Two empirical papers on agglomeration

Motoaki Takahashi

Graduate Urban Economics at the University of Osaka

Introduction

- So far we discussed quantitative spatial models (and the Rosen-Roback model).
- Now we study two empirical papers on agglomeration (or spatial wage gaps).
- Agglomeration of factories: Greenstone et al. (2010).
- Agglomeration (vs sorting, skill composition) of workers: Combes et al. (2008).

Agglomeration: an idea (1)

- In short, agglomeration is increasing returns to scale.
- To fix the idea, we consider two situations.
- The easiest possible example is the production function

$$F(L_i) = A_i L_i^{\alpha},$$

  where
  - A_i is the parameter of productivity in city i,
  - L_i is the labor input in city i,
  - $\alpha > 1$ is the parameter of (increasing) returns to scale.
- Here we assumed that each city has one (representative) firm which hires all workers there.
- But this production function is not strictly concave.
- So the first order condition does not characterize the firm's optimization.

Agglomeration: an idea (2)

- Another approach is that each infenitesimal firm exhibits decreasing or constant returns to scale, but a city as a whole exhibits increasing returns to scale.
- Suppose that there are N_i mass of firms.
- All firms have the same productivity in city i but produce different varieties.
- The production function of firm $\varphi \in [0, N_i]$ is

$$F_\varphi(l_\varphi) = A(L_i) \cdot l_\varphi^\beta,$$

  where
  - l_φ is firm φ's labor input,
  - $A(\cdot)$ is the productivity as a function of city-level labor $L_i = \int_0^{N_i} l_\varphi d\varphi$,
  - $\beta \leq 1$ is the parameter of (constant or decreasing) returns to scale.
- Now the first order condition characterizes firms' optimization.

Caveats

- The two examples above are just to set up a basic idea.
- These are not directly related to the two empirical papers we will look at.

Greenstone, Hornbeck, and Moretti (2010)

Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings

Introduction

- The authors want to estimate causal effects of a large manufacturing plant on incumbent plants nearby.
  - The main outcome variable is total factor productivity (TFP).
- The ideal experiment would be that they build large manufacturing plants in randomly selected U.S. counties.
- And they compare factories' TFP between selected and non-selected counties.
- This is obviously impossible.
- What they compare is "winning" counties that attracted a large manufacturing plant with "losing" counties that were the new plant's runner-up choice.
  - They found a magazine that tells these winning and losing counties.
- Winning and losing counties have similar trends in TFP prior to the new plant opening.

Case study: BMW

- Bavarian Motor Works (BMW) considered 250 potential sites for its new plant before 1991.
- BMW announced that it had narrowed the list to 20 U.S. counties in 1991.
- Six months later, BMW announces that two finalists in the competition were Greenville-Spartanburg, South Carolina, and Omaha, Nebraska.
- In 1992, BMW announced that it would site the plant in Greenville-Spartanburg.

Winners and losers

- The authors mainly collect data from the corporate real estate journal *Site Selection*.
- Each issue of this journal includes an article titled "Million Dollar Plants."
- These articles describe how a large plant decided where to locate.
- That is, these articles report the county that the plant chose ("winners") and the runner-up counties ("losers").
  - In the BMW example, the winner is reenville-Spartanburg, South Carolina, and I guess the loser is Omaha, Nebraska.

Data sources

- The "Million Dollar Plants" articles reveal the "winner" county and one or two runner-up ("loser") county.
- The authors indentify the MDPs in the Standard Statistical Establishment List (SSEL).
  - The Census Bureau's "most complete, current, and consistent data for U.S. business establishment."
- And they matched the plants to the Annual Survey of Manufactures (ASM) and the Census of Manufactures (CM) from 1973–98.
- They use the ASM and CM for information on incumbent establishments in winner and loser counties, too.

Summary statistics on winner and loser counties

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

In short, losing counties are a lot similar to winning counties than "average" counties.

Econometric model (1)

- Assume the following Cobb-Douglas production function

$$Y_{pijt} = A_{pijt} L_{pijt}^{\beta_1} (K_{pijt}^{B})^{\beta_2} (K_{pijt}^{E})^{\beta_3} M_{pijt}^{\beta_4}, \tag{1}$$

where

  - p: plant, i: industry, j: case (winner and losers for a MDP), t: year,
  - Y_{pijt}: the total value of shipments minus changes in inventories,
  - A_{pijt}: TFP,
  - L_{pijt}: total labor hours of production,
  - K_{pijt}^{B}: building capital stock,
  - K_{pijt}^{E}: machinery and equipment capital stock,
  - M_{pijt}: the dollar value of material.

Econometric model (2)

- The authors impose the following functional form on TFP

$$
\begin{aligned}
\ln(A_{pijt}) =& \delta 1(\text{Winner})_{pj} + \psi \text{Trend}_{jt} + \Omega[\text{Trend}_{jt} \times 1(\text{Winner})_{pj}] \\
& + \kappa 1(\tau \geq 0)_{jt} + \gamma[\text{Trend}_{jt} \times 1(\tau \geq 0)_{jt}] \\
& + \theta_1[1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt}] \\
& + \theta_2[\text{Trend}_{jt} \times 1(\text{Winner})_{pj} \times 1(\tau \geq 0)_{jt}] \\
& + \alpha_p + \mu_{it} + \lambda_j + \epsilon_{pijt},
\end{aligned} \tag{2}
$$

  where
  - $1(\text{Winner})_{pj}$ is a dummy equal to one if plant p is in a winner county,
  - τ denotes years and is normalized so that $\tau = 0$ is a MDP's opening,
  - Trend_{jt} is just a (case-specific) time trend.
- The most important paramter in our context is θ_1.

Models 1 and 2

- The authors sometimes make restrictions that

$$\psi = \Omega = \gamma = \theta_2 = 0.$$

  - This restriction rules out differential trends.
  - It is a difference-in-difference estimator.
  - This is "model 1."
- The full model characterized by (1) and (2) is called "model 2."

Yet another model

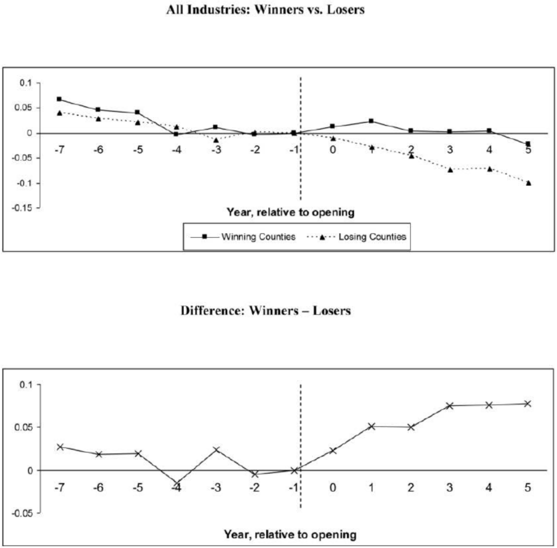
- But, it seems that the authors have another econometric model.
- Basically they regress the log of output on the log of inputs, the plant, sector-year, case fixed effects, and the event time indicators from $\tau = -7$ to $\tau = 5$.
- The coefficients on the event time indicators mean TFP in winning and losing counties relative to the year before the MDP opened.

Table 4

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		

This is from the third model.

Figure 1



This is from the third model.

Table 5

	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

This is from models 1 and 2.

Interpretation of table 5

- Model 1 implies an increase in TFP of roughly 4.8 percent in incumbent plants in winning counties.
- Model 2 implies that the MDP opening is associated with an approximately 12 percent increase in TFP 5 years later.

Combes, Duranton, and Gobillon (2008)

Spatial wage disparities: Sorting matters!

- Core question: Why do wages differ across places?
- Three channels: (i) **skills/sorting**, (ii) **endowments**, (iii) **local interactions/agglomeration**.
- Main message: **sorting matters a lot**; ignoring worker heterogeneity biases "agglomeration" estimates upward.

Conceptual framework: people vs places

- Spatial wage gaps can reflect:
  1. **Skills composition** (workers sort across areas)
  2. **Endowments** (place productivity)
  3. **Interactions** (agglomeration)
- Unified reduced form: wage of worker i in area a and industry k combines place productivity and skills.
- Empirical strategy: two-stage decomposition to separate (i) worker heterogeneity/sorting from (ii) area productivity drivers.

Stage 1 (individual data): estimate "net" local wage indices

First-stage wage equation (area-year FE + worker FE + industry FE)

$$\log w_{i,t} = \beta_{a(i,t),t} + \mu_{k(i,t)} + \tilde{I}_{a(i,t),k(i,t),t}\gamma_{k(i,t)} + \tilde{X}_{i,t}\varphi + \delta_i + \varepsilon_{i,t}.$$

- δ_i: worker fixed effect (unobserved ability); $\beta_{a,t}$: area-year "place" index net of worker/industry.
  - Description of the other variables are in the next page.
- Construct **net wage** for an "average worker in an average industry":

$$\log w_{a,t}^{net} \equiv W_t + \hat{\beta}_{a,t}.$$

- Key finding: net-wage disparities are much smaller than mean-wage disparities $\Rightarrow$ **skills explain 40–50%** of spatial wage gaps (Table 3).
- Strong sorting: corr(mean worker FE in area, detrended area FE) ≈ 0.29.

Table 2

Effect of	Std. dev.	Simple correlation with:		
		log w	δ	$\beta - \theta$
log real wage (log w)	0.367	1.00	0.78	0.26
residuals (ϵ)	0.166	0.45	0.00	0.00
worker effects ($\delta + X\varphi$)	0.294	0.80	0.98	0.09
worker fixed effects (δ)	0.284	0.78	1.00	0.10
age ($X\varphi$)	0.058	0.23	0.08	0.00
industry fixed effects (μ)	0.043	0.25	0.16	0.05
within-industry interactions ($\widetilde{I}_k \gamma_k$)	0.024	−0.01	0.00	−0.45
within-industry share of professionals	0.011	0.16	0.12	0.29
within-industry establishments	0.019	−0.13	−0.08	−0.62
specialisation	0.017	0.03	0.02	−0.13
area fixed effects (β)	0.140	0.34	−0.05	0.55
de-trended area fixed effects ($\beta - \theta$)	0.065	0.26	0.10	1.00
time (θ)	0.118	0.26	−0.11	0.10

Table 3

	Mean wage	Net wage
(Max – Min)/Min	0.74	0.38
(P90 – P10)/P10	0.21	0.14
(P75 – P25)/P25	0.11	0.06
Coefficient of variation	0.08	0.05

Stage 2 (area panel): what drives place effects?

Second-stage regression on estimated area effects

$$\beta_{a,t} = w_0 + \theta_t + I_{a,t}\gamma + E_{a,t}\alpha + \upsilon_{a,t}.$$

- Between-industry interactions include (notably) **employment density** and **market potential**.
- Variance decomposition: interactions explain most of the systematic variation in $\beta_{a,t}$; endowments/amenities weak.
- Preferred magnitude: elasticity of wages w.r.t. **employment density** about **3%** (Table 6: ≈ 0.03 in 2SLS).
- Market potential matters but less robustly; amenity/endowment proxies (sea/mountain/lake/heritage) are statistically present yet small overall.

Table 5

Effect of	Std. dev.	Simple correlation with:		
		$\log w$	δ	$\beta - \theta$
between-industry interactions ($I\gamma$)	0.077	0.22	0.12	0.90
density	0.067	0.20	0.12	0.84
land area	0.024	–0.15	–0.08	–0.62
diversity	0.002	–0.04	–0.06	–0.31
market potential	0.036	0.19	0.08	0.78
amenities ($E\alpha$)	0.011	–0.10	–0.06	–0.48
residuals (η)	0.029	0.04	–0.08	0.03

Table 6

Regression	(1) Levels OLS 1	(2) Levels FGLS	(3) Levels OLS 2	(4) Levels 2SLS	(5) First-Dif. OLS	(6) First-Dif. 2SLS
log *Density*	0.0371[a]	0.0357[a]	0.0322[a]	0.0302[a]	0.0349[a]	0.0289
	(0.0008)	(0.0010)	(0.0007)	(0.0063)	(0.0043)	(0.0175)
log *Area*	0.0113[a]	0.0106[a]	0.0218[a]	0.0041	–	–
	(0.0014)	(0.0016)	(0.0013)	(0.0154)		
log *Diversity*	0.0020	0.0006	−0.0046[b]	−0.0407[c]	−0.0047	−0.0296
	(0.0023)	(0.0025)	(0.0020)	(0.0208)	(0.0032)	(0.0200)
log *Potential*			0.0351[a]	0.0244[a]	0.1385[a]	0.1427[c]
			(0.0014)	(0.0042)	(0.0474)	(0.0715)
Sea			0.0111[a]	0.0004	–	–
			(0.0033)	(0.0046)		
Mountain			0.0333[a]	0.0209[a]	–	–
			(0.0032)	(0.0041)		
Lake			−0.0254[a]	−0.0263[a]	–	–
			(0.0054)	(0.0088)		
Heritage			−0.0091[b]	−0.0202[a]	–	–
			(0.0043)	(0.0068)		
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (within time)	60%	–	72%	–	–	–

Implications: why "sorting matters"

- Spatial wage gaps = **place productivity differences magnified by sorting**: high-δ_i workers concentrate in high-$\beta_{a,t}$ areas.
- If you ignore worker fixed effects (use aggregate or limited controls), you attribute worker composition to agglomeration:
  - Density coefficient roughly **doubles** in aggregate-style estimates (e.g., $\sim$6% vs $\sim$3%).
- Bottom line:
  1. **Big role for sorting/skills** (40–50% of disparities).
  2. **Moderate role for agglomeration** (density $\sim$3%).
  3. **Small role for endowments** (in their measures).

References I

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.