Economic Development and the Spatial Distribution of Income in Cities

Peter Deffebach David Lagakos Yuhei Miyauchi Eiji Yamada

NBER Summer Institute, Urban Economics - July 26th, 2024

Cities in Developed vs. Developing World

- How are they different?
- Literature has focused mainly on city-level aggregate statistics
  - Developing-world cities are e.g. slower, denser on average
- Limited cross-country evidence on **internal city structure**

What We Do

- Build a new internationally-comparable dataset of internal city structure
  - Micro-data of travel surveys from 32 cities
  - Neighborhood income data from 140 cities
  - Natural amenities and infrastructure

What We Do

- Build a new internationally-comparable dataset of internal city structure
  - Micro-data of travel surveys from 32 cities
  - Neighborhood income data from 140 cities
  - Natural amenities and infrastructure
- Document in developing countries, poor households more likely to live
  - Further from city center
  - Close to natural amenities (hills, rivers)
  - Mostly opposite patterns in developed countries

What We Do

- Build a new internationally-comparable dataset of internal city structure
  - Micro-data of travel surveys from 32 cities
  - Neighborhood income data from 140 cities
  - Natural amenities and infrastructure
- Document in developing countries, poor households more likely to live
  - Further from city center
  - Close to natural amenities (hills, rivers)
  - Mostly opposite patterns in developed countries
- Use quantitative model to assess role of residential and transport infrastructure in explaining these patterns

Data Sources

- Microdata of household travel surveys
  - 32 cities in developing world from Japan International Cooperation Agency (JICA)
  - Income, employment, demographics
  - Travel diaries (time, duration, purpose, mode) at fine spatial resolution
- Aggregate commuting flows across census tracts in the U.S.
- Census tract income: Brazil, USA, France, Spain, and UK
- Natural amenities (hills, rivers, coast): Open Street Maps
- Residential infrastructure (e.g., electricity, sewage): Bangladesh, Sri Lanka, Brazil
- City centers and boundaries

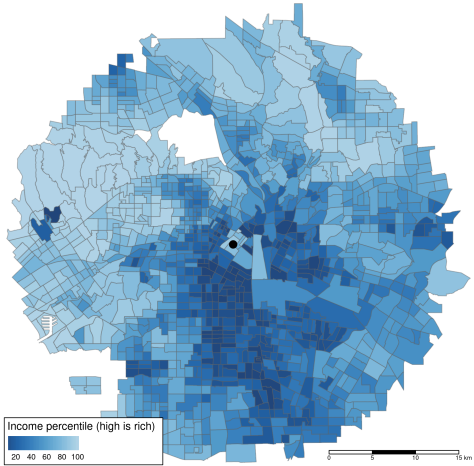
Additional details

Data Sources

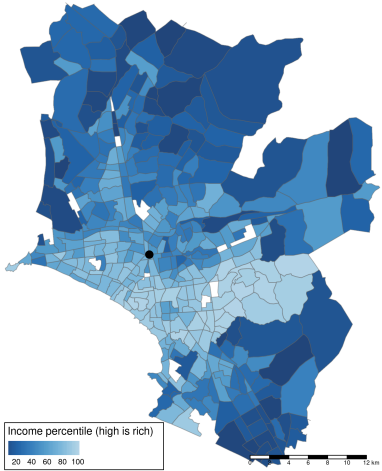
Data Source	Num. cities	Avg. km^2	Income	Res. Infrastructure	Commuting flows
JICA Surveys	32	4.9	×		×
USA ACS	55	4.0	×		
Spanish Tax Data	8	0.8	×		
UK Census	11	8.0	×		
French Tax Data	8	1.0	×		
Brazilian census	26	0.2	×	×	
USA LODES	55	4.0			×
Sri Lankan Census	1	1.2		×	
Bangladesh Census	1	1.0		×	

- 82 “developed” cities : USA, France, and Spain, UK
- 58 “less-developed” cities: Asia (19), Africa/Middle East (9), Latin America (29), Europe (1)
- 118,000 total neighborhoods, average 1.23 squared kilometers Map

Income by Neighborhood in Los Angeles and Lima

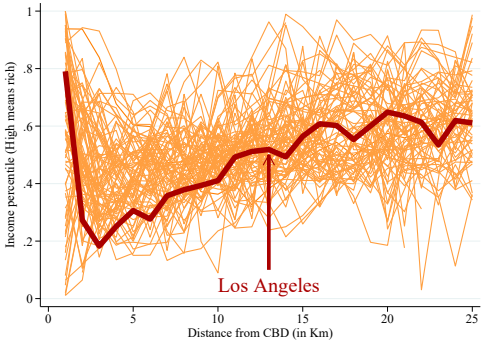


Los Angeles

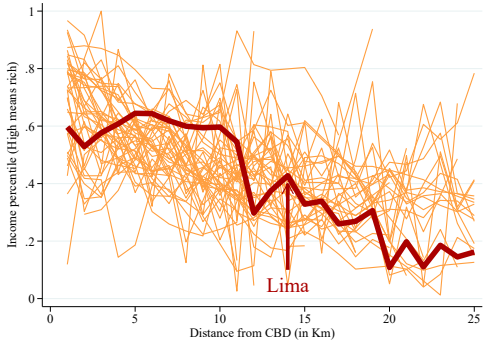


Lima

Neighborhood Income and Distance from City Center

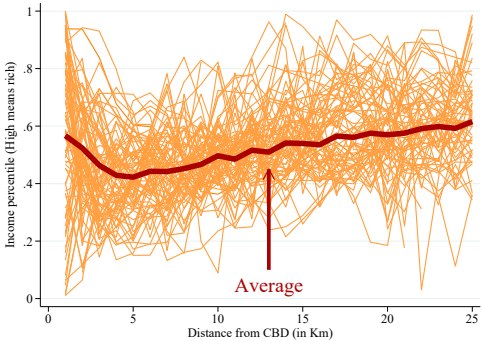


Cities in Developed Countries

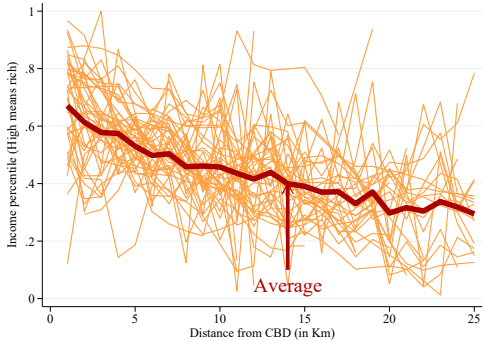


Cities in Less Developed Countries

Neighborhood Income and Distance from City Center

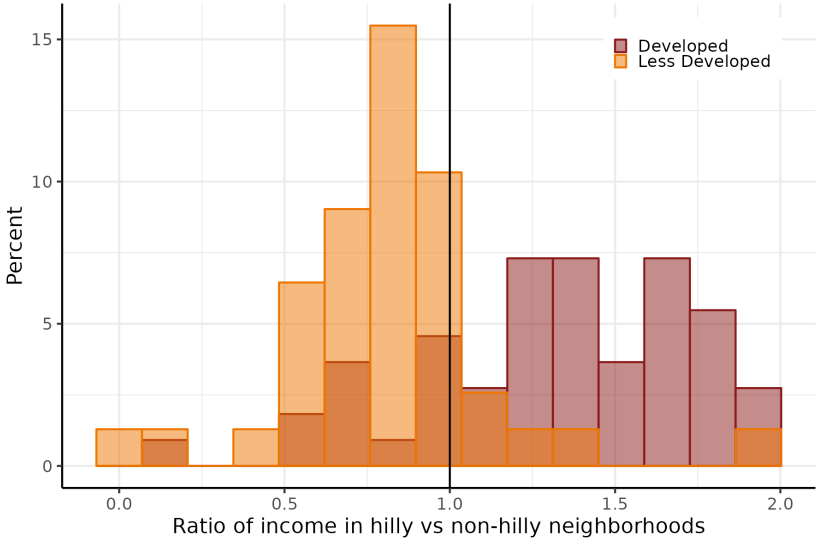


Cities in Developed Countries

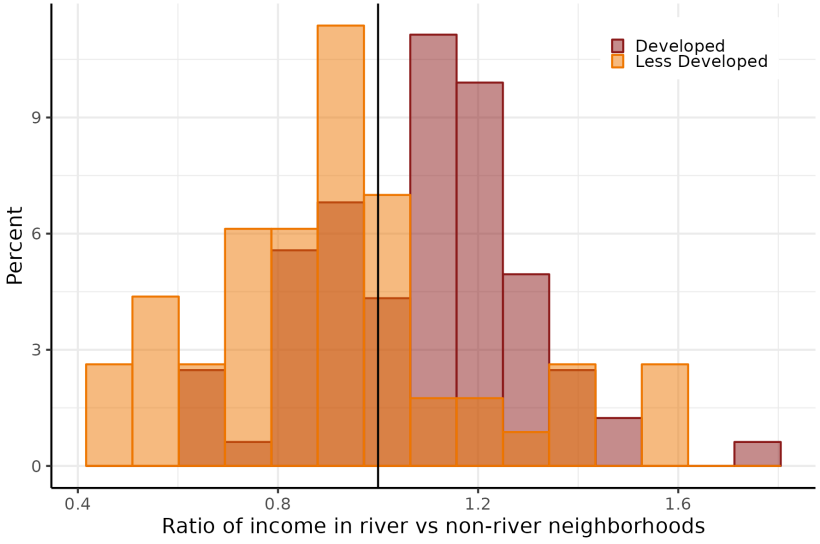


Cities in Less Developed Countries

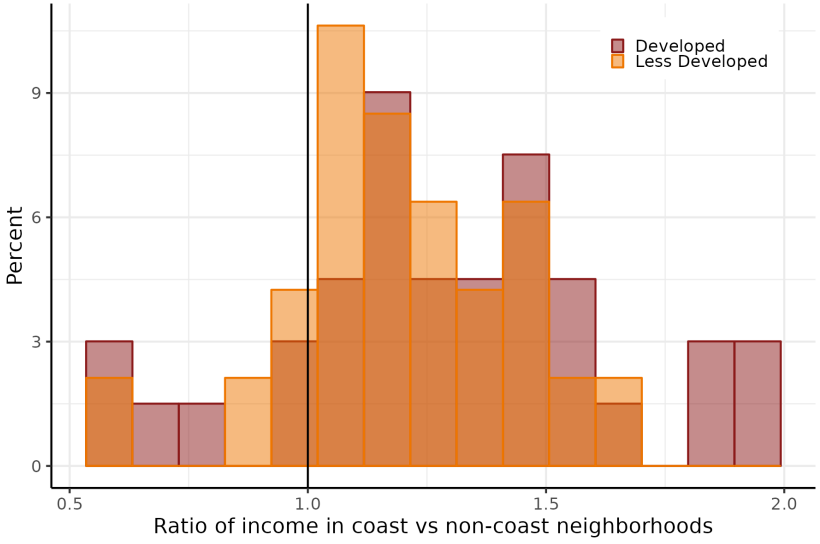
Relative Income of Hilly Neighborhoods



Relative Income of Neighborhoods Near Rivers



Relative Income of Neighborhoods Near Coasts



Regression Results: Distance, Hills, Rivers, Coasts

Dependent Variable:	Income percentile (high is rich)				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Developed × Log dist. from center	3.80***				3.72***
	(1.23)				(1.24)
Less Developed × Log dist. from center	-11.2***				-10.9***
	(1.23)				(1.21)
Developed × Hilly		13.5***			13.2***
		(3.32)			(3.18)
Less Developed × Hilly		-8.43***			-6.68***
		(1.76)			(1.87)
Developed × < 100m from river			3.86***		3.37***
			(1.14)		(1.10)
Less Developed × < 100m from river			-6.19***		-5.59***
			(1.79)		(1.57)
Developed × < 100m from coast				12.5***	12.2***
				(3.38)	(3.17)
Less Developed × < 100m from coast				6.99***	6.10**
				(2.46)	(2.41)
Difference in slopes: Developed vs Developing					
Log dist. from center	-15.0***				-14.7***
	(1.74)				(1.73)
Hilly		-21.9***			-19.9***
		(3.76)			(3.69)
< 100m from river			-10.1***		-8.95***
			(2.13)		(1.92)
< 100m from coast				-5.56	-6.10
				(4.18)	(3.98)
Weight	Cities equal	Cities equal	Cities equal	Cities equal	Cities equal

Weighting schemes | Controls | Samples | Income definitions | Distance definitions

Explaining These Patterns

- Hypothesis: Lack of **residential and transportation infrastructure** in locations further from city centers and with natural amenities in developing countries
- Assess using additional new evidence + quantitative urban model
- Quantify how much differential infrastructure provision explains these patterns

Model Environment

- Many heterogeneous neighborhoods $i \in N$ within a city
- Neighborhoods differ in:
  - Productivity, natural amenities, housing supply
  - Residential and transportation infrastructure
- Households with different earning potential ("type") $s \in \{L, H\}$ decide residences and workplaces

Households: Preferences

- Utility of individual ω of type s living in i and working in j:

$$U_{ij}^{s}(h,c) \times \epsilon_{ij\omega}^{s} = \left(\frac{h-\xi}{\alpha}\right)^{\alpha} \left(\frac{c}{1-\alpha}\right)^{1-\alpha} \left(\tau_{ij}^{s}\right)^{-1} B_{i}^{s} \epsilon_{ij\omega}^{s}$$

where

- h is consumption of housing
- ξ is "subsistence" housing need
- c is consumption of the final good
- τ_{ij}^{s} is the utility commuting cost
- B_{i}^{s} is the residential amenity
- $\epsilon_{ij\omega}^{s}$ is idiosyncratic preference shock

Households: Residence and Work Location Choice

- Assume $\epsilon^s_{ij\omega}$ follows i.i.d., Frechet distribution with shape parameter θ
- Choice probability of residence i and work j by type s is given by

$$\pi^s_{ij} = \frac{V(r_i, \tau^s_{ij}, w^s_j, B^s_i, \epsilon^s_{ij})^\theta}{\sum_{i',j'} V(r_{i'}, \tau^s_{i'j'}, w^s_{j'}, B^s_{i'}, \epsilon^s_{i'j'})^\theta}$$

  - V is indirect utility

Amenities, Commuting Costs, and Infrastructure

- Residential amenity

$$B_i^s = K_i^R \zeta_i g^s(\{I_i^s\}_s)$$

  - K_i^R is residential infrastructure (i.e. piped water)
  - ζ_i is a natural amenity (i.e. hills)
  - $g^s(\{I_i^s\}_s)$ captures endogenous amenities (i.e. restaurants)

- Commuting cost

$$\tau_{ij}^s = f^s(D_{ij}, K_{ij}^T)$$

  - D_{ij} is exogenous geographic friction (i.e. distance, hills)
  - K_{ij}^T is transportation infrastructure (i.e. roads, public transit)

Housing

- Inverse housing supply function

$$r_i = \psi H_i^{\kappa}$$

  - r_i is housing rent
  - H_i is housing stock
  - $1/\kappa$ is housing supply elasticity

- Housing markets clear when

$$H_i = \sum_{s \in \{L,H\}} \sum_j l_{ij}^s h_{ij}^s$$

Consumption Goods Production + Equilibrium

- Production function

$$Y_j = A_j^L l_j^L + A_j^H l_j^H$$

- Under perfect competition, wages are given by

$$w_j^s = A_j^s$$

- Equilibrium: Households make optimal consumption and location decisions, markets clear

Model Mechanisms: Infrastructure and Neighborhood Income

- Higher level of residential infrastructure i attracts higher-income residents:

$$\uparrow K_i^R \Rightarrow \uparrow r_i \Rightarrow \uparrow I_i^H / I_i^L$$

- Higher level of transportation infrastructure from i to various destinations j:
  - attacts higher-income residents:

$$\uparrow \{K_{ij}^T\}_j \Rightarrow \uparrow r_i \Rightarrow \uparrow I_i^H / I_i^L$$

  - enables commuting to high-wage workplaces

$$\uparrow \{K_{ij}^T\}_j \Rightarrow \uparrow \sum\nolimits_j w_j^s \pi_{ij}^s \quad \text{for each } s \in \{L, H\}$$

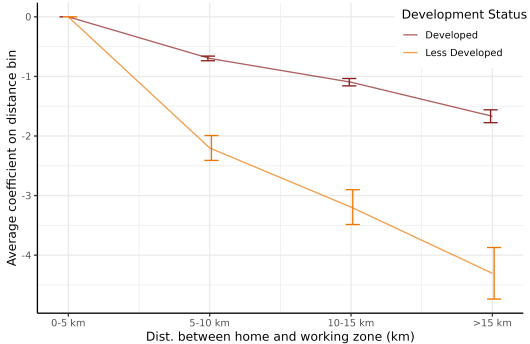
Evidence on Residential Infrastructure in Developing Countries

Dependent Variables:	Exclusive toilet connected to sewage lines	Exclusive electricity access	Tap water	
Model:	(1)	(2)	(3)	(4)
Variables				
Log dist. from center	-17.8***	-4.23***	-15.7***	-18.3***
	(2.32)	(0.781)	(3.31)	(1.97)
Hilly	-4.72**	-5.50***		
	(1.73)	(0.689)		
$<$ 100m from river	-7.87***	-5.18***	-26.3***	-0.299
	(2.64)	(0.987)	(4.96)	(3.07)
$<$ 100m from coast	6.94**	-0.006		5.63
	(3.08)	(0.650)		(4.51)
Weight	Cities equal	Cities equal	Cities equal	Cities equal
Subset	Brazil	Brazil	Dhaka	Colombo
Dependent variable mean	78.1	87.0	69.6	61.2

- In progress: Measuring residential infrastructure with **computer vision**

Higher Commuting Frictions in Developing Countries

$$\log \mathbb{E}[\pi_{ij}] = \sum_{s} \delta_s \mathbf{1}[D_{ij} = s] + \eta_i + \xi_j$$



- Suggest that transportation cost is higher in less developed cities
- Ongoing: Spatial variation in commuting patterns and modes

Additional Channels

- We (plan to) test and quantify:
  - Income-specific commuting frictions τ_{ij}^s (e.g., different transportation modes)
    Gravity by Income
  - Gap in spatial distribution of skill premium w_j^H / w_j^L
  - Income-specific endogenous amenities (e.g. travel to restaurants) $g^s(\{l_i^s\}_s)$
- Alternative un-modeled mechanisms:
  - Transition dynamics during city expansion. Poor migrants settle on outskirts of city (Gonzalez-Navarro and Udurraga, 2023)
  - Historical shocks, housing policy, race

Quantitative Analysis

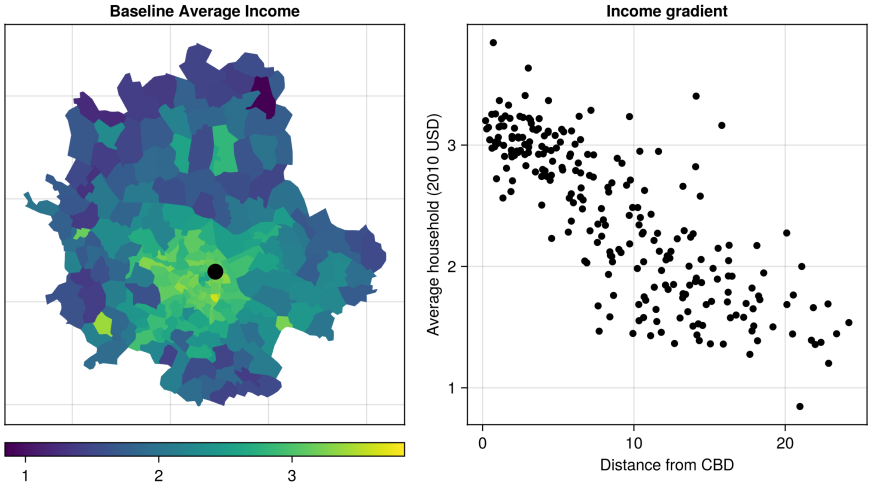
Q: How much can we explain the differences in developed vs less-developed city structure by infrastructure differences?

1. Calibrate model to each less-developed city
2. Change $\{K_i^R\}$ and $\{K_{ij}^T\}$ to the level and patterns of developed cities. Infer:
   - $\{K_{ij}^T\}$ from commuting flows (and mode choice)
   - $\{K_i^R\}$ from residential location choice (assume $K_i^R = K^R$ in developed cities)

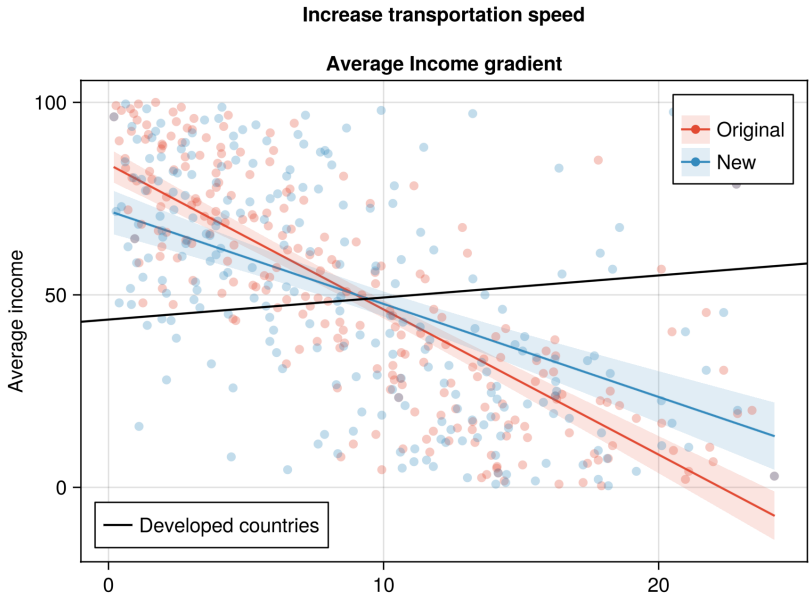
Today: Proof of concept using Hanoi, Vietnam calibration details

- Set $\{\hat{K}_{ij}^T\} \approx$ observed differences in commuting gravities

Hanoi: Average Income by Neighborhood and Income-Distant Gradient



Changes in Average Incomes from Transportation Infrastructure Provision



Conclusion

- We compile internationally comparable data on city structure to document:
  - In developing countries, poorer households live further from city centers, closer to hills, rivers, coasts
  - Mostly opposite patterns in developed country cities
- Development process turns areas with natural amenities from undesirable to desirable areas
- Work in progress: quantifying role of residential and transportation infrastructure

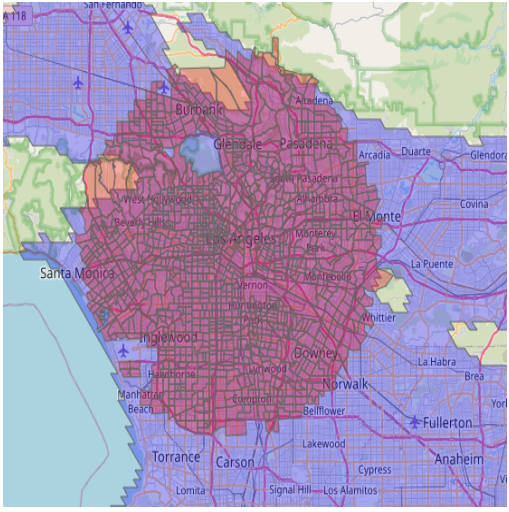
Additional details on Data

Back

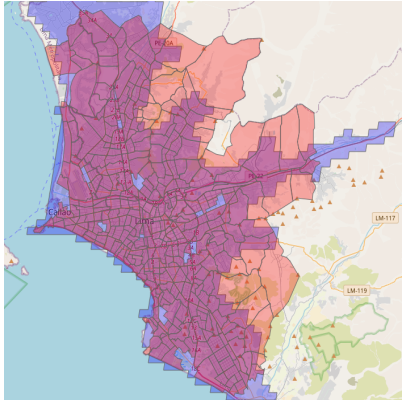
- What is a city?
  - World Settlement Footprint's "Built Up Areas" dataset (Florczyk et al., 2019)
  - Within 25km of city center
- What is a city center?
  - Get 4 city centers candidates, each with their own flaws
    - Weighted center nightlights from VIIRS
    - Weighted center of global population distribution from Landscan
    - Weighted center of road density network from Open Street Maps
    - City center according to Open Street Maps
  - Choose "most central" candidate (closest to all others)
- What is a hill?
  - Global elevation raster from Amazon Web Services (AWS) Terrain Tiles. 30m by 3m resolution
  - Average slope greater than 5 degrees (6% of USA zones hilly)

Neighborhoods (red) and Built up Area (blue) in Los Angeles and Lima

Back



Los Angeles

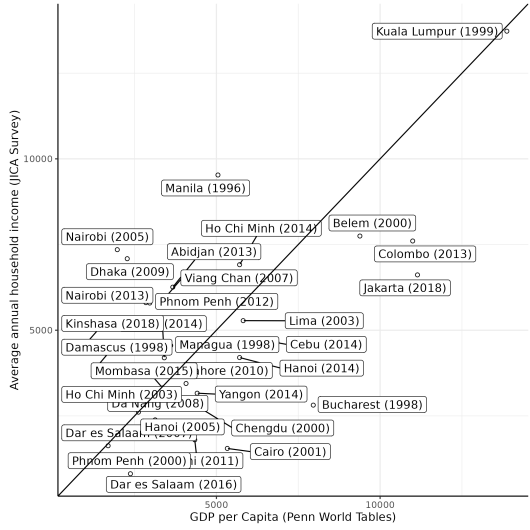


Lima

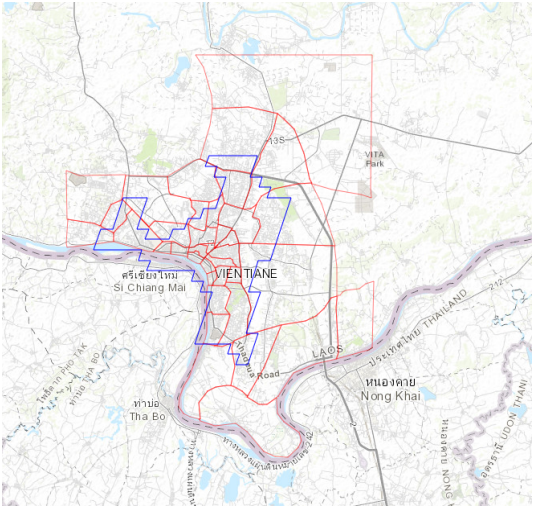
Location of Cities in Data

Back

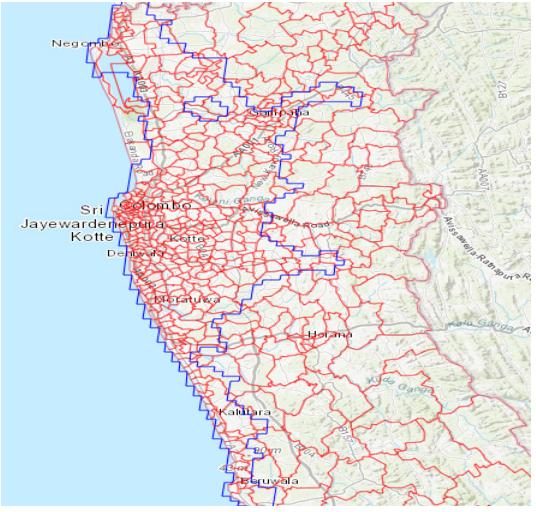
Comparison of Income: JICA Survey Vs. GDP Per Capita



Neighborhoods (red) and Built up Area (blue) in Vianghan and Colombo

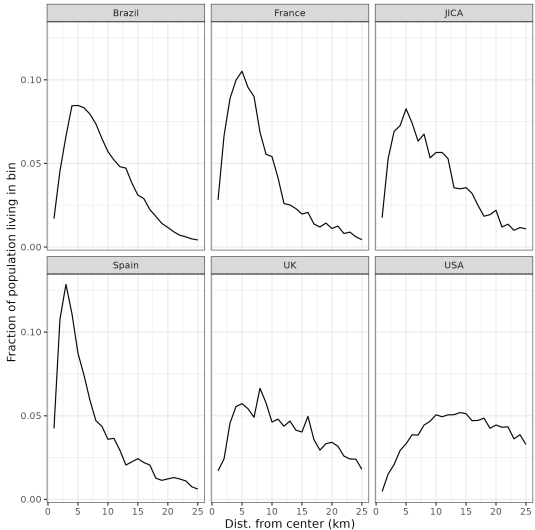


Viangchang, Laos



Colombo, Sri Lanka

Population distribution with distance from center



Neighborhood-level regressions: Inclusion of controls

Back

Dependent Variable:	Income percentile (high is rich)	
Model:	(1)	(2)
Variables		
Developed × Log dist. from center	4.12***	2.39*
	(1.30)	(1.24)
Less Developed × Log dist. from center	-10.6***	-10.4***
	(1.31)	(1.13)
Developed × Hilly	10.1***	12.1***
	(2.90)	(3.24)
Less Developed × Hilly	-6.75***	-8.15***
	(1.90)	(1.37)
Developed × < 100m from river	3.43***	1.37
	(1.00)	(1.14)
Less Developed × < 100m from river	-4.17***	-4.53***
	(1.55)	(1.44)
Developed × < 100m from coast	9.75***	11.4***
	(2.63)	(3.32)
Less Developed × < 100m from coast	5.13**	5.37**
	(2.37)	(2.19)
Difference in slopes: Developed vs Developing		
Log dist. from center	-14.7***	-12.8***
	(1.84)	(1.68)
Hilly	-16.8***	-20.3***
	(3.47)	(3.52)
< 100m from river	-7.59***	-5.90***
	(1.85)	(1.84)
< 100m from coast	-4.62	-5.98
	(3.54)	(3.98)
Weight	Cities equal	Cities equal

Neighborhood-level regressions: Alternative weighting schemes

Back

Dependent Variable:	Income percentile (high is rich)		
Model:	(1)	(2)	(3)
Variables			
Developed × Log dist. from center	4.36***	-0.313	0.418
	(1.33)	(1.35)	(1.35)
Less Developed × Log dist. from center	-11.8***	-7.08***	-8.58***
	(1.28)	(1.31)	(1.14)
Developed × Hilly	13.2***	5.31	5.39
	(3.17)	(4.12)	(3.97)
Less Developed × Hilly	-7.15***	-4.12	-6.86*
	(2.13)	(3.32)	(3.83)
Developed × < 100m from river	3.49***	3.96***	3.89**
	(1.10)	(1.46)	(1.56)
Less Developed × < 100m from river	-3.72**	-5.72***	-3.68*
	(1.48)	(1.99)	(1.88)
Developed × < 100m from coast	11.6***	8.78**	8.17**
	(2.97)	(4.22)	(3.88)
Less Developed × < 100m from coast	7.02***	3.57	6.05*
	(2.42)	(3.17)	(3.33)
Difference in slopes: Developed vs Developing			
Log dist. from center	-16.2***	-6.77***	-9.00***
	(1.85)	(1.88)	(1.76)
Hilly	-20.4***	-9.43*	-12.2**
	(3.82)	(5.30)	(5.51)
< 100m from river	-7.22***	-9.67***	-7.57***
	(1.84)	(2.46)	(2.44)
< 100m from coast	-4.59	-5.21	-2.11
	(3.83)	(5.28)	(5.12)
Weight	Cities equal, Nbhd pop	Countries equal	Countries equal, Nbhd pop

Neighborhood-level regressions: Alternative samples

Back

Dependent Variable:	Income percentile (high is rich)			
Model:	(1)	(2)	(3)	(4)
Variables				
Developed × Log dist. from center	0.065	2.22*	1.39	3.72***
	(1.24)	(1.23)	(1.27)	(1.24)
Less Developed × Log dist. from center	-9.71***	-10.5***	-10.2***	-8.26***
	(1.30)	(1.26)	(1.12)	(1.71)
Developed × Hilly	9.80***	11.6***	9.62**	13.2***
	(3.57)	(3.35)	(3.83)	(3.18)
Less Developed × Hilly	-6.68***	-6.68***	-8.51***	0.264
	(1.91)	(1.91)	(1.35)	(4.40)
Developed × < 100m from river	3.37**	3.65***	3.02**	3.37***
	(1.31)	(1.17)	(1.25)	(1.10)
Less Developed × < 100m from river	-6.05***	-5.78***	-6.22***	-5.41***
	(1.63)	(1.62)	(1.50)	(1.96)
Developed × < 100m from coast	12.9***	12.3***	11.6**	12.2***
	(3.57)	(3.21)	(4.45)	(3.18)
Less Developed × < 100m from coast	6.43***	6.09**	6.35**	5.24
	(2.32)	(2.43)	(2.77)	(3.26)
Difference in slopes: Developed vs Developing				
Log dist. from center	-9.78***	-12.7***	-11.5***	-12.0***
	(1.80)	(1.76)	(1.69)	(2.12)
Hilly	-16.5***	-18.3***	-18.1***	-13.0**
	(4.05)	(3.85)	(4.06)	(5.43)
< 100m from river	-9.42***	-9.43***	-9.25***	-8.78***
	(2.09)	(2.00)	(1.95)	(2.25)
< 100m from coast	-6.43	-6.23	-5.26	-6.96
	(4.26)	(4.03)	(5.24)	(4.55)
Weight	Cities equal	Cities equal	Cities equal	Cities equal
Subset	¡15km from CBD	¡20km from CBD	Only fully urban areas	No Brazil
Fixed-effects				
City	Yes	Yes	Yes	Yes
Fit statistics				
Observations	84,893	104,964	111,660	40,442
R^2	0.065	0.054	0.055	0.037
Within R^2	0.043	0.050	0.042	0.037

Neighborhood-level regressions: Alternative income measures

Back

Dependent Variables:	Log nbhd average income	Ratio to mean income	Income z-score
Model:	(1)	(2)	(3)
Variables			
Developed × Log dist. from center	0.055***	0.022	0.050
	(0.019)	(0.019)	(0.040)
Less Developed × Log dist. from center	-0.248***	-0.304***	-0.388***
	(0.031)	(0.038)	(0.041)
Developed × Hilly	0.214***	0.250***	0.538***
	(0.047)	(0.056)	(0.120)
Less Developed × Hilly	-0.178***	-0.133***	-0.093*
	(0.044)	(0.048)	(0.054)
Developed × < 100m from river	0.056***	0.060***	0.126***
	(0.015)	(0.016)	(0.036)
Less Developed × < 100m from river	-0.092***	-0.079**	-0.154***
	(0.026)	(0.033)	(0.053)
Developed × < 100m from coast	0.199***	0.238***	0.471***
	(0.048)	(0.059)	(0.116)
Less Developed × < 100m from coast	0.207***	0.375***	0.424***
	(0.055)	(0.097)	(0.098)
Difference in slopes: Developed vs Developing			
Log dist. from center	-0.304***	-0.326***	-0.438***
	(0.0362)	(0.0423)	(0.0573)
Hilly	-0.391***	-0.383***	-0.631***
	(0.0647)	(0.0737)	(0.131)
< 100m from river	-0.147***	-0.139***	-0.280***
	(0.0307)	(0.0360)	(0.0642)
< 100m from coast	0.00854	0.137	-0.0470
	(0.0724)	(0.113)	(0.151)
Weight	Cities equal	Cities equal	Cities equal

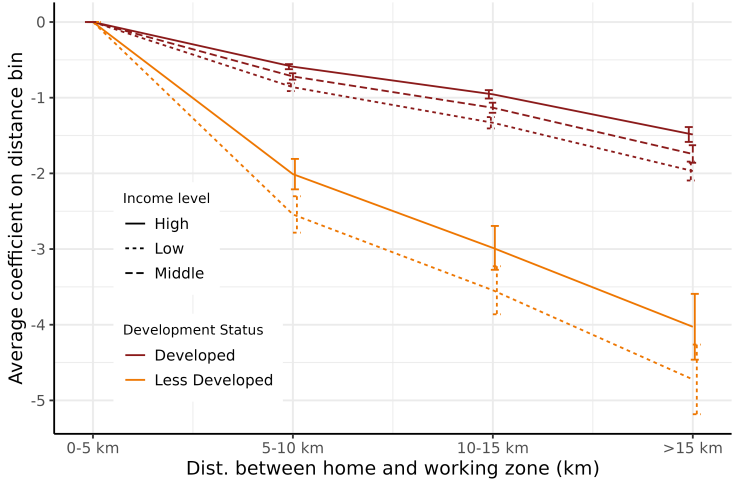
Neighborhood-level regressions: Alternative distance measures

Back

Dependent Variable:	Income percentile (high is rich)		
Model:	(1)	(2)	(3)
Variables			
Developed × Dist. to center, neighborhood percentile rank	0.140***		
	(0.031)		
Less Developed × Dist. to center, neighborhood percentile rank	-0.308***		
	(0.033)		
Developed × Dist. from center, population percentile rank		0.143***	
		(0.031)	
Less Developed × Dist. from center, population percentile rank		-0.315***	
		(0.031)	
Developed × Dist. from center (km)			0.728***
			(0.146)
Less Developed × Dist. from center (km)			-1.78***
			(0.195)
Difference in slopes: Developed vs Developing			
Dist. to center, neighborhood percentile rank	-0.448***		
	(0.0447)		
Dist. from center, population percentile rank		-0.458***	
		(0.0433)	
Dist. from center (km)			-2.51***
			(0.244)
Weight	Cities equal	Cities equal	Cities equal

Gravity coefficients by income

Back



Calibrated Parameters

- From data
  - Skill groups populations μ_L, μ_H : 67% "low-skill", 33 % "High skill" based on education levels
  - Wages at each location w_j^s
    - Average household income by work location and skill group
    - On average 2000 low-wkill, 3,400 high-skill (2010 USD)
  - Population flows π_{ij}^s from travel survey
  - No rent data yet. Assume $r_i = 1$ for all home locations
- Parametric assumptions
  - Housing share α and subsistence housing requirement ξ such that
    - Match average housing expenditure share
    - Match reasonable ratio of housing expenditure for skill groups. 40 % H, 45 % L.
    - Chose $\alpha = 0.25$, $\xi = 500.0$.
    - Model's housing expenditure shares: 44% for L, 36% for H
  - Choose inverse housing elasticity $\kappa = 1.0$. Construction not sensitive to prices.

Simulating Improvements in Transportation Infrastructure

- We parametrize the changes in commuting cost as:

$$\hat{\tau}_{ij}^{s} = \exp(-D_{ij}\Delta)$$

  where Δ captures the improvement of commuting technology in terms of semi-elasticity in travel distance, which we set $\Delta = 0.042$

- Amounts to 80% reduction in the commuting cost (in the unit of indirect utility) for the longest possible trip within the city (similar to the observed gap between developed and developing cities)

[ ]A.J. Florczyk, M. Melchiorri, C. Corbane, M. Schiavina, M. Maffenini, M. Pesaresi, P. Politis, S. Sabo, S. Freire, D. Ehrlich, T. Kemper, P. Tommasi, D. Airaghi, and L. Zanchetta. Description of the GHS Urban Centre Database 2015. *Publications Office of the European Union*, 2019.

[ ]Macro Gonzalez-Navarro and Raimundo Udurraga. Immigration and Slums. *Working paper*, 2023.