Infrastructure resilience against environmental shocks and economic effects of transport disruptions

Evidence from India

Utkarsh Kumar

Columbia University

August 9, 2023

Motivation

- Transport infrastructure is essential for economic development
- Developing world spending heavily on transport infrastructure
    - India to spend USD 3 trillion by 2030 (Economic Syrvey 2017-18)
- However.....
- Transport infrastructure lacks resilience in developing countries
    - Poor quality roads → slow speeds, accidents
    - Poor railway transport → mean delays + uncertainty
- Little research on economic effects of transport quality

Motivation II

- Poorly functioning infrastructure can:
  - Reduce acces to work, goods and services
  - Create uncertainty in movement of people, goods and services
- Labor market effects
  - Larger time costs
  - Uncertainty in access to urban labor/goods markets
- Research Question:
  - How resilient is transport infrastructure in developing countries to environmental shocks?
  - How do transport disruptions affect economic outcomes?

Preview of methods and results

- I combine data on several environmental variables with novel data on Indian railways
  - Air pollution (PM2.5),relative humidity and visibility
  - Satellite data on agricultural crop fires in north India
  - Web-scraped Indian railway data on travel times for millions of trips

Preview of methods and results

- I combine data on several environmental variables with novel data on Indian railways
  - Air pollution (PM2.5),relative humidity and visibility
  - Satellite data on agricultural crop fires in north India
  - Web-scraped Indian railway data on travel times for millions of trips
- I find that India's transport network is not resilient to environmental shocks
  - Air pollution from agricultural burning in North India leads to 30% more delays over scheduled travel time on the entire railway network
  - The effect is much larger in region close to agricultural fire burning
  - Uncertainty in transport (measured by standard deviation of train delays) increases by 5%-10%

Preview of methods and results

- I combine data on several environmental variables with novel data on Indian railways
  - Air pollution (PM2.5),relative humidity and visibility
  - Satellite data on agricultural crop fires in north India
  - Web-scraped Indian railway data on travel times for millions of trips
- I find that India's transport network is not resilient to environmental shocks
  - Air pollution from agricultural burning in North India leads to 30% more delays over scheduled travel time on the entire railway network
  - The effect is much larger in region close to agricultural fire burning
  - Uncertainty in transport (measured by standard deviation of train delays) increases by 5%-10%
- Transport disruptions adversely affect India's rural workers
  - Monthly savings decrease as expenditure increases
  - A large portion of increase in expenditure comes from fuel and transport categories

Related literature

- Infrastructure quality, economic activity and welfare
  - Gertler et al (2019), Firth (2019)
- Uncertainty in trade and transport
  - Brooks and Donovan (2020), Allen and Atkin (2021)
- Effct of transport infrastructure on economic activity
  - Redding and Sturm (2008); Donaldson (2018), Bartelme (2015); Donaldson and Hornbeck (2016); Alder (2019); Balboni(2019)
- Non-health related effects of air pollution
  - Khanna et al. (2021); Hanna and Oliva (2015); He et al (2019); Graff Zivin and Neidell (2012)

Roadmap

Roadmap

Research Setting

- Heavy air pollution in North India (October-February) IGP Map
  - Sandwiched between himalayan nad vindya ranges
  - Crop burning; solid fuel based cooking and heating
  - 23 million tons of rice-residue burnt annually
- PM2.5 + Humidity → Dense smog → Low visibliity Evidence
  - Compounded by Winter Inversion
  - Environmental trends Visualization

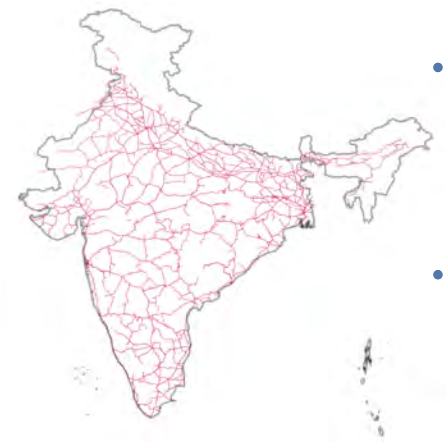
Heavy smog affects railways...

Railways officials said the smog in Delhi has delayed all trains to Chandigarh.

...And road transport

Drivers were blinded by the thick smog plaguing parts of North India (Image: Newsflare)

Rail and road network



- Indian Railways are 3rd largest in the world
  - 2019 - 8 billion passenger trips; 1.2 bn tonnes of freight
  - Highly subsidized mode of transport or even free of cost
  - Freight segment cost effective for bulky goods
- Railway network heavily concentrated in Indo-Gangetic Plains
  - Highway network broadly similar

Roadmap

Data I

- Train Delays: (September - February)
  - Daily station-station arrival delays and scheduled arrival times
  - 2744 super fast trains
  - District-district delays; Destination district level delays
- Air Pollution data and weather data
  - Daily PM2.5 data from CPCB India
  - Visibility, relative humidity (Iowa Environmental Mesonet)
- Fires data (mostly crop fires (Singh et al 2021)
  - Fire Info. for Resource Management System (FIRMS)
  - Fire activity at a fine resolution with confidence
  - Sum “expected fire” in each pixel in (Singh et al 2021)

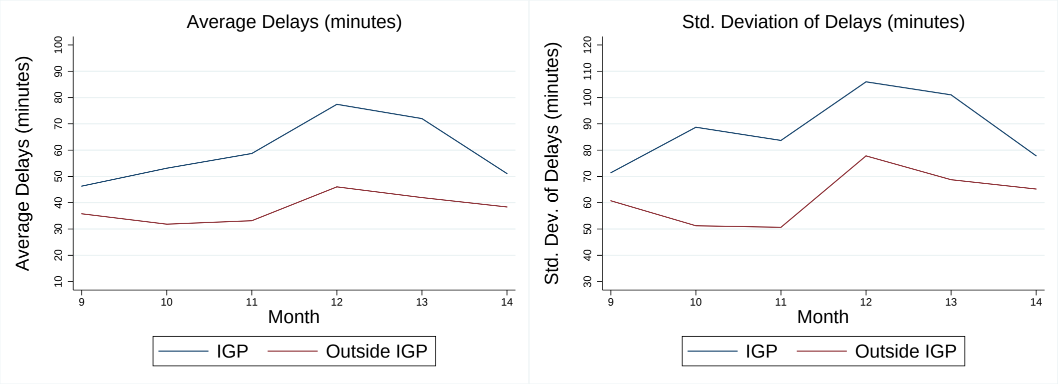
Trains Data Statistics

Table: Data Summary

Variable	Value
Total Trains	2744
Total Trains Crossing IGP	1644
Total Trains Outside IGP	1100
Total Routes	35911
Total Districts	484
Fraction of GDP in Data (Nightlights)	0.864
Average Districts on a Train Route	22
Observations	24373036

Raw Data

- Rise in delays and uncertainty in rail operations
  - Does not include train cancellations
- Delays propagate on the network (Gaurav and Srivastava '18)



Air Pollution Affects Indian Trains

Specification:

$$PropDelays_{i,od,t} = \beta_0 + \beta_1 log(pm2.5_t) + \beta_2 Humidity_t + \beta_3 Humidity_t^2 + \delta_i + \gamma_{od} + \eta_{dow} + \tau_y + \epsilon_{i,od,t}$$

- $PropDelays_{i,od,t}$ - (Delays/Scheduled travel time) for train i on route od on day t
- $pm2.5_t$ - average PM2.5 in the IGP on day t
- $Humidity_t$ - average humidity in the IGP on day t
- δ_i, γ_{od}, η_{dow}, τ_y - train, route, day of week and year fixed effects resp.

Air pollution causes transport delays

Table: Effect of PM2.5 on Train Delays

	Prop. Delay (1)	Prop. Delays (2)	Prop. Delays (3)	Prop. Delays (4)	Prop. Delay (5)
log(PM 2.5)	**0.042*****	**0.051*****	**0.06*****	**0.03*****	**0.135*****
	[0.00]	**[0.001]**	**[0.001]**	**[0.001]**	**[0.002]**
Rel. Humidity		0.00***	-0.06***	0.01***	0.002***
		[0.00]	[0.001]	[0.00]	[0.00]
$Rel. Humidity^2$			0.00***		
			[0.00]		
N	19745693	19699501	19699501	8102401	8617493
Fixed Effects	X	X	X	X	X
Control Mean	0.15	0.15	0.15	0.10	0.28
Outside IGP				Yes	
Cross IGP					Yes

Notes: Sample runs from October through January. Each column represents a fixed effects regression with train fixed effects, route fixed effects, year fixed effects and day of week fixed effects. Outside IGP means sample is restricted to trains that do not cross IGP at all where cross IGP means the sample was restricted to trains that cross IGP at least once. Standard errors are clustered at the origin district and destination district level.

Interpretation · DiD Specification · IV Specification

Air pollution causes transport uncertainty

Table: First Stage: Effect of PM2.5 on Uncertainty in Train Delays

	(1)	(2)	(3)	(4)	(5)
	Std. Dev. Delays	Std. Dev. Delays	Std. Dev. Delays	Std. Dev. Delays	Std. Dev. Delays
log(PM 2.5)	**5.55*****	**5.34*****	**3.87*****	**1.14***	**8.1*****
	[0.39]	**[0.3871]**	**[0.3876]**	**[0.58]**	**[0.53]**
Rel. Humidity		0.20***	-38.24***	-0.04	0.41***
		[0.02]	[0.67]	[0.03]	[0.03]
$Rel.Humidity^2$			0.26***		
			[0.01]		
N	273529	273529	273529	123356	149142
Fixed Effects	X	X	X	X	X
Control Mean	55	55	55	41	71
Outside IGP				Yes	
Cross IGP					Yes

Notes: Sample runs from October through February. Each column represents a fixed effects regression with route fixed effects and year fixed effects. Outside IGP means sample is restricted to trains that do not cross IGP at all. Standard errors are clustered at the origin district and destination district level.

Roadmap

Data II

- Center for Monitoring Indian Economy (CMIE)
  - Monthly panel of Indian households
  - Information on
    - Income
    - Expenditure (with categories)
    - Occupation
- Delays aggregated at a weekly level
  - Average arrival delays at district level
  - Standard Deviation of arrival delays at district level
- Instruments State Ranks IV First Stage
  - PM 2.5; Relative Humidity

Correlation b/w Train delays and time taken to travel

Table: Correlation between train delays and time taken to travel

	Delay (1)	Std. Dev. Delay (2)
Travel Time (10 kms)	0.1987*** [0.0039]	0.1553*** [0.0028]
N	1545511	1545379
Fixed Effects	X	X

Notes: Sample runs from September through February. Each column represents a fixed effects regression with year fixed effects and household fixed effects.

Identification

Specification:

$$log(y_{imt}) = \beta_0 + \beta_1 log(delay_{d,w,mt}) + \Gamma X_{it} + \gamma_i + \delta_m + \sigma_t + \epsilon_{imt}$$

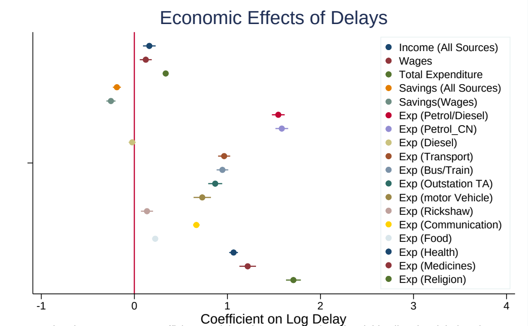
Assumptions:

- Air pollution in IGP affects economic outcomes outside IGP only through transport disruptions
- Average train delays captures overall transport disruption for workers

Threats:

- Labor market integration beyond districts
- Goods market channel
  - For ex. complementarities in labor productivity inside and outside IGP

Economic Effects of Delays



Each point represents a coefficient on delays from a regression of variables listed as labels using PM 2.5 and relative humidity as instruments for Delays and standard deviations in delays.
Sample is restricted to 18-65 year old rural men and women.
Year, month and District FE included along with individual level controls.

Rank 2

Conclusion

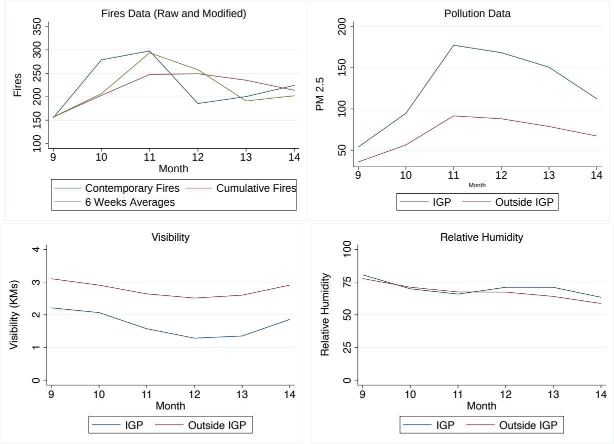
- Indian transport infrastructure faces substantial disruption due to agricultural crop burning in Norther India
  - Spike in particulate pollution leads to substantial delays and uncertainty in Indian Railway operations
- Delays in transport are associated with decline in household savings for rural households
  - This is due to increased expenses from substituting towards private modes of transport
- Next Steps
  - Firm level analysis using India's firm level panel dataset
  - Incorporate data from road transport
  - Estimate aggregate effects of transport disruptions in India using a model

Thank You!

uk2154@columbia.edu

APPENDIX

Environment Data Statistics



Back to Setting | Back to Data

Brooks and Donovan 2020 (Discussion)

- Bridges improve connectivity between rural and urban areas
  - Faster transport and more certainty during floods
- Differences:
  - Construction of bridges is about quantity of infrastructure
    - I study effect of quality
  - BD 2020 study the effect of uncertainty using a structural model
    - I use direct data on uncertainty
  - Construction of bridges is a much larger shock
    - Intuitively effects of quantity >> effects of quality
    - But....Cost of quantity >> cost of quality

Back

Interpretation of magnitudes

- Air pollution in IGP leads to a 3%-13.5% increase in delays (lower bounds)
  - Does not account for cancellations
- Some figures on freight transport (Gupta et al. 2020)
  - Average haulage period is 5.32 days
  - Freight trains are slow: travel 204 kms in 1 day (2017)
  - Low priority on the network
- Hummels and Schaur (2013): Using US data
  - Each day in transit $\approx$ 0.6-2.1% Ad-Valorem tariff (US)

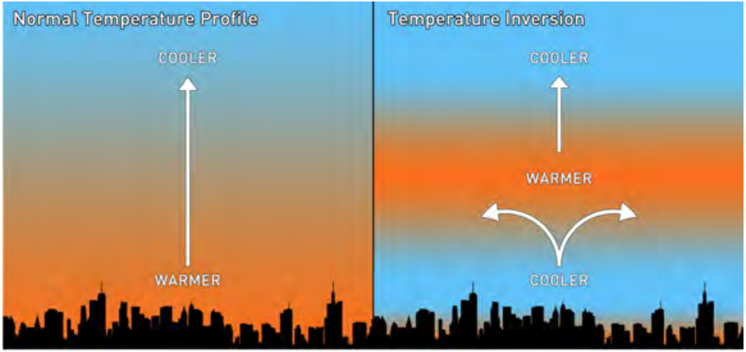
Back

Visibility

- Ansari et al. (2020)
  - PM2.5 from crop residue burning combined with local temperature, weaker winds cause stagnant air pollution and affect visibility
- Banerjee et al. (2017)
  - Visibility inversely associated with PM2.5 and PM10 in post-monsoon months
- Dey and Singh (2012)
  - Visibility most sensitive to PM pollution and RH and not dust

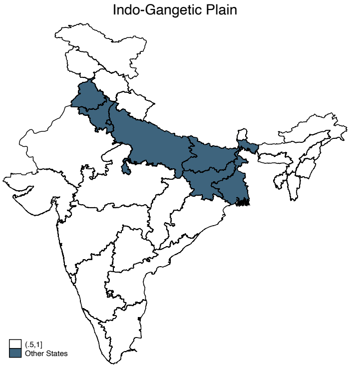
Back

Winter Inversion



Back

Indo-Gangetic Plain (IGP)

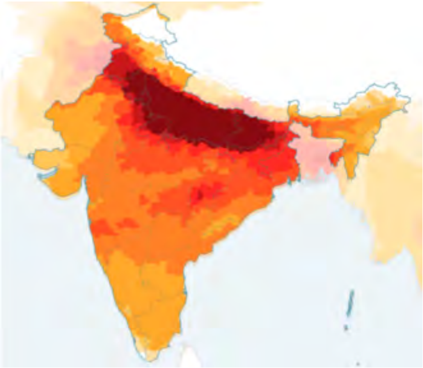


- IGP is home to 450 million people (6% global population)
- Mostly rural

Back

Air pollution map

- Railway network heavily concentrated in this region (Open Railway Map)



EPIC, UChicago

Back

Diff-in-Diff Specification

Table: Effect of PM2.5 on Train Delays (in minutes)

	Delay (1)	Delay (2)	Delay (3)	Delay (4)
Outside September			-0.80***	6.91***
			[0.15]	[0.16]
PM 2.5	-0.02***	0.00	-0.13***	0.00*
	[0.00]	[0.004]	[0.00]	[0.00]
Outside September $\times PM2.5$			**0.24*****	**0.11*****
			[0.003]	**[0.00]**
Rel. Humidity	0.97***	6.00***	0.59***	-4.84***
	[0.00]	[0.17]	[0.00]	[0.04]
$Rel.Humidity^2$		-0.03***		0.04***
		[0.00]		[0.00]
N	4627105	4627105	24326793	24326793
September	Yes	Yes		
Fixed Effects	X	X	X	X

Notes: Outside September is a dummy that takes a value 1 when month is not September. Each column represents a fixed effects regression with train fixed effects, route fixed effects, year fixed effects and day of week fixed effects. Standard errors are clustered at the origin district and destination district level.

Back

IV Specification

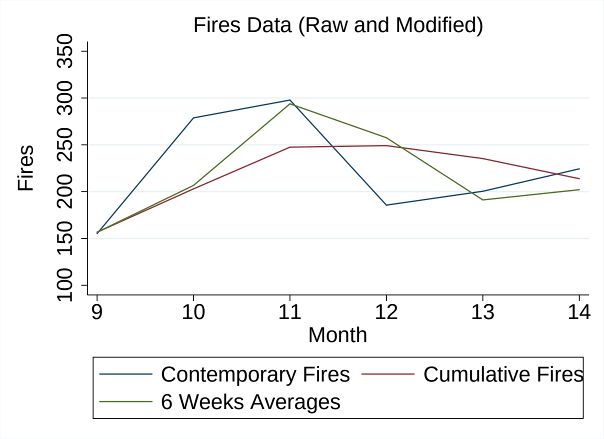
Table: First Stage (IV): Effect of PM2.5 on Train Delays

	1st Stage (1) PM 2.5	2nd Stage (2) Delay	1st Stage (3) PM 2.5	2nd Stage (4) Delay	1st Stage (5) PM 2.5	2nd Stage (6) Delay
Current Fires	-0.09***					
	[0.00]					
Rel. Humidity	-3.30***	15.72***	-0.99***	0.76***	-2.84***	1.23***
	[0.00]	[0.47]	[0.01]	[0.01]	[0.00]	[0.01]
PM2.5		4.8***		**0.25*** **		**0.40*** **
		[0.14]		**[0.00]**		**[0.00]**
Cumulative Fires			15.80***			
			[0.03]			
Fires (6 week Av.)					6.64***	
					[0.001]	
N	24326793	24326793	23652744	23654718	23654718	23654718
F-Stat	1433.870		1.79e+07		1.72e+06	
Fixed Effects	X	X	X	X	X	X

Notes: Each column represents a fixed effects regression with train fixed effects, route fixed effects, year fixed effects, month fixed effects and day of week fixed effects. Standard errors are clustered at the origin district and destination district level.

Fires Data Back

Fires


Fires Data (Raw and Modified)
Fires
100
150
200
250
300
350
9
10
11
12
13
14
Month
Contemporary Fires
Cumulative Fires
6 Weeks Averages

Back

Trade Costs (Interpretation and Assumptions)

- Every winter, transport network experiences a shocks
  - Proxy this shock by delays (of fastest trains)
    - Plan to support this assumption using driving times

Trade Costs (Interpretation and Assumptions)

- Every winter, transport network experiences a shocks
  - Proxy this shock by delays (of fastest trains)
    - Plan to support this assumption using driving times
  - Transport costs go up while presumably loading and unloading costs remain unchanged
    - Travel time commonly used to proxy for cost of transport/trade
    - Assumption: holds for short term shocks too
    - Anecdotes: extra payments to drivers for extra days taken to transport goods

Trade Costs (Interpretation and Assumptions)

- Every winter, transport network experiences a shocks
  - ◇ Proxy this shock by delays (of fastest trains)
    - ◇ Plan to support this assumption using driving times
  - ◇ Transport costs go up while presumably loading and unloading costs remain unchanged
    - ◇ Travel time commonly used to proxy for cost of transport/trade
    - ◇ Assumption: holds for short term shocks too
    - ◇ Anecdotes: extra payments to drivers for extra days taken to transport goods
  - ◇ Assumption: These costs show up in market prices (model)
  - ◇ Limitation: Railway's share of freight: ~ 45%
- Model does not have missed work days
- Model does not account for inventories

First Stage For IV Regressions

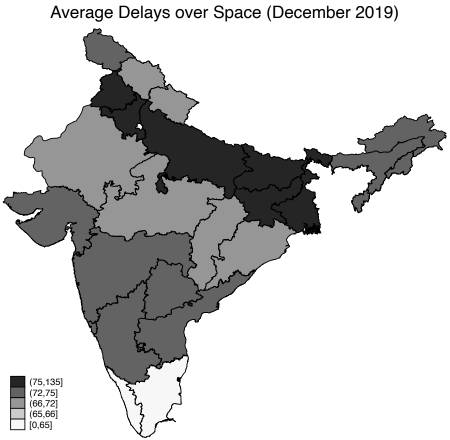
Table: IV First Stage: Polution and Humidity on delays and standard deviations

	log(Delay) (1)	log(Delay) (2)	log(Delay) (3)	log(Delay) (4)	log(Delay) (5)
Panel A: Full Sample					
log(Rel. Humidity)	0.39***	0.42***	0.17***	0.06*	0.44
	[0.05]	[0.06]	[0.05]	[0.03]	[0.38]
log(PM 2.5)	0.12***	0.06***	-0.00	0.01*	-0.06
	[0.01]	[0.01]	[0.01]	[0.01]	[0.05]
N	12082283	5917475	9096896	2524784	362555
Fixed Effects	X	X	X	X	X
	log(Sd Delay)	log(Sd Delay)	log(Sd Delay)	log(Sd Delay)	log(Sd Delay)
Panel A: Full Sample					
log(Rel. Humidity)	0.44***	0.33***	0.12***	0.17***	0.23
	[0.04]	[0.06]	[0.05]	[0.02]	[0.35]
log(PM 2.5)	0.13***	0.05***	0.00	0.04***	-0.15
	[0.01]	[0.01]	[0.01]	[0.01]	[0.09]
N	12054478	5864924	9078416	2524784	323588
Fixed Effects	X	X	X	X	X

Notes: Sample runs from October through February. Each column represents a fixed effects regression with year, month and household fixed effects.

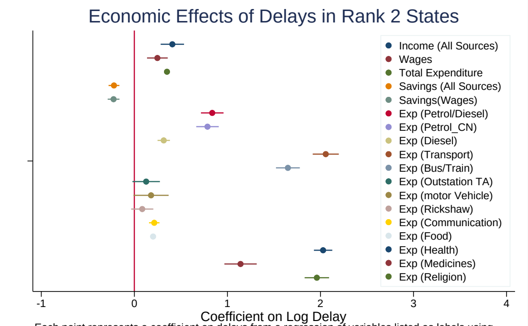
Back

Research Design



Back

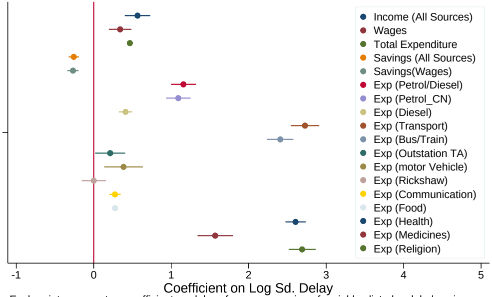
Economic Effects of Delays



Each point represents a coefficient on delays from a regression of variables listed as labels using PM 2.5 and relative humidity as instruments for Delays and standard deviations in delays.
Sample is restricted to 18-65 year old rural men and women.
Year, month and District FE included along with individual level controls.

Back

Economic Effects of Sd. Delays in Rank 2 States



Each point represents a coefficient on delays from a regression of variables listed as labels using PM 2.5 and relative humidity as instruments for Delays and standard deviations in delays.
Sample is restricted to 18-65 year old rural men and women.
Year, month and District FE included along with individual level controls.

Back

Determinants of speed-Variables by destination district

Table: Determinants of Speed of Fastest Indian Trains

	Log(speed) (1)	Log(speed) (2)	Log(speed) (3)	Log(speed) (4)
Ln(Distance)	0.16***	0.16***	0.18***	0.18***
	[0.01]	[0.01]	[0.01]	[0.01]
Ln(Planned Speed)	0.86***	0.85***	0.81***	0.81***
	[0.01]	[0.02]	[0.02]	[0.02]
Morning	0.01**	0.01**	0.00	0.00
	[0.00]	[0.00]	[0.00]	[0.00]
ln(Total Stops)	-0.07***	-0.08***	0.06***	0.10***
	[0.01]	[0.01]	[0.02]	[0.02]
ln(# Stops District)		0.02***	-0.01***	-0.01
		[0.01]	[0.00]	[0.00]
ln(# districts)			-0.18***	-0.18***
			[0.02]	[0.03]
IGP				-0.19***
				[0.04]
# Stops IGP				-0.00***
				[0.00]
N	43874112	43874112	43874112	43874112
District Fixed Effects	X	X	X	X
Month Fixed Effects	X	X	X	X

Notes: Standard errors are clustered at the destination state level.

Determinants SD. Speed-Variables by destination district

Table: Determinants of Variance in Speed of Fastest Indian Trains

	Log(Sd Speed) (1)	Log(Sd Speed) (2)	Log(Sd Speed) (3)	Log(Sd Speed) (4)
Ln(Distance)	-0.32***	-0.30***	-0.32***	-0.32***
	[0.03]	[0.03]	[0.03]	[0.03]
Ln(Planned Speed)	1.33***	1.27***	1.19***	1.19***
	[0.05]	[0.04]	[0.04]	[0.04]
ln(time)	0.33***	0.32***	0.42***	0.42***
	[0.04]	[0.04]	[0.04]	[0.04]
Morning	0.03***	0.03***	0.00	0.00
	[0.01]	[0.01]	[0.01]	[0.01]
ln(Total Stops)	0.21***	0.19***	0.70***	0.70***
	[0.02]	[0.02]	[0.03]	[0.03]
ln(# Stops District)		0.09***	-0.05**	-0.05**
		[0.02]	[0.02]	[0.02]
ln(# districts)			-0.68***	-0.68***
			[0.03]	[0.03]
IGP				0.29
				[0.19]
# Stops IGP				-0.00
				[0.00]
N	2009168	2009168	2009168	2009168
Month Fixed Effects	X	X	X	X
District Fixed Effects	X	X	X	X

Notes: Standard errors are clustered at the destination state level.