# 14.582: International Trade II — Lecture 17: Economic Geography (Empirics II)

# Plan for Today's Lecture

1. First of two lectures on estimating magnitude of agglomeration externalities

# Why is output so agglomerated?

Three broad explanations:

1. Some production input is exogenously agglomerated.
   - Natural resources (as in the wine industry in EG (1997))
   - Institutions ("exogenous"?)
2. Some consumption amenity is exogenously or endogenously agglomerated
   - Nice places to live (for place-based amenities that are non-tradable)
   - People (i.e. workers) just like to live near each other
   - Some non-tradable amenities that are endogenously provided but with IRTS in those goods' production functions (e.g. opera houses)
3. Some production input agglomerates endogenously
   - Some positive externality (i.e. spillover) that depends on proximity. This almost surely explains Silicon Valley, Detroit, Boston biotech, carpets in Dalton, etc.
   - This is what is usually meant by the term, 'agglomeration economies'
   - This source of agglomeration has attracted the greatest interest among economists.

# What are sources of possible (production-side) agglomeration economies?

- The literature on this is enormous.
  - Probably begins in earnest with Marshall (1890)
  - Survey in Duranton and Puga (2004, *Handbook of Urban and Regional Econ*)
- Typically 3 forces for potential agglomeration economies:
  1. Thick markets (reduce search costs and idiosyncratic risk) for imperfectly tradable inputs (e.g. workers)
  2. Increasing returns to scale combined with trade costs (on either inputs or outputs) that increase with distance
  3. Knowledge spillovers that decrease on distance

# Empirical work on the causes of agglomeration

- Recent surveys on this in:
  - Redding (2010, J Reg. Sci. survey)
  - Rosenthal and Strange (2004, *Handbook of Urban and Regional Econ*)
  - Head and Mayer (2004, *Handbook of Urban and Regional Econ*)
  - Overman, Redding and Venables (2004, *Handbook of International Trade*)
  - Combes et al textbook, *Economic Geography*
  -

- Broadly, three approaches:
  1. Estimating agglomeration economies directly
  2. Estimating agglomeration economies from the extent of agglomeration in an observed spatial equilibrium.
  3. Testing for multiple equilibria (which is often a consequence of agglomeration economies)

# 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 producers
- 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 with this approach is an analogy to the challenge that faces 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?
- Greenstone, Hornbeck and Moretti (JPE, 2010) consider a natural experiment approach to this question.
  - See also Greenstone and Moretti (2004) on how the same natural experiment affected total county land values (i.e. a measure of the welfare effects of agglomeration economies).

# 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).

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 |

Note.—For each case to be weighted equally, counties are weighted by the inverse of their number per case. Similarly, plants are weighted by the inverse of their number per county multiplied by the inverse of the number of counties per case. The sample includes all plants reporting data in the ASM for each year between the MDP opening and 8 years prior. Excluded are all plants owned by the firm opening an MDP. Also excluded are all plants from two uncommon two-digit SIC values so that subsequently estimated clustered variance matrices would always be positive definite. The sample of all U.S. counties excludes winning counties and counties with no manufacturing plant reporting data in the ASM for 9 consecutive years. These other U.S. counties are given equal weight within years and are weighted across years to represent the years of MDP openings. Reported *t*-statistics are calculated from standard errors clustered at the county level. *t*-statistics greater than 2 are reported in bold. All monetary amounts are in 2006 U.S. dollars.

# Greenstone, Hornbeck and Moretti (2010)

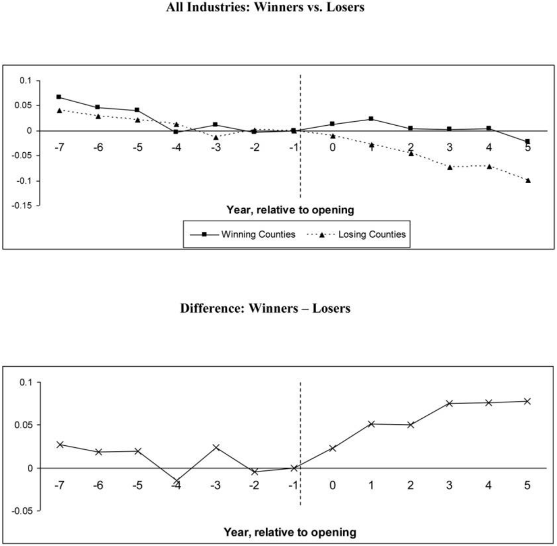



FIG. 1.—All incumbent plants' productivity in winning versus losing counties, relative to the year of an MDP opening. These figures accompany table 4.

# Greenstone, Hornbeck and Moretti (2010)

TABLE 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 |

NOTE.—The table reports results from fitting several versions of eq. (8). Specifically, entries are from a regression of the natural log of output on the natural log of inputs, year by two-digit SIC fixed effects, plant fixed effects, and case fixed effects. In model 1, two additional dummy variables are included for whether the plant is in a winning county 7 to 1 years before the MDP opening or 0 to 5 years after. The reported mean shift indicates the difference in these two coefficients, i.e., the average change in TFP following the opening. In model 2, the same two dummy variables are included along with pre- and post-trend variables. The shift in level and trend are reported, along with the pre-trend and the total effect evaluated after 5 years. In cols. 1, 2, and 5, the sample is composed of all manufacturing plants in the ASM that report data for 14 consecutive years, excluding all plants owned by the MDP firm. In these models, additional control variables are included for the event years outside the range from $\tau = -7$ through $\tau = 5$ (i.e., −20 to −8 and 6 to 17). Column 2 adds the case fixed effects that equal one during the period that $\tau$ ranges from −7 through 5. In cols. 3 and 4, the sample is restricted to include only plants in counties that won or lost an MDP. This forces the industry by year fixed effects to be estimated solely from plants in these counties. For col. 4, the sample is restricted further to include only plant by year observations within the period of interest (where $\tau$ ranges from −7 to 5). This forces the industry by year fixed effects to be estimated solely on plant by year observations that identify the parameters of interest. In col. 5, a set of 47 plant openings in the entire country were randomly chosen from the ASM in the same years and industries as the MDP openings (this procedure was run 1,000 times, and reported are the means and standard deviations of those estimates). For all regressions, plant by year observations are weighted by the plant's total value of shipments 8 years prior to the opening. Plants not in a winning or losing county are weighted by their total value of shipments in that year. All plants from two uncommon two-digit SIC values were excluded so that estimated clustered variance-covariance matrices would always be positive definite. In brackets is the value in 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. 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.

# 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.

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 |

NOTE.—The table reports results from fitting three regressions. In panel A, the dependent variables are the log of number of establishments and the log of total manufacturing output in the county, based on data from the Census of Manufactures. Controls include county, year, and case fixed effects. Reported are the county-level difference-in-difference estimates for receiving an MDP opening. Because data are available every 5 years, depending on the census year relative to the MDP opening, the sample years are defined to be 1–5 years before the MDP opening and 4–8 years after the MDP opening. Thus, each MDP opening is associated with one earlier date and one later date. The col. 1 model is weighted by the number of plants in the county in years −6 to −10, and the col. 2 model is weighted by the county's total manufacturing output in years −6 to −10. In panel B, the dependent variable is log wage and controls include dummies for age by year, age squared by year, education by year, sex by race by Hispanic by citizen, and case fixed effects. Reported is the county-level difference-in-difference estimate for receiving an MDP opening. Because data are available every 10 years, the sample years are defined to be 1–10 years before the MDP opening and 3–12 years after the MDP opening. As in panel A, each MDP opening is associated with one earlier date and one later date. The sample is restricted to individuals who worked more than 26 weeks in the previous year, usually work more than 20 hours per week, are not in school, are at work, and work for wages in the private sector. The number of observations reported refers to unique individuals: some Integrated Public Use Microdata Series county groups include more than one Federal Information Processing Standard (FIPS), so all individuals in a county group were matched to each potential FIPS. The same individual may then appear in more than one FIPS, and observations are weighted to give each unique individual the same weight. 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.

# Kline and Moretti (QJE, 2014)

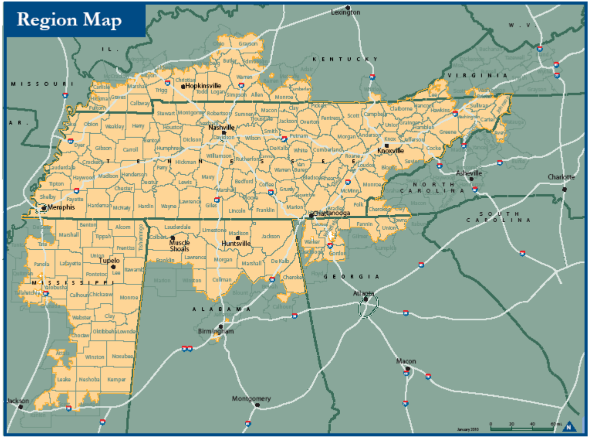
- A different identification strategy for estimating agglomeration externalities: the 1933-onwards Tennessee Valley Authority
  - Enormous "place-based policy"
  - Perhaps one of the best examples of a "big push" policy ever tried
  - Famous episode in post-Depression (New Deal, FDR, etc) history
- KM (2014) uses this policy to generate quasi-experimental variation in local size, and hence to estimate agglomeration externalities. But, importantly, they also:
  - Of course estimate the direct effect of the TVA, which is of substantial independent interest
  - Also ask whether the 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

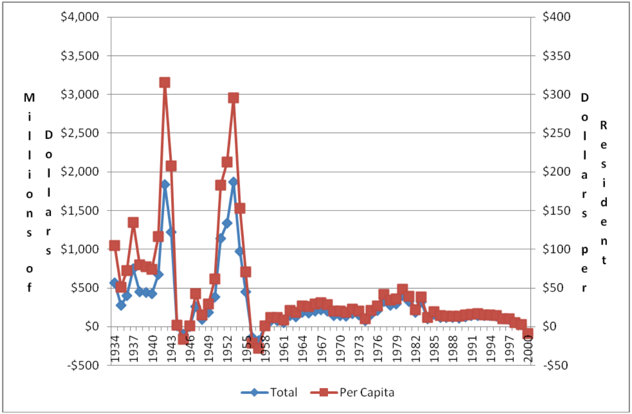
**Figure I: The TVA Service Area**




Notes: Figure depicts TVA service area as of 2010.

# TVA spending over time

**Figure II: Federal Transfers to TVA by Year (2000 Dollars)**




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

# TVA Covariates (i.e. "Balance Table")

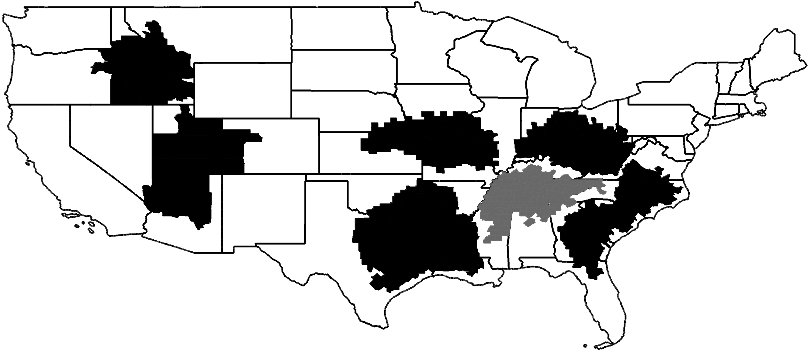
**Table I: Summary Statistics**

| | Overall | | | | Trimmed Sample | |
|---|---|---|---|---|---|---|
| | TVA | Non-TVA | Non-TVA South | Non-TVA Proposed Authorities | Non-TVA | Non-TVA South |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 1930 Characteristics | | | | | | |
| Log Population | 9.991 | 9.977 | 9.989 | 9.940 | 9.905 | 9.979 |
| Log Employment | 8.942 | 8.967 | 8.959 | 8.908 | 8.881 | 8.947 |
| Log # of Houses | 8.445 | 8.508 | 8.455 | 8.466 | 8.442 | 8.445 |
| Log Average Manufacturing Wage | 1.406 | 1.802 | 1.545 | 1.685 | 1.728 | 1.538 |
| Manufacturing Employment Share | 0.075 | 0.090 | 0.080 | 0.077 | 0.080 | 0.078 |
| Agricultural Employment Share | 0.617 | 0.455 | 0.541 | 0.510 | 0.487 | 0.547 |
| % White | 0.813 | 0.885 | 0.722 | 0.830 | 0.863 | 0.724 |
| % Urbanized | 0.153 | 0.280 | 0.233 | 0.216 | 0.242 | 0.215 |
| % Illiterate | 0.088 | 0.045 | 0.092 | 0.060 | 0.051 | 0.092 |
| % of Whites Foreign Born | 0.002 | 0.059 | 0.013 | 0.020 | 0.030 | 0.011 |
| Log Average Farm Value | 5.252 | 5.646 | 5.386 | 5.552 | 5.579 | 5.370 |
| Log Median Housing Value | 9.271 | 9.581 | 9.360 | 9.452 | 9.516 | 9.358 |
| Log Median Contract Rent | 8.574 | 9.030 | 8.679 | 8.834 | 8.934 | 8.672 |
| % Own Radio | 0.079 | 0.296 | 0.114 | 0.210 | 0.256 | 0.112 |
| Max Elevation (meters) | 1576.190 | 2364.531 | 1068.943 | 1758.893 | 2044.656 | 1070.334 |
| Elevation Range (Max-Min) | 1127.761 | 1521.322 | 712.336 | 1083.293 | 1251.074 | 715.253 |
| % Counties in South | 1.000 | 0.342 | 1.000 | 0.554 | 0.447 | 1.000 |
| Changes 1920-1930 | | | | | | |
| Log Population | 0.051 | 0.049 | 0.067 | 0.004 | 0.037 | 0.060 |
| Log Employment | 0.082 | 0.096 | 0.111 | 0.045 | 0.083 | 0.103 |
| Log # of Houses | 0.078 | 0.092 | 0.108 | 0.046 | 0.078 | 0.100 |
| Log Average Manufacturing Wage | 0.117 | 0.217 | 0.108 | 0.172 | 0.197 | 0.103 |
| Manufacturing Employment Share | -0.010 | -0.035 | -0.018 | -0.018 | -0.026 | -0.018 |
| Agricultural Employment Share | -0.047 | -0.036 | -0.047 | -0.046 | -0.042 | -0.047 |
| % White | 0.012 | -0.011 | -0.010 | 0.000 | -0.006 | -0.004 |
| % Urbanized | 0.047 | 0.064 | 0.080 | 0.042 | 0.054 | 0.069 |
| % Illiterate | -0.030 | -0.014 | -0.029 | -0.019 | -0.015 | -0.028 |
| % of Whites Foreign Born | -0.001 | -0.023 | -0.016 | -0.012 | -0.015 | -0.012 |
| Log Average Farm Value | -0.013 | -0.076 | 0.025 | -0.182 | -0.102 | 0.013 |
| # of Observations | 163 | 2326 | 795 | 828 | 1744 | 779 |
| # of States | 6 | 46 | 14 | 25 | 43 | 14 |

Notes: The unit of observation is a county. The trimmed sample is obtained by dropping control counties which, based on their pre-program characteristics, have a predicted probability of treatment in the bottom 25 percent. All monetary values are in constant 2000 dollars. Data are from the 1920 and 1930 Census of Population and Housing, with the exception of farm value data, which are from the 1920 and 1930 Agricultural Census, and elevation data, which were collected by Fishback, Haines, and Kantor (2011). Manufacturing wage is obtained by dividing the total annual wage bill in manufacturing by the estimated number of workers in the industry. Details on data construction and limitations are provided in the online Appendix.

# Alternative "Valley Authorities" (Placebo)

**Figure A2: Map of Proposed Authorities**



Notes: The map displays in black the six proposed authorities: the Atlantic Seaboard Authority, the Great Lakes-Ohio Valley Authority, the Missouri Valley Authority, the Arkansas Valley Authority, the Columbia Authority, and the Western Authority. The TVA region is displayed in gray.

# Estimation Part I: “Reduced-Form” effects of TVA

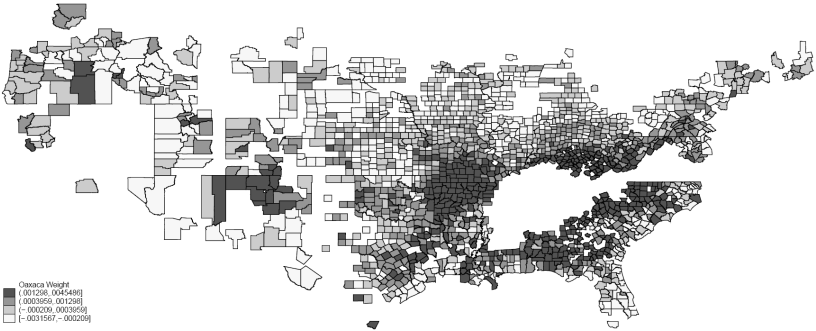
- Given the imbalance seen in Table 1, clearly important to control for baseline differences in TVA and non-TVA regions
- KM (2014) do this via Oaxaca-Blinder regressions (similar—indeed, in some sense isomorphic to) propensity-score reweighted regressions:
  - Estimate following regression on all non-TVA counties:

$$y_{it} - y_{it-1} = \beta X_i + (\epsilon_{it} - \epsilon_{it-1})$$

  - Where $X_i$ is a set of pre-program characteristics (38 economic, social, demographic and geographical variables measured in 1920 and 1930)
  - Then use $\widehat{\beta}$ estimate to construct counterfactual mean for TVA regions (and drop from subsequent regressions all non-TVA counties with very different values of $\widehat{\beta} X_i$ than TVA counties).
  - See Figure 3 for map of implicit weights (and dropped counties)

# Oaxaca weights

**Figure III: Weight on Untreated Counties**




Notes: In a Oaxaca-Blinder regression, each control county is implicitly assigned a weight: counties that look more similar to TVA counties in the years before TVA receive more weight. The weight, which may be negative, is proportional to an estimate of the odds of treatment. See Kline (2011) for discussion.

# Kline and Moretti (2014): Check for pre-trends

**Table II: Decadalized Growth Rates in TVA Region vs. Conterfactual Regions 1900-1940**

| | Outcome | Point Estimate (Unadjusted) (1) | Clustered S.E. (2) | Point Estimate (Controls) (3) | Clustered S.E. (4) | Spatial HAC (5) | N (6) |
|---|---|---|---|---|---|---|---|
| | **PANEL A: TVA Region vs Rest of US** | | | | | | |
| (1) | Population | 0.007 | (0.016) | 0.010 | (0.012) | (0.016) | 1776 |
| (2) | Total Employment | -0.009 | (0.016) | 0.005 | (0.013) | (0.016) | 1776 |
| (3) | Housing Units | -0.006 | (0.015) | 0.007 | (0.011) | (0.013) | 1776 |
| (4) | Average Manufacturing Wage | 0.009 | (0.018) | 0.010 | (0.021) | (0.016) | 1428 |
| (5) | Manufacturing Share | 0.007* | (0.004) | 0.005 | (0.004) | (0.005) | 1776 |
| (6) | Agricultural Share | -0.007* | (0.004) | -0.001 | (0.005) | (0.005) | 1776 |
| (7) | Average Agricultural Land Value | 0.078*** | (0.021) | 0.025 | (0.018) | (0.018) | 1746 |
| | **PANEL B: TVA Region vs. U.S. South** | | | | | | |
| (1) | Population | -0.018 | (0.018) | 0.003 | | (0.016) | 850 |
| (2) | Total Employment | -0.028 | (0.018) | 0.001 | | (0.016) | 850 |
| (3) | Housing Units | -0.025 | (0.016) | 0.005 | | (0.013) | 850 |
| (4) | Average Manufacturing Wage | 0.001 | (0.015) | 0.001 | | (0.016) | 687 |
| (5) | Manufacturing Share | 0.005 | (0.005) | 0.005 | | (0.005) | 850 |
| (6) | Agricultural Share | 0.003 | (0.004) | -0.002 | | (0.005) | 850 |
| (7) | Average Agricultural Land Value | -0.009 | (0.020) | -0.007 | | (0.017) | 839 |
| | **PANEL C: TVA Region vs. Proposed Authrorities** | | | | | | |
| (1) | Population | 0.026 | (0.019) | 0.011 | | (0.016) | 926 |
| (2) | Total Employment | -0.012 | (0.017) | 0.006 | | (0.015) | 926 |
| (3) | Housing Units | -0.014 | (0.016) | 0.006 | | (0.013) | 926 |
| (4) | Average Manufacturing Wage | 0.012 | (0.015) | 0.008 | | (0.017) | 734 |
| (5) | Manufacturing Share | 0.007 | (0.006) | 0.005 | | (0.006) | 926 |
| (6) | Agricultural Share | -0.005 | (0.006) | 0.004 | | (0.006) | 926 |
| (7) | Average Agricultural Land Value | 0.080*** | (0.026) | 0.017 | | (0.018) | 908 |

Notes: Column (1) gives the unconditional difference between TVA and non-TVA counties in the 1900-1940 change in the log of the relevant outcome divided by four (shares not converted to logs). Column (3) adjusts for pre-program differences between TVA counties and controls via a Oaxaca-Blinder regression as in Kline (2011). Covariates include time invariant geographic characteristics and levels and trends in pre-program industrial mix, population, and demographic characteristics (see Section III.A for full list of covariates). Clustered S.E. columns provide standard errors estimates clustered by state. Spatial HAC column provides standard error estimates based upon technique of Conley (1999) using bandwidth of 200 miles. Stars based upon clustered standard errors. Legend: * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

# Kline and Moretti (2014): Reduced-form estimates

**Table III: Decadalized Impact of TVA on Growth Rate of Outcomes (1940-2000)**

| | Outcome | Point Estimate (Unadjusted) | Clustered S.E. | Point Estimate (Controls) | Clustered S.E. | Spatial HAC | N |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| | **PANEL A: TVA Region vs Rest of US** | | | | | | |
| (1) | Population | 0.004 | (0.021) | 0.007 | (0.020) | (0.018) | 1907 |
| (2) | Average Manufacturing Wage | 0.027*** | (0.006) | 0.005 | (0.004) | (0.005) | 1172 |
| (3) | Agricultural Employment | -0.130*** | (0.026) | -0.056** | (0.024) | (0.027) | 1907 |
| (4) | Manufacturing Employment | 0.076*** | (0.013) | 0.059*** | (0.015) | (0.023) | 1907 |
| (5) | Value of Farm Production | -0.028 | (0.028) | 0.002 | (0.032) | (0.026) | 1903 |
| (6) | Median Family Income (1950-2000 only) | 0.072*** | (0.014) | 0.021 | (0.013) | (0.011) | 1905 |
| (7) | Average Agricultural Land Value | 0.066*** | (0.013) | -0.002 | (0.012) | (0.016) | 1906 |
| (8) | Median Housing Value | 0.040** | (0.017) | 0.005 | (0.015) | (0.015) | 1906 |
| | **PANEL B: TVA Region vs. U.S. South** | | | | | | |
| (1) | Population | -0.007 | (0.018) | 0.014 | | (0.019) | 942 |
| (2) | Average Manufacturing Wage | 0.003 | (0.006) | 0.001 | | (0.005) | 610 |
| (3) | Agricultural Employment | -0.097*** | (0.030) | -0.051* | | (0.027) | 942 |
| (4) | Manufacturing Employment | 0.079*** | (0.023) | 0.063*** | | (0.024) | 942 |
| (5) | Value of Farm Production | -0.005 | (0.025) | -0.006 | | (0.026) | 939 |
| (6) | Median Family Income (1950-2000 only) | 0.041*** | (0.012) | 0.024** | | (0.011) | 942 |
| (7) | Average Agricultural Land Value | 0.031* | (0.018) | -0.003 | | (0.017) | 942 |
| (8) | Median Housing Value | 0.019 | (0.017) | 0.007 | | (0.016) | 942 |
| | **PANEL C: TVA Region vs. Proposed Authrorities** | | | | | | |
| (1) | Population | 0.011 | (0.018) | 0.001 | | (0.017) | 991 |
| (2) | Average Manufacturing Wage | 0.018*** | (0.007) | 0.005 | | (0.006) | 618 |
| (3) | Agricultural Employment | -0.101*** | (0.029) | -0.071*** | | (0.027) | 991 |
| (4) | Manufacturing Employment | 0.066*** | (0.024) | 0.053** | | (0.024) | 991 |
| (5) | Value of Farm Production | 0.002 | (0.026) | 0.011 | | (0.035) | 989 |
| (6) | Median Family Income (1950-2000 only) | 0.060*** | (0.012) | 0.025** | | (0.011) | 991 |
| (7) | Average Agricultural Land Value | 0.060*** | (0.019) | -0.003 | | (0.016) | 991 |
| (8) | Median Housing Value | 0.033** | (0.016) | 0.009 | | (0.016) | 991 |

Notes: Point estimates obtained from regression of 1940-2000 change in outcomes divided by six on TVA dummy. All outcomes besides share variables are transformed to logarithms before taking difference. In specification titled controls, counterfactual change in TVA sample computed via Oaxaca-Blinder regression as in Kline (2011). Covariates include time invariant geographic characteristics and levels and trends in pre-program industrial mix, population, and demographic characteristics (see Section III.A for full list of covariates). Clustered S.E. column provides standard errors estimates clustered by state. Spatial HAC column provides standard error estimates based upon technique of Conley (1999) using bandwidth of 200 miles. Stars based upon clustered standard errors. Legend: * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

# Kline and Moretti (2014): Reduced-form estimates

**Table IV: Decadalized Impact of TVA on Growth Rate of Outcomes Over Two Sub-Periods**

| | | Entire U.S. | | South | | Proposed Authorities | |
|---|---|---|---|---|---|---|---|
| | | 1940-1960 | 1960-2000 | 1940-1960 | 1960-2000 | 1940-1960 | 1960-2000 |
| | Outcome | (1) | (2) | (3) | (4) | (5) | (6) |
| (1) | Population | 0.037 | -0.008 | 0.042 | -0.000 | 0.028 | -0.013 |
| (2) | Average Manufacturing Wage | -0.005 | 0.014* | -0.003 | 0.010 | 0.007 | 0.012 |
| (3) | Agricultural Employment | 0.106*** | -0.134*** | 0.106*** | -0.130*** | 0.119*** | -0.166*** |
| (4) | Manufacturing Employment | 0.114*** | 0.033** | 0.116*** | 0.035* | 0.097** | 0.032** |
| (5) | Value of Farm Production | 0.076* | -0.030 | 0.081** | -0.044 | 0.118** | -0.033 |
| (6) | Median Family Income | N/A | 0.017 | N/A | 0.016 | N/A | 0.019* |
| (7) | Average Agricultural Land Value | 0.027 | -0.017 | 0.018 | -0.015 | 0.029 | -0.021 |
| (8) | Median Housing Value | 0.019 | -0.003 | 0.010 | 0.005 | 0.020 | 0.003 |

Notes: Full set of controls included in all specifications. Point estimates obtained from Oaxaca-Blinder regression of 1940-1960 or 1960-2000 change in log outcomes divided by two or four respectively on TVA dummy and interacted controls as in Kline (2011). Covariates include time invariant geographic characteristics and levels and trends in pre-program industrial mix, population, and demographic characteristics (see Section III.A for full list of covariates). Stars based on standard errors clustered by state (entire U.S.) or spatial HAC estimates (South and Proposed Authorities) using technique of Conley (1999) with bandwidth of 200 miles. Legend: * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

# Estimation Part 2: "Structural" estimates of agglomeration externalities

- Now introduce a simple model of spatial equilibrium (a la Rosen-Roback; Roback, 1982). Ingredients:
  - Production function $Y_{it} = A_{it} K_{it}^{\alpha} F_i^{\beta} L_{it}^{1-\alpha-\beta}$, 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
  - TFP given by:

$$\ln A_{it} = g\left(\frac{L_{it-1}}{R_i}\right) + \delta_t D_i + \eta_i + \gamma_t + \varepsilon_{it}$$

  - Where $g(\cdot)$ is the agglomeration function (assumed to be a function of *lagged* population density; $R_i$ is area of county), and $D_i$ is TVA treatment dummy
- So TVA has direct effects (that vary over time, $\delta_t$, to potentially match the RF findings of Table 4) and potentially also indirect effects via agglomeration (i.e. via attracted $L_{it-1}$ and $g'(\cdot) > 0$).

# More structural estimation details

- Letting $\varepsilon_{it}$ be a unit root (Blanchard and Katz, 1992) with $\Delta\varepsilon_{it} = \lambda X_i + \nu_{it}$, can write SR labor demand function as (with "tilde" denoting original variables divided by $\beta$):

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

- In practice:
  - Proxy $g(\cdot)$ with three-knot spline
  - IV for each spline term $k$ with lagged instrument: $Z_{it}^{(k)} \equiv g_k\left(\frac{L_{it-2}}{R_i}\right) - g_k\left(\frac{L_{it-3}}{R_i}\right)$
  - Calibrate (SR) LD elasticity $-\frac{1-\alpha}{\beta}$ from labor literature (Hammermesh 1993 chapter, values: 1-1.5)

# Kline and Moretti (2014): Structural estimates

**Table VI: Structural Estimates of Agglomeration Function (log basis)**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | OLS | OLS | OLS | 2SLS | 2SLS | 2SLS |
| *Change in Log Manufacturing Density Spline Components:* | | | | | | |
| | 0.173 | 0.147 | 0.146 | 0.443 | 0.400 | 0.396 |
| Low | (0.037) | (0.037) | (0.037) | (0.102) | (0.108) | (0.107) |
| | | | | [177.17] | [159.14] | [157.20] |
| | 0.221 | 0.227 | 0.226 | 0.456 | 0.440 | 0.438 |
| Medium | (0.045) | (0.044) | (0.045) | (0.124) | (0.123) | (0.124) |
| | | | | [106.74] | [109.55] | [110.13] |
| | 0.143 | 0.151 | 0.141 | 0.466 | 0.467 | 0.453 |
| High | (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 | .2483 | .1298 | .1038 | .9545 | .669 | .7171 |
| P-value slopes equal zero | 1.9e-07 | 5.1e-07 | 6.9e-07 | 1.5e-04 | 7.4e-04 | .001 |
| N | 6057 | 6057 | 6057 | 5935 | 5935 | 5935 |

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-stat 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.

# What do these estimates imply?

- With K-share in manufacturing of $\alpha = 0.3$ and LD elasticity of 1.5, $\beta = 0.47$. So then pooled agglomeration elasticity is about 0.2.

- Can show that in steady-state, if each location has own agglomeration elasticity $\sigma_i$, then effect of TVA investments on aggregate worker utility is:

$$\frac{d\bar{u}}{d\delta} = \left(\frac{1}{1-\alpha}\right) \frac{\sum_i \frac{D_i L_i}{\beta - \sigma_i}}{\sum_i \frac{L_i}{\beta - \sigma_i}}$$

- So therefore if (as per Table 6) elasticities look pretty constant ($\sigma_i = \sigma, \forall i$) then we have

$$\frac{d\bar{u}}{d\delta} = \left(\frac{1}{1-\alpha}\right) \frac{\sum_i D_i L_i}{\sum_i L_i}$$

- Using annual real discount rate of 3%, this implies NPV TVA benefits of \$23.8-\$36.5B (compared to NPV of federal transfers of \$17.3B)