



Prime

Locations

Gabriel M. Ahlfeldt.

Thilo Albers.

Kristian Behrens.

Marcelo Torres

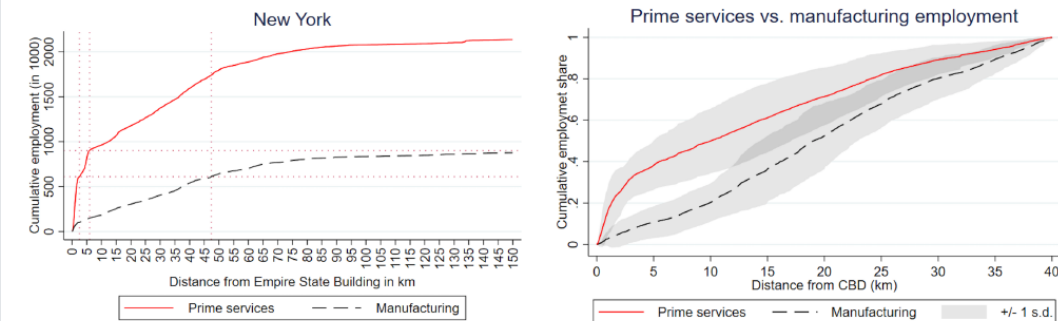
125698



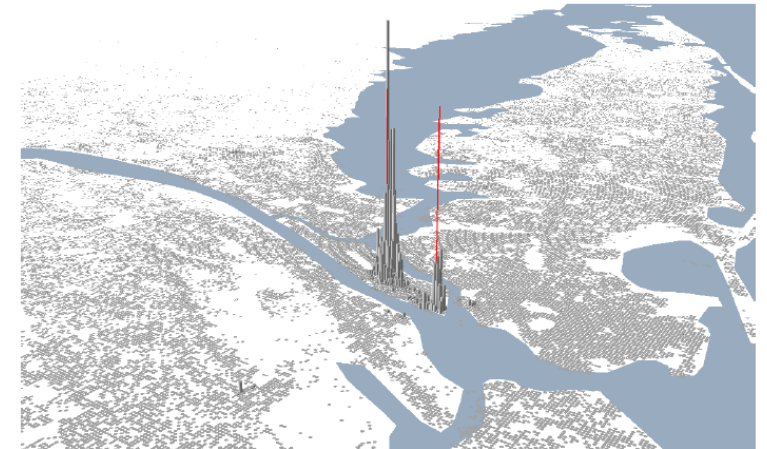
Motivación

- Hechos estilizados:
 1. Concentración geográfica de *prime services* (servicios intercambiables basados en conocimiento como finanzas, seguros, bienes raíces y servicios profesionales). **38% del empleo en prime services está aglomerado en alrededor de 0.7% del área desarrollable de las ciudades.**
 2. Las ciudades más grandes en 1900 tenían más probabilidades de desarrollar sistemas de transporte, que anclaban los CBDs.
 3. Ciudades que eran más grandes tenían ya distritos de CBD, al aparecer los efectos de aglomeración relevantes, la actividad se fue aglomerando a su alrededor.

Figure 2: Observed geographic concentration of prime services in cities



(a) Distribution of prime services within cities by distance from the CBD



(b) Geographic concentration of prime service employment in NY



Resumen de la investigación

- Investiga la concentración de empleo en servicios basados en conocimiento.
- Aportes:
 - + 1/ Técnica para detectar *prime locations*.
 - + 2/ Modelo de agentes que racionaliza los datos
 - + 3/ Evidencia causal de las relaciones del modelo con shocks de la historia de las ciudades y un instrumento innovador: *subway potential*.
- Efectos de aglomeración e infraestructura temprana explican por qué las ciudades más grandes en 1900 tuvieron menos cambios en las *prime locations* que las chicas durante el siglo XX. Ciudades grandes en 1900 y ciudades que desarrollaron un sistema de transporte tuvieron una geografía más estable de *prime services*.



Datos

Table 1: Data summary

Disaggregated data, address- or grid-level		Urban biographies
Economic data	Physical geography	Historical data
Core data	Grid level	Address-level
Central banks *	Land cover	First settlement *
Consultancy firms *	Elevation	First political institution *
Co-working spaces *	Water cover	City hall 1900 *
Insurance firms *		City hall 2000 *
Investment banks *	City level (built from grid)	
Law firms *	Developable land †	
Stock exchanges *	Irregular shape index †	City level
	Fragmentation index †	Colonial occupation in 1800 *
Employment weighting	Caloric potential of hinterland	Government type (1800-2000) *
Scott's Business Directories (P)		Manmade disasters *
NETS National Establishment		Natural disasters *
Time Series (P)		Population (1800-2000) *
		Rapid transit openings †
Validation data		Subway potential (1900) †
Geotagged photos		Market potential (1900) †
Geotagged Twitter tweets		
Co-working spaces *		Validation data
Grade-A office buildings - SNL (P)		1880 US census (10% sample) †
Starbucks establishments		Member list of International
Emporis tall buildings (P)		Chamber of Commerce (1922) *
US County Business Patterns		NPS historic buildings †

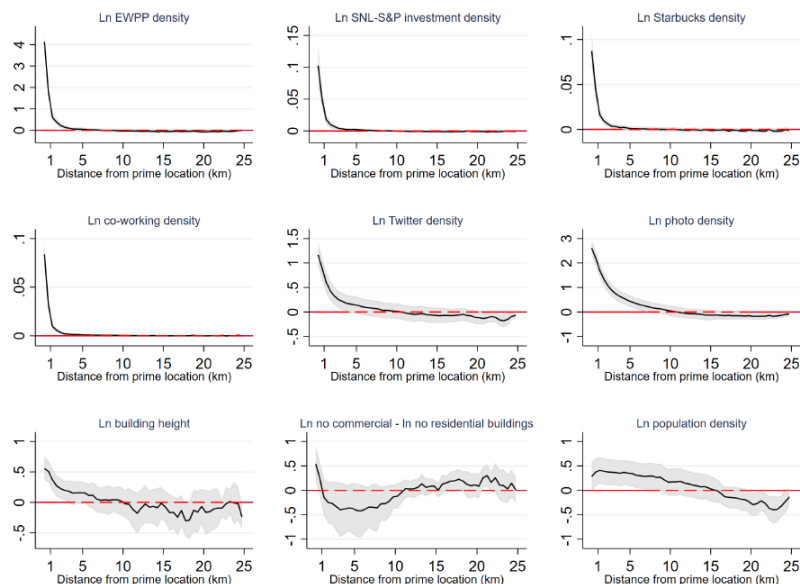
Notes: All economic data are at least at the address-level (many of them are geocoded at the rooftop level). * indicates primary data in the sense that we produced them from scratch. † indicates data for which we could rely on previous work and data, but that involved either substantial additional archival research (e.g., rapid transit openings) or substantial additional own calculations (e.g., developable land measure). (P) marks proprietary data. No sign means that the data were simply matched to our city dataset, such as caloric potential from [Galor and Özak \(2016\)](#). All sources are documented in the accompanying Global Cities dataset Appendix ([Ahlfeldt et al., 2020b](#)).

- 125 Ciudades globales.
- Dos fuentes:
 - + Datos micro de ubicación de firmas y big data (*scraped* de Google + proceso de asignación de pesos de trabajo *employment-weighted prime points* EWPPs)
 - + Datos históricos de las ciudades (población, ubicación del ayuntamiento, rentas, etc. -en 1900 y 2000-)



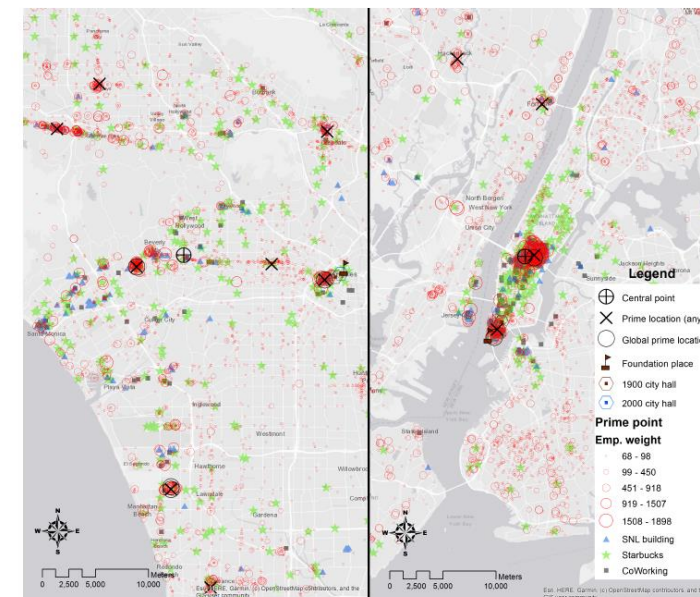
Datos

Figure 6: Prime location distance gradients



Notes: The gradients are averaged across up to 125 global cities, depending on data availability. Underlying each panel is a regression of an outcome measure against 500m-distance-bin effects and city effects at the city-grid cell level. Solid black lines are the distance-bin point estimates, gray-shaded areas are the respective 95% confidence intervals. For a description of the underlying data, see the Global Cities dataset Appendix (Ahlfeldt et al., 2020b).

Figure 5: Prime locations in Los Angeles (left panel) and New York (right panel).



Notes: This figure illustrates the output of our cluster algorithm that identifies focal points as the most clustered prime points and prime locations as the centers of clustered focal points. Own illustration based on OpenStreetMap background map. Complete city profiles, including similar maps for all 125 cities, are available in the Global Cities dataset appendix (Ahlfeldt et al., 2020b).

- Algoritmo que usa los EWPPs para obtener el CBD.
- + Calcula la suma de EWPPs en un radio de cada punto, mantiene los más densos (percentil 99), y busca los dominantes dentro de los que forman *clusters*.
- CBDs encontrados tienen sentido comparándolos con los gradientes de otras variables.

Datos

Table 2: Determinants of prime locations

	(1) FP	(2) CH1900	(3) Global PL	(4) PL	(4a) PL	(4b) PL	(4c) PL	(4d) PL
	full sample				early skyscraper		early subway	
					yes	no	yes	no
Water in own or adjacent cell	1.634*** (0.23)	1.320*** (0.23)	0.831*** (0.18)	0.558*** (0.14)	0.422** (0.18)	0.681*** (0.22)	0.622+ (0.39)	0.549*** (0.15)
Average elevation	-12.130** (5.01)	-7.171*** (1.79)	-1.515+ (0.98)	-1.453 (1.05)	-1.706 (1.92)	-1.223 (1.29)	-5.774 (6.79)	-1.334 (1.07)
Average slope	10.332** (5.21)	2.243 (2.30)	-1.529 (1.70)	-4.207** (1.73)	-1.037 (3.42)	-6.127*** (1.91)	9.640 (7.79)	-4.916*** (1.78)
Bare ground	-1.860+ (1.18)	-1.368* (0.77)	0.397 (0.70)	0.015 (0.55)	-10.714*** (0.79)	-0.041 (0.57)	(.)	0.001 (0.55)
Foundation place in cell		5.599*** (0.24)						
Near 1900 city hall			4.098*** (0.12)	3.392*** (0.09)	3.730*** (0.13)	3.023*** (0.14)	3.634*** (0.27)	3.364*** (0.10)
Cities	125	125	125	125	70	55	14	111
N (cells)	111,114	111,114	111,114	111,114	62,459	48,655	12,119	98,995
McFadden's Pseudo R ²	.07	.24	.21	.12	.15	.09	.16	.12

Notes: Unit of observation are 1.5×1.5 km grid cells. Dependent variables: FP = Foundation place (in cell); Global PL = Global prime location; PL = prime location; CH1900 = 1900 city hall. Units of independent variables: water: dummy (1,0); elevation: km; slope: %, bare ground: dummy (1,0); Foundation place: dummy (1,0); Near 1900 city hall: (queen) contingency to 1900 city hall or city hall in own cell dummy (1,0). All regressions include metro fixed effects. Where no city hall in 1900 exists, we take the first city hall instead. In the rare cases where FP or CH1900 (2) lie outside our grid, we assign a 1 to the closest grid cell. Columns (4a-d) split the sample by metro-wide attributes. All models are estimated using a Poisson maximum likelihood estimator with standard errors clustered by metro. Using interactions between the 'Near 1900 city hall' variable with a dummy for the early skyscraper and early subway samples (rather than splitting the sample) yields differences of .66 (p-value=.00) and .35 (p-value: .18) respectively. For a description of the variables and the underlying data, see the Global Cities dataset Appendix (Ahlfeldt et al., 2020b). + p<0.15, * p < 0.1, ** p < 0.05, *** p < 0.01.

- *Path dependency*: CDBs cercanos a *Historic Cores* de las ciudades.
- A partir de 1900 los efectos de aglomeración son más relevantes para determinar la ubicación del CBD.
- Determinantes geográficos son relevantes para la ubicación de los *Historic Cores*. Pero pierden importancia con el tiempo.
- Los *Historic Cores* dependían más de características geográficas (agua, altitudes bajas, valles, terreno plano).



Modelo

- *Agent-based model of internal city structure (ABM)*
 - + Explica concentraciones extremas (hecho estilizado 1)
 - + Permite distintos equilibrios espaciales (hechos estilizados 2 y 3)
- **Locations l .** Cada ubicación tiene un centroide, un área, productividad B_l , parámetros relacionados con bienes raíces, $N(l)$ es el set de vecinos de l .
- **Workers w .** Preferencias Cobb-Douglass sobre *housing* h , bien de consumo g de precio 1. Productividad θ_w . Salario neto: $w_{w\varphi}^{\text{net}} = w_{w\varphi} \times e^{-\gamma d_{\ell j}}$ donde d es la distancia entre la ubicación l donde vive y j donde trabaja, para la firma φ , **esta varía de ciudad a ciudad dependiendo de su transporte público.**
- **Firmas φ .** Producen $\mu_{w\varphi\ell} = B_{\ell\varphi} \times \theta_w \times \theta_\varphi$ con su productividad θ_φ . Efectos de aglomeración:

$$B_{\ell\varphi} = B_\ell \times Y_{\ell\varphi}, \quad \text{with} \quad Y_{\ell\varphi} = \frac{1}{2} \times \left[(1 + E_{\ell \setminus \varphi})^\epsilon + \left(1 + \sum_{j \in \mathcal{N}(\ell)} E_j \right)^{\epsilon_n} \right]$$

ϵ es un efecto de la fuerza del efecto de la aglomeración



Modelo

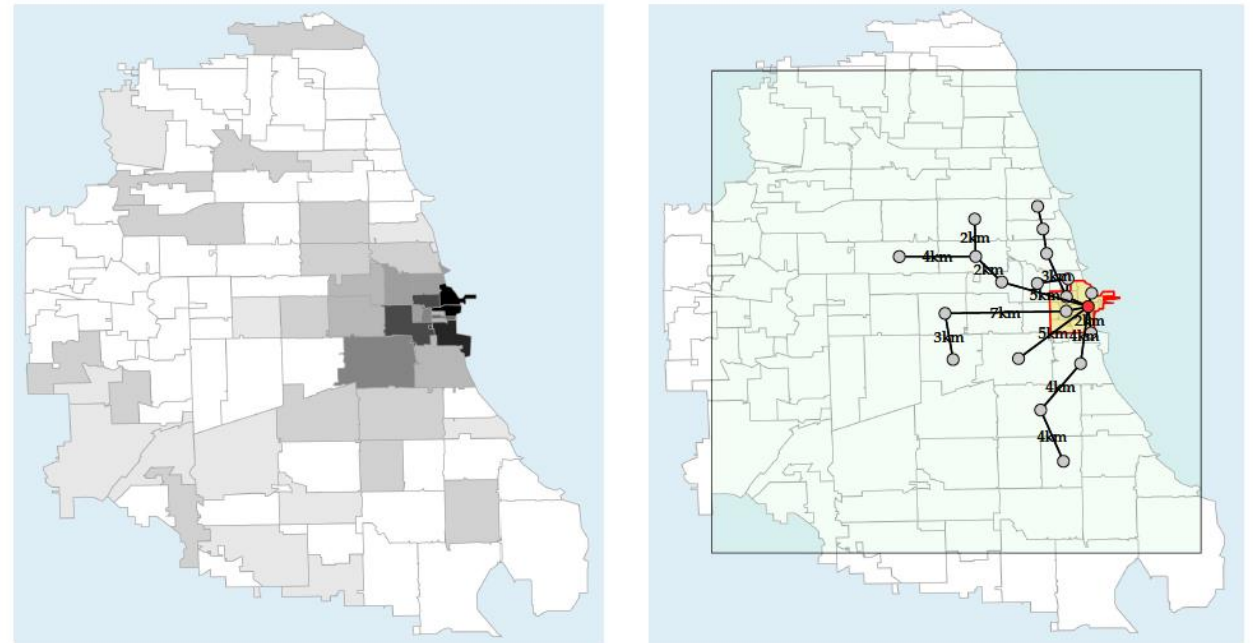
- Para limitar el tamaño de las empresas: un número de trabajadores no genera costos administrativos SOC $\text{SOC}(|N_\varphi|) = \begin{cases} 0 & \text{if } |N_\varphi| \leq \bar{R}_\varphi \\ |N_\varphi|^\gamma - \bar{R}_\varphi^\gamma & \text{if } |N_\varphi| > \bar{R}_\varphi \end{cases}$, y a partir de un umbral \bar{R}_φ hay costos.
- Ganancias de las empresas: $\pi_\varphi = B_{\ell\varphi} \times \theta_\varphi \times \left(\sum_{\omega \in N_\varphi} \theta_\omega \right) - \text{SOC}(|N_\varphi|) - \xi \times |N_\varphi| \times r_\ell$
- Los trabajadores y las firmas comparten las ganancias.
- **Constructoras.** Combinan capital K con el área de cada ubicación l para producir espacio de *housing* o de oficinas.



Modelo

- Simulaciones Monte Carlo para encontrar los equilibrios en tres pasos:
 - + Modelo sin externalidades de aglomeración ni efectos de transporte. -> **Monocéntrico histórico.**
 - + Modelo con externalidades de aglomeración y efectos de transporte con shocks al centro de la ciudad. -> **Nuevo equilibrio de corto plazo.**
 - + Modelo donde se disipan los shocks. -> **Nuevo equilibrio de largo plazo.**
- Proceso: firmas escogen ubicación -> trabajadores y firmas hacen *rematches* cuando se mueven las firmas -> los precios se ajustan para vaciar los mercados
- Trabajadores más productivos trabajan en empresas más productivas.
- Las empresas más productivas se benefician más de los efectos de aglomeración.

Figure 8: Initial equilibrium and endogenously developed transport network



Notes: Darker colors in the left panel denote more employment; white ZIP Code Tabulation Areas have zero employment. In the right panel, the red node denotes the prime location of the city. The length of the links between two nodes is indicated for all links of more than 1 kilometre. The green rectangle corresponds to the empirical extent of the grid that we use for Chicago in our sample of 125 cities.



Modelo

- Shocks: reubicación forzada de firmas
- Sistemas de transporte: impuesto a firmas y trabajadores para financiarlos. Costo hundido grande.
- 10 mil simulaciones.
- 6 medidas de resumen para cada ciudad
 - + el log de la distancia promedio entre EWPPs;
 - + el log de la distribución acumulada de las distancias entre EWPPs a 1.5km;
 - + el log de la distancia promedio entre PLs
 - + el log de la distancia promedio de los EWPPs al ayuntamiento de 1900;
 - + el gradiente de distancia a los EWPPs desde el *Historic Core* de 1900;
 - + el log de la distancia promedio de las PLs al *Historic Core* de 1900;



Modelo

- Regresiones de los efectos del tamaño de la ciudad, la concentración geográfica inicial, los shocks, y el desarrollo endógeno de los sistemas de transporte sobre éstas 6 medidas:

$$\ln C_i^2 = \ln \tilde{\alpha} + \zeta \ln C_i^1 + \tilde{\beta} \ln P_i^1 + \gamma D_i^2 + \epsilon_i,$$

Donde P es la población, C es la medida de resumen, D son los shocks (proporción de reubicaciones forzosas).

- Regresiones de los efectos del tamaño de la ciudad, la concentración geográfica inicial, los shocks, y el desarrollo endógeno de los sistemas de transporte sobre éstas 6 medidas:

$$\ln C_i^2 = \ln \tilde{\alpha} + \zeta \ln C_i^1 + \tilde{\beta} \ln P_i^1 + \gamma D_i^2 + \tilde{\delta} M_i^2 + \epsilon_i.$$

Donde M es la indicadora de tener sistema de transporte.



Resultados

Table 4: Agglomeration vs. transport-induced path dependency

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Log city size	0.134*** (0.01)	0.015*** (0.00)	0.013* (0.01)	0.084*** (0.01)	0.125*** (0.01)	0.072*** (0.01)	0.184*** (0.02)
Shock intensity (share)	-0.059*** (0.00)	-0.048*** (0.00)	-0.036*** (0.00)	-0.127*** (0.00)	-0.290*** (0.00)	-0.583*** (0.01)	-0.589*** (0.01)
Network (0,1)	0.165*** (0.00)	0.033*** (0.00)	0.008*** (0.00)	0.000 (0.00)	0.127*** (0.00)	0.079*** (0.01)	0.084*** (0.01)
Initial concentration	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,000	8,000	8,000	8000	8,000	8,000	8,000
R ²	.615	.515	.0193	.264	.639	.638	.495

Notes: Standard errors in parentheses. CH = city hall, EWPP = employment-weighted workplaces, PL = prime location. Unit of observation is the output of one run of the ABM simulation. We multiply all distance measures by -1 so that the estimated coefficients for all our dependent variables have the same sign. 50% of the runs do not allow for endogenous subway development to emulate exogenous variation. Ln city size is the log of the randomly assigned number of workers. Shock intensity is the share of employment in the initial prime location temporarily displaced by the shock. Initial CH location is determined as the location that minimizes the (pre-shock) employment-weighted distance to all workplaces. 50% of the runs exclude shocks to introduce extensive-margin variation in shock exposure. + $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: City size and shock effects on spatial concentration measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Log city size	0.308*** (0.01)	0.050*** (0.00)	0.021*** (0.01)	0.085*** (0.01)	0.260*** (0.01)	0.155*** (0.01)	0.272*** (0.01)
Shock intensity (share)	-0.061*** (0.00)	-0.049*** (0.00)	-0.036*** (0.00)	-0.127*** (0.00)	-0.292*** (0.00)	-0.584*** (0.01)	-0.590*** (0.01)
Initial concentration	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,000	8,000	8,000	8000	8,000	8,000	8,000
R ²	.376	.369	.0184	.264	.569	.629	.484

Notes: Standard errors in parentheses. CH = city hall, EWPP = employment-weighted workplaces, PL = prime location. Unit of observation is the output of one run of the ABM simulation. We multiply all distance measures by -1 so that the estimated coefficients for all our dependent variables have the same sign. 50% of the runs do not allow for endogenous subway development to emulate exogenous variation. Ln city size is the log of the randomly assigned number of workers. Shock intensity is the share of employment in the initial prime location temporarily displaced by the shock. Initial CH location is determined as the location that minimizes the (pre-shock) employment-weighted distance to all workplaces. 50% of the runs exclude shocks to introduce extensive-margin variation in shock exposure. + $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$



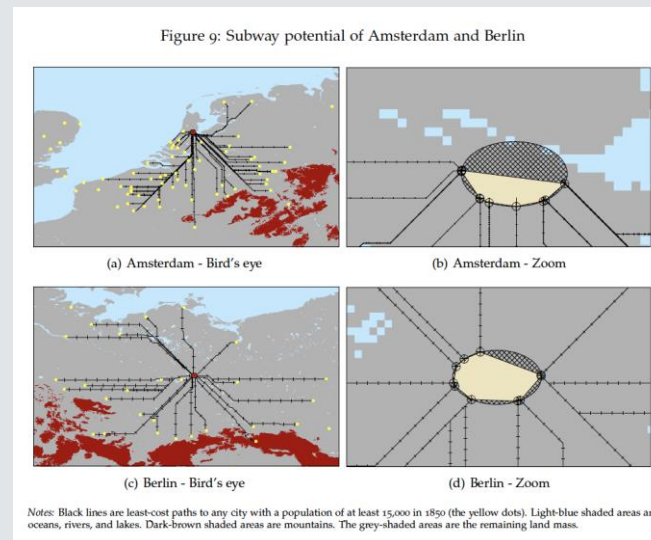
Resultados (*Evidencia empírica causal*)

- Instrumentos: potencial calórico agrícola, ocupación colonial en 1900 y *subway potential* para el problema de correlación de características de las ciudades X con la población inicial P y el término de error en las ecuaciones estimadas:

$$\ln C_i^2 = \ln \alpha + \zeta \ln(C_i^1) + \tilde{\beta} \ln P_i^1 + \theta^2 \ln P_i^2 + \gamma D_i^2 + \tilde{b} X_i + e_i$$

$$\ln C_i^2 = \ln \alpha + \zeta \ln(C_i^1) + \tilde{\beta} \ln P_i^1 + \delta^1 M_i^1 + \theta^2 \ln P_i^2 + \gamma D_i^2 + \tilde{b} X_i + \tilde{\delta}^2 M_i^2 + e_i$$

- Instrumentos correlacionados en 1900, no al final del siglo XX.
- El *subway potential* está construido con conexiones en tren posibles a otras ciudades en 1900: a partir de las estaciones predichas se construye el potencial con el área del polígono que las conecta en relación con el tamaño a 4 km del *Historic Centre*.



Resultados (Evidencia empírica causal)

- Desastres naturales son shocks que disminuyen la concentración y a largo plazo permiten múltiples equilibrios.
- Población original aumenta la concentración (*path dependency*).
- Haber desarrollado un sistema de transporte temprano aumenta la concentración de *prime services*.

Table 6: Agglomeration- vs. transport-induced path dependency

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Ln population 1900	0.097*** (0.03)	0.207*** (0.05)	0.215*** (0.05)	0.088*** (0.03)	0.144*** (0.03)	0.291*** (0.07)	0.320*** (0.09)
Subway in 1910	0.465*** (0.14)	0.465** (0.23)	0.724*** (0.27)	0.526*** (0.14)	0.507*** (0.15)	0.572* (0.29)	0.535+ (0.34)
Disasters since 1900	-0.103*** (0.04)	-0.122** (0.06)	-0.190*** (0.06)	-0.128*** (0.04)	-0.114*** (0.04)	-0.184*** (0.06)	-0.084 (0.08)
Ln population 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subway 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market access	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1900 pop. & subway IV	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage
Kleinb.-Paap F (p-val.)	0	0	0	0	0	0	0
Observations	125	125	125	125	125	125	125

Notes: Unit of observation is cities. Columns (1) and (5) employ weighted distances. 2SLS estimates. CD = cumulative density at 750 m; CH = city hall; EWPP = employment-weighted prime points; PL = prime locations. We multiply all distance measures by -1 so that the estimated coefficients for all our dependent variables have the same sign. Geographic controls include: Developed area 2000, irregular shape index, fragmentation index, share of land not developable within 5km of 1900 city hall, distance from 1900 city hall to water, share of land with steep slope. See the Global Cities dataset Appendix (Ahlfeldt et al., 2020b) for sources and data construction. Instruments for 1900 population and 1900 subway are two variables predicted in two zero-stage locally weighted regressions using population in 1900 and an indicator for subways in 1910 as dependent variables and non-parametric interactions of caloric potential, the colonial territory (in 1800) indicator, and subway potential as predictors. For details on the construction of the instrumental variables, see Appendix A.10. Robust standard errors in parentheses. + p < 0.15, * p < 0.1, ** p < 0.05, *** p < 0.01



Conclusiones

- La estructura de las ciudades es afectada por *path dependency*. Difícil de afectar por política pública forzada, como creación artificial de PLs.
- Ciudades grandes en 1900 y ciudades que desarrollaron un sistema de transporte tuvieron una geografía más estable de *prime services*.
 - + Efectos de aglomeración.
 - + Desarrollo endógeno de sistemas de transporte.
- Múltiples equilibrios posibles en el desarrollo de la estructura de *PLs*. (Por ejemplos con o sin shocks)

