Introduction

- Economic activity spatially concentrated
  - Not just due to natural advantage (Ellison and Glaeser 1999)
  - Does locating near other firms create productivity advantages, i.e., agglomeration (Marshall 1890)?
- Agglomeration would answer big questions
  - Why do cities exist?
  - Why do firms with nationally traded goods produce in high-cost locations like NYC or San Francisco?
  - Why are cars in Detroit and technology in Silicon Valley?
- Practically relevant
  - Governments compete to subsidize industrial plants (Ossa 2017)

Introduction

- Goals of this paper
  - Quantify agglomeration spillovers by estimating how openings of large plants affect TFP of incumbent plants
  - Study mechanisms: input-output links, labor flows, technology links
  - Ask how productivity gains translate into local factor prices
- Identification challenging
  - Firms choose locations to maximize profits and consider local transportation, local factor supply, labor force quality, intermediate goods suppliers, etc.

Introduction

- Research design: use firms' reported location rankings
  - *Site Selection*, a corporate real estate journal, has "Million Dollar Plant" section describing location choices of large new plants
  - It reports winning county alongside runner-up counties (losers)
- Results
  - Similar pre-trends in TFP in winning and losing counties
  - After million dollar plant opens, incumbent plants in winning counties experience 12 percent TFP gain after 5 years
  - On average, worth about $430 million in output
  - Mechanisms: larger spillovers in industries that share labor pools with the million dollar plant, or share technological linkages.
  - Mechanisms: little evidence for importance of input-output links

Introduction

- Interpretation: Roback
  - Empirical support: increases in wages and firm entry in winning counties

Theory: Possible Reasons for Agglomeration

1. Size of the labor market
   - Improves match quality if labor market has search frictions and firms/workers are heterogeneous. This implies denser areas have more productive firms.
   - Insurance against idiosyncratic shocks (e.g., laid off workers find new jobs, firms can easily fill openings). Implies dense areas have firms w same productivity, but different wages (sign ambiguous)
2. Transportation costs for local services and local intermediate goods.
   - Denser areas-> cheaper and faster delivery
   - e.g., high tech firm needing a specialized technician finds on in Silicon Valley but not Nevada desert
   - This implies firms in denser areas have same productivity but lower production costs
3. Knowledge spillovers
   - Faster innovation
   - Faster technology adoption (social learning, Griliches 1958)

Theory: Possible Reasons for Agglomeration

4. Local amenities
   - Skilled workers may prefer certain amenities, so firms employing skilled workers concentrate near these amenities
   - Implies firm productivity unrelated to density, but wages do relate to density
5. Natural advantage or productive amenities.
   - Oil industry is in states with oil fields, wine in California, ports on coast

Theory

- Now simple model to formalize
- Consider factor-neutral spillovers, homogenous firms
- Incumbents choose labor L, capital K, and land T to maximize

$$\max_{L,K,T} f(A, L, K, T) - wL - rK - qT$$

- Factor prices w, r, q, productivity $A(N)$ depends on number of local firms (N)
- Agglomeration means $\partial A/\partial N > 0$

Theory

- Let $L^*(w, r, q)$ be optimal labor inputs given factor costs, similarly K^* and T^*
- In equilibrium, L^*, K^*, T^* set so marginal product = price
- Assume capital internationally traded but land and labor prices depend on local conditions
- Profits

$$
\begin{aligned}
\Pi^* &= f\left[A(N), L^* w(N), r, q(N)), K^*(w(N), r, q(N)), T^*(\cdot)\right] \\
&\quad - w(N)L^*(w(N), r, q(N)) - rK^*(w(N), r, q(N)) \\
&\quad - q(N)T^*(w(N), r, q(N))
\end{aligned}
$$

- Explicitly show dependence of TFP, wages, and land prices on number of firms N

Theory

- Total derivative of incumbents' profits with respect to change in number of firms:

$$
\begin{aligned}
\frac{d\Pi^*}{dN} = \; & \left(\frac{\partial f}{\partial A} \times \frac{\partial A}{\partial N}\right) \\
& + \frac{\partial w}{\partial N}\left\{\left[\frac{\partial L^*}{\partial w}\left(\frac{\partial f}{\partial L} - w\right) - L^*\right] + \left[\frac{\partial K^*}{\partial w}\left(\frac{\partial f}{\partial K} - r\right)\right]\right\} \\
& + \frac{\partial w}{\partial N}\left\{\frac{\partial T^*}{\partial w}\left(\frac{\partial f}{\partial T} - q\right)\right\} \\
& + \frac{\partial q}{\partial N}\left\{\left[\frac{\partial L^*}{\partial q}\left(\frac{\partial f}{\partial L} - w\right)\right] + \left[\frac{\partial K^*}{\partial q}\left(\frac{\partial f}{\partial K} - r\right)\right]\right\} \\
& + \frac{\partial q}{\partial N}\left\{\left[\frac{\partial T^*}{\partial q}\left(\frac{\partial f}{\partial T} - q\right) - T^*\right]\right\}
\end{aligned}
$$

Theory

- If firms are price takers and factors are paid their marginal products, this becomes

$$\frac{d\Pi^*}{dN} = \left(\frac{\partial f}{\partial A} \times \frac{\partial A}{\partial N}\right) - \left(\frac{\partial w}{\partial N} L^* + \frac{\partial q}{\partial N} T^*\right)$$

- Competing effects. Positive spillovers -> productivity of all factors increases (first term)
- But density increases land and labor prices (second term).

Theory

- If $d\Pi^*/dN \leq 0$, then agglomeration spillover is smaller than increase in production costs.
- If $d\Pi^*/dN > 0$, then agglomeration spillover is larger than increase in production costs.
- These are short-run. In long run, Roback equilibrium returns (firms/workers indifferent where to locate, productivity capitalizes into land prices)
- Predictions: (1) large spillovers will increase entry; (2) prices of locally traded inputs will rise.
- If incumbents are heterogeneous: then expect larger spillovers from incumbents economically closer to the new plant.

Theory

- Summary of four empirical predictions from the theory:
  1. Opening a new plant increases TFP of incumbent plants
  2. TFP increase is larger for firms economically closer to new plant
  3. Density of economic activity increases as firms enter
  4. Price of locally supplied factors increases

Research Design

- Comparing TFP of incumbent plants in counties with and without openings suffers from omitted variables bias
- Case study: Bavarian Motor Works (BMW)
- BMW considered 250 sites then in 1991 announced 20 counties they were considering
- After 6 months, announced two finalists: Greenville-Spartanburg, South Carolina; and Omaha, Nebraska
- 1992: BMW chooses Greenville-Spartanburg, supported by $115 million package from state/local governments
- Key considerations: expected production costs (low union density; supply of qualified workers; 58 nearby German companies; transportation infrastructure); value of subsidy
- Local government estimated 5-year impact of $2 billion, including 2,000 direct jobs and 2,000 indirect jobs

Research Design

- Site Selection. Each issue has "Million Dollar Plants" article.
- Article reports chosen county and runner-up county or counties
- Research design: incumbent firms in losing counties are counterfactual for incumbents in winning counties.

Data

- Million Dollar Plant (MDP) articles
- Census Standard Statistical Establishment List (SSEL), Annual Survey of Manufactures (ASM), Census of Manufactures (CM), 1973-98
- Paper uses 47 million dollar plant openings
- Requirements for using an opening
    - New plant in manufacturing sector, owned by reported firm, in SSEL between 2 years before and 3 years after publication of MDP article
    - SSEL plant in same county as MDP plant
    - Had incumbent plants in winning and losing counties for each of previous 8 years
- Main ASM sample: plants continuous present in 8 years before MDP opening plus year of opening
- Exclude plants owned by same firm as MDP

Data

- Measures of economic distance
  1. input-output links
  2. CPS worker transitions
  3. Pairwise patent citation

Data

TABLE 1
THE MILLION DOLLAR PLANT SAMPLE

	(1)
Sample MDP openings:[a]	
Across all industries	47
Within same two-digit SIC	16
Across all industries:	
Number of loser counties per winner county:	
1	31
2+	16
Reported year − matched year:[b]	
−2 to −1	20
0	15
1 to 3	12
Reported year of MDP location:	
1981–85	11
1986–89	18
1990–93	18
MDP characteristics, 5 years after opening:[c]	
Output ($1,000s)	452,801
	(901,690)
Output, relative to county output 1 year prior	.086
	(.109)
Hours of labor (1,000s)	2,986
	(6,789)

Data

TABLE 2
SUMMARY STATISTICS FOR MEASURES OF INDUSTRY LINKAGES

MEASURE OF INDUSTRY LINKAGE	DESCRIPTION	MEAN All Plants	MEAN Only 1st Quartile	MEAN Only 4th Quartile	STANDARD DEVIATION
Labor market pooling:					
CPS worker transitions	Proportion of workers leaving a job in this industry that move to the MDP industry (15 months later)	.119	.002	.317	.249
Intellectual or technology spillovers:					
Citation pattern	Percentage of manufactured industry patents that cite patents manufactured in MDP industry	.022	.001	.057	.033
Technology input	R&D flows from MDP industry, as a percentage of all private-sector technological expenditures	.022	.000	.106	.084
Technology output	R&D flows to MDP industry, as a percentage of all original research expenditures	.011	.000	.042	.035
Proximity to customers and suppliers:					
Manufacturing input	Industry inputs from MDP industry, as a percentage of its manufacturing inputs	.017	.000	.075	.061
Manufacturing output	Industry output used by MDP industry, as a percentage of its output to manufacturers	.042	.000	.163	.139

Data

TABLE 3

County and Plant Characteristics by Winner Status, 1 Year Prior to a Million Dollar Plant Opening

	All Plants					Within Same Industry (Two-Digit SIC)				
	Winning Counties (1)	Losing Counties (2)	All U.S. Counties (3)	t-Statistic (Col. 1 − Col. 2) (4)	t-Statistic (Col. 1 − Col. 3) (5)	Winning Counties (6)	Losing Counties (7)	All U.S. Counties (8)	t-Statistic (Col. 6 − Col. 7) (9)	t-Statistic (Col. 6 − Col. 8) (10)
A. County Characteristics										
No. of counties	47	73				16	19			
Total per capita earnings ($)	17,418	20,628	11,259	**−2.05**	**5.79**	20,230	20,528	11,378	−.11	**4.62**
% change, over last 6 years	.074	.096	.037	−.81	1.67	.076	.089	.057	−.28	.57
Population	322,745	447,876	82,381	−1.61	**4.33**	357,955	504,342	83,430	−1.17	**3.26**
% change, over last 6 years	.102	.051	.036	**2.06**	**3.22**	.070	.032	.031	1.18	1.63
Employment-population ratio	.535	.579	.461	−1.41	**3.49**	.602	.569	.467	.64	**3.63**
Change, over last 6 years	.041	.047	.023	−.68	**2.54**	.045	.038	.028	.39	1.57
Manufacturing labor share	.314	.251	.252	**2.35**	**3.12**	.296	.227	.251	1.60	1.17
Change, over last 6 years	−.014	−.031	−.008	1.52	−.64	−.030	−.040	−.007	.87	**−3.17**
B. Plant Characteristics										
No. of sample plants	18.8	25.6	7.98	−1.35	**3.02**	2.75	3.92	2.38	−1.14	.70
Output ($1,000s)	190,039	181,454	123,187	.25	**2.14**	217,950	178,958	132,571	.41	1.25
% change, over last 6 years	.082	.082	.118	.01	−.97	−.061	.177	.182	−1.23	**−3.38**
Hours of labor (1,000s)	1,508	1,168	877	1.52	**2.43**	1,738	1,198	1,050	.92	1.33
% change, over last 6 years	.122	.081	.115	.81	.14	.160	.023	.144	.85	.13

Econometrics

$$Y_{pijt} = A_{pijt} L_{pijt}^{\beta_1} K_{pijt}^{B\beta_2} K_{pijt}^{E\beta_3} M_{pijt}^{\beta_4}$$

- Plant p,industry i, case j, year t.
- Y value of shipments minus inventories; TFP A; labor hours of production L; building capital K^B; machinery and equipment capital K; materials M

$$\ln(A_{pijt}) = \alpha_p + \mu_{it} + \lambda_p + \varepsilon_{pijt} + A(N_{pijt})$$

- TFP: Permament component α_p; case component λ_p [*sic*? should be case j]; time-varying TFP shocks μ_{it}; stochastic error ε_{pijt}

Econometrics

- Further specifics:

$$
\begin{aligned}
\ln(A_{pijt}) &= \delta 1(Winner)_{pj} + \psi Trend_{jt} \\
&\quad + \Omega[Trend_{jt} \times 1(Winner)_{pj}] \\
+\kappa 1(\tau &\geq 0)_{jt} \\
+\gamma[Trend_{jt} \times 1(\tau &\geq 0)_{jt}] \\
+\theta_1[1(Winner)_{pj} \times 1(\tau &\geq 0)_{jt}] \\
+\theta_2[Trend_{jt} \times 1(Winner)_{pj} \times 1(\tau &\geq 0)_{jt}] \\
&\quad + \alpha_p + \mu_{it} + \lambda_p + \varepsilon_{pijt}
\end{aligned}
$$

- $1(Winner)_{pt}$ dummy for located in winning county; τ event time; *Trend* time trend

Econometrics

- Combining gives regression equation:

$$
\begin{aligned}
\ln(Y_{pijt}) &= \beta_1 \ln(L_{pijt}) + \beta_2 \ln(K^B_{pijt}) \\
&\quad + \beta_3 \ln(K^E_{pijt}) + \beta_4 \ln(M_{pijt}) \\
&\quad + \delta 1(Winner)_{pj} + \psi Trend_{jt} \\
&\quad + \Omega[Trend_{jt} \times 1(Winner)_{pj}] \\
+\kappa 1(\tau &\geq 0)_{jt} + \gamma[Trend_{jt} \times 1(\tau \geq 0)_{jt}] \\
+\theta_1[1(Winner)_{pj} \times 1(\tau &\geq 0)_{jt}] \\
+\theta_2[Trend_{jt} \times 1(Winner)_{pj} \times 1(\tau &\geq 0)_{jt}] \\
&\quad + \alpha_p + \mu_{it} + \lambda_p + \varepsilon_{pijt}
\end{aligned}
$$

- Main parameters of interest: θ_1, θ_2 (mean shift; trend break)
- Also estimate simpler mean shift model (diff-in-diff; "Model 1"), assuming $\psi = \Omega = \gamma = \theta_2 = 0$. Identifying assumption:

$$E[[1(Winner)_{pj} \times 1(\tau \geq 0)_{jt}]\varepsilon_{pijt} | \alpha_p, \mu_{it}, \lambda_p] = 0$$

- "Model 2" : full specification

Econometrics

- Sensitivity analyses: translog; industry-specific β from cost shares; Olley-Pakes, Levinson-Petrin, ACF; different levels of fixed effects; allow effects of inputs to differ by winner and post-million dollar plant status.
- Some estimates use all plants nationally, or only plants in willing/losing counties for $\tau = -8$ to $\tau = 0$.
- Main estimates: plants in all years $\tau = -7$ to $\tau = 5$.
- Also consider: unobserved changes in inputs; sample attrition; analysis of factor input demands
- Standard errors clustered by county
- Main results weighted by total value of shipments at $\tau = -8$

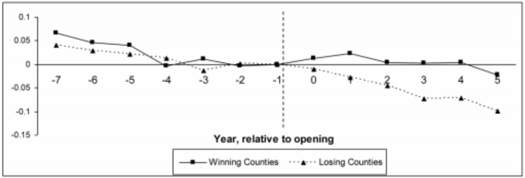
Results

TABLE 4
INCUMBENT PLANT PRODUCTIVITY, RELATIVE TO THE YEAR OF AN MDP OPENING

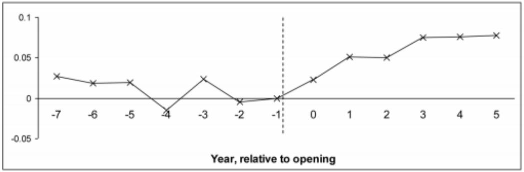
Event Year	In Winning Counties (1)	In Losing Counties (2)	Difference Col. 1 − Col. 2 (3)
$\tau = -7$	.067	.040	.027
	(.058)	(.053)	(.032)
$\tau = -6$	.047	.028	.018
	(.044)	(.046)	(.023)
$\tau = -5$	.041	.021	.020
	(.036)	(.040)	(.025)
$\tau = -4$	−.003	.012	−.015
	(.030)	(.030)	(.024)
$\tau = -3$	.011	−.013	.024
	(.022)	(.022)	(.021)
$\tau = -2$	−.003	.001	−.005
	(.027)	(.011)	(.028)
$\tau = -1$	0	0	0
$\tau = 0$	.013	−.010	.023
	(.018)	(.011)	(.019)
$\tau = 1$	.023	−.028	.051**
	(.026)	(.024)	(.023)
$\tau = 2$	.004	−.046	.050
	(.036)	(.046)	(.033)
$\tau = 3$	.003	−.073	.076*
	(.047)	(.057)	(.043)
$\tau = 4$	.004	−.072	.076**
	(.053)	(.062)	(.033)
$\tau = 5$	−.023	−.100	.077**
	(.069)	(.067)	(.035)
R^2	.9861		
Observations	28,732		

Results

All Industries: Winners vs. Losers



Difference: Winners – Losers



Results

TABLE 5
CHANGES IN INCUMBENT PLANT PRODUCTIVITY FOLLOWING AN MDP OPENING

	ALL COUNTIES: MDP WINNERS − MDP LOSERS		MDP COUNTIES: MDP WINNERS − MDP LOSERS		ALL COUNTIES: RANDOM WINNERS
	(1)	(2)	(3)	(4)	(5)
	A. Model 1				
Mean shift	.0442*	.0435*	.0524**	.0477**	− 0.0496***
	(.0233)	(.0235)	(.0225)	(.0231)	(.0174)
				[$170 m]	
R^2	.9811	.9812	.9812	.9860	~0.98
Observations (plant by year)	418,064	418,064	50,842	28,732	~400,000
	B. Model 2				
Effect after 5 years	.1301**	.1324**	.1355***	.1203**	−.0296
	(.0533)	(.0529)	(.0477)	(.0517)	(.0434)
				[$429 m]	
Level change	.0277	.0251	.0255	.0290	.0073
	(.0241)	(.0221)	(.0186)	(.0210)	(.0223)
Trend break	.0171*	.0179**	.0183**	.0152*	− 0.0062
	(.0091)	(.0088)	(.0078)	(.0079)	(.0063)
Pre-trend	−.0057	−.0058	−.0048	−.0044	−.0048
	(.0046)	(.0046)	(.0046)	(.0044)	(.0040)
R^2	.9811	.9812	.9813	.9861	~.98
Observations (plant by year)	418,064	418,064	50,842	28,732	~400,000
Plant and industry by year fixed effects	Yes	Yes	Yes	Yes	Yes
Case fixed effects	No	Yes	Yes	Yes	NA
Years included	All	All	All	$-7 \leq \tau \leq 5$	All

Results

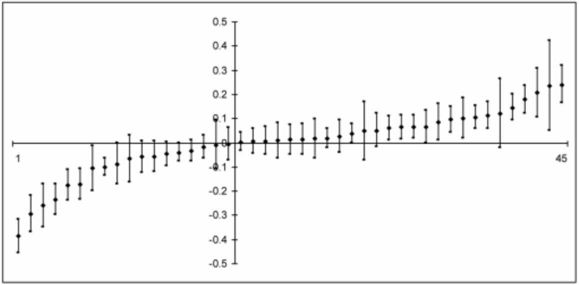


FIG. 2.—Distribution of case-specific mean shift estimates, following an MDP opening. The figure reports results from a version of model 1 that estimates the parameter θ_1 for each of the 47 MDP cases. The figure reports only 45 estimates because two cases were excluded for Census confidentiality reasons.

Results

TABLE 6
CHANGES IN INCUMBENT PLANT OUTPUT AND INPUTS FOLLOWING AN MDP OPENING

	Output (1)	Worker Hours (2)	Machinery Capital (3)	Building Capital (4)	Materials (5)
Model 1: mean shift	.1200***	.0789**	.0401	.1327*	.0911***
	(.0354)	(.0357)	(.0348)	(.0691)	(.0302)
Model 2: after 5 years	.0826*	.0562	−.0089	−.0077	.0509
	(.0478)	(.0469)	(.0300)	(.0375)	(.0541)

Results

TABLE 7
CHANGES IN INCUMBENT PLANT PRODUCTIVITY FOLLOWING AN MDP OPENING FOR INCUMBENT PLANTS IN THE MDP'S TWO-DIGIT INDUSTRY AND ALL OTHER INDUSTRIES

	All Industries (1)	MDP's Two-Digit Industry (2)	All Other Two-Digit Industries (3)
		A. Model 1	
Mean shift	.0477**	.1700**	.0326
	(.0231)	(.0743)	(.0253)
	[$170 m]	[$102 m]	[$104 m]
R^2	.9860	.9861	
Observations	28,732	28,732	
		B. Model 2	
Effect after 5 years	.1203**	.3289	.0889*
	(.0517)	(.2684)	(.0504)
	[$429 m]	[$197 m]	[$283 m]
Level change	.0290	.2814***	.0004
	(.0210)	(.0895)	(.0171)
Trend break	.0152*	.0079	.0147*
	(.0079)	(.0344)	(.0081)
Pre-trend	−.0044	−.0174	−.0026
	(.0044)	(.0265)	(.0036)
R^2	.9861	.9862	
Observations	28,732	28,732	

Results

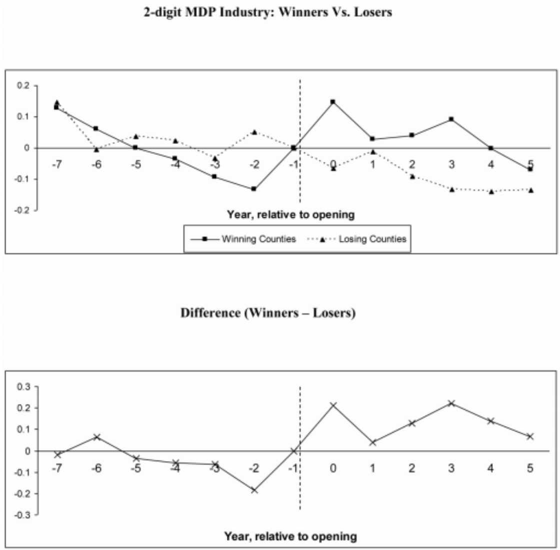
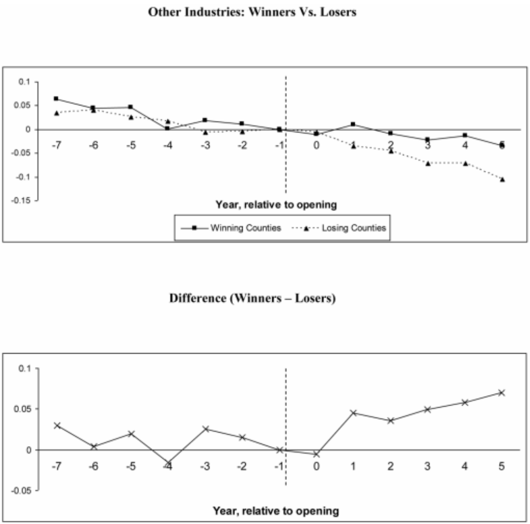


FIG. 3.—Incumbent plants' productivity in the MDP's two-digit industry, winning versus losing counties, relative to the year of an MDP opening. These figures accompany table 7, column 2 (MDP's two-digit industry).

Results


Other Industries: Winners Vs. Losers
0.1
0.05
0
-0.05
-0.1
-0.15
-7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5
Year, relative to opening
Winning Counties
Losing Counties
Difference (Winners – Losers)
0.1
0.05
0
-0.05
-7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5
Year, relative to opening

Results

TABLE 8

Changes in Incumbent Plant Productivity Following an MDP Opening, by Measures of Economic Distance between the MDP's Industry and Incumbent Plant's Industry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPS worker transitions	.0701***						.0374
	(.0237)						(.0260)
Citation pattern		.0545***					.0256
		(.0192)					(.0208)
Technology input			.0320*				.0501
			(.0173)				(.0421)
Technology output				.0596***			.0004
				(.0216)			(.0434)
Manufacturing input					.0060		−.0473
					(.0123)		(.0289)
Manufacturing output						.0150	−.0145
						(.0196)	(.0230)
R^2	.9852	.9852	.9851	.9852	.9851	.9852	.9853
Observations	23,397	23,397	23,397	23,397	23,397	23,397	23,397

Results

TABLE 9
CHANGES IN COUNTIES' NUMBER OF PLANTS, TOTAL OUTPUT, AND SKILL-ADJUSTED WAGES FOLLOWING AN MDP OPENING

	A. CENSUS OF MANUFACTURES		B. CENSUS OF POPULATION
	Dependent Variable: Log(Plants) (1)	Dependent Variable: Log(Total Output) (2)	Dependent Variable: Log(Wage) (3)
Difference-in-difference	.1255**	.1454	.0268*
	(.0550)	(.0900)	(.0139)
R^2	.9984	.9931	.3623
Observations	209	209	1,057,999

Results

TABLE 10

CHANGES IN INCUMBENT PLANT PRODUCTIVITY FOLLOWING AN MDP OPENING, ROBUSTNESS TO DIFFERENT SPECIFICATIONS

	Baseline Specification (1)	Translog Functional Form (2)	Input-Industry Interactions (3)	Input-Winner, Input-Post (4)	Region-Year Fixed Effects (5)	Region-Year-Industry Fixed Effects (6)	Unweighted (7)
Model 1: mean shift	.0477**	.0471**	.0406*	.0571**	.0442*	.0369*	.0146
	(.0231)	(.0226)	(.0220)	(.0245)	(.0230)	(.0215)	(.0107)
Model 2: after 5 years	.1203**	.1053*	.0977**	.1177**	.1176**	.0879**	.0065
	(.0517)	(.0535)	(.0487)	(.0538)	(.0520)	(.0442)	(.0281)

Comments

- Demonstrates role of research design (why some people say "designed based" instead of "reduced form")
- Also value of non-standard data like trade journals (Site Selection)
- Urban assumes agglomeration effects and spatial models estimate them; here a direct test of them
- No paper is perfect
- Value of clear graphs
- Role of simple v. detailed estimators (e.g., Cobb-Douglas versus ACF)
- Value of memorable setting/title? ("Million Dollar Plants")
- Also value of taking simple setting (plant openings and agglomeration) to learn as much as possible (e.g., test factor prices, mechanisms for spillovers)

Comments

- No effects unweighted
- This is revenue TFP, not quantity TFP. How are markups changing? If it's markups and not physical productivity, this would change the story
- External validity limited, 47 openings/closings
- e.g., hard to learn from this what a county should bid as subsidy for plant opening
- Key question in many urban settings is shape of agglomeration function, not just sign.

Introduction

- Many economic questions depend on housing supply and demand elasticities
  - 2000s housing bubble
  - Welfare analysis of policies
  - Great Divergence between skilled and unskilled cities
    - Inequality
  - Mobility costs
    - Adjustment to local economic shocks
  - Shape of cities

Introduction

- What is the elasticity of housing supply?
  - When housing demand has a positive shock, how do home prices change?
- What determines the elasticity of housing?
- Standard answer: economic variables
  - Construction costs
  - Population density
  - Zoning
- This paper: what is role of geography?

Introduction

- Basic idea
  - Can't build house on water or slopes above 15%

Geography Dataset

- Use USGS Digital Elevation Model (DEM) at 90-m resolution: slope maps for continental U.S.
  - Within 50km radius of each metro center city, find area which has land above 15% slope
- Example: Los Angeles
  - Very High Housing Values
  - Strong incentive to build on open land
  - 6,456 block groups lie within 50km of Los Angeles centroid
  - Steep-slope block groups have more than half their area above 15% incline
  - Steep-slope block groups are 48% of land area within 50km of LA center but only 4% of population

Geography Dataset

- How much area within 50km of city center is wetlands, lakes, rivers, other water?
- 1992 USGS National Land Cover Dataset
  - Satellite-based
  - 30-by-30m cell data
  - Calculate by census tract
- Digital contour maps show area within 50-km radii lost to oceans and Great Lakes
- First Measure of Undevelopable Area
  - Based on ex ante constraints, not ex post (e.g., density)

Geography Dataset

TABLE I
PHYSICAL AND REGULATORY DEVELOPMENT CONSTRAINTS (METRO AREAS WITH POPULATION > 500,000)

Rank	MSA/NECMA name	Undevelopable area (%)	WRI	Rank	MSA/NECMA name	Undevelopable area (%)	WRI
1	Ventura, CA	79.64	1.21	26	Portland–Vancouver, OR–WA	37.54	0.27
2	Miami, FL	76.63	0.94	27	Tacoma, WA	36.69	1.34
3	Fort Lauderdale, FL	75.71	0.72	28	Orlando, FL	36.13	0.32
4	New Orleans, LA	74.89	−1.24	29	Boston–Worcester–Lawrence, MA–NH	33.90	1.70
5	San Francisco, CA	73.14	0.72	30	Jersey City, NJ	33.80	0.29
6	Salt Lake City–Ogden, UT	71.99	−0.03	31	Baton Rouge, LA	33.52	−0.81
7	Sarasota–Bradenton, FL	66.63	0.92	32	Las Vegas, NV–AZ	32.07	−0.69
8	West Palm Beach–Boca Raton, FL	64.01	0.31	33	Gary, IN	31.53	−0.69
9	San Jose, CA	63.80	0.21	34	Newark, NJ	30.50	0.68
10	San Diego, CA	63.41	0.46	35	Rochester, NY	30.46	−0.06
11	Oakland, CA	61.67	0.62	36	Pittsburgh, PA	30.02	0.10
12	Charleston–North Charleston, SC	60.45	−0.81	37	Mobile, AL	29.32	−1.00
13	Norfolk–Virginia Beach–Newport News, VA–NC	59.77	0.12	38	Scranton–Wilkes-Barre–Hazleton, PA	28.78	0.01
14	Los Angeles–Long Beach, CA	52.47	0.49	39	Springfield, MA	27.08	0.72
15	Vallejo–Fairfield–Napa, CA	49.16	0.96	40	Detroit, MI	24.52	0.05
16	Jacksonville, FL	47.33	−0.02	41	Bakersfield, CA	24.21	0.40
17	New Haven–Bridgeport–Stamford, CT	45.01	0.19	42	Harrisburg–Lebanon–Carlisle, PA	24.02	0.54
18	Seattle–Bellevue–Everett, WA	43.63	0.92	43	Albany–Schenectady–Troy, NY	23.33	−0.09
19	Milwaukee–Waukesha, WI	41.78	0.46	44	Hartford, CT	23.29	0.49
20	Tampa–St. Petersburg–Clearwater, FL	41.64	−0.22	45	Tucson, AZ	23.07	1.52
21	Cleveland–Lorain–Elyria, OH	40.50	−0.16	46	Colorado Springs, CO	22.27	0.87
22	New York, NY	40.42	0.65	47	Baltimore, MD	21.87	1.60
23	Chicago, IL	40.01	0.02	48	Allentown–Bethlehem–Easton, PA	20.86	0.02
24	Knoxville, TN	38.53	−0.37	49	Minneapolis–St. Paul, MN–WI	19.23	0.38
25	Riverside–San Bernardino, CA	37.90	0.53	50	Buffalo–Niagara Falls, NY	19.05	−0.23

Geography Dataset

TABLE I
(CONTINUED)

Rank	MSA/NECMA name	Undevelopable area (%)	WRI	Rank	MSA/NECMA name	Undevelopable area (%)	WRI
51	Toledo, OH	18.96	−0.57	74	Dallas, TX	9.16	−0.23
52	Syracuse, NY	17.85	−0.59	75	Richmond–Petersburg, VA	8.81	−0.38
53	Denver, CO	16.72	0.84	76	Houston, TX	8.40	−0.40
54	Columbia, SC	15.23	−0.76	77	Raleigh–Durham–Chapel Hill, NC	8.11	0.64
55	Wilmington–Newark, DE–MD	14.67	0.47	78	Akron, OH	6.45	0.07
56	Birmingham, AL	14.35	−0.23	79	Tulsa, OK	6.29	−0.78
57	Phoenix–Mesa, AZ	13.95	0.61	80	Kansas City, MO–KS	5.82	−0.79
58	Washington, DC–MD–VA–WV	13.95	0.31	81	El Paso, TX	5.13	0.73
59	Providence–Warwick–Pawtucket, RI	13.87	1.89	82	Fort Worth–Arlington, TX	4.91	−0.27
60	Little Rock–North Little Rock, AR	13.71	−0.85	83	Charlotte–Gastonia–Rock Hill, NC–SC	4.69	−0.53
61	Fresno, CA	12.88	0.91	84	Atlanta, GA	4.08	0.03
62	Greenville–Spartanburg–Anderson, SC	12.87	−0.94	85	Austin–San Marcos, TX	3.76	−0.28
63	Nashville, TN	12.83	−0.41	86	Omaha, NE–IA	3.34	−0.56
64	Louisville, KY–IN	12.69	−0.47	87	San Antonio, TX	3.17	−0.21
65	Memphis, TN–AR–MS	12.18	1.18	88	Greensboro–Winston–Salem–High Point, NC	3.12	−0.29
66	Stockton–Lodi, CA	12.05	0.59	89	Fort Wayne, IN	2.56	−1.22
67	Albuquerque, NM	11.63	0.37	90	Columbus, OH	2.50	0.26
68	St. Louis, MO–IL	11.08	−0.73	91	Oklahoma City, OK	2.46	−0.37
69	Youngstown–Warren, OH	10.52	−0.38	92	Wichita, KS	1.66	−1.19
70	Cincinnati, OH–KY–IN	10.30	−0.58	93	Indianapolis, IN	1.44	−0.74
71	Philadelphia, PA–NJ	10.16	1.13	94	Dayton–Springfield, OH	1.04	−0.50
72	Ann Arbor, MI	9.71	0.31	95	McAllen–Edinburg–Mission, TX	0.93	−0.45
73	Grand Rapids–Muskegon–Holland, MI	9.28	−0.15				

Note. WRI = Wharton Regulation Index.

Geography Dataset

TABLE II
PARTIAL CORRELATES OF UNAVAILABLE LAND SHARE (50-KM RADIUS)

	Share of area unavailable for development	
	OLS-regional FE β	Adds coastal dummy β
	(1)	(2)
Log population in 2000	0.443	−0.01
	(0.336)	(0.364)
Log median house value in 2000	0.592	0.41
	(0.081)***	(0.085)***
ΔLog median house value (1970–2000)	0.240	0.122
	(0.054)***	(0.057)**
Log income in 2000	0.233	0.164
	(0.056)***	(0.060)***
ΔLog income (1990–2000)	−0.002	0.006
	(0.020)	(0.022)
ΔLog population (1990–2000)	−0.027	−0.043
	(0.027)	(0.029)
Immigrants (1990–2000)/population (1990)	0.009	−0.007
	(0.011)	(0.012)
Share with bachelor's degree (2000)	0.006	−0.004
	(0.020)	(0.022)
Share workers in manufacturing (2000)	−0.01	0.005
	(0.021)	(0.023)
Log(patents/population) (2000)	0.762	0.771
	(0.260)***	(0.287)***
January monthly hours of sun (average 1941–1970)	−3.812	−12.047
	(11.252)	(12.318)
Log tourist visits per person (2000)	0.493	0.719
	(0.261)*	(0.286)**

Notes. Standard errors in parentheses. Rows present the coefficients (β) and standard errors of separate regressions, where the variable described in the row is the dependent variable on the left-hand side and the unavailable land share (geographic constraint) is the explanatory variable on the right-hand side. The regressions in column (1) include regional fixed effects as controls, whereas those in column (2) also include a coastal dummy for metropolitan areas within 100 km of the oceans or Great Lakes (as defined in Rappaport and Sachs [2003]). * significant at 10%; ** significant at 5%; *** significant at 1%.

Geography Dataset

- Other data: 2005 Wharton Regulation Index (WRI)
  - Measures zoning, project approval policies that constrain new residential real estate development
  - Aggregate original municipal data to get MSA-area measures

Model

- Why should geography affect housing supply elasticity, rather than just price level?
- Model assumptions:
  - Developers: no market power (in input or output markets)
  - Consumers compete for locations
- Preferences in city k:

$$U\left(C_k\right) = \left(C_k\right)^{\rho}$$

- Consumption (C_k) combines amenities (A_k) and private goods.
- Private consumption: wages minus rents minus commuting costs
- All jobs in city center

Model

- One house per person: $H_k = POP_k$
- Utility:
$$U\left(C_k\right) = \left(A_k + w_k - \gamma r' - td\right)^\rho$$
- w_k is wage
- γ units of land/housing-space consumption
- r' for rent per unit of housing-space consumption
- t: monetary cost per distance commuted
- d: distance of consumer's residence to CBD.
- Nonarbitrage condition defines rents: everyone in a city gets same utility $\overline{U}_k$

Model

- One house per person: $H_k = POP_k$
- Utility:
$$U\left(C_k\right) = \left(A_k + w_k - \gamma r' - td\right)^{\rho}$$
- w_k is wage
- γ units of land/housing-space consumption
- r' for rent per unit of housing-space consumption
- t: monetary cost per distance commuted
- d: distance of consumer's residence to CBD.
- Nonarbitrage condition defines rents: everyone in a city gets same utility $\overline{U}_k$
- Total rent paid:
$$r\left(d\right) = r_0 - td$$

Model

- Circular city has radius Φ_h
  - Only share Λ_k is developable
  - Radius depends on households and land availability:

$$\Phi_k = \sqrt{\gamma H_k / \Lambda_k \pi}$$

- $P(d)$ home price equals land cost $LDC(d)$ plus construction cost CC

$$P(d) = CC + LC(d)$$

- Steady-state:

$$r(\Phi_k) = iCC$$

- Implication:

$$r_0 = iCC + t\sqrt{\gamma H_k / \Lambda_k \pi}$$

Model

- Mean rent $\widetilde{r}_k$ equals rent paid by home two-thirds distance from CBD to city edge:

$$\widetilde{r}_k = r\left(\frac{2}{3}\Phi_k\right)$$

- Final housing supply:

$$P_k^S = CC + \frac{1}{3i}t\sqrt{\frac{\gamma H_k}{\Lambda_k \pi}}$$

Model

- Final housing demand. Amenities are

$$A_k = \tilde{A}_k - \alpha\sqrt{POP_k}$$

- What does α reflect? What is its sign?
- Labor demand:

$$w_k = \tilde{w}_k - \psi\sqrt{POP_h}$$

- Implied housing demand:

$$\sqrt{H_k} = \frac{\tilde{A}_k + \tilde{w}_k}{(\psi + \alpha)} - \frac{i}{(\psi + \alpha)} P\left(0\right)$$

- Note: shocks to labor productivity or amenities $\left(\tilde{A}_k + \tilde{w}_k\right)$ shift city's demand curve up. Use this to identify supply elasticity later.

Model

- Equilibrium number of households in a city:

$$H_k^* = \left(\frac{\tilde{A}_k + \tilde{w}_k - iCC}{(\psi + \alpha) + t\sqrt{\frac{\gamma}{\Lambda_k \pi}}} \right)$$

- Amenities and wages must at least cover annuitized cost of construction for a site to be inhabitable.

Model

- Clearly more land availability increases supply: $\partial \tilde{P}^S / \partial \Lambda_k < 0$
- But does land affect supply elasticity?
- Define "city-specific supply inverse elasticity of average housing prices" [egh] as

$$\beta_k^S \equiv \frac{\partial \ln \tilde{P}_k^S}{\partial \ln H_k}$$

- Proposition 1: as land constraints increase, demand shocks imply stronger positive impacts on growth of housing values. (Opposite sign also holds.)
- Proof:

$$\beta_k^S = \frac{1}{2}\left[\frac{1}{3i} t \left(\frac{\sqrt{\gamma H_k / \Lambda_k \pi}}{\tilde{P}_k}\right)\right]$$

therefore

$$\frac{\partial \beta_k}{\partial \Lambda_k} = \frac{\partial^2 \ln \tilde{P}_k}{\partial \ln H_k \partial \Lambda_k} = -\frac{1}{4} \frac{\frac{1}{3i} t \sqrt{\frac{\gamma H_k}{\Lambda_k^3 \pi}} CC}{\left(\tilde{P}_k\right)^2} < 0$$

- [Does this give any intuition? Unclear.]

Model

- Why do areas with difficult housing supply conditions have any housing at all?
- Proposition 2: metro areas with low land availability are more productive or have more amenities
- Simple intuition: theory of compensating differences.
  - Positive association between area attractiveness and land constraints, given it's a metro area
- Proposition 3: if amenities have Pareto distribution across cities, population levels in metro areas is independent of degree of land availability.
- Summary of 3 propositions:
  - Metro areas that we observe are more productive or have better amenities; no correlation between land availability and population size; and lower land availability implies lower housing supply elasticities.

Geography and Home Price Elasticities

- Recall: $\tilde{P} = CC + LC(H_k)$
- Taking logs, totally differentiating, manipulating gives

$$d \ln \tilde{P}_k = \sigma_k d \ln CC + \beta_k^S d \ln H_k$$

- σ_k is initial share of construction costs CC in housing prices $\tilde{P}_k$
- β_k^S is inverse home supply elasticity
- Use long-differences (1970-2000) to focus on long-run effects
- Supply equation in changes:

$$\Delta \ln \tilde{P}_k = \sigma_k \Delta \ln CC_k + \beta_k^S \Delta \ln H_k + \sum R_k^j + \varepsilon_k$$

Geography and Home Price Elasticities

TABLE III
HOUSING SUPPLY: GEOGRAPHY AND LAND USE REGULATIONS

	$\Delta\log(P)$ (supply): 1970–2000					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta\log(Q)$	0.650	0.336	0.305	0.060		
	(0.107)***	(0.116)***	(0.146)***	(0.215)		
Unavailable land $\times\Delta\log(Q)$		0.560	0.449	0.511	0.516	−5.329
		(0.118)***	(0.140)***	(0.214)***	(0.116)***	(0.904)***
Log(1970 population) $\times$						0.481
unavailable land $\times\Delta\log(Q)$						(0.117)***
log(WRI)$\times\Delta\log(Q)$				0.237	0.268	0.301
				(0.130)*	(0.068)***	(0.066)***
$\Delta\log(Q)\times$ ocean			0.106			
			(0.065)			
Midwest	−0.099	−0.041	−0.022	−0.015	−0.009	0.002
	(0.054)*	(0.052)	(0.054)	(0.055)	(0.050)	(0.049)
South	−0.236	−0.170	−0.163	−0.129	−0.116	−0.115
	(0.065)***	(0.062)***	(0.062)***	(0.069)*	(0.050)**	(0.048)**
West	0.016	0.057	−0.022	0.059	0.069	0.035
	(0.076)	(0.072)	(0.054)	(0.072)	(0.063)	(0.046)
Constant	0.550	0.594	0.594	0.528	0.601	0.061
	(0.055)***	(0.052)***	(0.052)***	(0.058)***	(0.046)***	(0.045)***

Notes. Standard errors in parentheses. The table shows the coefficient of 2SLS estimation of a metropolitan housing supply equation. On the left-hand side, I try to explain changes in median housing prices by metro area between 1970 and 2000, adjusted for construction costs (see theory and text). On the right-hand side, the main explanatory endogenous variable is the change in housing demand [the log of the number of households – log(Q)] between 1970 and 2000. Some specifications interact that endogenous variable with the unavailable land share (due to geography) and the log of the Wharton Regulation Index (WRI), which we treat as exogenous in this table. The instruments used for demand shocks are a shift-share of the 1974 metropolitan industrial composition, the magnitude of immigration shocks, and the log of January average hours of sun. The identifying assumptions are that the covariance between the residuals of the supply equations and the instruments are zero. *significant at 10%; ** significant at 5%; ***significant at 1%.

Geography and Home Price Elasticities

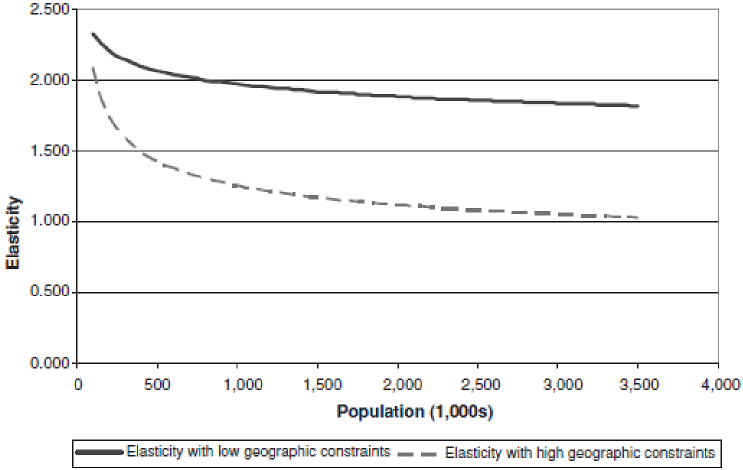


FIGURE I
Impact of Geography on Elasticities by Population

Indirect Effects of Geography

- Regulations are correlated with home price growth. Is this a causal relationship?
- Homevoter hypothesis: land use controls are tools for local homeowners to maximize land prices.
- Consider local supply equation:

$$\Delta \ln \tilde{P}_k = \beta_0 + \beta^{REG} \ln WRI_k \Delta \ln H_k + \bar{\beta} \Delta \ln H_k + \underbrace{\overline{\beta_k^{\delta}} \Delta \ln H_k + \eta_k}_{\xi_k}$$

- Does β^{REG} suffer from omitted variables bias? Depends on what determines it. Imagine

$$\ln WRI_k = \varphi_0 + \varphi_1 \overline{\beta_k^{\delta}} + \varphi_2 \overline{\beta_k^{\delta}} \Delta \ln H_k + \varphi_3 \ln \tilde{P}_k + \mu_k$$

- Probably. Example: higher prices lead to higher regulations ($\varphi_3 > 0$).

Indirect Effects of Geography

TABLE IV
ENDOGENEITY OF LAND USE REGULATIONS

	Log WRI			
	(1)	(2)	(3)	(4)
Unavailable land, 50-km radius	0.134 (0.067)**		−0.174 (0.125)	−0.241 (0.132)*
Unavailable land in growing cities (1940–1970)		0.165 (0.069)**		
Unavailable land in declining cities (1940–1970)		−0.054 (0.153)		
Declining cities dummy (1940–1970)		−0.076 (0.051)		
Unavailable land, 50-km radius × Δlog housing units (1970–2000)			0.451 (0.158)***	0.375 (0.154)**
ΔLog housing price (1970–2000) = log housing price (1970)				0.198 (0.088)**
Log (inspection expenditures/local tax revenues) (1982)	0.051 (0.015)***	0.047 (0.015)***	0.051 (0.015)***	0.041 (0.015)***
Share of Christian "nontraditional" denominations (1970)	−0.308 (0.086)***	−0.304 (0.084)***	−0.314 (0.084)***	−0.291 (0.090)***
Share with bachelor's degree in 1970	0.983 (0.332)***	0.538 (0.342)	0.867 (0.328)***	0.089 (0.404)
Non-Hispanic white share in 1980	0.036 (0.113)	0.08 (0.110)	−0.069 (0.116)	−0.017 (0.120)

Indirect Effects of Geography

TABLE IV
(CONTINUED)

	Log WRI			
	(1)	(2)	(3)	(4)
Midwest	−0.266	−0.307	−0.289	−0.266
	(0.039)***	(0.039)***	(0.039)***	(0.044)***
South	−0.19	−0.222	−0.261	−0.196
	(0.054)***	(0.053)***	(0.058)***	(0.066)***
West	−0.029	−0.08	−0.096	−0.088
	(0.050)	(0.050)	(0.054)*	(0.056)
Constant	1.425	1.471	1.578	−0.759
	(0.137)***	(0.135)***	(0.144)***	(1.055)
Observations	269	269	269	269
R^2	.43	.46	—	—
Method	FGLS	FGLS	2SLS	3SLS

Notes. Standard errors in parentheses. The dependent variable in all regressions is the log of the WRI for each metro area. To deal with heterogeneous sample sizes (strong correlation of WRI values within MSA) columns (1) and (2) use a Feasible Generalized Least Squares (FGLS) procedure, where each observation is weighted proportionally to the inverse of the square error of OLS estimates (which are actually always very close in magnitude and significance). In columns (3) and (4), changes in the log of local housing prices and quantities are instrumented using the demand shocks in Table III (industry shift-share, hours of sun, and immigrant shocks), plus the land unavailability variable. *significant at 10%; **significant at 5%; ***significant at 1%.

Indirect Effects of Geography

TABLE V

HOUSING SUPPLY: ENDOGENOUS REGULATIONS

	$\Delta\log(P)$ (supply)	
	(1)	(2)
Unavailable land $\times \Delta\log(Q)$	0.581	−5.260
	(0.119)***	(1.396)***
Log(1970 population) $\times$ unavailable land $\times \Delta\log(Q)$		0.475
		(0.119)***
Log(WRI) $\times \Delta\log(Q)$	0.109	0.280
	(0.078)*	(0.077)***
Midwest	−0.009	0.002
	(0.049)	(0.048)
South	−0.075	−0.109
	(0.049)	(0.049)**
West	0.149	0.059
	(0.063)	(0.065)
Constant	0.659	0.577
	(0.048)***	(0.048)***

Notes. Standard errors in parentheses. The table shows the coefficient of 2SLS estimation of a metropolitan housing supply equation. The specification and instruments used for demand shocks are as in Table III. Demand shocks are interacted with the unavailable land share (due to geography) and the log of the WRI. The latter variable is treated as endogenous using the share of local public expenditures on "protective inspections" and the share of nontraditional Christian denominations as instruments. Because we are instrumenting for log(WRI)$\times \Delta$ log(Q), I also include the interaction between the regulation and the demand instruments in the IV list. *significant at 10%; **significant at 5%; ***significant at 1%.

Indirect Effects of Geography

- Use Table 5, column (2) to estimate supply elasticities at metro area level.
- Mean elasticity is 1.75 (2.5 unweighted)

Indirect Effects of Geography

TABLE VI
Supply Elasticities (Metro Areas with Population > 500,000)

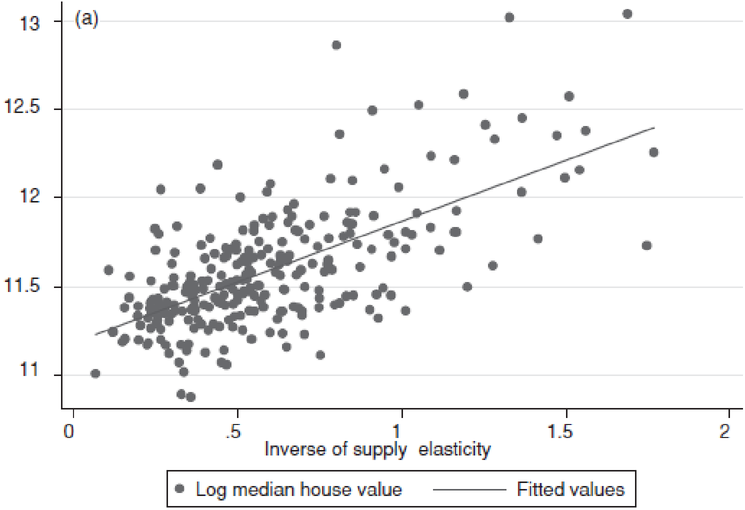
Rank	MSA/NECMA name	Supply elasticity	Rank	MSA/NECMA name	Supply elasticity
1	Miami, FL	0.60	26	Vallejo–Fairfield–Napa, CA	1.14
2	Los Angeles–Long Beach, CA	0.63	27	Newark, NJ	1.16
3	Fort Lauderdale, FL	0.65	28	Charleston–North Charleston, SC	1.20
4	San Francisco, CA	0.66	29	Pittsburgh, PA	1.20
5	San Diego, CA	0.67	30	Tacoma, WA	1.21
6	Oakland, CA	0.70	31	Baltimore, MD	1.23
7	Salt Lake City–Ogden, UT	0.75	32	Detroit, MI	1.24
8	Ventura, CA	0.75	33	Las Vegas, NV–AZ	1.39
9	New York, NY	0.76	34	Rochester, NY	1.40
10	San Jose, CA	0.76	35	Tucson, AZ	1.42
11	New Orleans, LA	0.81	36	Knoxville, TN	1.42
12	Chicago, IL	0.81	37	Jersey City, NJ	1.44
13	Norfolk–Virginia Beach–Newport News, VA–NC	0.82	38	Minneapolis–St. Paul, MN–WI	1.45
14	West Palm Beach–Boca Raton, FL	0.83	39	Hartford, CT	1.50
15	Boston–Worcester–Lawrence–Lowell–Brockton, MA–NH	0.86	40	Springfield, MA	1.52
16	Seattle–Bellevue–Everett, WA	0.88	41	Denver, CO	1.53
17	Sarasota–Bradenton, FL	0.92	42	Providence–Warwick–Pawtucket, RI	1.61
18	Riverside–San Bernardino, CA	0.94	43	Washington, DC–MD–VA–WV	1.61
19	New Haven–Bridgeport–Stamford–Danbury–Waterbury, CT	0.98	44	Phoenix–Mesa, AZ	1.61
20	Tampa–St. Petersburg–Clearwater, FL	1.00	45	Scranton–Wilkes-Barre–Hazleton, PA	1.62
21	Cleveland–Lorain–Elyria, OH	1.02	46	Harrisburg–Lebanon–Carlisle, PA	1.63
22	Milwaukee–Waukesha, WI	1.03	47	Bakersfield, CA	1.64
23	Jacksonville, FL	1.06	48	Philadelphia, PA–NJ	1.65
24	Portland–Vancouver, OR–WA	1.07	49	Colorado Springs, CO	1.67
25	Orlando, FL	1.12	50	Albany–Schenectady–Troy, NY	1.70

Indirect Effects of Geography

TABLE VI
(CONTINUED)

Rank	MSA/NECMA name	Supply elasticity	Rank	MSA/NECMA name	Supply elasticity
51	Gary, IN	1.74	74	Atlanta, GA	2.55
52	Baton Rouge, LA	1.74	75	Akron, OH	2.59
53	Memphis, TN–AR–MS	1.76	76	Richmond–Petersburg, VA	2.60
54	Buffalo–Niagara Falls, NY	1.83	77	Youngstown–Warren, OH	2.63
55	Fresno, CA	1.84	78	Columbia, SC	2.64
56	Allentown–Bethlehem–Easton, PA	1.86	79	Columbus, OH	2.71
57	Wilmington–Newark, DE–MD	1.99	80	Greenville–Spartanburg–Anderson, SC	2.71
58	Mobile, AL	2.04	81	Little Rock–North Little Rock, AR	2.79
59	Stockton–Lodi, CA	2.07	82	Fort Worth–Arlington, TX	2.80
60	Raleigh–Durham–Chapel Hill, NC	2.11	83	San Antonio, TX	2.98
61	Albuquerque, NM	2.11	84	Austin–San Marcos, TX	3.00
62	Birmingham, AL	2.14	85	Charlotte–Gastonia–Rock Hill, NC–SC	3.09
63	Dallas, TX	2.18	86	Greensboro–Winston–Salem–High Point, NC	3.10
64	Syracuse, NY	2.21	87	Kansas City, MO–KS	3.19
65	Toledo, OH	2.21	88	Oklahoma City, OK	3.29
66	Nashville, TN	2.24	89	Tulsa, OK	3.35
67	Ann Arbor, MI	2.29	90	Omaha, NE–IA	3.47
68	Houston, TX	2.30	91	McAllen–Edinburg–Mission, TX	3.68
69	Louisville, KY–IN	2.34	92	Dayton–Springfield, OH	3.71
70	El Paso, TX	2.35	93	Indianapolis, IN	4.00
71	St. Louis, MO–IL	2.36	94	Fort Wayne, IN	5.36
72	Grand Rapids–Muskegon–Holland, MI	2.39	95	Wichita, KS	5.45
73	Cincinnati, OH–KY–IN	2.46			

Indirect Effects of Geography



Indirect Effects of Geography

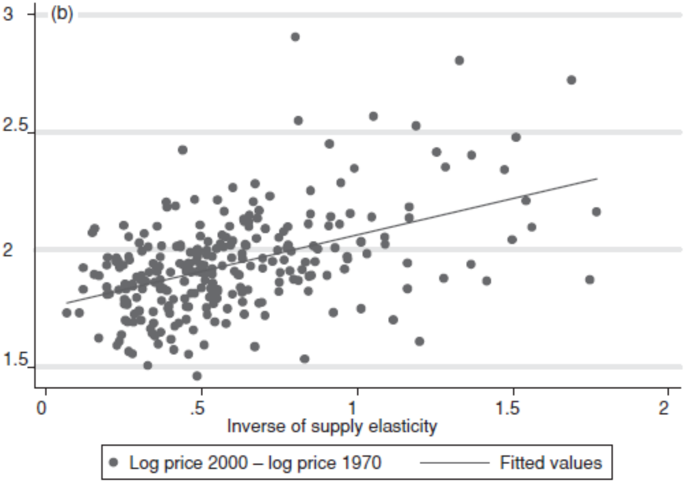


FIGURE II
Estimated Elasticities and Home Values (2000)
(a) Levels, (b) changes.

Conclusions

- Geography matters for urban growth

Introduction

- Urban economics a sibling field
  - Public finance, applications also involve labor, trade

Introduction

- Motivating Fact: population of US central cities fell by 17 percent between 1950 and 1990
  - Central city measured by 1950 geography
  - Metropolitan areas as a whole had 72 percent population growth
- This paper's claim: interstate highway system played a large role
- Theoretical framework: monocentric city model (Alonso)

Introduction

- Results
  - Each highway passing through a central city decreases its population by 18 percent
  - In absence of interstate highways, central city population would have grown by 8 percent
  - Highways can explain a third of the change in central city population relative to metro area population

Introduction

- Research design
  - Instrument actual highways with 1947 plan
  - Intuition: reflects national and aggregate concerns, not local suburbanization supply or demand

Introduction

TABLE I

AGGREGATE TRENDS IN SUBURBANIZATION, 1950–1990

	1950	1960	1970	1980	1990	Percent change 1950–1990
Panel A: Large MSAs						
MSA population	92.9	115.8	134.0	144.8	159.8	72
Total CC population	44.7	48.5	51.3	49.2	51.0	14
Constant geography CC population	44.7	44.2	42.6	37.9	37.1	−17
N for constant geog. CC population	139	132	139	139	139	
Panel B: Large Inland MSAs						
MSA population	39.2	48.9	57.0	65.0	73.5	88
Total CC population	16.8	19.7	22.1	22.1	23.2	38
Constant geography CC population	16.8	16.5	15.4	13.3	12.5	−26
N for constant geog. CC population	100	94	100	100	100	
Total U. S. population	150.7	178.5	202.1	225.2	247.1	64

Notes: All populations are in millions. CC stands for central city. The sample includes all metropolitan areas (MSAs) of at least 100,000 people with central cities of at least 50,000 people in 1950. The sample in Panel B excludes MSAs with central cities located within 20 miles of a coast, major lake shore, or international border. MSA populations are for geography as of year 2000. Constant geography central city population uses 1950 central city geography. Census tract data are not available to build constant geography central city populations for some small cities in 1960. These cities are assigned a population of 0 for constructing the aggregates. Reported total U. S. population excludes Alaska and Hawaii.

Introduction

- Notes on Table 1
  - Central city decline more rapid for inland cities
    - Implies space is important to explain falling urban density
  - Constant-population geography cities shrank faster than moving-boundary cities
    - Implies space is important to explain falling urban density

Background: Highways and Instruments

- Federal Aid Highway Act of 1944
  - Required plan for national highways. Goal:
  - "[C]onnect by routes as direct as practicable, the principal metropolitan areas, cities, and industrial centers, to serve the national defense, and to connect at suitable border points with routes of continental importance in the Dmoinion of Canada and the Republic of Mexico..."
  - Variables considered (in order): nationwide distribution of population; manufacturing activity; agricultural production; location of post-World War II employment; strategic highway network drawn by War Department in 1941; location of military and naval establishments; interregional traffic demand.
  - Does not metion commuting anywhere

Background: Highways and Instruments

- Process
  - States submit proposals for their portion of interstate highways
  - States negotiate with federal roads administration
  - By 1947 plan: 37,324 miles proposed
  - Most/much was built
- Most funding from Federal Aid Highway and Highway Revenue Acts of 1956 and thereafter
- 1956 Interstate Highway Act: expand to 41,000 miles, federal government pays 90 percent
  - Incorporatea highways primarily for commuting
    - Those are bad for the research design
- By 1990: 43,420 miles
  - 1950-1990, miles in main sample of 139 metro areas grew 161 to 16,716

Background: Highways and Instruments

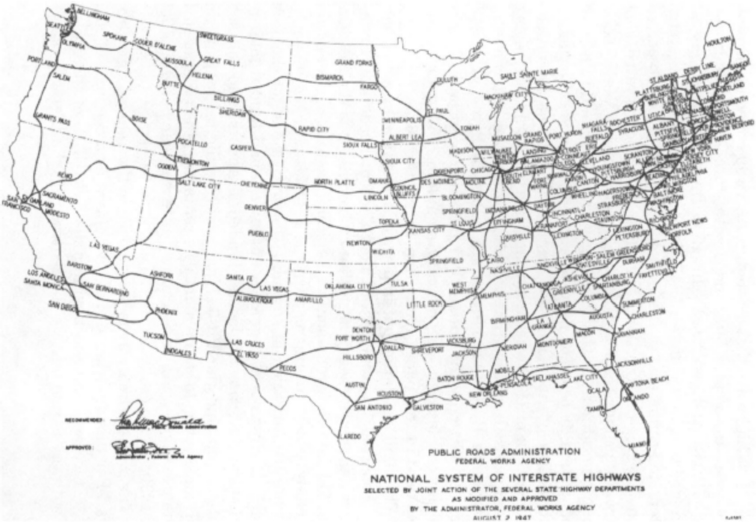


FIGURE I
The Projected System of Interstate Highways in 1947

Background: Highways and Instruments

- Each month, states reported completion of segments via "PR-511" data
- These data also explain funding source (1956 Federal Highway Act, other federal, local, or toll segments)
- Combine with digitalmap of interstate system
- Key measure of highways: rays
  - Ray means: highway connects central city with region outside the central city

Background: Highways and Instruments

- Definition of ray
  - If straight line through city, counts as two rays
  - If terminates near central city, counts as one
  - Must pass within one mile of CBD
  - Must serve a significant area of MSA outside the 1950-definition central city
  - Highway that splits near border of central city: multiple rayw
  - Two highways that pass within one mile of CBD then converge: one ray in direction of convergence
  - Limited-access expressways not part of interstate: count as rays
  - Ray exists in an MSA-year if at least one mile was open to traffic
- Why so much detail? [discuss]

Background: Highways and Instruments

- Summary statistics on rays
  - Median #rays among MSAs: 0 in 1950 and 1960, 2 thereafter
  - Mean #rays among MSAs: 2.34 in 1990
- Robustness check: highways within four miles of CBD
  - Correlation between growth of two measures 1950-1990: 0.76
  - Correlation between growth in interstate and total rays: 0.91

Background: Highways and Instruments

- How to use different highway segments? Some reflect reverse causality:
  - State and local funding of highways probably responds to local commuting demand
  - States build urban sections earlier than rural sections
  - Long-difference regressions: 1947 number of rays instruments for actual 1950-1990 change in rays.
  - Panel regressions: instrument for highways with number of rays in a year if states allocated federal funds uniformly across planned highways

Background: Highways and Instruments

- Is the 1947 plan a valid instrument?
  - [valid instrument]: Requires that 1947 plan developed for trade and national defense, not metro area development
  - [conditionally valid instrument]: Also requires data on [all] variables that affect suburbanization and are correlated with the plan
    - [examples?]
  - Requires a strong first-stage: planned rays must be good predictor of actual rays

Background: Highways and Instruments

- Support for instrument validity
  - Regressing 1950-1990 change in rays on 1940-1950 change in log MSA population gives coefficient of 1.85 (se 90.95)
  - Uing planned rays as dependent variable gives coefficient of -0.43 (se = 0.87).
  - 1950 percent high school graduate, 1950 non-white population don't predict highway construction or planned rays
  - [nice to show table making this comprehensive and clear]
- Level of MSA population does predict rays
  - Larger cities had more rays and more planned rays
- Summary: government planned more highway construction in large cities, not fast-growing cities or cities with pressure for segregation
  - Important to control for MSA population in regressions
  - [Careful way of dealing with instrument validity]

Background: Highways and Instruments

TABLE II
FIRST STAGE RESULTS
LARGE MSAs IN 1950

Panel A: Long difference 1950–1990			
	Change in Rays		
	1	2	3
Rays in the plan	.677	.526	.510
	(.074)**	(.076)**	(.074)**
1950 Central city radius		.325	.306
		(.071)**	(.072)**
Change in simulated log income			−.939
			(1.819)
			.856
Change in log of MSA population			(.279)**
Constant	.866	.218	.463
	(.218)**	(.247)	(1.231)
Observations	139	139	139
R-squared	.38	.47	.50

Background: Highways and Instruments

Panel B: Full panel 1950–1990				
	Rays		Smoothed Rays	
	1	2	4	5
Smoothed rays in plan	.826	.448	.833	.712
	(.029)**	(.040)**	(.023)**	(.030)**
Log simulated income		1.563		.198
		(.234)**		(.179)
Log MSA population		.170		.591
		(.200)		(.154)**
MSA Fixed effects	Yes	Yes	Yes	Yes
Groups	132	132	132	132
R-squared	.55	.84	.67	.88

Notes: Panel A shows the first stage results for the regressions in Table IV. Panel B shows the first stage results for the regressions in Table VI. Listed specification numbers match those in the corresponding second stage tables. "Smoothed rays" is calculated by multiplying the stock of rays in 1990 in each MSA by the fraction of these rays' mileage that is completed at each point in time. "Smoothed rays in the plan" is calculated by multiplying the number of rays in the 1947 plan by the fraction of federally funded highway mileage in the 1956 Federal Aid Highway Act completed at each point in time. All coefficients remain significant when estimated using the more selected sample in Table I Panel B.

Data

- [This paper is a good example: it's a benefit to a paper to be
  - Methodical in definitions and data construction
  - Transparent in explaining step-by-step what was done so it can be replicated
  - Explain why choices were made when decisions must be made
  - As comprehensive as humanly possible
  - Goal is to get the most accurate possible answer to the question, done in the best way possible]

Data

- Central city data from City and County Data Books, aggregated to MSAs
- Hold MSA geography constant at 2000 definitions
  - Minutae: combine PMSAs with no central city t adjacent PMSAs with a central city
  - In New England, use census-defined NECMAs
  - ...
- Central city is largest central city as of 2000
- Other central cities relegated to suburbs
- Half of central cities annexed land over time
  - [yuck]
  - Use 1950, 1970, 1980 printed maps to match central cities over time
  - Identify census tracts within 1950 central city
  - Neighborhood Change Database for 1970-1990

Data

- Use 2004 Road Atlas to identify highway rays
  - Interstate or non-interstate highways within 1 mile of central business district
- Form PR-511 Database: opening dates of each segment of interstate highway system
  - State, route number, milepost, segment legnth, funding source, date of completing construction
  - Digital version gone so scanned and hand-edited
  - Includes all federal roads and state portions of roads built by 1956
  - Half of locally funded mileage built after 1956
- 1982 Economic Census Geographic Reference Manual: central business district definitions
- 1950-1990 PUMS microdata: skill prices (wages)
  - [Why not use county-year mean income?]

Data

Appendix Table:
Summary Statistics

Variable	1950	1990	1990–1950	Min	Max
Fraction of population living inside 1950 geography central city	.47	.22	−.25	−.51	−.03
	(.16)	(.11)	(.10)		
Fraction of population living inside of central city	.47	.35	−.12	−.42	.33
	(.16)	(.19)	(.14)		
Log population of 1950 geography central city (millions)	−1.82	−2.10	−.28	−.81	.72
	(.94)	(.99)	(.28)		
Log population of politically defined central city (millions)	−1.82	−1.64	.18	−.81	2.22
	(.94)	(1.00)	(.57)		
Log MSA population (millions)	−.99	−.43	.56	−.21	1.95
	(.91)	(.98)	(.40)		
Total number of rays through central city	.07	2.66	2.59	0	7
	(.33)	(1.70)	(1.66)		
Number of interstate rays through central city	.07	2.34	2.27	0	7
	(.33)	(1.58)	(1.54)		
Number of planned rays	2.55	2.55	–	0	7
	(1.51)	(1.51)			
Total rays in the MSA	.14	4.38	4.24	0	14
	(.50)	(2.40)	(2.36)		
Central city radius (miles)	3.18	5.04	1.86	1.16[a]	11.98[a]
	(1.63)	(2.77)	(2.27)		
MSA radius (miles)	25.99	25.99	–	8.48	93.17
	(10.42)	(10.42)			
Simulated mean log income using 1940 employment distribution	9.61	10.27	.66	.56	.89
	(.12)	(.05)	(.06)		
Gini coefficient of simulated income distribution	.33	.37	.04	−.02	.07
	(.03)	(.01)	(.02)		
Median log household income	9.99	10.60	.60	.27	.99
	(.16)	(.15)	(.16)		

Notes: The sample size is 139 and consists of the same MSAs used in the regressions in Table IV. Standard deviations are in parentheses. Min and max refer to the 1950–1990 difference unless the variable does not change over time or has a superscript.

a. This number refers to the min or max of the 1950 level, not the change from 1950 to 1990.

Land Use Model

- Alonso-Mills-Muth (1960s)
- Monocentric city
- Identical agents
- Rental rate of land adjusts with distance from city center to compensate for commute times
- Implications of adding highway:
  - For a given commute time, supply of land shifts out, so city-wide land rental prices fall
    - Decreases population density through price effect
  - Income rises, so if space is a normal good, population density further falls

Land Use Theory

- Implication #1: population in metro areas should spread out along highways
- Implication #2: central city population should increase with metro population and with radius of central city
- Implication #3: central city population should decrease with number of rays

Land Use Theory

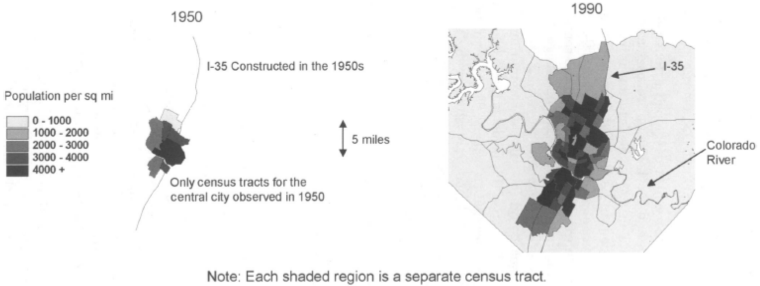


FIGURE II
Development Patterns in Austin, TX.

Land Use Theory

- Austin a simple example: large city with two rays and few political or natural boundaries
  - I-35 opened in 1950s
  - Metro area population grew rapidly 1950-1980
  - Central city population declined 1950-1980 then grew slightly
  - Most residential development 1950-1990 was in North-South direction, not East-West

Land Use Theory

- Regression approach:

$$\log P_{ij} = \alpha_i + \beta dis_{ij}^{cbd} + c dis_{ij}^{hwy} + \varepsilon_{ij}$$

- P_{ij}: population per square mile in tract j of MSA i
- dis_{ij} distance from tract centroid to central business district
- dis^{hwy}: distance to nearest highway

Land Use Theory

TABLE III
THE SPATIAL DISTRIBUTION OF METROPOLITAN AREA POPULATIONS

Panel A: 1970 and 1990 Cross-Sections

Sample		Log population density 1970	Log population density 1990
Large MSAs in 1950 (36,250 tracts, 139 MSAs)	Distance to CBD	−.132 (.001)**	−.114 (.001)**
	Distance to highway	−.014 (.002)**	−.019 (.002)**
Large MSAs in 1950 with central cities at least 20 miles from a coast or border (17,336 tracts, 100 MSAs)	Distance to CBD	−.134 (.002)**	−.117 (.001)**
	Distance to highway	−.055 (.003)**	−.054 (.003)**

Land Use Theory

Panel B: Evolution between 1970 and 1990		
Sample		ΔLog population density
Large MSAs in 1950 (36,250 tracts, 139 MSAs)	Distance to CBD	.021 (.000)**
	ΔDistance to highway	−.015 (.002)**
Large MSAs in 1950 with central cities at least 20 miles from a coast or border (17,336 tracts, 100 MSAs)	Distance to CBD	.021 (.001)**
	ΔDistance to highway	−.008 (.003)**

Notes: Each pair of entries lists coefficients and standard errors from a regression of log population density on the listed variables at the census tract level. All regressions include MSA fixed effects. Regressions in Panel B also include the distance to the nearest highway in 1970. Estimated coefficients on distance to the nearest highway in 1970 are between −0.002 and 0.004. Regressions using the distance to planned highways as an instrument for the distance to observed highways yield similar results. When standard errors are clustered by MSA, results for the larger sample in Panel B and results for the smaller sample in Panel A remain significant at the 5 percent level. Other results are not statistically significant with clustering. Regressions are weighted by the fraction of MSA population that is represented in the tract. Analogous unweighted regressions produce highway distance coefficients that are larger in absolute value. All distances are in miles.

Main Results

$$\Delta \ln N_i^c = \delta_0 + \delta_1 \Delta ray_i + \delta_2 r_{ci} + \delta_3 \Delta w_i + \delta_4 \Delta \ln N_i^{MSA} + \delta_5 \Delta G_i + \varepsilon_i$$

- N^c: 1950-definition central city population
- r_c: central city radius
- N^{MSA}: MSA population
- w_i: mean log annual personal income
- G_i: Gini coefficient for income distribution of MSA i
- Instrument income distribution with 1940 employment shares across one-digit industries times national skill prices in each industry

Main Results

TABLE IV
LONG-DIFFERENCE REGRESSIONS OF THE DETERMINANTS OF CONSTANT GEOGRAPHY CENTRAL CITY POPULATION GROWTH, 1950–1990

Large MSAs in 1950						
	Change in log population in constant geography central cities					
	OLS3	IV1	IV2	IV3	IV4	IV5
Change in number of rays	−.059 (.014)**	−.030 (.022)	−.106 (.032)**	−.123 (.029)**	−.114 (.026)**	−.101 (.046)*
1950 central city radius	.080 (.014)**		.111 (.023)**	.113 (.023)**	.106 (.023)**	.125 (.021)**
Change in simulated log income	.084 (.378)			.048 (.417)	−6.247 (6.174)	−.137 (.480)
Change in log of MSA population	.363 (.082)**			.424 (.094)**	.374 (.079)**	.405 (.108)**
Change in Gini coeff of simulated income					−23.416 (23.266)	
Log 1950 MSA population						−.062 (.062)
Constant	−.640 (.260)*	−.203 (.078)*	−.359 (.076)**	−.588 (.281)*	4.580 (5.091)	−.611 (.265)*
Observations	139	139	139	139	139	139
R-squared	.39	.00	.01	.30	.33	.37

Notes: In columns IV1–IV5, the number of rays in the 1947 plan instruments for the change in the number of rays. Standard errors are clustered by state of the MSA central city. Standard errors are in parentheses. ** indicates significant at the 1 percent level, * indicates significant at 5 percent level. Summary statistics are in the Appendix Table. First stage results are in Table II.

Main Results

TABLE V
COEFFICIENTS FROM LONG-DIFFERENCE REGRESSIONS, 1950–1990

	Census division fixed effects		Specification 1	2	3	4	5
All large MSAs in 1950	No	OLS	.005	−.032	−.059	−.055	−.040
			(.016)	(.015)*	(.014)**	(.014)**	(.016)*
	Yes	OLS	−.012	−.039	−.052	−.051	−.024
			(.014)	(.014)*	(.014)**	(.014)**	(.014)
	Yes	IV	−.026	−.080	−.104	−.104	−.065
			(.020)	(.023)**	(.021)**	(.022)**	(.031)*
All large MSAs in 1950 with central city at least 20 miles from an international border or coast	No	OLS	−.000	−.040	−.062	−.063	−.032
			(.018)	(.019)*	(.017)**	(.016)**	(.019)
	No	IV	−.014	−.101	−.142	−.132	−.098
			(.026)	(.030)**	(.038)**	(.030)**	(.110)
	Yes	OLS	−.019	−.044	−.053	−.053	−.018
			(.017)	(.018)*	(.016)**	(.016)**	(.017)
	Yes	IV	−.028	−.086	−.134	−.130	−.058
			(.023)	(.027)**	(.030)**	(.028)**	(.065)

Notes: Each entry in this table represents the coefficient on rays from a separate regression. The number of rays in the 1947 plan instruments for the change in the number of rays in the IV regressions. The square root of 1950 central city area enters as a control in Specifications 2–5; the change in simulated log income and the change in log MSA population enter in Specifications 3–5; the change in the Gini coefficient enters in Specification 4; log 1950 MSA population enters in Specification 5. Sample sizes are 139 for the top block and 100 for the bottom block. Standard errors are clustered by state of the MSA central city. Standard errors are in parentheses. ** indicates significant at the 1 percent level, * indicates significant at 5 percent level.

Main Results

- OLS coefficient on rays is -0.05, IV coefficient is -0.10
  - This happens in many (most?) published papers: IV$>>$OLS
  - Why?

Main Results

- OLS coefficient on rays is -0.05, IV coefficient is -0.10
  - This happens in many (most?) published papers: IV$>>$OLS
  - Why?
    - Omitted variables
    - Measurement error
    - Publication bias

Main Results

- OLS coefficient on rays is -0.05, IV coefficient is -0.10
  - If problem was reverse causality (cities build highways where suburbanization demand is greatest), IV would be smaller than OLS.
  - So problem is "misspecification of the transportation infrastructure"
    - i.e., measurement error
    - If we include suburban and urban rays, OLS is -0.044 and IV is -0.69.

Main Results

- Benefits of 1950-1990 long difference than of annual panel
  - Long difference can address measurement error
  - Better data in 1950 and in 1990 than in intervening years
  - Short-run versus long-run effects
- Benefits of also looking at annual panel
  - More statistical power
  - Additional evidence from (slightly) different research design

Panel Results

- Decadal data, 1950-1990

TABLE VI
PANEL IV REGRESSIONS OF THE DETERMINANTS OF CONSTANT GEOGRAPHY CENTRAL CITY POPULATION, 1950–1990

	Large MSAs in 1950					
	Log central city population					
	1	2	3	4	5	6
Number of rays	−0.111	−0.142	−0.140			
	(0.016)**	(0.026)**	(0.028)**			
(1990 Rays) × (Fraction of Ray miles completed at t)				−0.097	−0.089	−0.086
				(0.016)**	(0.012)**	(0.013)**
Log simulated income		−0.083	−0.061		−0.288	−0.229
		(0.117)	(0.109)		(0.075)**	(0.077)**
Log MSA population		0.266	0.263		0.294	0.286
		(0.104)*	(0.105)*		(0.100)**	(0.098)**
Simulated Gini coefficient			−0.623			−1.415
			(1.106)			(0.847)
MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.20	0.22	0.22	0.14	0.56	0.57

Notes: The instrument used is (rays in the plan) × (MSA mileage of highways running through the central city at time t)/(MSA mileage of highways running through the central city in 1990). Standard errors are clustered by the state of the central city. Standard errors are in parentheses. ** indicates significant at the 1 percent level, * indicates significant at the 5 percent level. First stage results are in Table II. Each regression includes 132 MSAs with five observations each, one for each year 1950–1990. There are fewer MSAs in this sample than that in Table IV because of lack of census tract data for seven MSAs in 1960.

Other Sensitivity Analysis

- Did 1947 plan respond to early 20th century decline in cities?
  - This would invalidate the instrument
- Relate 1910-1950 population trends to the 1947 instrument
  - Pretrend regression has statistically insignificant coefficient of -0.003
  - [What is standard error? This has 63 cities only.]
- If highway rays increase metro area population, then controlling for it understates effects of highways on central cities
  - So in long-difference regressions, instrument for metro area population growth with Jan temperature and precipitation, or with Bartik (shift-share instrument)
  - In panel regressions, instrument for annual population with its lag
    - [gulp]
  - These give point estimates within 0.01 of main results.

Other Sensitivity Analysis

- Could also measure suburbanization as log central city population share
  - Downsides [small]:
    - Harder to connect estimates to motivating summary stats
    - Less simple interpretation of coefficient on rays
  - Gives slightly larger estimates: -0.10 in main regressions and -0.14 in IV

Conclusions

- Why did cities hollow out in the 1970s? Many explanations
  - School busing and desegregation
  - Central city riots
  - Subsidies to suburbs
  - Highways

Conclusions

- Results
  - Each ray decreases central city population by 9 percent
  - Central city population 1950-1998 fell 28 percent, but without highways would have increased by 8 percent

Conclusions

- Future research: mechanisms
  - Firm suburbanization?
    - 1950: 20 of 40 million MSA jobs were in central cities
    - By 1990, only 27 of 87 million jobs were in central cities
    - Due to workers or shipping costs?
- Other future research: what else caused suburbanization?

Other Comments

- No welfare analysis here. Why?
- This instrument is getting used in lots of papers
- Why is the effect of a ray -0.10? Is there an urban land use model which would rationalize this fact based on transportation costs, housing supply elasticity, urban size, agglomeration elasticity, etc.?
- Main regression is differences on levels.
  - Uncommon but makes sense here.