Shagor Rahman

Computer Science Department Ranking &

Urban Growth

1. Outline

Much has been written about skilled labor and the impact this has on urban growth (Glaeser 2003). Additionally there is increasing debate about the cost and benefit of higher education. This analysis seeks to isolate and determine whether a single higher educational ranking, i.e. the ranking of a University’s Computer Science Department (CS), is a causal attribute that alone contributes to growth. Cities such as Austin, TX, and Durham, NC, have recently seen an influx in high tech businesses flocking to attract the highly valued skills the CS programs are producing. This has seemingly created a dynamic growth environment in cities that had been in decline, or in some cases missed previous boons generally.

The change is visible, the impact on municipalities via tax revenues and employment growth is paramount. In other words, if we can predict that a University with a highly ranked CS program is predictive in urban growth, would not a municipality give serious consideration to make the required investment in time and resources?

1. Overview

I have created a few models in order to test this hypothesis. Given the various ways of measuring urban growth I thought it important that a high CS ranking be significant across several measures before warranting a full rejection of the Null Hypothesis. Which here, is that coefficient for the variable = 0.

The dependent variables used here will be GDP, Income, and Employment. All data will be collected at the ‘Metropolitan Statistical Area’ (MSA) level. The models are based off the Arellano-Bond estimation approach, a Generalized Method of Moments model specific to models containing dynamic panel data. The model is such that includes a lag of the dependent variable as a regressor, and then leverages a lag (at T-2) of both the dependent variable observations and the regressors as instrumental variables to address both heteroscedasticity and autocorrelation.

Where the \* indicates the first difference.

The estimator of the parameter vector is a matrix weighted average of the T-2 period specific two stage least square estimates (Greene). The models in all cases here, using the PLM package, will leverage the ‘Twoway’ effect, which leverages the first differences, and also includes time dummies (Croissant and Millo). In order to reject the Null Hypothesis that the coefficient for CS Ranking = 0, we will seek a 95% confidence interval.

1. Data and Model description

Data for this analysis was aggregated from three separate and reputable sources. All were collected annually from 2010 to 2015.

Cencus.gov – Specifically the ‘American FactFinder’ which collects survey data on the United States and other areas. This data is broken down at various levels, including MSA. I collected General Population Data, Income Per Capita, demographics to calculate the percent of working age population (age 18-64), Educational Attainment, Native Born Percentage, and industry group classification. In the last category, it became clear that the industry group ‘Services’ offered the most significant coefficient so the model isolated this variable.

Bureau of Economic Analysis – From the US Department of Commerce, this data set details Gross Domestic Product per MSA and an aggregate level. I used the Total Population to calculate per capita equivalents.

Shanghai Rankings – Shanghai Rankings offers current and historical rankings of global universities, both generally and department level. Its methodology is transparent and recognized as ‘Scientifically sound’ by international observers. I extracted rankings for both Universities overall as well as rankings specific for Computer Science programs.

Specifically, I extracted ratings from 2015 going back to 2010. I used ‘200-X’ as the ‘magnitude’ rank. Where X = school rank for the given year. I then added in the MSA code for each school using the Geocoding system from the ‘Federal Financial Institution Examination Council.’ Finally, I aggregated across unique MSA code and year, so if two schools were located in the same MSA and were both ranked for either category (CS or General) then scores summed.

Variable Table

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | ID | GEO.id2 | V1 | | Year | Year | V2 | | MSA Code \_ Year | MSA\_YR | V3 | | MSA Name | GEO.display-label | V4 | | Total Population | EST\_VC03 | V5 | | Labor Force1 | EST\_VC18 | V6 | | HS\_ED | EST\_VC142 | V11 | | Native? | EST\_VC197 | V13 | | Employed\_% | EST\_VC240 | V15 | | Poverty | EST\_VC366 | V16 | | CS Ranking | CS\_MAG | V36 | | General Ranking | GM\_MAG | V37 | | Income Per Capita | INC\_PP | V38 | | GDP Per Capita | GDP\_PP | V39 | | Ages 18-64 | LABOR\_FORCE | V40 | |  |  |
|  |  |  |
| Models |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| As is shown above, the Dependent variables in all cases were lagged twice (where available, i.e. years 2015,2014,2013), Labor Force, and both ‘General’ and ‘CS’ were lagged once.  In each case the fitted values were made ‘Robust’ separately.  I will rely on the Z and P values for V36 to ultimately determine whether to soundly reject the Null that a CS ranking is not a deterministic growth variable.  Additionally, I will evaluate the Sargan (/J), Autocorrelation, and Wald Test to first determine if the model fits the data appropriately.   1. Conclusions   Evaluating first the ‘validity’ of the model. We can see that two of the three models passed in terms of validity, i.e. using GDP Per Capita (V39) and Employment (V15) were both well fit given the data, however Income per capita was not. (See the below Tables and attached ‘OUTPUT.xlsx’).  Income Per Capita (V38) failed the ‘Sargan Test,’ where the Null Hypothesis states the model is valid. The P Value was lower than the 5% significant level. Additionally, it failed for both Autocorrelation tests, and finally, While the Wald test passed for ‘time dummies’, i.e. they were not all = 0, the Wald test for coefficients failed. This may be due to the ‘Income Measurement Error in Surveys,’ as reported by the census itself (Moore, Stinson, and Welniak). People tend to report their income with cognizant and unintended errors.  The use of V39 and V15 as dependent variables were validated by test statistics. That is, structured as a ‘Null Hypothesis’ that the test is valid. The P Values for both the Sargon and Autocorrelation in both cases were not significant to dismiss the model’s validity. Additionally, the Wald Tests for both Coefficients and Time Dummies exhibited P values low enough to reject both, indicating alleviation from multi collinearity.  Unfortunately, in both of these cases, with a valid model, the test does not reveal a statistically significant result for the CS variable. In all cases the P values specific to ‘V36 – The Computer Science Ranking Variable, was above the 5% significance level to validly reject the Null Hypothesis. The coefficient for a University’s Computer Science do not differ from 0 significantly based on this analysis.  Further analysis can still explore this relationship. After all, rankings in theory, would have a delay of several years before impacting growth. Perhaps rankings can be lagged 6-7 years (to give students a chance to graduate, join the work force, etc.) before being leveraged as a regressor to predict growth. Additionally, few variables exhibited a P Value below the significance level indicating they differ from 0. This may simply follow the intuition that growth variables are difficult to isolate given the range of inputs that lead to local growth. |  |  |
|  |  |  |
|  |  |  |
| Attachments and Tables  GDP\_CITY.R – Final R Script  Variables:  Gdp.gmm  Inc.gmm  Emp.gmm  Gdp.gmm.r  Inc.gmm.r  Emp.gmm.r  Data Files – ‘CITY.CSV’  ‘Labels’ – ‘CITY\_LABELS.xlsx’  Output of Test Statistics - OUTPUT.xlsx:  Tabs – Arranged by Variable Names above.  References  Arellano M, Bond S (1991). \Some Tests of Specication for Panel Data : Monte Carlo  Evidence and an Application to Employment Equations." Review of Economic Studies,  58(2), 277{297. |  |  |
| Croissant, Millo, “Panel Data Econometrics in R: The plm Package. |  |  |
| Glaeser, Edward. “The Rise of the Skilled City.”  <http://whartonrealestate.org/research/papers/full/446.pdf> 2003 |  |  |

Greene W (2003). Econometric Analysis. 5th edition. Prentice Hall.

Moore, Stinson, Welniak. “Income Measurement Error in Surveys: A Review.”

https://www.census.gov/srd/papers/pdf/sm97-05.pdf