MIT 14.582: PhD International Economics II
Sp 2026, Lectures 14-15: Economic Geography and Urban Economics (Empirics)

Dave Donaldson

Agglomeration externalities

- In the previous lecture we saw the propensity for economic activity to agglomerate (or "cluster") spatially
  - Both as a whole, and within industries (with within pairs of industries)
- One reason for such phenomena might be *agglomeration externalities*—a causal effect running from local density to local aggregate productivity or utility
  - Synonymous with local aggregate increasing returns to scale

Agglomeration externalities in production

- Letting A_f denote the TFP of some firm f in a given location $i(f)$, and $\{L\}_{i(f)}$ denote the set of all factors employed in location $i(f)$ other than inside firm f, we could write potential agglomeration externality effects as:

$$A_f = g_f(\{L\}_{i(f)})$$

- This could allow for externalities that depend:
  - On the firms (or industries) doing the "sending" (e.g. same industry, cross-industry, benefits of diversity, etc.)
  - On the firm (or industry) doing the "receiving"
  - On the types of factors that generate externalities (e.g. high-skilled labor)
- Other directions that appear occasionally in the literature include:
  - Non-Hicks-neutral externalities (e.g. high-skilled labor benefits more)
  - Externalities that affect nearby locations too (possibly with a spatial decay)
  - Externalities that work with temporal lags

Estimating agglomeration externalities

- How large are agglomeration externalities in production?
  - That will be the goal of today's and next lecture
- Some surveys on this theme:
  - Rosenthal and Strange (2004, *Handbook of Urban and Regional Econ*)
  - Combes and Gobillon (2015, *Handbook of Urban and Regional Econ*)
- Broadly, three approaches:
  1. Estimating agglomeration economies "directly": treatment effect of density on TFP
  2. Estimating agglomeration economies from the extent of agglomeration in an observed spatial equilibrium. (Can think of EG (1997) like this.)
  3. Testing for multiple equilibria and/or multiple steady-states (which are often a consequence of positive agglomeration economies)
- For now, our main focus will be #1
- NB: like most of the literature, we won't have much to say about mechanisms for *why* agglomeration happens

Estimating agglomeration economies directly

- A large literature has argued that if agglomeration economies exist then units of production (and factors) should be more productive if they are surrounded by other producer. Three nice examples:
  - Henderson (2003, JUE) on across-firm (within-location) externalities
  - Moretti (2004, AER) on local (within-city) human capital externalities
  - Arzaghi and Henderson (2008, REStud) on Manhattan's advertising industry
- A central challenge in the literature analogous to that faced in the 'peer effects' literature (e.g. Manski, 1993): does one unit actually affect a proximate unit, or are proximate units just similar on unobservable dimensions?
- We will discuss three recent studies that offer natural experiment approaches to this question.
  - Greenstone, Hornbeck and Moretti (JPE, 2010)
  - Kline and Moretti (QJE, 2014)
  - Peters (ECMA, 2022)
  - (And you saw Faber and Gaubert (AER, 2019) in a recitation.)
- Will focus on "regional" scale for now and turn to "city" scale (neighborhood-level) effects later

Greenstone, Hornbeck and Moretti (2010)

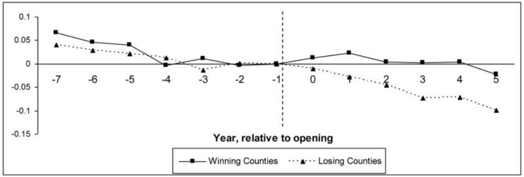
- GHM look at the effect that 'million dollar plants' (huge industrial plants) have on incumbent firms in the vicinity of the new MDP

- Consider the following example (from paper):
  - BMW did worldwide search for new plant location in 1991. 250 locations narrowed to 20 US counties. Then announced 2 finalists: Omaha, NB and Greenville-Spartanburg, SC. Finally, chose Greenville-Spartanburg.
  - Why? BMW says:
    - Low costs of production: low union density, supply of quality workers, numerous global firms in area (including 58 German companies), good transport infrastructure (rail, air, highway, port access), and access to key local services.
    - Subsidy ($115 million) received from local government.

- GHM obtain list of the winner and loser counties for 82 MDP openings and compare winners to losers (rather than comparing winners to all 3,000 other counties, or to counties that look similar on observables).

Greenstone, Hornbeck and Moretti (2010)

All Industries: Winners vs. Losers



Difference: Winners – Losers

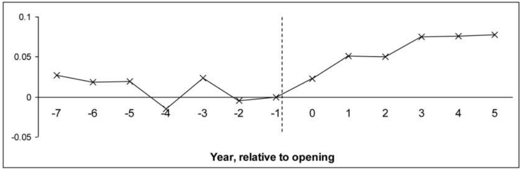


FIG. 1.—All incumbent plants' productivity in winning versus losing counties, relative to the year of an MDP opening. These figures accompany table 4.

Greenstone, Hornbeck and Moretti (2010)

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)

NOTE.—The table reports results from fitting versions of eq. (8) for each of the indicated outcome variables (in logs). See the text for more details. Standard errors clustered at the county level are reported in parentheses.

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Greenstone, Hornbeck and Moretti (2010)

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	

Note.—The table reports results from fitting versions of eq. (8). As a basis for comparison, col. 1 reports estimates from the baseline specification for incumbent plants in all industries (baseline estimates for incumbent plants in all industries, col. 4 of table 5). Columns 2 and 3 report estimates from a single regression, which fully interacts the winner/loser and pre/post variables with indicators for whether the incumbent plant is in the same two-digit industry as the MDP or a different industry. Reported in parentheses are standard errors clustered at the county level. The numbers in brackets are the value (2006 U.S. dollars) from the estimated increase in productivity: the percentage increase is multiplied by the total value of output for the affected incumbent plants in the winning counties.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

Greenstone, Hornbeck and Moretti (2010)

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

NOTE.—The table reports results from fitting versions of eq. (9), which is modified from eq. (8). Building on the model 1 specification in col. 4 of table 5, each column adds interaction terms between winner/loser and pre/post status with the indicated measures of how an incumbent plant's industry is linked to its associated MDP's industry (a continuous version of results in table 7). These industry linkage measures are defined and described in table 2, and here the measures are normalized to have a mean of zero and a standard deviation of one. The sample of plants is that in col. 4 of table 5, but it is restricted to plants that have industry linkage data for each measure. For assigning this linkage measure, the incumbent plant's industry is held fixed at its industry the year prior to the MDP opening. Whenever a plant is a winner or loser more than once, it receives an additive dummy variable and interaction term for each occurrence. Reported in parentheses are standard errors clustered at the county level.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

Implications for agglomeration externalities

- Clearly the functional form we started with ($A_f = g_f(\{L\}_{i(f)})$) is too high-dimensional for realistic empirical work
- A highly simplified and parameterized version that is very popular is the isoelastic form (where $L_{i(f)}$ denotes the total population of location $i(f)$):

$$\ln A_f = \kappa_f + \alpha \ln L_{i(f)}$$

- Here, the agglomeration elasticity α is the object of much interest
- If we interpret GHM's effects as working this way then could deduce:
  - By 5 years, local manuf. plants (0.27 share of local L) see TFP rise by 14%
  - Greenstone and Moretti (2004) use the same natural experiment and find (at 5 years) a 5% increase (though not very precisely estimated) in county-wide employment
  - So even if TFP effect is confined to non-MDP manuf. plants, could expect $\alpha \approx (0.27)(0.14)/(0.05) = 0.76$
- But not at all obvious this is the right functional form $g_f(\cdot)$ to have in mind

Nice follow-on work by Giroud et al (2022)

- Giroud et al (2022) revisit the GHM (2010) findings and document three interesting findings:

1. Re-do baseline GHM estimation on sample of plants that are active in at least one year before and after the MDP opening (GHM required being open for 8 consecutive years prior) and find smaller (but still very non-trivial) estimates
2. Estimate spillovers at different distances from the MDP (rather than "in county")
3. Estimate spillovers onto other plants owned by MDP county firms (regardless of location of those plants)

Giroud et al (2022): Findings #1 and #2

TABLE II

LOCAL PRODUCTIVITY SPILLOVER.

	TFP		
		Unweighted	Distance
	(1)	(2)	(3)
MDP	0.040	0.038	
	(0.016)	(0.014)	
MDP × (<50 miles)			0.043
			(0.015)
MDP × (50 to 100 miles)			0.027
			(0.014)
MDP × (100 to 250 miles)			0.011
			(0.010)
Plant FE	Yes	Yes	Yes
Industry × year FE	Yes	Yes	Yes
Case FE	Yes	Yes	Yes
R-squared	0.88	0.82	0.86
Observations	157,000	157,000	2,209,000

Note: The dependent variable is TFP at the plant level. MDP is an indicator for the winner county that is 1 from the year of the MDP opening onward. In column (3), (<50 miles), (50 to 100 miles), and (100 to 250 miles) are indicators for whether a plant lies within 50 miles, between 50 and 100 miles, and between 100 and 250 miles, respectively, of the MDP. Only the main coefficients of interest are shown. Except for column (2), observations are weighted by plant-level employment. Standard errors are double clustered at the county and year level. The sample period is from 1977 to 1998.

- Analog of column (1) specification in GHM was 0.048
- Column (3) suggests that the spillovers decay spatially. But is this evidence for “weak” spatial spillovers?

Giroud et al (2022): Finding #3

TABLE III

GLOBAL PRODUCTIVITY SPILLOVER.

	TFP		
	(1)	(2)	(3)
MDP	0.018 (0.007)	0.020 (0.008)	0.018 (0.008)
Plant FE	Yes	Yes	Yes
Industry × year FE	Yes	-	-
Industry × county × year FE	-	Yes	Yes
Case FE	Yes	-	Yes
Control group	Plants of runner-up firms	Plants of MC firms in same county	Plants of runner-up firms in same county
R-squared	0.87	0.86	0.88
Observations	1,407,000	1,046,000	423,000

Note: The dependent variable is TFP at the plant level. MDP is an indicator for whether the plant's parent firm has a plant in the winner county before and after the MDP opening. The indicator is 1 from the year of the MDP opening onward. In column (1), the control group consists of all plants of runner-up firms. In column (2), the control group consists of all plants of MC firms in the same county as the treated plant. In column (3), the control group consists of all plants of runner-up firms in the same county as the treated plant. In all three columns, the sample is restricted to MC plants outside the winner and runner-up counties. Only the main coefficients of interest are shown. Observations are weighted by plant-level employment. Standard errors are double clustered at the county and year level. The sample period is from 1977 to 1998.

- Note similar results across various control groups
- Even though these estimates are about half as big as those in Table 2, the typical MDP county firm has 6.3 plants in other counties (conditional on having at least one plant in another county). So this effect is big.

Giroud et al (2022): Finding #3

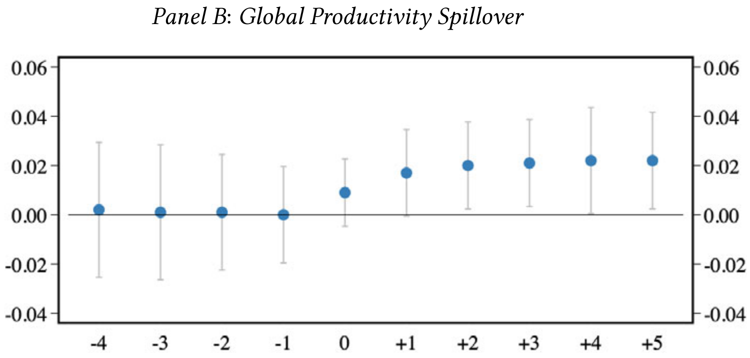


FIGURE 1.—Treatment effect dynamics. This figure shows the coefficient estimates from columns (2) and (4) in Table A.II of Supplemental Appendix A along with 95% confidence intervals. The estimates are obtained using the imputation estimator of Borusyak, Jaravel, and Spiess (2023). The base year is $\tau = -5$.

Further comments on GHM (2010)

- Policy implications (more on this sort of thing in Lecture #23):
  - Was it worth it for a typical county to (use subsidies to) attract an MDP?
  - Greenstone and Moretti (2004) use the same natural experiment as in GHM (2010)—but not the Giroud et al (2022) extension—to estimate the effect on total county land values. It's positive (and pretty big).
  - To the extent that all factors other than land are fully mobile, this should capitalize all of the local net benefits of the MDP.
  - (But could also include fiscal revenues from non-local taxes.)

- Subsequent findings:
  - Monte et al (AER, 2018): use the MDP shock to estimate a commuting elasticity parameter (propensity to commute to a given workplace location if the wage there rises) that we will see in future lectures (on within-city settings)

Kline and Moretti (QJE, 2014)

- A different identification strategy for estimating agglomeration externalities: the 1933-onwards Tennessee Valley Authority
  - Enormous "place-based policy" (again, more in Lecture #23)
  - Perhaps one of the best examples of a "big push" policy ever tried
  - Famous episode in post-Depression (New Deal, etc.) history

- KM (2014) uses this policy (partially) to generate quasi-experimental variation in local size, and hence to estimate agglomeration externalities. But, importantly, they also:
  - Estimate the direct effect of the TVA, which is of substantial independent interest
  - Ask whether the estimated agglomeration externalities take the form that is required for TVA to have an additional impact on national welfare through the fact that it promoted agglomeration.

What exactly was the TVA?

- A big (see Fig 1), ongoing (see Fig 2) place-based policy
- Key components:
  - Lots of public investment in infrastructure—hydroelectric dams, 650-mile navigation canal (1939-45), extensive road network (mostly done by 1950s), new schools, flood-control systems
  - Electricity sold inside TVA at reduced rates
- Which counties were selected into TVA?
  - See Table 1
  - Also potential additional "valley authorities" discussed in Congress in 1940s/1950s but never authorized. KM (2014) construct map of these regions (see Fig A2) based on their reading of the written proposals. Treat this as a placebo.

TVA region

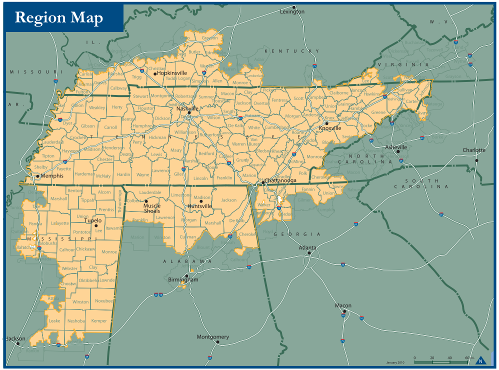


FIGURE I

The TVA Service Area (as of 2010)

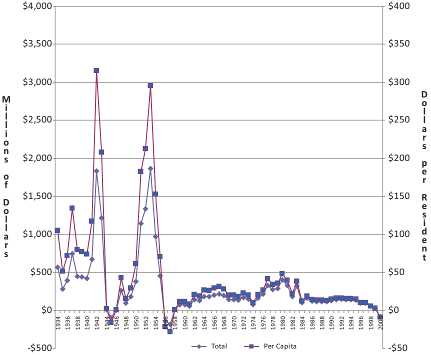


FIGURE II

Federal Transfers to TVA by Year (2000 Dollars)

Federal transfers defined as net federal outlays plus property transfers minus repayments (see Online Appendix for sources).

Estimates of agglomeration externalities

- KM introduce a simple model of spatial equilibrium (a la Rosen-Roback; Roback, 1982; see Glaeser-Gottlieb, 2009 survey covered in recitation). Ingredients:
  - Production function $Y_{it}^M = A_{it}^M (K_{it}^M)^\alpha (F_i^M)^\beta (L_{it}^M)^{1-\alpha-\beta}$ for manufacturing in each US county i and decade t, where capital K and labor L are assumed to be freely mobile and fixed factor F_i is not.
  - Output of this good is freely traded globally. But locations offer exogenous amenities $\bar{u}_{it}$ so labor mobility implies $\bar{u}_{it} w_{it}$ is equalized across all i within each t.
  - TFP given by:
  $$\ln A_{it}^M = g\left(\frac{L_{it-1}^M}{R_i}\right) + \delta_t D_i + \eta_i + \gamma_t + \varepsilon_{it}$$
  where $g(\cdot)$ is the agglomeration function (assumed to be a function of *lagged* manufacturing employment L^M density; R_i is area of county), and D_i is TVA treatment dummy
- So TVA has direct effects (that vary over time, δ_t) and potentially also indirect effects via agglomeration (i.e. via attracted L_{it-1}^M and $g'(\cdot) > 0$).

More structural estimation details

- Letting $\Delta\varepsilon_{it} = \lambda X_i + \nu_{it}$ for pre-determined controls X_i, can write short-run labor demand function as (with "tilde" denoting original variables divided by β):

$$\begin{aligned}
\ln L_{it}^M - \ln L_{it-1}^M &= -\frac{1-\alpha}{\beta}(\ln w_{it} - \ln w_{it-1}) + (\widetilde{\delta}_t - \widetilde{\delta}_{t-1})D_i \\
&+ \frac{1}{\beta}\left[g\left(\frac{L_{it-1}^M}{R_i}\right) - g\left(\frac{L_{it-2}^M}{R_i}\right)\right] + \widetilde{\lambda}X_i \\
&+ \widetilde{\gamma}_t - \widetilde{\gamma}_{t-1} + \widetilde{\nu}_{it}
\end{aligned}$$

- In practice:
  - Estimate on sample of all US counties and four decadal changes, 1960-2000
  - Assume D_i conditionally exogenous (established fairly convincingly, e.g. using neighbors to TVA counties or placebo VAs, in earlier reduced-form part of paper)
  - Proxy $g(x)$ with three-knot spline in $\ln(x)$
  - IV for each spline term k with lagged instrument: $Z_{it}^{(k)} \equiv g_k\left(\frac{L_{it-2}^M}{R_i}\right) - g_k\left(\frac{L_{it-3}^M}{R_i}\right)$
  - Calibrate (SR) LD elasticity $-\frac{1-\alpha}{\beta}$ from labor literature (Hammermesh 1993 chapter, values: 1-1.5). And use labor share $\beta = 0.47$.

KM (2014): Estimates of $\frac{1}{\beta} g(\cdot)$

TABLE VI

STRUCTURAL ESTIMATES OF AGGLOMERATION FUNCTION (LOG BASIS)

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
Change in log manufacturing density spline components:						
Low	0.173	0.147	0.146	0.443	0.400	0.396
	(0.037)	(0.037)	(0.037)	(0.102)	(0.108)	(0.107)
				[177.17]	[159.14]	[157.20]
Medium	0.221	0.227	0.226	0.456	0.440	0.438
	(0.045)	(0.044)	(0.045)	(0.124)	(0.123)	(0.124)
				[106.74]	[109.55]	[110.13]
High	0.143	0.012	0.141	0.466	0.467	0.453
	(0.051)	(0.050)	(0.050)	(0.150)	(0.150)	(0.151)
				[206.66]	[204.69]	[200.36]
Log manufacturing wages	−1.5	−1.5	−1.5	−1.5	−1.5	−1.5
TVA	0.007	0.012	0.008	−0.003	0.002	−0.002
	(0.014)	(0.014)	(0.014)	(0.012)	(0.013)	(0.012)
Regional trends	no	no	yes	no	no	yes
1940 manufacturing density	no	yes	yes	no	yes	yes
Decade effects	yes	yes	yes	yes	yes	yes
Controls for 1920 and 1930 characteristics	yes	yes	yes	yes	yes	yes
p-value equal slopes	0.2483	0.1298	0.1038	0.9545	0.6695	0.7171
p-value slopes equal 0	0.0000	0.0000	0.0000	0.0002	0.0007	0.0016
N	6,057	6,057	6,057	5,935	5,935	5,935

Notes. Dependent variable is change in log county manufacturing employment. Manufacturing density is manufacturing employment per square mile. Standard errors clustered by state in parentheses. Angrist-Pischke cluster robust first stage *F*-statistic in brackets. All estimates weighted by 1950 county population. "Low" refers to spline component corresponding to log density below 60th percentile of 1980 distribution, "Medium" to log density between 60th and 85th percentile of 1980 distribution, and "High" to log density above 85th percentile of 1980 distribution. Spline coefficients give the elasticity of labor demand with respect to lagged manufacturing density over the relevant range. The instruments are changes in the spline components of log manufacturing density lagged by two decades.

- So implied function $g(\cdot)$ is pretty log-linear: vindication for canonical isoelastic form $\ln A_i = \kappa + \alpha \ln L_i$
- Implied (col 4, medium estimate) value of $\alpha \approx \beta \times 0.456 = 0.214$
- Fact that $g(\cdot)$ is isoelastic implies that marginal product of labor equalized everywhere (in KM's model) so no misallocation, despite externalities
  - So if direct TVA effect D_i is same everywhere then wouldn't matter where it was put—and no sense of a "Big Push" after all
  - (But more on this in Lectures #21-23)

Peters (2022)

- Examines the post-WWII forced resettlement to West Germany of ethnic Germans living in Eastern Europe
  - Happened mostly from 1945-48. By 1950, 8 million were resettled (20% of W. German pop.)
  - "One of largest forced population movements in world history"
  - Argues that allocation of refugees to W. German counties was as-good-as-random: driven by housing availability as assessed by UK/US governments, refugees were told where to go (and held there by mobility restrictions)

- Did this raise Y/L in areas of West Germany that saw inflows of refugees?

Peters (2022): The Setting

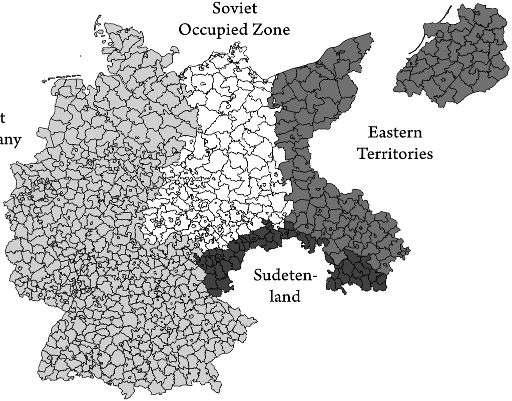


FIGURE 1.—The German Reich in 1939. *Note*: The figure shows the German Reich in the boundaries of 1939. The light-gray shaded part in the west is the area of to-be West Germany. The medium-gray shaded parts in the east are the Eastern Territories of the German Reich. The dark shaded area in the southeast is the Sudetenland. The white shaded part is the area of the Soviet Occupied Zone. The intra-regional spatial units are counties.

Peters (2022): Sources of pop. changes in W. Germany

TABLE III

THE POPULATION OF WEST GERMANY: 1939–1950.

	Population Losses 1939–1950				Population Gains 1939–1950			
Population 1939	Military Losses	Civilian Losses	Non-military Deaths	Others	Refugees	Inflows from SOZ	Births	Population 1950
39.3 m	2 m	0.4 m	5.2 m	0.5 m	7.9 m	1.5 m	7 m	47.6 m

Note: The table reports aggregate population trends in West Germany between 1939 and 1950. “Inflows from SOZ” are individuals who fled the Soviet Occupied Zone. Source: Edding (1951, p. 2).

Peters (2022): Regional variation within West Germany

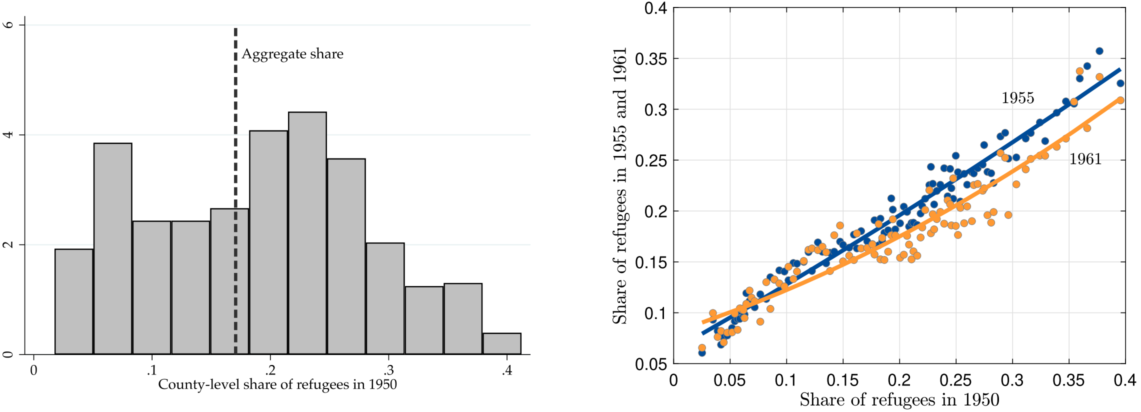


FIGURE 2.—The Heterogeneity and Persistence of Refugee Inflows. *Note*: The left panel shows the distribution of the share of refugees in 1950 across counties. The right panel shows the correlation between the share of refugees in 1950 and 1955 (blue) and 1961 (orange) as binned scatter plots for 100 percentiles of the refugee share in 1950.

Peters (2022): Characteristics of refugees

TABLE IV

CHARACTERISTICS OF REFUGEES AND NATIVES.

	Male Share	Age Distribution			Educational Attainment			
		<15	20–65	65+	Elem. School	High School	Vocational School	College
Natives	46.5	20.4	68.5	11.1	66.8	26.3	4.9	1.9
Refugees	46.9	21.9	68.5	9.7	67.4	25.6	4.9	2.1

Note: The first panels report the share of males and the age distribution in 1958. The last panel reports the distribution of educational attainment of the cohort born before 1920 as observed in the 1970 census. These individuals were at least 25 years old in 1945 and hence completed their educational attainment prior to the expulsion. Source: Besser (2007).

Peters (2022): Effects on Y/L and L

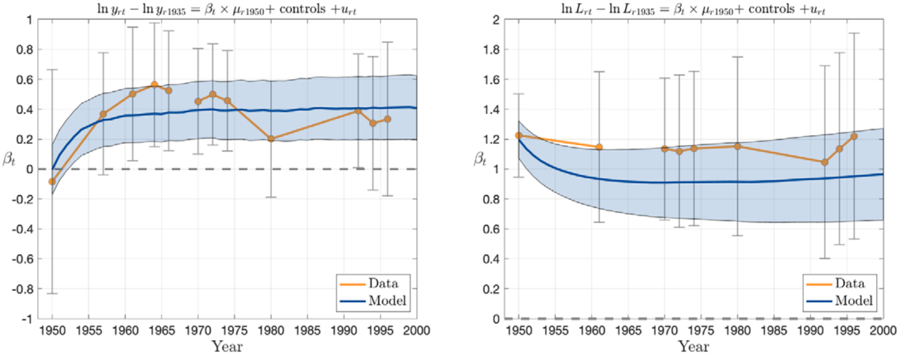


FIGURE 4.—Model Fit: the Dynamic Effects on Income and Population Growth. *Note*: The figures report the coefficient β of the regression $y_{rt} = \beta_t \mu_{r1950} + x'_{rt}\gamma + u_{rt}$ for different time horizons and for income growth (left panel) and population growth (right panel) as dependent variables. The vector x_{rt} controls for state fixed effects, population density in 1939, war-time destruction, log income per capital (population) in 1939, and the distance to the inner German border (see columns 2 or 6 of Table VI). For both the model and the data, I also report 95% confidence intervals.

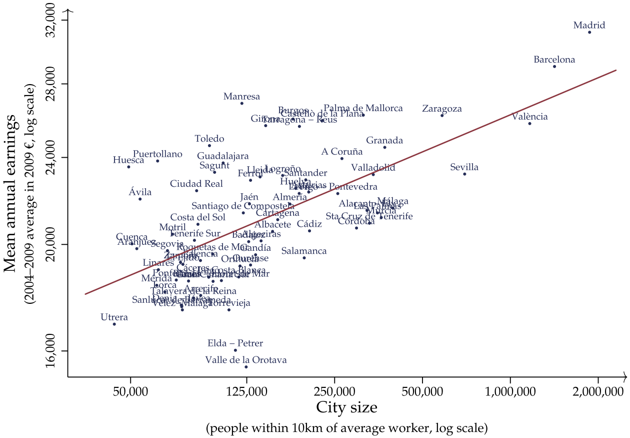
$\Rightarrow$ implied "α" is $\approx (0.6 \text{ to } 0.2)/(1.2 \text{ to } 1.1) = 0.17 \text{ to } 0.55$

Complementary evidence from “mover designs”

- The research designs we have studied so far have involved changing the attributes of a “place” and then observing the impact on economic outcomes
- An alternative is to track how individual-level outcomes (e.g. wages) change as individuals change their place
- This is the strategy pursued in several recent papers
  - Glaeser and Mare (JoLE, 2001)
  - Combes, Duranton and Gobillon (JUE, 2018)
  - de la Roca and Puga (REStud, 2017)
  - Card, Rothstein and Yi (2021)
- We will focus on the last two (which extend ideas in the earlier studies)

de la Roca and Puga (2017): Nominal wages are higher (on average) in big cities

(Also replicated many times in many countries; see recitation for Glaeser and Gottlieb's (JEL 2009) treatment for US cities.)



Mover designs

- The basic idea (following the influential analog for estimating firm-specific effects on individual-level wages, in Abowd, Kramarz and Margolis (ECMA, 1999)) is to estimate

$$\ln w_{ict} = \alpha_c + \mu_i + \beta x_{it} + \varepsilon_{ict}$$

- Where
  - w_{ict} is the *nominal* wage of individual i who works in city c at time t
  - α_c is a city-level fixed effect
  - μ_i is an individual-level fixed effect
  - x_{it} is a vector of time-varying *observed* components of worker "skill" (age, experience, perhaps education if we can observe it, etc.)

- If our goal is to identify α_c (e.g. to then project these on city sizes), which individuals contribute to its identification? (See Kline, 2026, *Handbook of Labor Economics* chapter.)

- Under what assumptions could we hope to obtain consistent (for fixed C, but large N and/or T) estimates of α_c? (Ditto.)

de la Roca and Puga (2017): “Event Study” figure

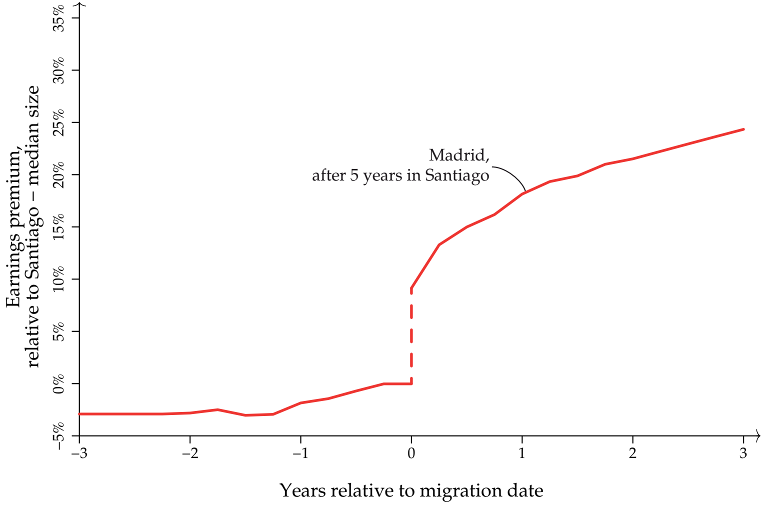


FIGURE 6

Non-parametric pre- and post-migration earnings profile relative to median-sized city

de la Roca and Puga (2017): Controlling for individual-level heterogeneity

TABLE 1
Estimation of the static city size earnings premium

	(1)	(2)	(3)	(4)
Dependent variable	Log earnings	City indicator coefficients column (1)	Log earnings	City indicator coefficients column (3)
Log city size		0.0455 (0.0080)***		0.0241 (0.0058)***
City indicators	Yes		Yes	
Worker fixed effects	No		Yes	
Experience	0.0319 (0.0005)***		0.1072 (0.0018)***	
Experience^2	−0.0006 (0.0000)***		−0.0014 (0.0000)***	
Firm tenure	0.0147 (0.0006)***		0.0042 (0.0004)***	
Firm tenure^2	−0.0005 (0.0000)***		−0.0003 (0.0000)***	
Very-high-skilled occupation	0.7752 (0.0062)***		0.2350 (0.0057)***	
High-skilled occupation	0.4976 (0.0046)***		0.1758 (0.0040)***	
Medium-high-skilled occupation	0.2261 (0.0031)***		0.0873 (0.0029)***	
Medium-low-skilled occupation	0.0542 (0.0021)***		0.0152 (0.0019)***	
University education	0.2014 (0.0037)***			
Secondary education	0.1084 (0.0022)***			
Observations	6,263,446	76	6,263,446	76
R^2	0.4927	0.2406	0.1144	0.1422

(This drop by about 50% is common in other settings too.)

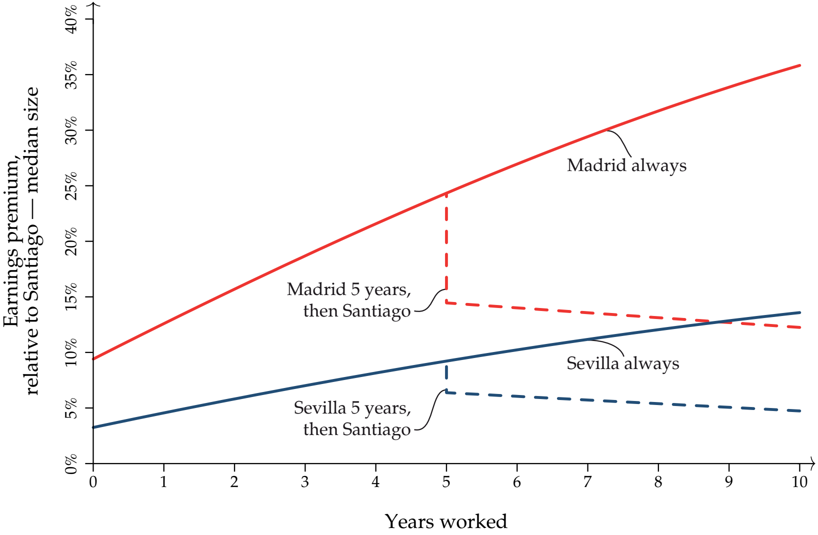
Extending the mover design

- de la Roca and Puga (2017) actually focus most on a dynamic specification:

$$\ln w_{ict} = \alpha_c + \mu_i + \sum_j \delta_{jc} e_{ijt} + \beta x_{it} + \varepsilon_{ict} \tag{1}$$

- Where
  - e_{ijt} is the number of years that individual i has spent working in city j up to date t
  - δ_{jc} is a city pair-specific return-to-experience parameter (how much wages when working in c are affected by prior experience in any city j
- Why do this?

de la Roca and Puga (2017): "Learning by working in big cities"



Panel (a) Profiles allowing for learning benefits of bigger cities

de la Roca and Puga (2017)

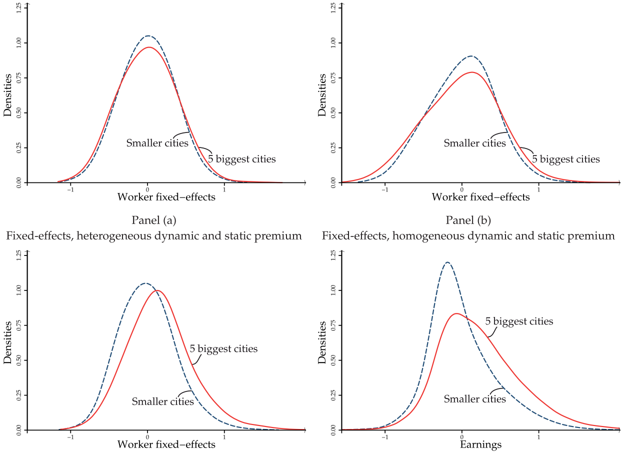
TABLE 2

Estimation of the dynamic and static city size earnings premia

	(1)	(2)	(3)
Dependent variable	Log earnings	Initial premium (city indicator coefficients column (1))	Medium-term premium (initial + 7.7 years local experience)
Log city size		0.0223	0.0510
		(0.0058)***	(0.0109)***
City indicators	Yes		
Worker fixed effects	Yes		

(So the 50% drop we saw earlier wasn't due to sorting, but due to omitting the city-specific experience effects.)

de la Roca and Puga (2017): Surprisingly little sorting (once you adjust for city-specific learning slopes)



Panel (a)
Fixed-effects, heterogeneous dynamic and static premium

Panel (b)
Fixed-effects, homogeneous dynamic and static premium

Panel (c)
Fixed-effects, static premium

Panel (d)
Earnings

FIGURE 8
Comparisons of worker fixed-effects distributions across cities

de la Roca and Puga (2017): Do workers learn (from big cities) at heterogeneous rates?

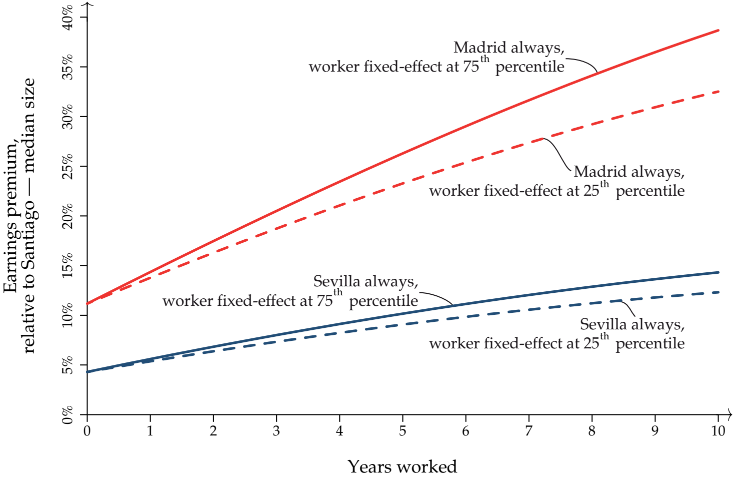


FIGURE 7

Earnings profiles relative to median-sized city, high- and low-ability worker

US evidence from Card, Rothstein and Yi (AEJ Applied, 2025)

- CRY use US employer-employee matched data (LEHD) from 2010-18 (quarterly) to estimate extended versions of de la Roca and Puga (2017)'s model
- One main point they make is that an alternative model that may be of interest is the original AKM (firm-based) model (with $f(i,t)$ the worker's firm):

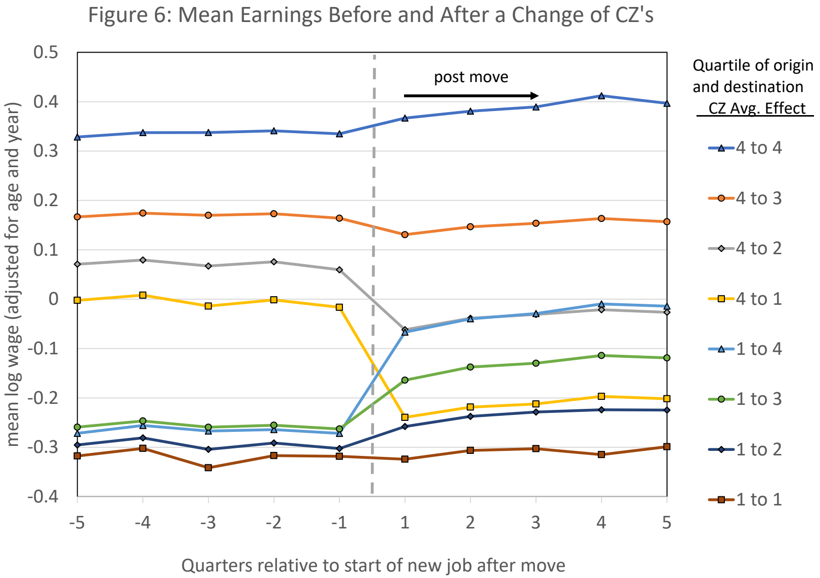
$$\ln w_{ict} = \alpha_{f(i,t)} + \mu_i + \beta x_{it} + \varepsilon_{it}$$

- In this model, our earlier model of equation (1), with α_c instead of $\alpha_{f(i,t)}$, is misspecified. In particular, if you define the city wage premium as

$$\Psi_c \equiv \frac{\sum_{c(f)=c} N_f \alpha_f}{\sum_{c(f)=c} N_f}$$

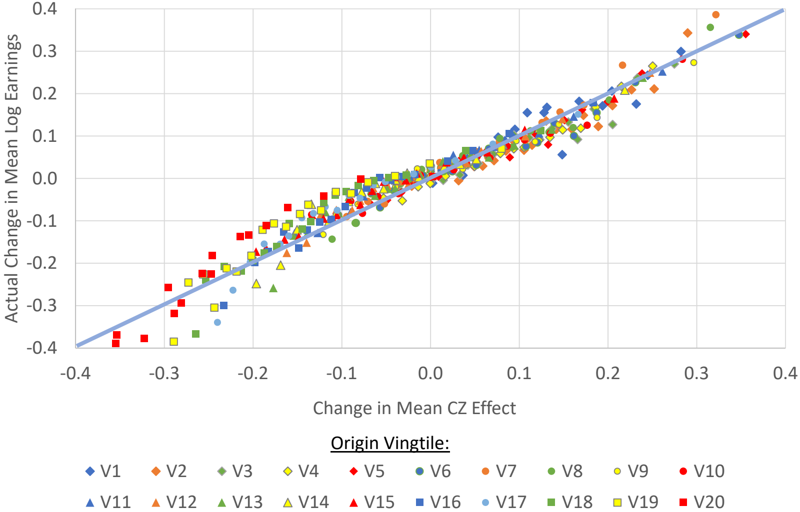
  then α_c will not be an unbiased estimator of Ψ_c unless people move to/from randomly chosen firms within each city
- Instead, CRY document a "local ladder" effect: people moving from low-wage city to a high-wage city tend to be moving from a relatively good (within the origin city) firm to a relatively bad (within the destination) city.

Card et al (2021): "Event Studies" figure

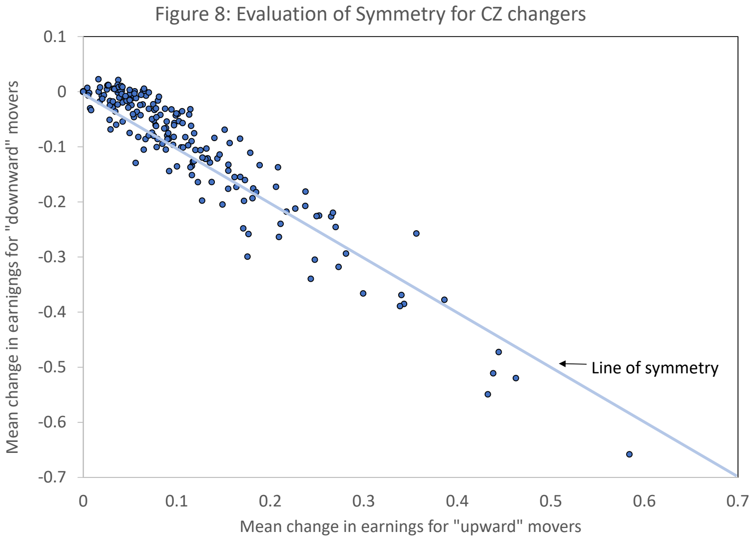


Card et al (2021): Test of AKM-like functional form

Figure 7: Predicted and Actual Changes in Wages for CZ Movers,
by Origin and Destination Vingtile of Average CZ Effect



Card et al (2021): Test for symmetry of AKM-like functional form



More on Card et al (2021)

- CRY extend the de la Roca and Puga (2017) model to model the "city" (c) effect, α_c, to actually be a city-times-industry effect, α_{ck}
- They then use cross-industry-and-city movers to test (and find numerous pieces of evidence consistent with) a model

  $$\alpha_{ck} = \alpha_c + \alpha_k + \epsilon_{ck}$$

  where ϵ_{ck} is as good as random
- In other words: NYC pays more than NOLA, not because it has more workers in finance (an industry with a relatively high wage nationally), but because workers in all of its industries (not just finance) are on average paid more than they are in the same industry in NOLA
- However, relative to de la Roca and Puga (2017), CRY do not estimate city-specific learning functions
  - So not clear whether these are rejected by the event study figures in CRY (i.e. learning by working in big cities more important in Spain than in US) or just not studied in CRY

Other Literature

- Firm-level heterogeneity across cities:
    - Combes, Duranton and Gobillon (ECMA, 2012)
    - Gaubert (AER, 2017)
- Consumption-side externalities:
    - Diamond (AER, 2016)
    - Handbury and Weinstein (REStud, 2015)

Ideas for Further Research

- Still need more/better natural experiments for estimating agglomeration functions
- Endogenous amenities/social spillovers have received much less attention than endogenous productivity
- Extensions of agglomeration functions that need more exploration:
    - More tests of the functional form beyond KM
    - Industry-specific functions
    - Factor-specific functions (E.g. skill-biased agglomeration effects? See Baum-Snow, Pavan and Freedman, 2018)
    - Different spatial scales
- Comparing micro- (e.g. worker-level) and macro- approaches (see Combes and Gobillon, 2015 survey)
- Is (all of) what we're identifying really an externality?
- Which mechanisms are more important for driving agglomeration externalities?
- Dynamics of agglomeration functions (e.g. Allen and Donaldson, 2019)