The Geographic Diversity of U.S. Nonmetropolitan Growth Dynamics: A Geographically Weighted Regression Approach

by

Mark Partridge, Ohio State University
Dan Rickman, Oklahoma State University
Kamar Ali & Rose Olfert, University of Saskatchewan

Prepared for presentation at the Department of Agriculture Economics, Purdue University

October 19, 2007

Available at: http://aede.osu.edu/programs/Swank/

Mark Partridge
Swank Professor in Rural-Urban Policy
Swank Program in Rural-Urban Policy
Motivation

• Heterogeneous preferences and land use suggest different marginal growth effects
  – Persistent decline in resource based communities
  – High amenity areas are experiencing rapid growth
  – Exurbia faces sprawl and managing growth
  – Rust Belt cities to sprawling Sunbelt cities.
• University attainment (Simon, 1998+)
  – Amenity driven migration to “high quality of life” locales
    • Richard Florida creative class effects (Adamson et al. 2004)
    • Could create a Brain Drain in low-amenity areas.
    • Life cycle effects and retirement migration.
    • Amenities are usually superior goods.
  – Knowledge economy effects (endogenous growth)

• Other examples:
  • Immigration—rural vs. urban effects
    – (Florida-like diversity effects)
  • Distance to the urban center—terrain or multiple urban centers.
  • Climate (Deller et al., 2001)

• New Economic Geography & Regional Science—space matters
• Global regression techniques (OLS or SEM) assume common responses or β
• Misleading if regression coefs β are hetero
  – OLS coef could be + (sign) but some locations have – (sign)
  – OLS could be insign, but half the sample is positive (sign) and the other half neg (sign)
  – At the very least, it would be helpful to have a way to guide global interactions that aren’t ad hoc.
  • E.g., (say) interact Midwest dummy with (say) education.
Underlying Model

• Structural supply and demand model of job growth
  – Solve for underlying reduced-form model
• Firms maximize $\pi$ in location $i$
• $\Pi_i = TR_i - TC_i$
  – Profits vary due to differing agglomeration econ. vs congestion effects and other factors
Solve for the indirect profit function:

\[ \Pi_i = F_i(w_i, \text{DIST}_{ij}, K_H_i, N_i, U_S_i, \cdot) \]

- Wages, distance within urban hierarchy, human capital (KH), population density (N), U.S. trends.
- Greater productivity in \( i \) increases a firm’s willingness to locate there.
- In a Roback (1982) sense, perfect mobility, greater productivity implies firms are willing to pay higher \( w_i \) and \( r_i \).
Regional differentials in $\Pi$ induce shifts in firm location and employment. (emp is a function of firms)

$$\Delta EMP^D_i = f_i(\Pi^k_i - \Pi^kUS), \quad f''_{\Pi ki}( ) > 0$$
• Household location (Roback, 1982):
  – Tradeoff higher wages & rents for amenities ($S_i$) and endog. urban amenities/congestion from pop. density $N_i$

• Maximize indirect utility:
  – $V_i = G_i(w_i, \text{DIST}_{ij}, N_i, S_i, \cdot)$
    • Households are willing to tradeoff higher $r_i$ and lower $w_i$ for better combination of amenities
Regional differentials in $V$ induce shifts in household location and employment to equalize $V$ across regions.

$$\Delta EMP^S_i = g_i(V_i - V_{US}), \quad g'_{Vil}( ) > 0.$$
• Solve for reduced form in a dynamic sense:
• set change in LD = change in LS
  – Assume long-run equil regional unemployment rates.
• Ala Blanchard and Katz (1992), etc.
• \( EmpGr_i = H_i(DIST_{ij}, N_i, S_i, KH_i, \cdot) \)
• Note that the function H varies across all \( i \)
Empirical Model

Global OLS—common a’s (or β’s) across all \( i = 1, 2, ..., n \)

\[ y_i = a_0 + a_1x_{i1} + a_2x_{i2} + \cdots + a_kx_{ik} + \varepsilon_i \]
Spatial Econometrics

- SEM model uses a weight matrix to control for the spatial correlation in the residuals.
- \( y = X\beta + u, \quad u = \lambda Wu + \varepsilon \)
- \( W \) is a \( n \times n \) row standardized connectivity matrix
- \( w_{ij} = 1/d_{ij}^2 \) where \( d_{ij} \) is the distance between \( i \) and \( j \).
- But still a common \( \beta \) vector across space
• GWR Model
  – Allows different regression coefficients for all the \( i \) observations.

\[
y_i = \alpha_{i0} + \alpha_{i1} x_1 + \alpha_{i2} x_2 + \cdots + \alpha_{ik} x_k + \epsilon_i \\
a_i^* = (X'W_i X)^{-1} X'W_i Y
\]

  • \( i = 1, 2, \ldots, n \), \( n \) = number of observations

  • \( W_i = \text{diag}(w_{i1}, w_{i2}, \ldots, w_{in}) \)

    \[
    w_{ij} = e^{-\left(\frac{d_{ij}}{h}\right)^2}
    \]
  – \( d_{ij} \leq h \) and \( h \) is the bandwidth
  – Bandwidth could be a fixed distance or optimal number of neighbors chosen by AIC.
    • \( h \) = nearest 846 nonmetro neighbors in nonmetro sample
    • \( h \) = nearest 893 metro neighbors in metro sample

Swank Program in Rural-Urban Policy
• Advantages of the GWR model
  – Heterogeneity in the regression coef
  – Accounts for spurious “spatial error” corr. that arises due to corr. of regress coef and X values.
  – Multicollinearity can be averaged across 1,000s of regressions.
  – Has fixed effects due to indiv constant terms
  – Can map the coefficients

• Weaknesses include less efficiency due to smaller sample than global approaches
  – Overwhelming amount of output! $n$ regressions
Empirical Model

- Dependent variable is 1990-2004 total county employment growth
  - 2003 U.S. Census metro definitions
  - 1972 nonmetro counties, 1,057 metro counties

- Explanatory $X$ are measured in 1990 to avoid endogeniety
  - Initial educational attainment shares (college grads!)
  - Climate and topography measures
  - “Disequilibrium” economic controls including industry mix
Several spatial measures that reflect proximity to cities higher in the urban hierarchy (land use)

- Rural counties: distance in km to nearest urban center of any size
- Micropolitan and MA counties: distance in km to center of urban core if multi-county, 0 for single county urban area;
• Spatial measure—continued
  – incremental dist (kms) from the cnty to reach an MA
  – the incremental distances to reach urban centers of at least 250,000, at least 500,0000, and at least 1.5 million population
  – if nearest area is higher-tiered, remaining tier distances set equal to 0
– Example 1: Rawlins, Wyoming
– Rawlins Census 2000 population: 8,538, rural county

  • **100** miles to Laramie WY: distance to nearest urban area (MICRO)
  • 121 miles to Casper, WY: incremental distance to MA: **21** miles
  • 190 miles to Ft. Collins, CO: incremental dist to MA>250,000: **69** miles
  • 243 miles to Denver, CO: incremental dist to MA>500,000: **53** miles;
    • since Denver is 2.5 mill.>1.5million, the other incremental dist var. for >1.5 million = **0**.
      – We convert this to kms
Rural Gilmer County, West Virginia: Glenville is the county seat
- 55 miles to Clarksburg (urban area micropolitan: 91,509)
- 5 incremental miles (60 total-55) to Parkersburg (nearest MA: 161,907)
- 27 incremental miles (87 total-60) to Charleston (nearest MA>250K: 307,689)
- 70 incremental miles (157 total-87) to Pittsburgh (nearest MA>500K: 2.5mill.)
- 0 incremental miles to reach MA of at least 1.5 million

These capture potential distance penalties for access or growth shadow effects
Empirical Results

- Nonmetropolitan results
  - OLS, SEM, median GWR coefs are similar
  - GWR is a significant statistical improvement
    - Spatial heterogeneity! One Size Does NOT Fit All!
    - 27 of 37 coefs sign vary across the sample
  - Global coefficients can be misleading
    - E.g.: avg Jan Sun Hours is insign in OLS & SEM
    - Yet, there are positive & negative GWR coefs that are statistically different. The average effect $\approx$ zero
    - Global models incorrectly imply little or no effect
• Sensitivity analysis shows results are robust
  • 1999 vs. 2003 MA boundaries
  • 1990-2000 emp. growth vs 1990-2004

• Maps of key results to illustrate spatial hetero.
Nonmetro Employment Change 1990-2004

1a: Variations in the Coeff. of January Temperature

Note: 2003 MA boundary definitions used
Nonmetro Employment Change 1990-2004

1b: Variations in the Coeff. of % Water Area

Note: 2003 MA boundary definitions used
Nonmetro Employment Change 1990-2004

1c: Variations in the Coeff. of Typography

Note: 2003 MA boundary definitions used
Nonmetro Employment Change 1990-2004

1d: Variations in the Coeff. of % Immigrated 1985-90

Note: 2003 MA boundary definitions used
1e: Variations in the Coeff. of College Graduates

Note: 2003 MA boundary definitions used
Metro Results

- Not emphasized (test robustness)
- The expected pattern of sprawl related growth
  - dist to center of MA is positively related to growth
- MA results have the expected spatial hetero
  - But not as much diversity as for nonmetro results.
  - Only 20 of 37 coefs significantly vary spatially
- Maps of key variables with spatial hetero in $\beta$
Metropolitan Employment Change 1990-2004

2a: Variations in the Coeff. of January Temperature

Note: 2003 MA boundary definitions used
Metro Employment Change 1990-2004

2b: Variations in the Coeff. of Typography

Note: 2003 MA boundary definitions used
Metro Employment Change 1990-2004

2c: Variations in the Coeff. of % Immigrated 1985-90

Note: 2003 MA boundary definitions used
Metro Employment Change 1990-2004

2d: Variations in the Coeff. of % College Graduates

Note: 2003 MA boundary definitions used
Policy Implications/Summary

- Global regression approaches obscure important spatial heterogeneity
  - Global approaches can obscure differences at the tails. Can be very misleading.
    - Spatial negative and positive effects offset each other

- Regional Science models often do not consider this type of hetero
  - Why are there geographical “cusps” where the X effects shift so suddenly?
• Empirical work: Global approaches give avg effect (a benchmark), but GWR helps inform the nuances of local policy.
  • Don’t abandon OLS or global approaches, but better recognize its deficiencies.

• For Policy: **One Size Does NOT Fit All!**
  • Education has mixed effects and evidence of brain drains
  • A cold winter is not uniformly a negative factor
  • Old expression: “when you seen one rural community, you only seen one rural community.”