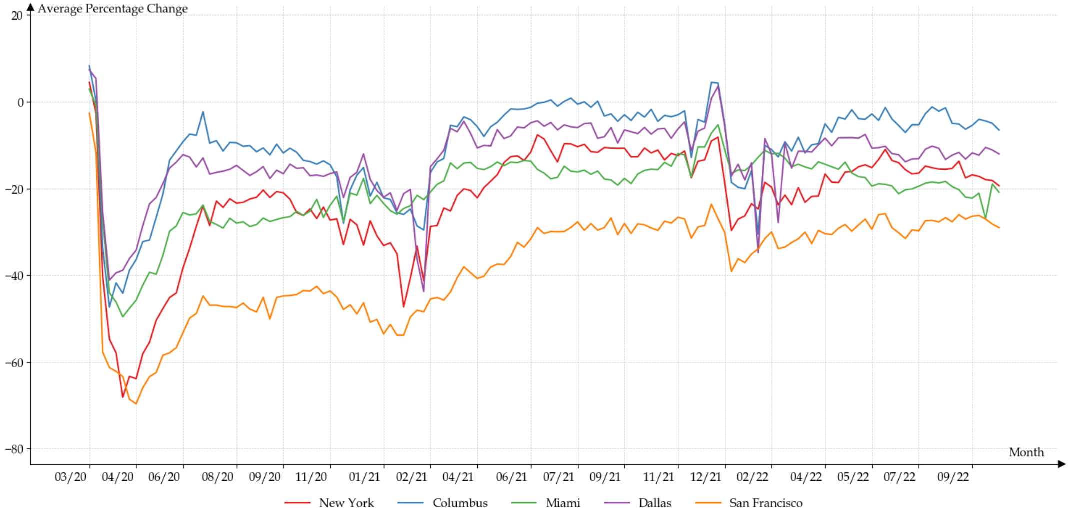
The city transportation equilibrium



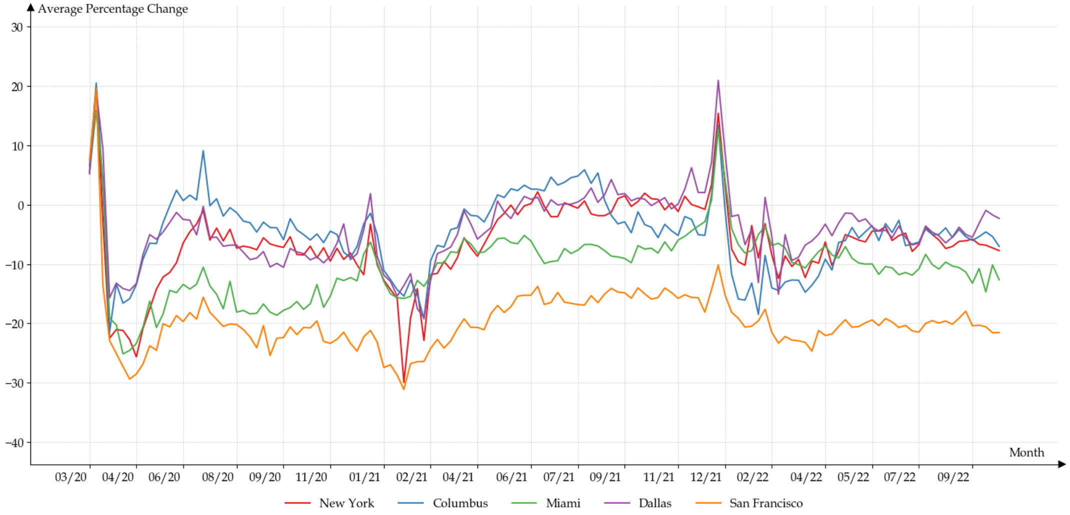
Travel quantities (GCCMR): Commuting trips for five cities



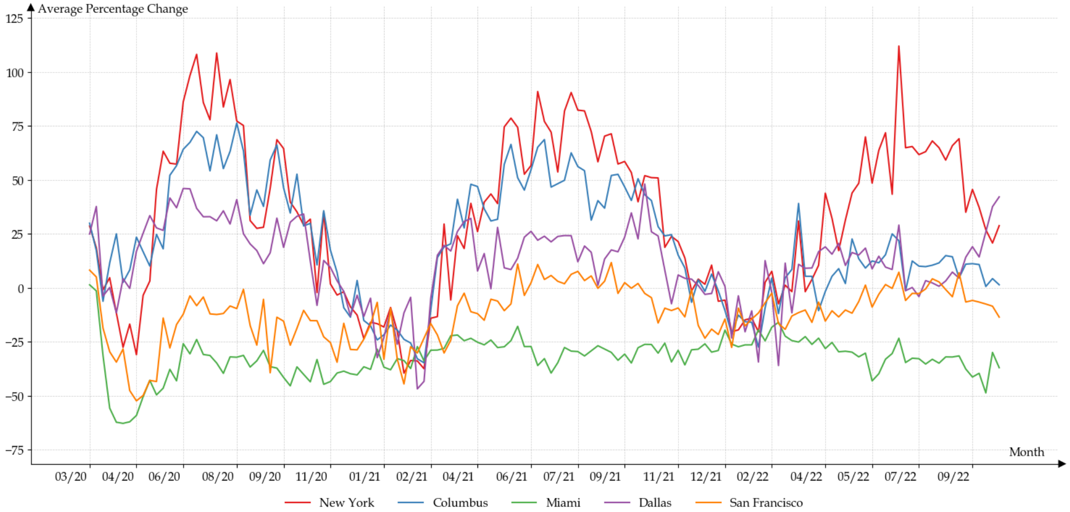
Trips to retail and recreation for five cities



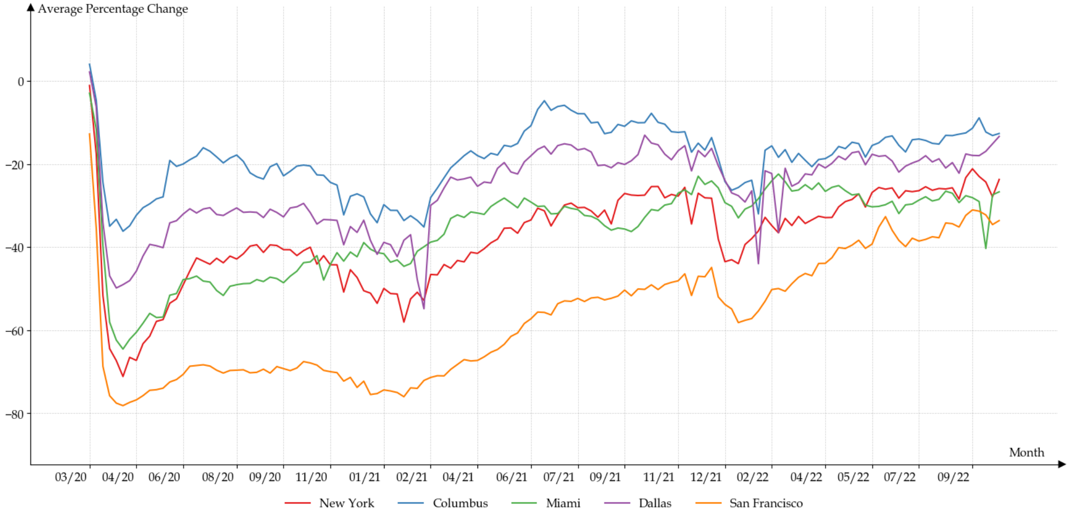
Trips to grocery and pharmacy for five cities



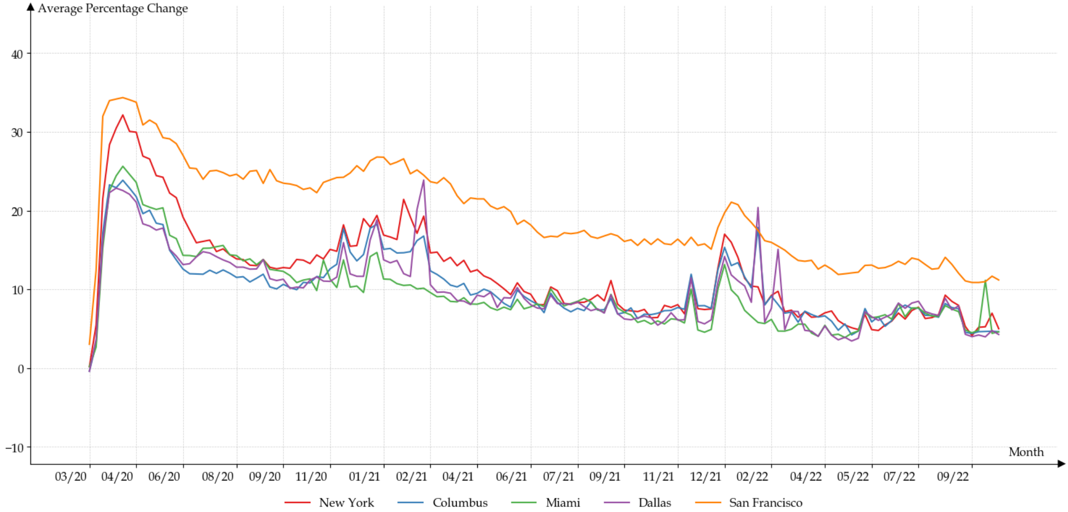
Trips to parks for five cities



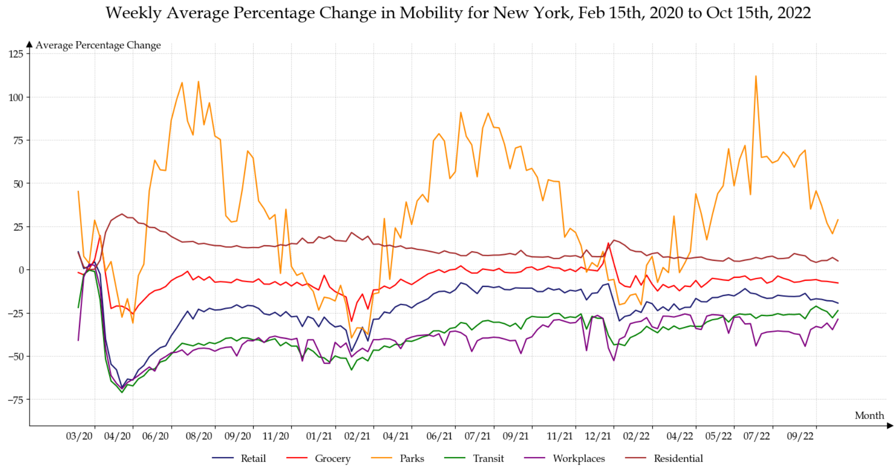
Trips to transit stations for five cities



Residential trips (stays) for five cities



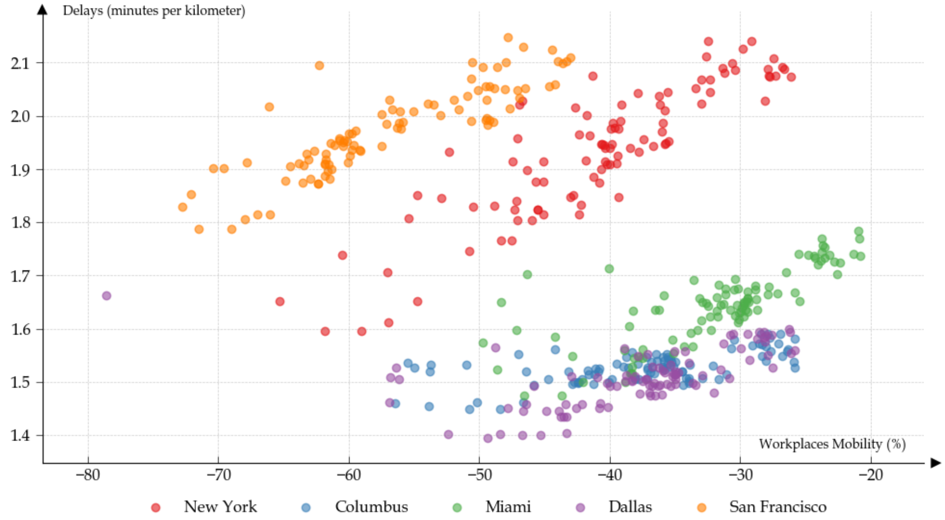
Quantities of trips for New York



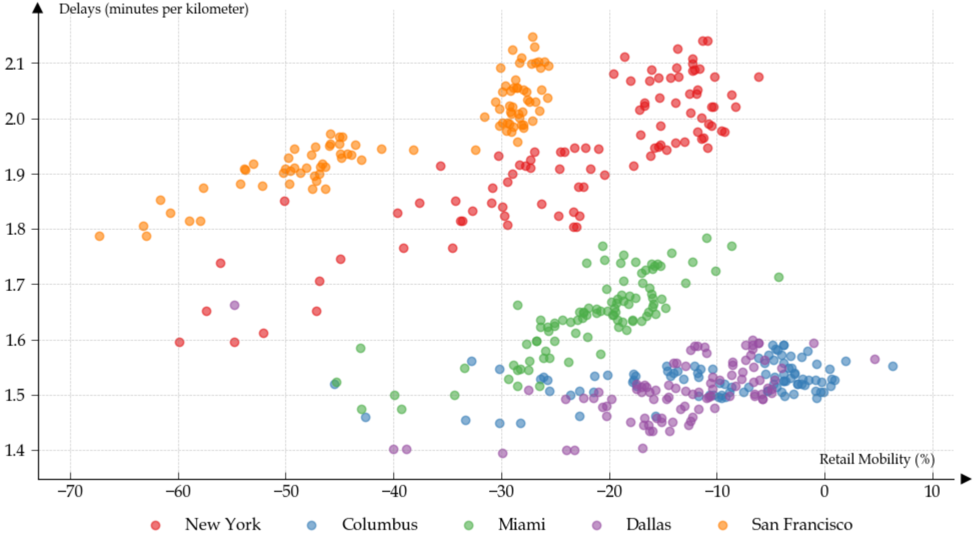
For our main data, let us temporarily consider "Delay" (in minutes per kilometer) rather than "Speed" (in kilometers per hour), a rescaled inverse that captures the cost of travel

In what follows, each dot is a city by week

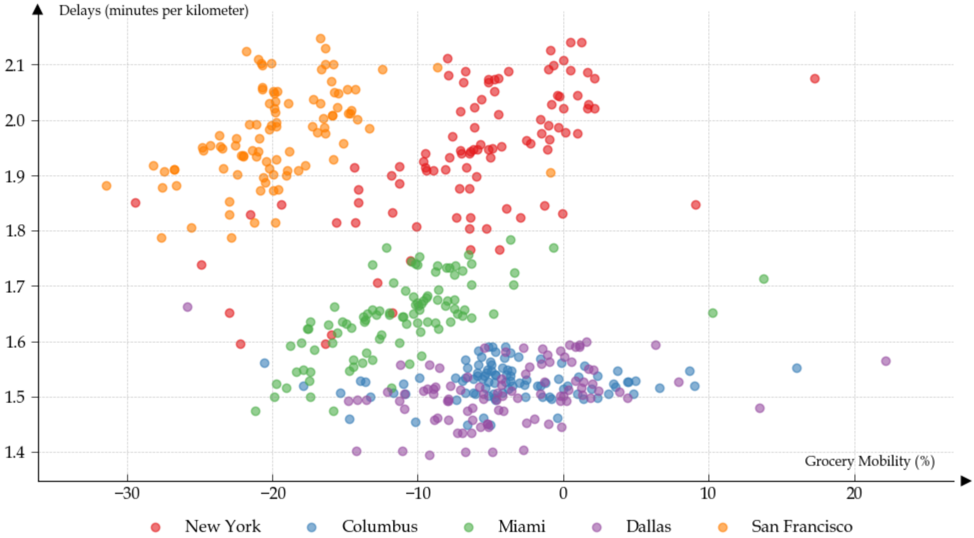
Prices and quantities: Commuting trips at peak hours



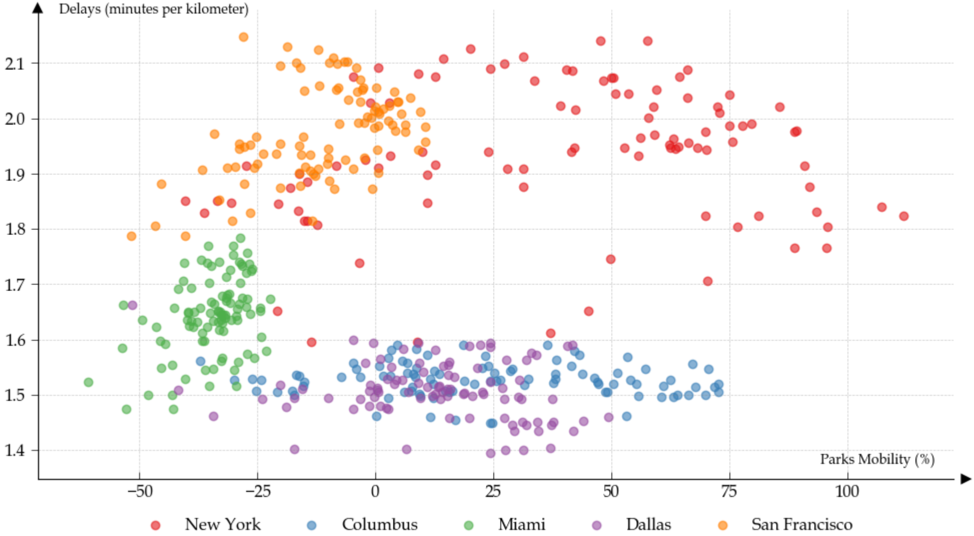
Prices and quantities: Retail and recreation at peak hours



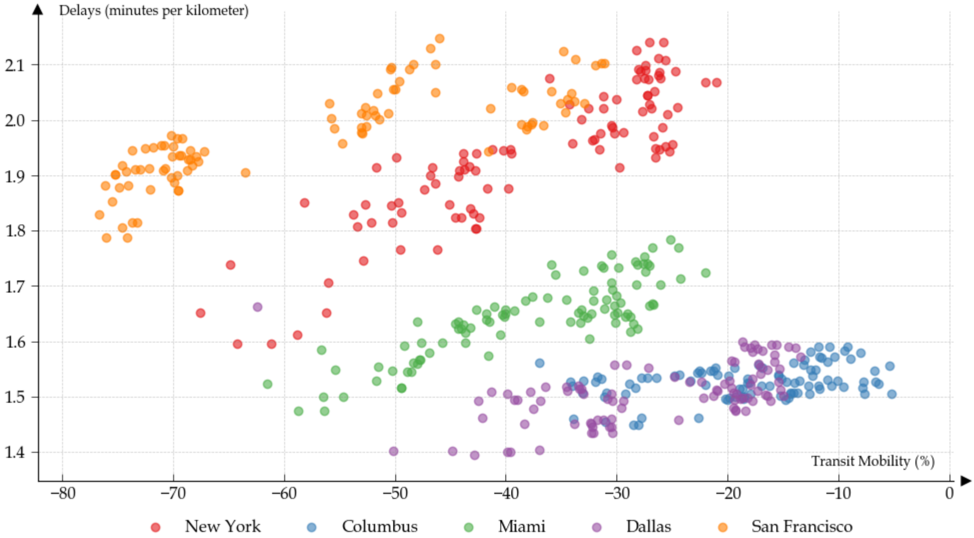
Prices and quantities: Grocery and pharmacy at peak hours



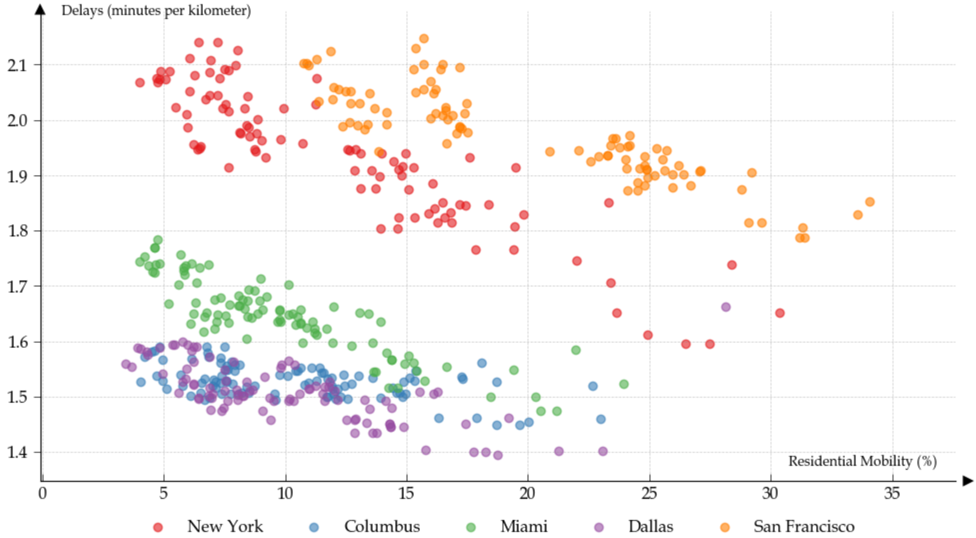
Prices and quantities: Parks at peak hours



Prices and quantities: Transit stations at peak hours

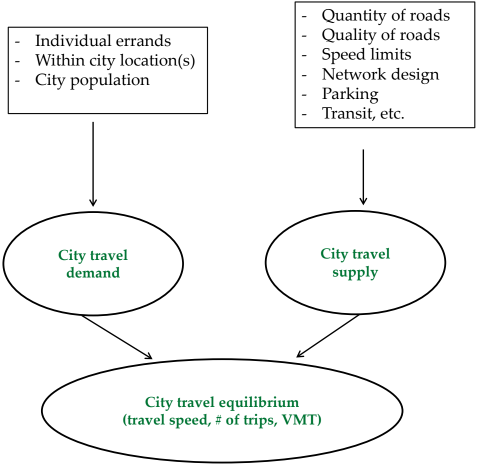


Prices and quantities: Residential 'trips' (stays) at peak hours

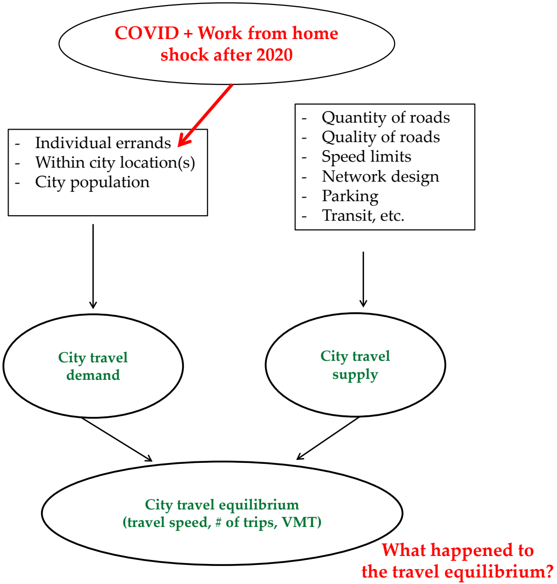


However tempting, we cannot interpret these relationships between the number of trips and delays (speed) as supply curves reflecting congestion effects. We need to isolate demand shifts

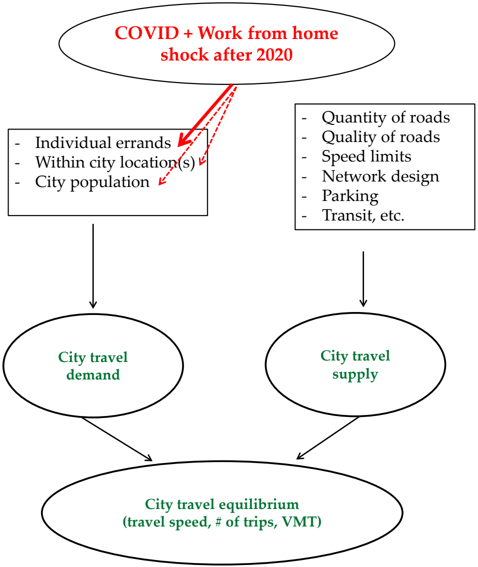
The city transportation equilibrium



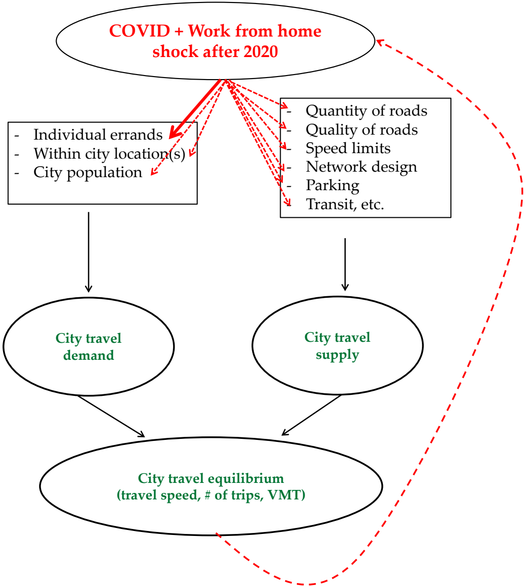
The city transportation equilibrium



The city transportation equilibrium



The city transportation equilibrium



Estimating congestion effects

We are now trying to estimate the effect of the number of trips from GCCMR data on travel speed

- Our first strategy exploits high-frequency variations in COVID cases

$\Rightarrow$ First stage:

$$\Delta\, log\ Trips_{cpg} = \rho\, WFHpropensity_c \times COVIDcases_{cp} + \alpha_c + \delta_p + \mu_{cp} \qquad (2)$$

where:

- $\Delta log\ Trips_{cpg}$ is the log difference in the number of trips of group g relative to early 2020 in city c and week p (and $\Delta\ Tripindex$ is for a weighted index for all trips)
- $WFHpropensity_c$ is a Dingel and Neiman (2020) predictor of WFH based on the local structure of occupations
  Note: With city fixed effects, this variable only rescales the instrument

Estimating congestion effects

- Second stage:

$$\Delta log\ Speed_{ckp} = \sigma \widehat{\Delta\ Trips}_{cpg} + \alpha_c + \delta_p + \epsilon_{ckp} \tag{3}$$

- Exclusion restriction: weekly COVID cases interacted with WFH propensity are uncorrelated with ϵ, conditionally on fixed effects for weeks and cities.
  Threats:
  - Correlated changing road availability (can be measured: OSM + trip distance)
  - Changing road quality (uncongested speed and trip distance)
  - Changing cities, eg population growth or decline (high frequency population data)

Weekly COVID cases in five cities

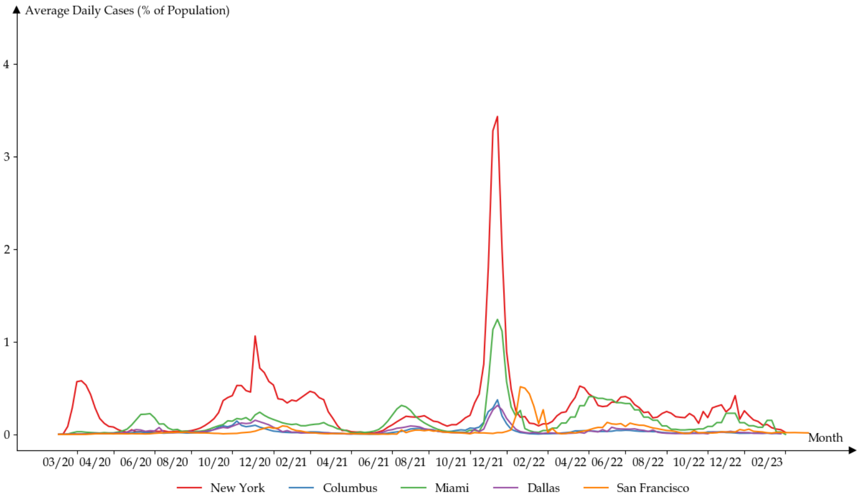


Figure: Weekly COVID Cases: January 2020 - March 2023

Dingel and Neiman (2020) WFH propensity

Table: Bottom 10 cities

City	WFH Predictor
Visalia	0.269
Cape Coral	0.281
Bonita Springs-Naples	0.281
Stockton	0.285
Youngstown	0.289
Bakersfield	0.291
Rockford	0.293
Lancaster	0.295
Grand Rapids	0.295
McAllen	0.297

Table: Top 10 cities

City	WFH Predictor
San Jose	0.511
Washington, D.C.	0.498
Trenton	0.494
Durham	0.46
Austin	0.455
Ann Arbor	0.449
San Francisco	0.448
Oakland	0.448
Antioch	0.448
Boston	0.444

OLS: Peak hour log speed

	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Retail/recreation trips)	-0.06759***					
	(0.0071)					
Δ log(Grocery trips)		-0.03218***				
		(0.0067)				
Δ log(Parks trips)			-0.00875***			
			(0.0011)			
Δ log(Trips to transit station)				-0.04894***		
				(0.0016)		
Δ log(Commutes)					-0.04744***	
					(0.0116)	
Δ log(Residential trips/stays)						0.45831***
						(0.0345)

OLS: Peak hour log speed

	(1)	(2)	(3)
	log night-hours speed	log off-peak hours speed	log peak hours speed
Δ log Trip index	0.01512***	-0.06089***	-0.03455***
	(0.0016)	(0.0031)	(0.0023)
Constant	3.80608***	3.67902***	3.70215***
	(0.0003)	(0.0005)	(0.0004)
N	13667	13472	13472
CityFE	✓	✓	✓
WeekFE	✓	✓	✓

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

First stage: Δ log trips on instrument

	(1) log(Δ Retail/recreation trips)	(2) Δ log(Grocery trips)	(3) Δ log(Parks trips)	(4) Δ log(Trips to transit station)	(5) Δ log(Commutes)	(6) Δ log(Residential trips/stays)
WFH X COVID	-1.4944***	-0.5572***	-5.7906***	-1.7708***	-0.4093***	0.6287***
	(0.1155)	(0.0688)	(0.4055)	(0.2502)	(0.0782)	(0.0309)
Constant	-0.1613***	-0.0422***	0.1879***	-0.2848***	-0.3423***	0.0667***
	(0.0011)	(0.0006)	(0.0034)	(0.0021)	(0.0007)	(0.0003)
N	13472	13472	13472	13472	13472	13472
CityFE	✓	✓	✓	✓	✓	✓
WeekFE	✓	✓	✓	✓	✓	✓

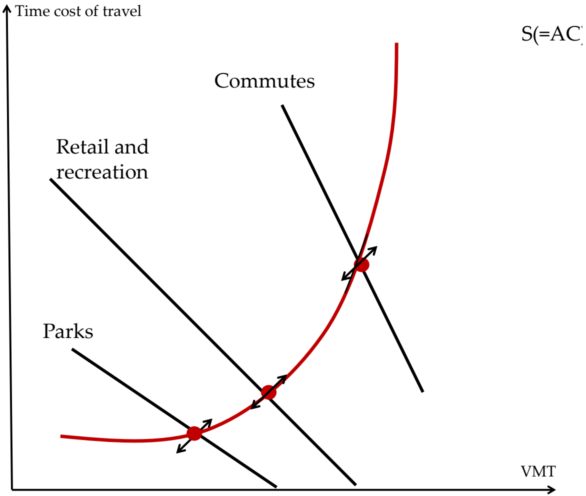
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Second stage: log peak hours speed

	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Retail/recreation trips)	-0.24024***					
	(0.0239)					
Δ log(Grocery trips)		-0.64427***				
		(0.0913)				
Δ log(Parks trips)			-0.06200***			
			(0.0063)			
Δ log(Trips to transit station)				-0.20273***		
				(0.0263)		
Δ log(Commutes)					-0.87720***	
					(0.1715)	
Δ log(Residential trips/stays)						0.57106***
						(0.0499)

Heterogeneous effects



Second stage: log off-peak hours speed

	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Retail/recreation trips)	-0.17224***					
	(0.0185)					
Δ log(Grocery trips)		-0.46190***				
		(0.0687)				
Δ log(Parks trips)			-0.04445***			
			(0.0046)			
Δ log(Trips to transit station)				-0.14535***		
				(0.0195)		
Δ log(Commutes)					-0.62890***	
					(0.1301)	
Δ log(Residential trips/stays)						0.40941***
						(0.0403)

Second stage: log night hours speed

	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Retail/recreation trips)	-0.06071***					
	(0.0121)					
Δ log(Grocery trips)		-0.16063***				
		(0.0365)				
Δ log(Parks trips)			-0.01571***			
			(0.0031)			
Δ log(Trips to transit station)				-0.05206***		
				(0.0115)		
Δ log(Commutes)					-0.21860***	
					(0.0597)	
Δ log(Residential trips/stays)						0.14430***
						(0.0272)
N	13667	13667	13667	13667	13667	13667
IVTest	170.044	68.885	208.806	50.111	28.165	417.991
IVPval	0.000	0.000	0.000	0.000	0.000	0.000
CityFE	✓	✓	✓	✓	✓	✓
WeekFE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Second stage: log trip index

	(1)	(2)	(3)
	log night-hours speed	log off-peak hours speed	log peak hours speed
Δlog Trip index	-0.04775***	-0.18836***	-0.13504***
	(0.0093)	(0.0177)	(0.0136)
N	13667	13472	13472
IVTest	268.821	264.212	264.212
IVPval	0.000	0.000	0.000
CityFE	✓	✓	✓
WeekFE	✓	✓	✓

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Robustness: Peaks hours speed, COVID only IV

	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Retail/recreation trips)	-0.23661***					
	(0.0224)					
Δ log(Grocery trips)		-0.64256***				
		(0.0880)				
Δ log(Parks trips)			-0.06963***			
			(0.0069)			
Δ log(Trips to transit station)				-0.20475***		
				(0.0258)		
Δ log(Commutes)					-0.72608***	
					(0.1192)	
Δ log(Residential trips/stays)						0.57668***
						(0.0477)

Robustness: Peaks hours speed, levels

	(1)	(2)	(3)	(4)	(5)	(6)
Retail/recreation trips	-0.11416***					
	(0.0105)					
Grocery trips		-0.27948***				
		(0.0381)				
Parks trips			-0.01475***			
			(0.0014)			
Trips to transit station				-0.09621***		
				(0.0110)		
Commutes					-0.49187***	
					(0.0845)	
Residential trips/stays						0.22640***
						(0.0183)
N	13472	13472	13472	13472	13472	13472
IVTest	224.943	71.008	285.162	79.608	37.059	392.582
IVPval	0.000	0.000	0.000	0.000	0.000	0.000
CityFE	✓	✓	✓	✓	✓	✓
WeekFE	✓	✓	✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Estimating the effects of work from home

- Our second estimation strategy relies on the cross-section of cities and the medium run effects of WFH in the Fall of 2022 (9/15 to 10/15, the end of GCCMR)
- First stage

$$\Delta\, Trips_c = \rho\, WFHpropensity_c + \mu_c \tag{4}$$

  where the WFH predictor is constructed from Dingel-Neiman interacted with local employment data or is a Bartik after reconstructing sector level WFH from city WFH
- Second stage:

$$\Delta log\ Speed_{ck} = \sigma \widehat{\Delta\ Trips}_c + \epsilon_{ck} \tag{5}$$

- Exclusion restriction: WFH predictor uncorrelated with ϵ. Threats:
  - Changing road availability correlated with WFH (can be measured: OSM + trip distance)
  - Changing road quality (uncongested speed and trip distance)
  - Changing city population (add late 2022 population)

WFH

	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Retail/recreation trips)	-13.77029					
	(7.2249)					
Δ log(Grocery trips)		-22.96483				
		(16.0299)				
Δ log(Parks trips)			4.67824			
			(3.0360)			
Δ log(Trips to transit station)				-6.36450		
				(3.4796)		
Δ log(Commutes)					-5.94352***	
					(1.7367)	
Δ log(Residential trips/stays)						27.53362***
						(7.7824)
N	130	130	130	130	130	130
IVTest	4.988	2.111	2.988	4.959	37.425	38.276
IVPval	0.031	0.139	0.095	0.041	0.000	0.000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Things yet to be done

- More descriptive work
- Modeling framework to combine the different elasticities
- Resolve tension about descriptives and identification
- Travel times rather than speed as dependent variable
- Use ACS measures of WFH in early 2024
- Alternative measures of COVID intensity
- Measuring changes in roads (supply effects)
- Recent developments in shift-share instruments and TWFE
- *Please add to my list...*

Conclusion

- Combine data about travel costs and number of trips over several years
- Show interesting facts about changes in urban travel
- Identify congestion elasticities associated with different types of trips using weekly variations
- Identify medium term effects of WFH