Can Tolling Help Everyone?

Jonathan Hall

University of Toronto

Traffic congestion is a major problem

Costs of traffic congestion

- 52 hr/commuter/yr in major urban areas
  (Schrank et al. 2012)
- 2.2% of annual gasoline consumption
  (Schrank et al. 2012; EIA 2012)
- Additional pollution more than 6 times the amount saved by current fleet of hybrid and electric vehicles
  (Samaras and Meisterling 2008; EPA 2011; Schrank et al. 2012; EIA 2013)
- Pollution responsible for 8,600 pre-term births
  (Currie and Walker 2011)

We know how to solve traffic congestion

Solution

- Tolls
- First proposed by Pigou in 1920

A barrier to congestion pricing is the belief that it hurts many road users

- Academics

  "First-best congestion pricing . . . introduces severe disparities in direct welfare impact." Small, Winston, and Yan, 2005

- Policy makers

  "[Congestion pricing is] unfair in terms of the economic impact." Maryland Gov. Parris Glendening

- Pundits

  "Exalted [toll] lanes leave the average Joe in the dust." Marc Fisher, The Washington Post

- Public

  "Turkeys don't vote for Christmas and motorists won't vote for more taxes to drive." Voter in Manchester, UK

KEY RESULT: A carefully designed toll on a portion of the lanes can yield a Pareto improvement before revenue spent

KEY RESULT: A carefully designed toll on a portion of the lanes can yield a Pareto improvement before revenue spent

Toll should be:

- Time-varying
- Collected electronically
- Set to maximize throughput, not profits or social welfare

KEY RESULT: A carefully designed toll on a portion of the lanes can yield a Pareto improvement before revenue spent

KEY RESULT: A carefully designed toll on a portion of the lanes can yield a Pareto improvement before revenue spent

- Give up some potential Kaldor-Hicks efficiency for a Pareto improvement

Key result: A carefully designed toll on a portion of the lanes can yield a Pareto improvement before revenue spent

- Give up some potential Kaldor-Hicks efficiency for a Pareto improvement
- If this allows us to overcome political opposition then we're trading potential efficiency gains for actual efficiency gains

KEY RESULT: A carefully designed toll on a portion of the lanes can yield a Pareto improvement before revenue spent

- Give up some potential Kaldor-Hicks efficiency for a Pareto improvement
- If this allows us to overcome political opposition then we're trading potential efficiency gains for actual efficiency gains
- What allows me to get this new result?
  - Identifying a second externality using insights from traffic engineering literature
  - Extend bottleneck model to include this externality

An additional driver can impose two externalities

1. Lengthen the line

2. Reduce throughput/reduce speed at which line moves

There are two ways congestion reduces throughput

- Once queue forms throughput at bottleneck drops
  - e.g. throughput on I-805N at 47th St. in San Diego regularly falls by 12% once a queue forms (Chung et al. 2007)
  - cf. Banks (1990), Hall and Agyemang-Duah (1991), Banks (1991), Persaud et al. (1998), Cassidy and Bertini (1999), Bertini and Malik (2004), Zhang and Levinson (2004), Bertini and Leal (2005), Cassidy and Rudjanakanoknad (2005), Rudjanakanoknad (2005), Chung et al. (2007), Guan et al. (2009), Oh and Yeo (2012), Srivastava and Geroliminis (2013)

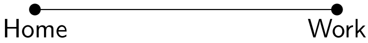
There are two ways congestion reduces throughput

- Once queue forms throughput at bottleneck drops
  - e.g. throughput on I-805N at 47th St. in San Diego regularly falls by 12% once a queue forms (Chung et al. 2007)
  - cf. Banks (1990), Hall and Agyemang-Duah (1991), Banks (1991), Persaud et al. (1998), Cassidy and Bertini (1999), Bertini and Malik (2004), Zhang and Levinson (2004), Bertini and Leal (2005), Cassidy and Rudjanakanoknad (2005), Rudjanakanoknad (2005), Chung et al. (2007), Guan et al. (2009), Oh and Yeo (2012), Srivastava and Geroliminis (2013)
- Queue behind bottleneck blocks upstream traffic
  - e.g. throughput on I-880N near San Francisco regularly falls by 25% due to queue spillovers from I-238 (Munoz and Daganzo 2002)

Extend work-horse model of dynamic congestion to capture additional externality

Bottleneck model (Vickrey, 1969; Arnott, de Palma, and Lindsey 1990)

- Road network

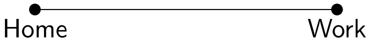


  - Can costlessly split this road into two routes: one priced and one free

Extend work-horse model of dynamic congestion to capture additional externality

Bottleneck model (Vickrey, 1969; Arnott, de Palma, and Lindsey 1990)

- Road network



  - Can costlessly split this road into two routes: one priced and one free
- Congestion
  - Only source of delay is a bottleneck of finite capacity
  - Only s^* vehicles can pass through bottleneck per minute
  - When queue $> \epsilon$ throughput falls to $s < s^*$

Agents want to arrive on time with no traffic or toll

Agent preferences

- trip cost $= \alpha_i$ (travel time $+ \delta_i$ time early $+ \xi\delta_i$ time late) $+$ toll
- Heterogeneity in three dimensions
  - Value of time – $\alpha_i > 0$ – cost of travel time in dollars
  - Inflexibility – $\delta_i \in [0, 1]$ – cost of time early in travel time
  - Desired arrival time – $t^* \sim$ Uniform $[t_s, t_e]$
- $\xi > 0$ – ratio of cost of being late to early

Agents want to arrive on time with no traffic or toll

Agent preferences

- trip cost $= \alpha_i$ (travel time $+ \delta_i$ time early $+ \xi\delta_i$ time late) $+$ toll
- Heterogeneity in three dimensions
  - Value of time – $\alpha_i > 0$ – cost of travel time in dollars
  - Inflexibility – $\delta_i \in [0, 1]$ – cost of time early in travel time
  - Desired arrival time – $t^* \sim \text{Uniform}\,[t_s, t_e]$
- $\xi > 0$ – ratio of cost of being late to early

Agents want to arrive on time with no traffic or toll

Agent preferences

- trip cost $= \alpha_i$ (travel time $+ \delta_i$ time early $+ \xi\delta_i$ time late) $+$ toll
- Heterogeneity in three dimensions
  - Value of time – $\alpha_i > 0$ – cost of travel time in dollars
  - Inflexibility – $\delta_i \in [0, 1]$ – cost of time early in travel time
  - Desired arrival time – $t^* \sim \text{Uniform}\,[t_s, t_e]$
- $\xi > 0$ – ratio of cost of being late to early

Agents want to arrive on time with no traffic or toll

Agent preferences

- trip cost $= \alpha_i\,(\text{travel time} + \delta_i\,\text{time early} + \xi\delta_i\,\text{time late}) + \text{toll}$
- Heterogeneity in three dimensions
  - Value of time – $\alpha_i > 0$ – cost of travel time in dollars
  - Inflexibility – $\delta_i \in [0, 1]$ – cost of time early in travel time
  - Desired arrival time – $t^* \sim \text{Uniform}\,[t_s, t_e]$
- $\xi > 0$ – ratio of cost of being late to early

Agents want to arrive on time with no traffic or toll

Agent preferences

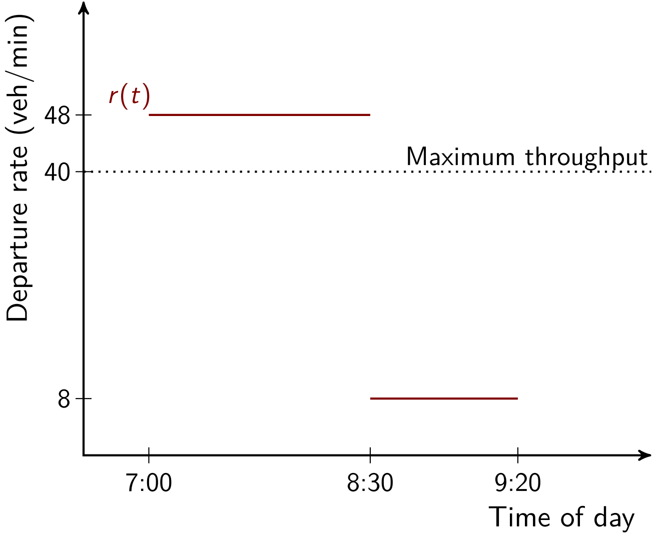
- trip cost $= \alpha_i$ (travel time $+ \delta_i$ time early $+ \xi\delta_i$ time late) $+$ toll
- Heterogeneity in three dimensions
  - Value of time – $\alpha_i > 0$ – cost of travel time in dollars
  - Inflexibility – $\delta_i \in [0, 1]$ – cost of time early in travel time
  - Desired arrival time – $t^* \sim$ Uniform $[t_s, t_e]$
- $\xi > 0$ – ratio of cost of being late to early

Agents want to arrive on time with no traffic or toll

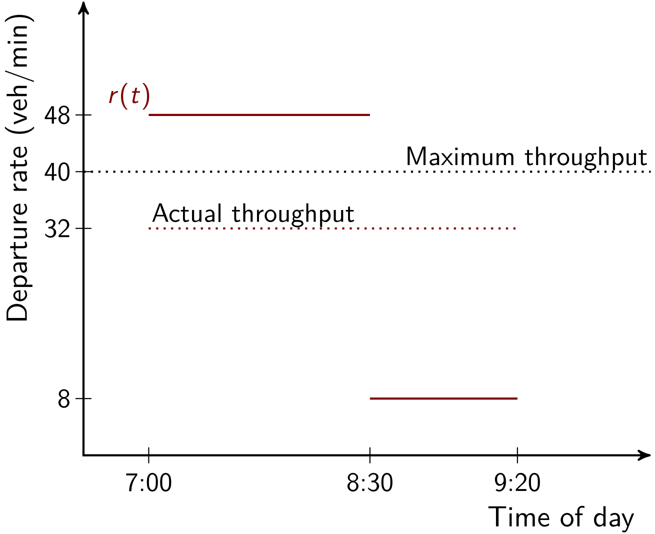
Agent preferences

- trip cost $= \alpha_i$ (travel time $+ \delta_i$ time early $+ \xi\delta_i$ time late) $+$ toll
- Heterogeneity in three dimensions
  - Value of time – $\alpha_i > 0$ – cost of travel time in dollars
  - Inflexibility – $\delta_i \in [0, 1]$ – cost of time early in travel time
  - Desired arrival time – $t^* \sim \text{Uniform}[t_s, t_e]$
- $\xi > 0$ – ratio of cost of being late to early
- Drivers choose
  - Time of departure $\in \mathbb{R}$
  - Route
- Perfectly inelastic demand

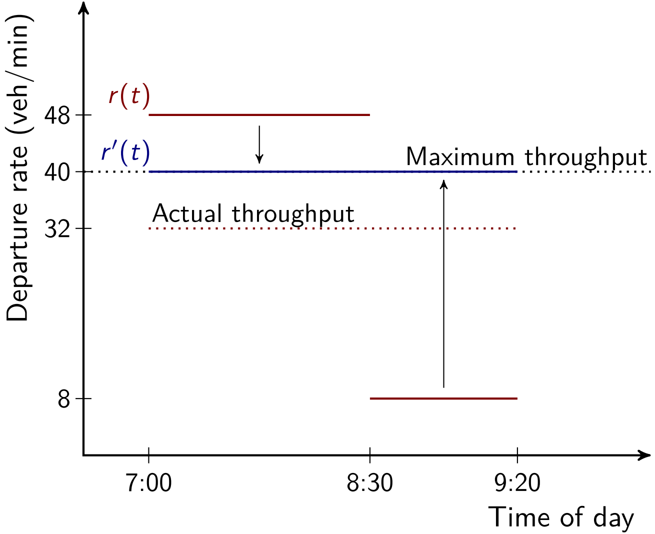
Not everyone can arrive on-time, and so queues form



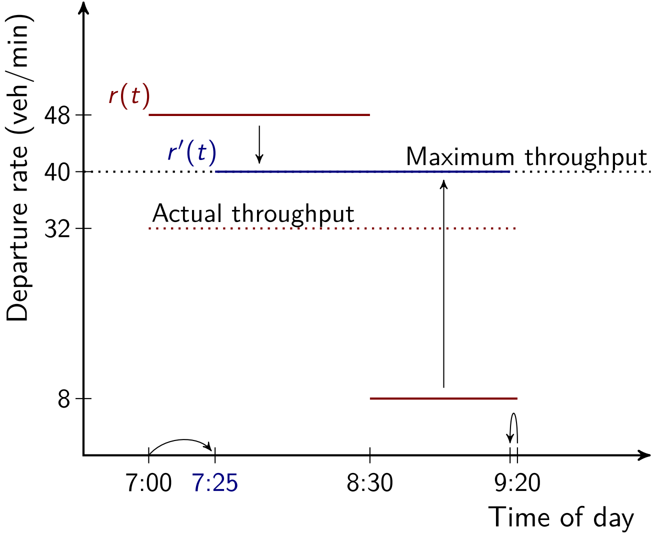
Throughput falls because of queuing



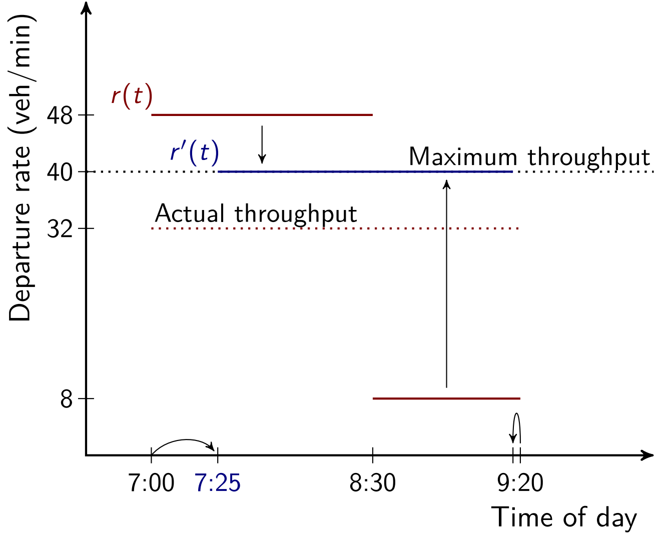
Use tolls to affect rate at which drivers depart



No queuing means higher throughput and shorter rush hour



No queuing means higher throughput and shorter rush hour



$\Rightarrow$ when agents are homogeneous pricing is a Pareto improvement

When there are rich and poor agents it is harder to make everyone better off

What happens when we price the entire road?

- Internalize externality
- Increase speeds and throughput
- Change currency from time to money

When there are rich and poor agents it is harder to make everyone better off

What happens when we price the entire road?

- Internalize externality
- Increase speeds and throughput
- Change currency from time to money

By only pricing a portion of the lanes we can still generate a Pareto improvement

Intuition for pricing a portion of the lanes

	Both lanes free	
	Lane 1	Lane 2
Pricing	Free	Free
Avg. queue length	long	long
Throughput	low	low
Travel time	long	long
Share of trips	50%	50%

By only pricing a portion of the lanes we can still generate a Pareto improvement

Intuition for pricing a portion of the lanes

	Both lanes free		Price one lane	
	Lane 1	Lane 2	Lane 1	Lane 2
Pricing	Free	Free	Toll	Free
Avg. queue length	long	long	0	↓
Throughput	low	low	↑	–
Travel time	long	long	↓	↓
Share of trips	50%	50%	↑	↓

Derive a simple sufficient condition for pricing a portion of the lanes to yield a Pareto improvement

What we learn from theory

- As long as some rich drivers traveling at the peak, then pricing some of the lanes yields a Pareto improvement
- Key parameters
  - Size of throughput drop
  - Correlation between value-of-time and inflexibility

Data from surveys and highway loop detectors

Data

- Caltrans Performance Measurement System
  - Use to estimate travel times for California State Route 91W from the center of Corona to I-605 junction
- California State Route 91 Impact Study
  - Surveys of drivers who use SR-91
  - Conducted between 1995 and 1999
- 2009 National Household Travel Survey
  - Use to confirm results are representative of other large MSAs

I estimate the joint distribution of inflexibility, value of time, and desired arrival time

Empirical overview

- Goal: Estimate joint distribution of agent preferences

I estimate the joint distribution of inflexibility, value of time, and desired arrival time

Empirical overview

- Goal: Estimate joint distribution of agent preferences
- Split population into two categories
  - Flexible
  - Inflexible

I estimate the joint distribution of inflexibility, value of time, and desired arrival time

Empirical overview

- Goal: Estimate joint distribution of agent preferences
- Split population into two categories
  - Flexible
  - Inflexible
- Within each category estimate the marginal distributions of
  - Value of time
  - Desired arrival time
  - Inflexibility

I estimate the joint distribution of inflexibility, value of time, and desired arrival time

Empirical overview

- Goal: Estimate joint distribution of agent preferences
- Split population into two categories
  - Flexible
  - Inflexible
- Within each category estimate the marginal distributions of
  - Value of time
  - Desired arrival time
  - Inflexibility
- Combine by assuming independence within each category

I estimate the joint distribution of inflexibility, value of time, and desired arrival time

Empirical overview

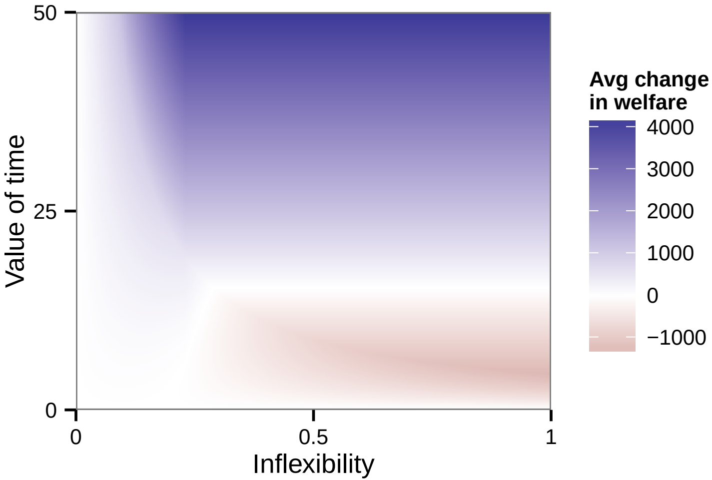
- Goal: Estimate joint distribution of agent preferences
- Split population into two categories
  - Flexible
  - Inflexible
- Within each category estimate the marginal distributions of
  - Value of time
  - Desired arrival time of inflexible agents
  - Inflexibility of flexible agents
- Combine by assuming independence within each category

We can use empirical results to evaluate counterfactuals

- Analytic solution for equilibrium on a route given agents on route
- Need to solve for which agents on which route numerically
  - Find which VOT indifferent between routes for each inflexibility and desired arrival time
- Simulate throughput drop of 10, 17.5, and 25%

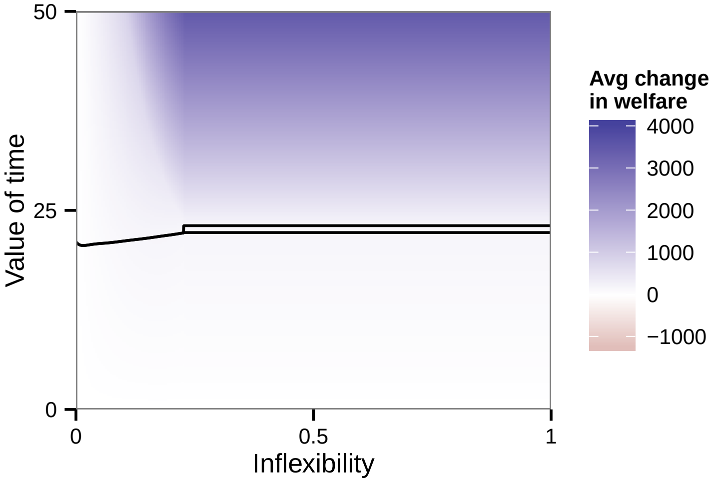
Pricing all of the road hurts the inflexible poor

Average annual change in welfare when pricing all lanes



Pricing 1/2 of lanes generates a Pareto improvement

Average annual change in welfare when pricing 1/2 of lanes



The welfare gains from pricing are large

Average annual welfare effects

Fraction of lanes priced	1	1/2
Largest welfare loss ($)	2,390	0
Welfare gains ($)		
Social	2,400	1,740
Private	1,080	760
Reduction in travel time (hours)	76.5	40.5
Value of reduction in travel time ($)	1,960	1390

The welfare gains from pricing are large

Average annual welfare effects

Fraction of lanes priced	1	1/2
Largest welfare loss ($)	2,390	0
Welfare gains ($)		
Social	2,400	1,740
Private	1,080	760
Reduction in travel time (hours)	76.5	40.5
Value of reduction in travel time ($)	1,960	1390

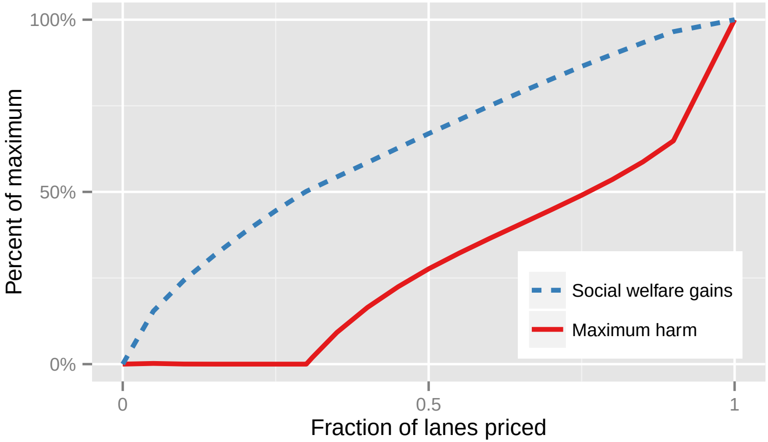
The welfare gains from pricing are large

Average annual welfare effects

Fraction of lanes priced	1	1/2
Largest welfare loss ($)	2,390	0
Welfare gains ($)		
Social	2,400	1,740
Private	1,080	760
Reduction in travel time (hours)	76.5	40.5
Value of reduction in travel time ($)	1,960	1390

If willing to relax requirement that pricing hurt no one, then can obtain a larger share of welfare gains

Max harm and social welfare gains with throughput drop of 10%



We can improve the welfare effects of congestion pricing

Things could add to analysis to help obtain a Pareto improvement

- Use of revenue
- Ways to let inflexible poor to pay with time to travel at peak
- Shocks to preferences–everyone has days they are inflexible

We can improve the welfare effects of congestion pricing

Things could add to analysis to help obtain a Pareto improvement

- Use of revenue
  - Negative tolls off peak
  - Cut sales tax
  - Expand highway
  - Subsidize public transit
- Ways to let inflexible poor to pay with time to travel at peak

- Shocks to preferences–everyone has days they are inflexible

We can improve the welfare effects of congestion pricing

Things could add to analysis to help obtain a Pareto improvement

- Use of revenue
  - Negative tolls off peak
  - Cut sales tax
  - Expand highway
  - Subsidize public transit
- Ways to let inflexible poor to pay with time to travel at peak
  - Public transit
  - Carpooling
- Shocks to preferences–everyone has days they are inflexible

Conclusion

- Congestion pricing can increase highway throughput
- Theoretically, pricing a portion of the lanes can help all road users, even before we use the revenue
- Empirically, pricing 1/2 of lanes on SR-91 will help all road users, with welfare gains of 3.5% median income

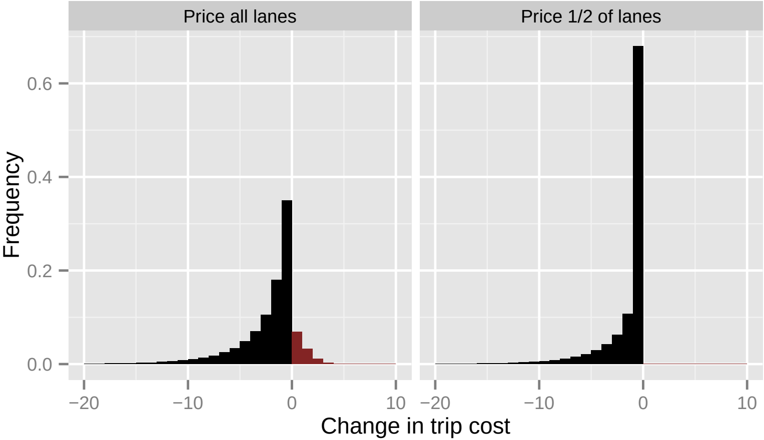


Figure: Distribution of changes in trip costs

About half of all trips are flexible

Estimating the fraction of drivers/trips that are flexible

		NHTS	
Fraction of . . .	SR-91 IS	Large MSAs	All
Drivers who leave early or late to avoid traffic	.57 [.55, .60]		
Workers who commute via interstate who can choose work arrival time	.50 [.47, .53]	.47 [.45, .49]	.44 [.43, .45]
Trips on interstate during morning that are flexible	.43 [.40, .47]	.35–.60 [.32, .62]	.30–.59 [.29, .60]

Next I estimate the distribution of value of time by category

1. Convert household income into value of time using USDOT formula
$$\text{VOT} = \frac{1}{2} \times \frac{\text{Income}}{2080}$$
2. Fit to log-normal distribution using maximum likelihood

This method gives similar results as more detailed studies

Comparison to Small et al. (2005)

	My estimates	Small et al. (2005)
Median	23.58	29.54
Interquartile range	17.06	10.47

This method gives similar results as more detailed studies

Comparison to Small et al. (2005)

	My estimates	Small et al. (2005)
Median	23.58	29.54
Interquartile range	17.06	10.47

- I need to separate by category
- My estimates
  - Undervalue time savings
  - Greater inequality $\Rightarrow$ makes it harder to find a Pareto improvement

The flexible tend to be richer than the inflexible

Distribution of value of time for morning highway users

	SR-91 IS		NHTS	
	All	All	Large MSAs	All
Definition of flexibility	1	3	3	3
Flexible				
Median	22.12	25.95	26.05	20.41
	(0.64)	(0.88)	(0.34)	(0.13)
Interquartile range	16.4	20.0	32.21	25.41
	(1.0)	(1.8)	(0.89)	(0.35)
N	413	303	7,059	21,342
Inflexible				
Median	22.71	22.16	22.52	19.02
	(0.73)	(0.56)	(0.27)	(0.11)
Interquartile range	16.4	15.19	24.55	18.95
	(1.3)	(0.88)	(0.59)	(0.19)
N	292	433	4,270	12,995
Rank correlation	-0.053	0.20***	0.157***	0.108***
	(0.059)	(0.057)	(0.037)	(0.028)

Next I estimate the distribution of desired arrivals

- Problem
  - Want to measure desired arrival time at highway exit
  - Observe actual arrival at destination

Next I estimate the distribution of desired arrivals

- Problem
  - Want to measure desired arrival time at highway exit
  - Observe actual arrival at destination
- Solution
  - Limit sample to those who must arrive on time
  - Assume distribution same for both categories
  - Recognize that true distribution will be smoothed version of observed distribution

The distribution of desired arrival times is not uniform

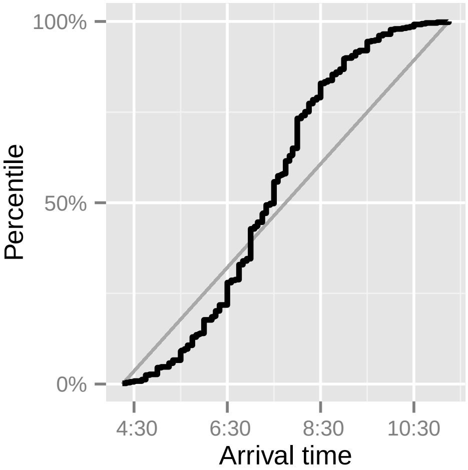


Figure: Cumulative distribution of desired arrival times on SR-91

But after trimming it is reasonably uniform

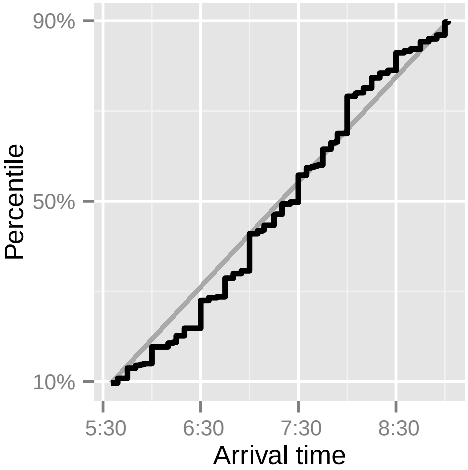


Figure: 10th–90th percentiles of cumulative distribution of desired arrival times

Estimate range of distribution by matching 10th and 90th percentiles to the expected value of their order statistics

Estimating parameters of distribution of desired arrival times

- Estimate parameters by matching extreme values of trimmed sample to the expected value of their order statistic
- Unbiased estimator of length of desired arrivals:

$$\widehat{LDA} = \frac{N+1}{m-n}\left(X_{(m)} - X_{(n)}\right)$$

  - N - # observations
  - m - order statistic of largest remaining observation
  - n - order statistic of smallest remaining observation

Estimate range of distribution by matching 10th and 90th percentiles to the expected value of their order statistics

Estimating parameters of distribution of desired arrival times

- Estimate parameters by matching extreme values of trimmed sample to the expected value of their order statistic
- Unbiased estimator of length of desired arrivals:

$$\widehat{LDA} = \frac{N+1}{m-n}\left(X_{(m)} - X_{(n)}\right)$$

  - N - # observations
  - m - order statistic of largest remaining observation
  - n - order statistic of smallest remaining observation
- Range of desired arrivals: 4.40 hours (std. err. 0.23)

Finally I structurally estimate the distribution of inflexibility and length of rush hour using observed travel times

- Model maps parameters to predictions of travel times
- Use GMM to find distribution of inflexibility that best fits data
- Moment condition:

$$\text{Theoretical travel time} = \text{average travel time in data}$$

  for 4:00, 4:05, 4:10, . . . , 10:00.
- Also estimate:
  - Length of rush hour
  - ξ - Ratio of cost of being late to cost of being early

Essentially estimating the distribution of the slope of the travel time profile

Intuition behind estimation

- Theory says for drivers arriving early or late

$$\frac{dT}{dt}(t) = \begin{cases} \delta & \text{if early} \\ -\xi \cdot \delta & \text{if late} \end{cases}$$

- Essentially estimating the distribution of the slope of the travel time profile

Problem: Data cannot tell us about the inflexibility of the inflexible agents

Assumptions on distribution of inflexibility

- Inflexibility of flexible agents $\sim$ Uniform on $\left[0, \bar{\delta}\right]$
- Inflexibility of inflexible agents $\sim$ Beta$(5, 0.5)$ on $\left[\bar{\delta}, 1\right]$
  - Modal inflexibility is 1

Model fits data well

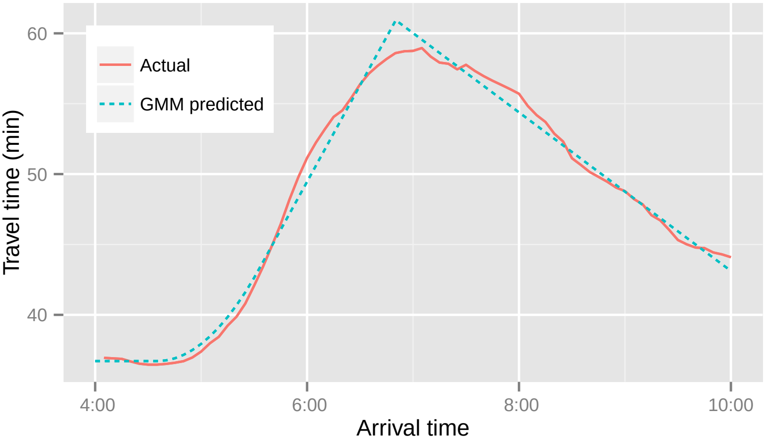


Figure: Actual vs. predicted travel times

As test of assumptions, estimate best non-parametric fit to travel time profile consistent with theory

Theory makes three restrictions on travel time profile

1. Travel times are positive
2. Travel times are increasing before the peak and decreasing after
3. Travel times are convex before the peak and convex after

GMM fits almost as well as non-parametric method

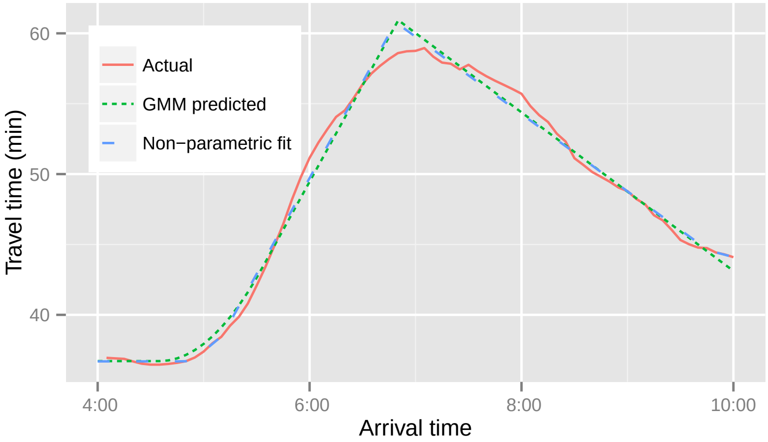


Figure: Actual vs. predicted travel times