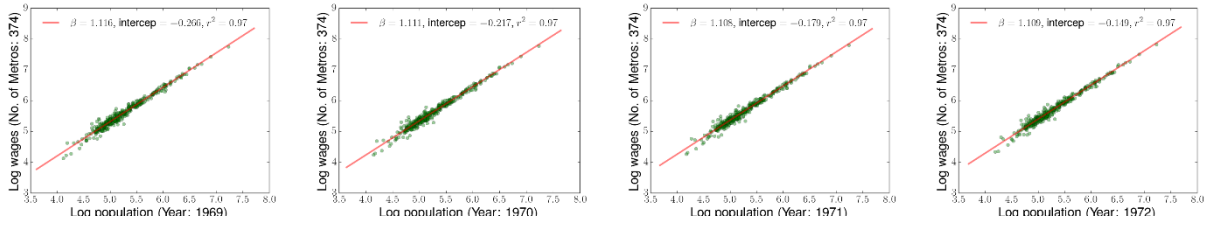
Scaling Analysis of Wages and GDP data for Metropolitan Areas in the USA 1969-2008

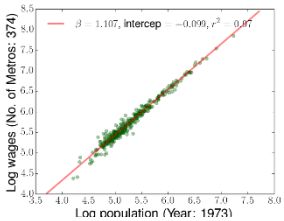
1. Clarify "Metropolitan Areas" and identify analysis objects

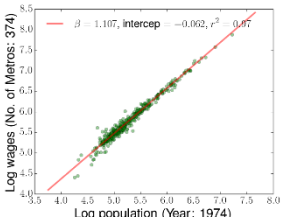
Metropolitan Areas, to be more specific, Metropolitan Statistical Areas, are defined by the U.S. Office of Management and Budget (OMB), and used by the U.S. Census Bureau and other federal government agencies for statistical purposes, and 381 MSAs has been defined. The GDP data consist only 366 MSAs, so I download the latest GDP data of 381 MSAs. However, the population data lack of seven MSAs, including 'Bloomington, IL', 'California-Lexington Park', 'Lafayette-West Lafayette, IN', 'Los Angeles-Long Beach-Anaheim, CA', 'Santa Maria-Santa Barbara, CA', 'Urban Honolulu, HI', 'Weirton-Steubenville, WV-OH'. Actually, I can find five of these names in population data, but with FIPSs different from these in GDP data, so I'm not sure whether they are the same statistical areas. So the scaling analysis in this article only include 374 MSAs. The population data is only from 1969 to 2008, so the Wages analysis is from 1969 to 2008, and GDP analysis is from 2001 to 2008.

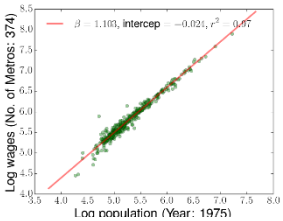
2. Log Wages and GDP, and make linear regressions

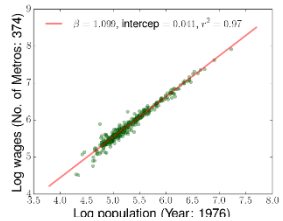
Wages:

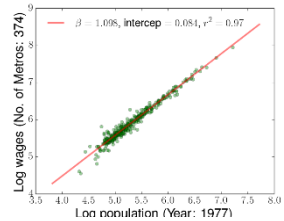


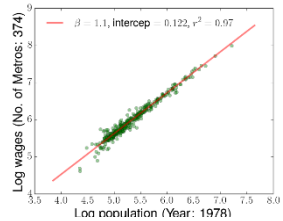

β = 1.107, intercep = −0.099, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1973)

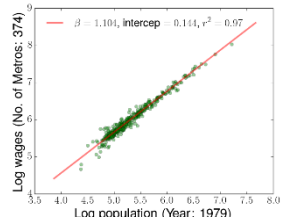

β = 1.107, intercep = −0.062, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1974)

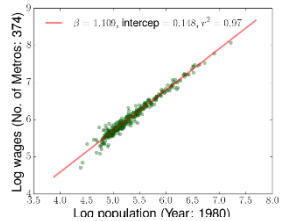

β = 1.103, intercep = −0.021, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1975)

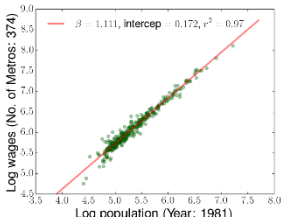

β = 1.099, intercep = 0.041, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1976)

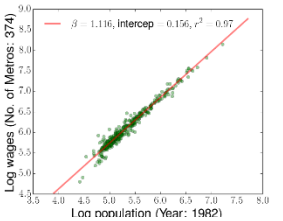

β = 1.098, intercep = 0.084, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1977)

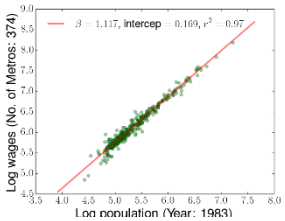

β = 1.1, intercep = 0.122, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1978)

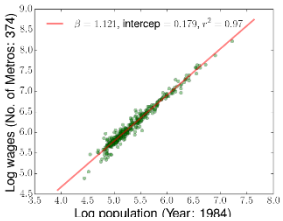

β = 1.104, intercep = 0.144, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1979)

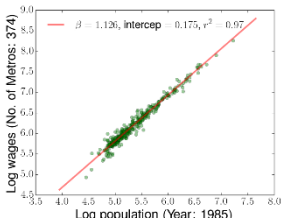

β = 1.109, intercep = 0.118, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1980)

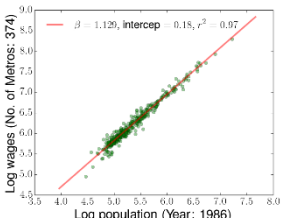

β = 1.111, intercep = 0.172, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1981)

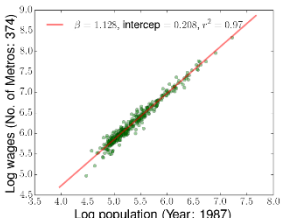

β = 1.116, intercep = 0.156, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1982)

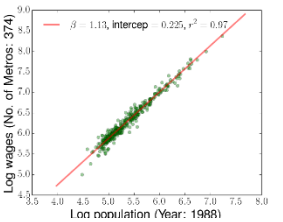

β = 1.117, intercep = 0.169, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1983)

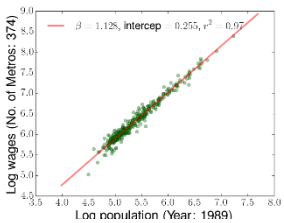

β = 1.121, intercep = 0.179, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1984)

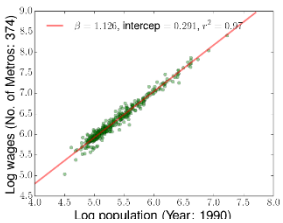

β = 1.126, intercep = 0.175, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1985)

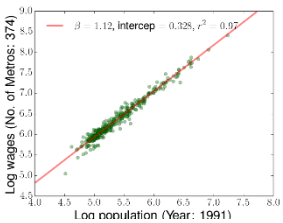

β = 1.129, intercep = 0.18, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1986)

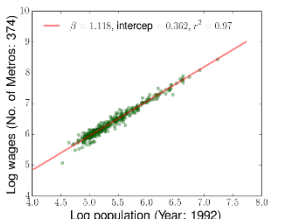

β = 1.128, intercep = 0.208, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1987)

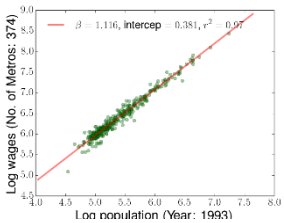

β = 1.13, intercep = 0.225, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1988)

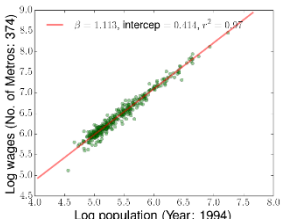

β = 1.128, intercep = 0.255, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1989)

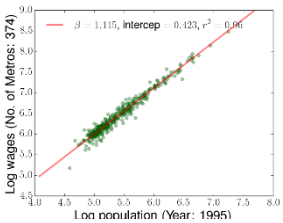

β = 1.126, intercep = 0.291, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1990)

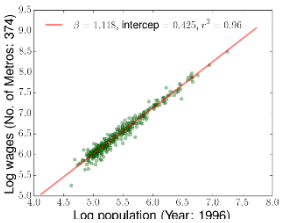

β = 1.12, intercep = 0.328, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1991)

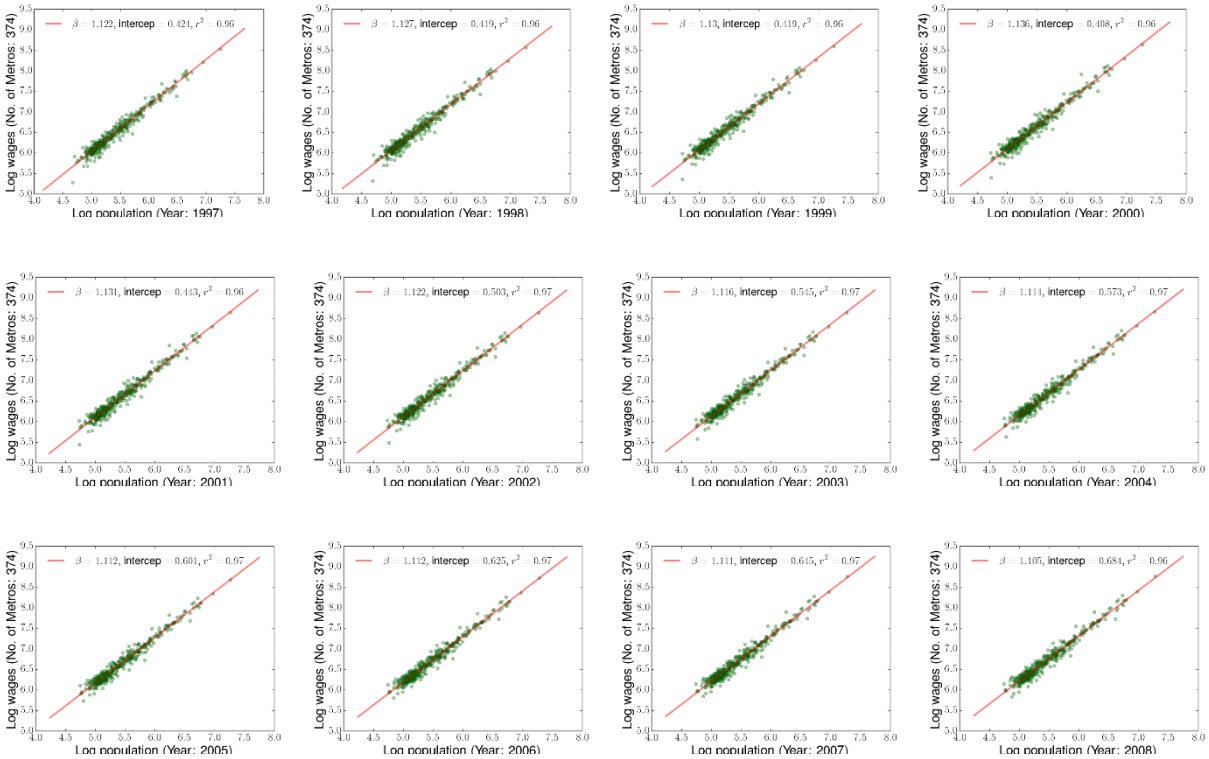

β = 1.118, intercep = 0.362, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1992)


β = 1.116, intercep = 0.381, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1993)

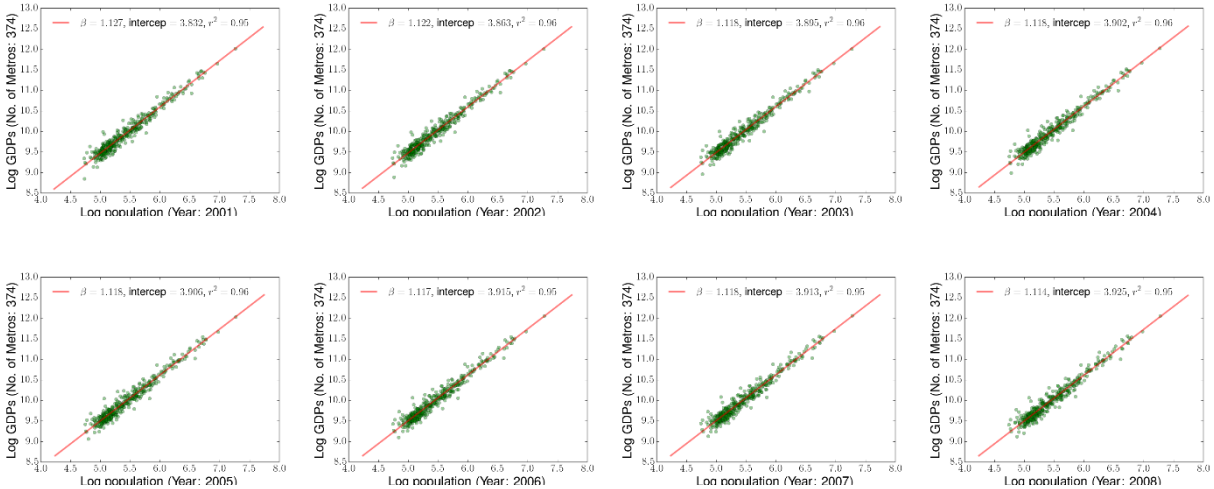

β = 1.113, intercep = 0.414, r² = 0.97
Log wages (No. of Metros: 374)
Log population (Year: 1994)


β = 1.115, intercep = 0.423, r² = 0.96
Log wages (No. of Metros: 374)
Log population (Year: 1995)

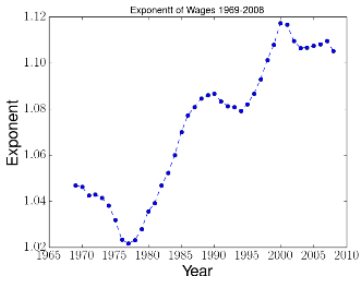

β = 1.118, intercep = 0.425, r² = 0.96
Log wages (No. of Metros: 374)
Log population (Year: 1996)

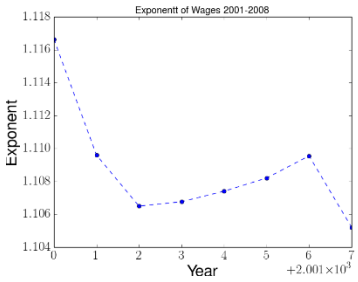


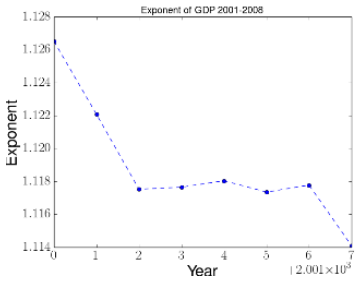
GDP:



Exponents of Wages and GDP:



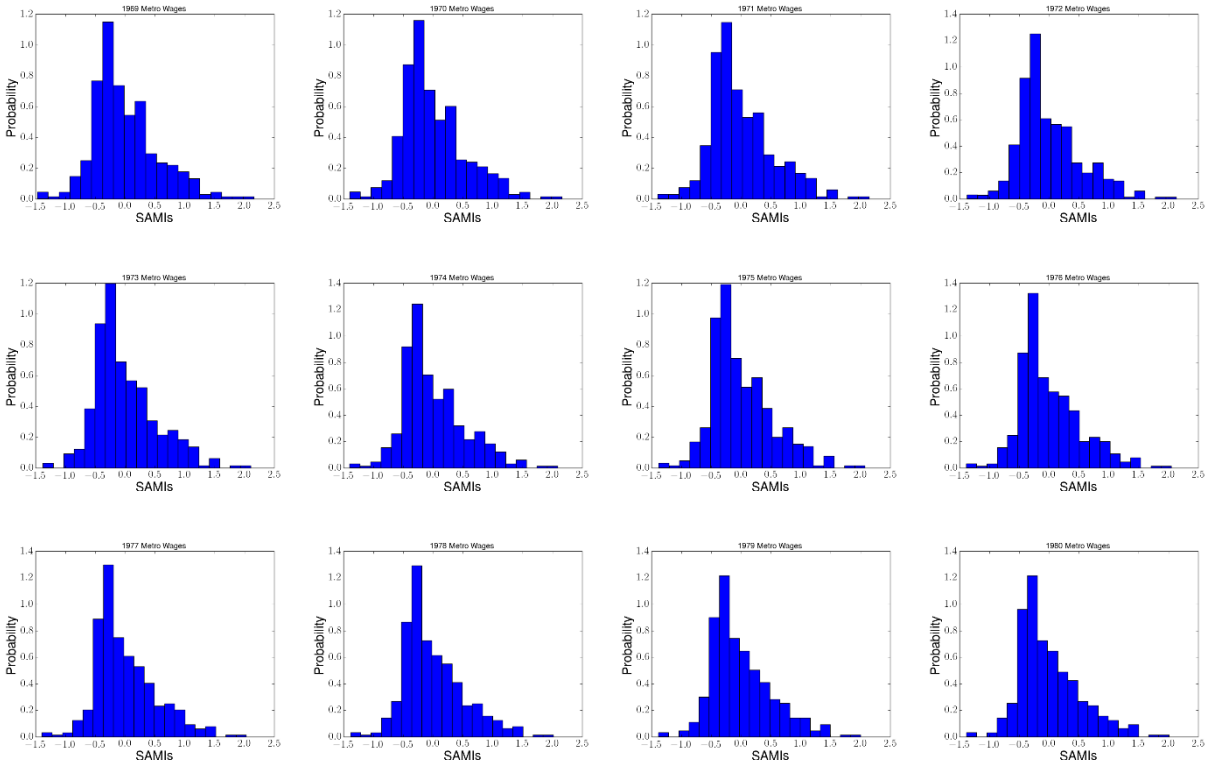


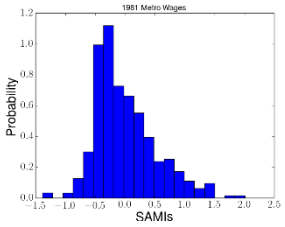


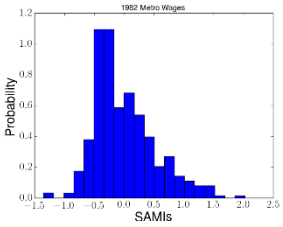
The Exponents of wages vary between 1.02 and 1.12 from 1969 to 2008, between 1.105 and 1.117 from 2001 to 2008 which well accorded with that of GDP from 2001 to 2008 except 2006. The observed range seems acceptable to the simplest model[1].

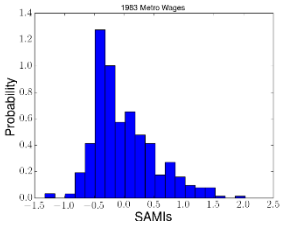
3. Extract residuals and make histograms

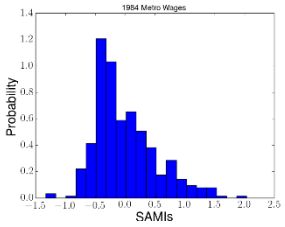
Wages:

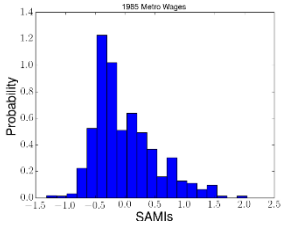


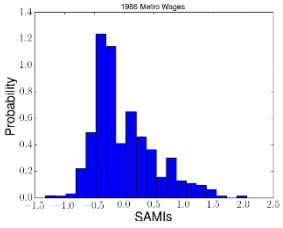

1981 Metro Wages
Probability
SAMIs

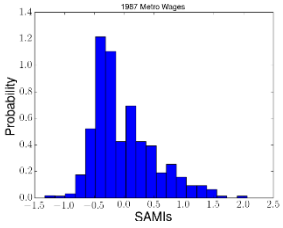

1982 Metro Wages
Probability
SAMIs

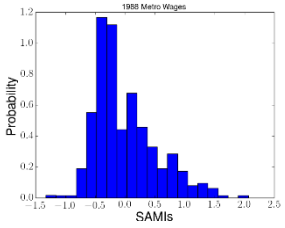

1983 Metro Wages
Probability
SAMIs

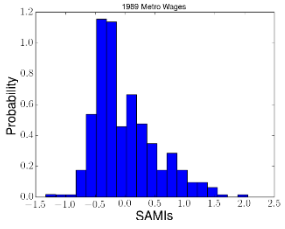

1984 Metro Wages
Probability
SAMIs

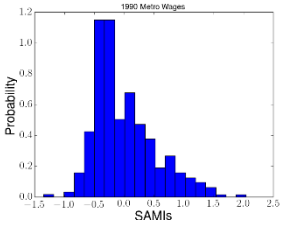

1985 Metro Wages
Probability
SAMIs

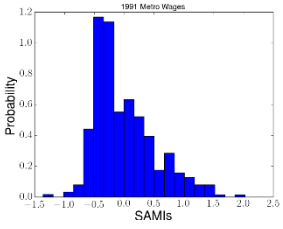

1986 Metro Wages
Probability
SAMIs

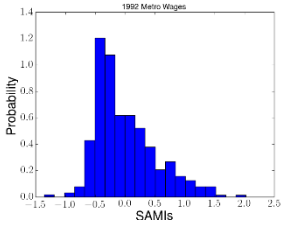

1987 Metro Wages
Probability
SAMIs

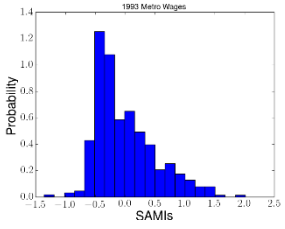

1988 Metro Wages
Probability
SAMIs

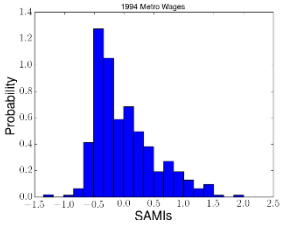

1989 Metro Wages
Probability
SAMIs

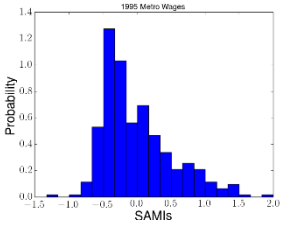

1990 Metro Wages
Probability
SAMIs

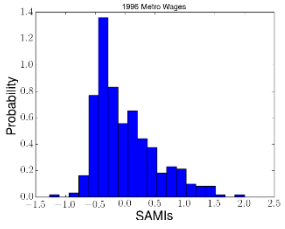

1991 Metro Wages
Probability
SAMIs

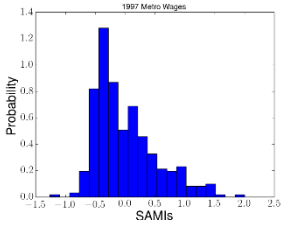

1992 Metro Wages
Probability
SAMIs

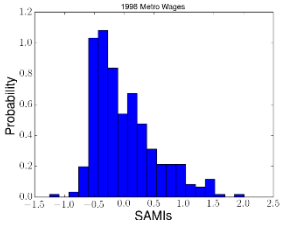

1993 Metro Wages
Probability
SAMIs

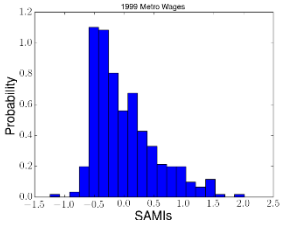

1994 Metro Wages
Probability
SAMIs

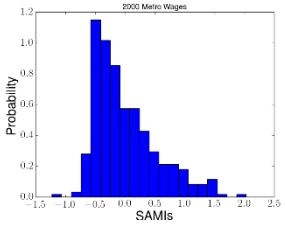

1995 Metro Wages
Probability
SAMIs

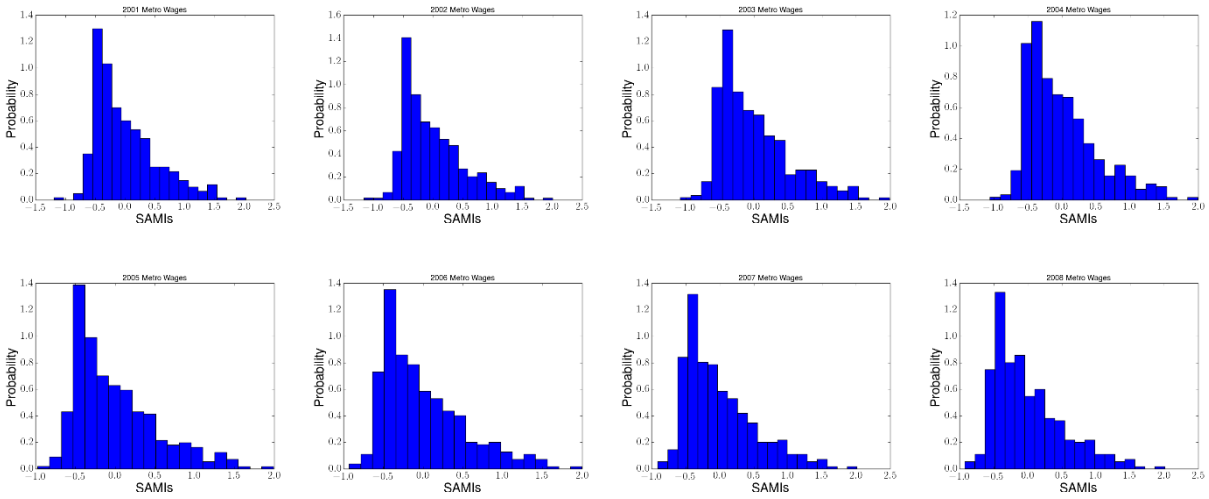

1996 Metro Wages
Probability
SAMIs


1997 Metro Wages
Probability
SAMIs

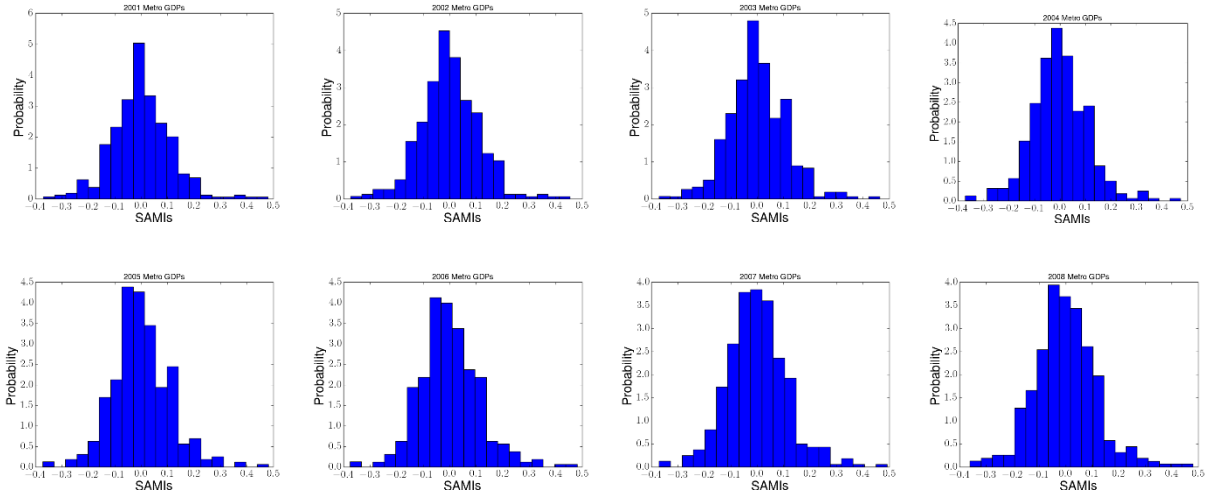

1998 Metro Wages
Probability
SAMIs


1999 Metro Wages
Probability
SAMIs


2000 Metro Wages
Probability
SAMIs



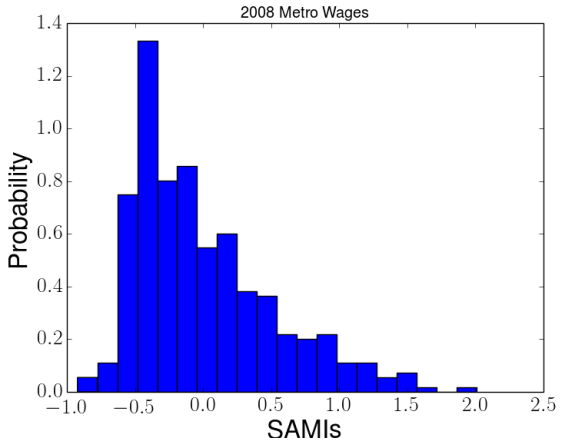
GDP:



It seems that not all the histograms well fit by a Normal bell curve (2002 Metro GDPs for example), some histograms show more like skewness (2008 Metro Wages for example) or student-t (2007 Metro GDPs for example) distributions. Although other curves might be better fit than the simple normal curve, I don't think these would be warranted, because the number of variables in the SAMI equation should be in accord with the number of variables in the fit curve, extra parameter in the curve won't be well explained through the SAMI equation.

4. Find the large and low residuals cities, and the position of New York

Take 2008 Metro Wages for example:



Top five: Trenton-Ewing NJ (0.3106), San Jose-Sunnyvale-Santa Clara CA (0.2944), Bridgeport-Stamford-Norwalk CT (0.2923), Columbus IN (0.2366), Boulder CO (0.2319).

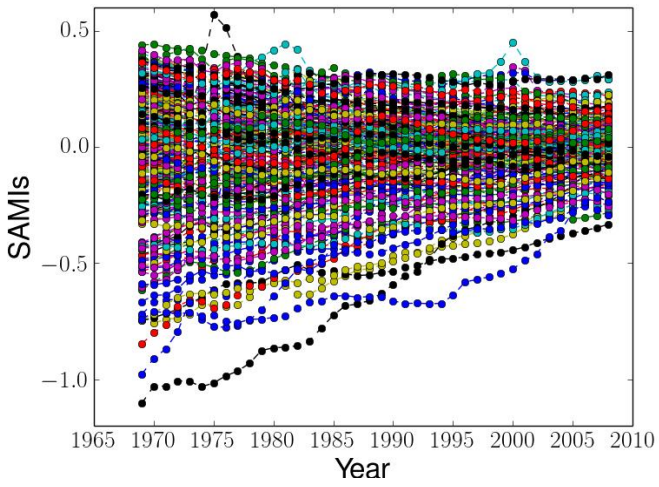
Bottom five: McAllen-Edinburg-Mission TX (-0.3318), Homosassa Springs FL (-0.2918), Lake Havasu City-Kingman AZ (-0.2900), Brownsville-Harlingen TX (-0.2811), Riverside-San Bernardino-Ontario CA (-0.2600).

New York is 0.0378, marginally higher wage than its size might predict.

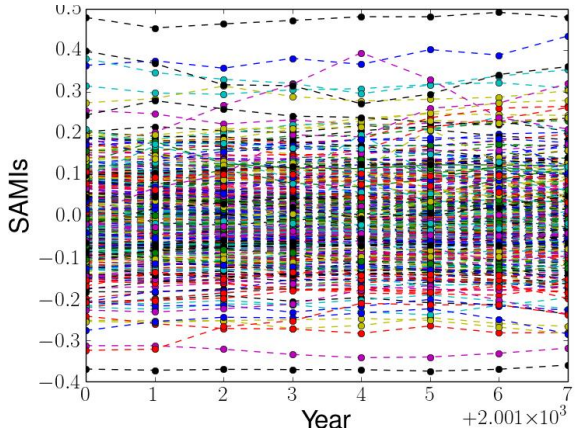
Other years' wages can be analyzed in the same way.

5. Residuals over time for cities. Which change quickly, which slowly

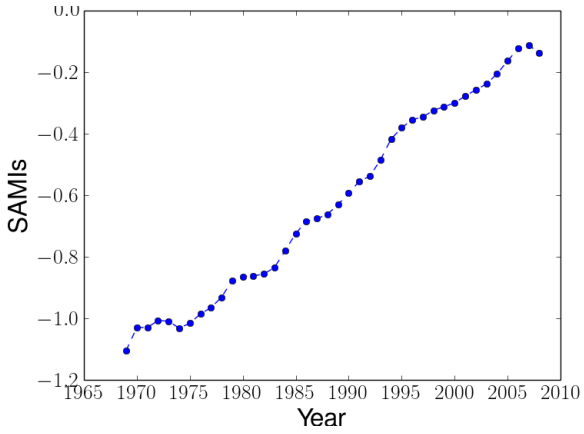
Wages (all the cities):



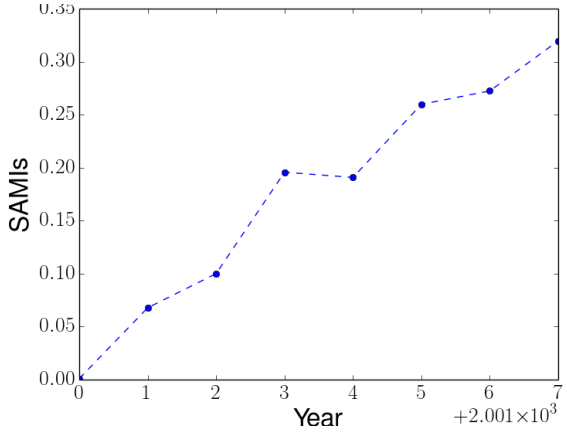
GDP (all the cities):



From Wages, a black line from the bottom change most quickly can be seen. It's Provo-Orem UT, its line chart is below. Most lines of cities are quiet smooth, and change slowly, provide the proof of the persistence of socio-economic processes[2], and thus identify the most slowly one is unnecessary.



From GDP, a purple line in the middle change most quickly can be observed. It's Columbus OH, its line chart is below. Most lines of cities are quiet smooth, and change slowly.



6. Which cities moved in similar ways, and speculate why

How to define "moved in the similar way" is essential. My understanding is that the cities have the same slope directions between every two points in their SAMIs lines above. In this view, the cities can be divided into different clusters as the following:

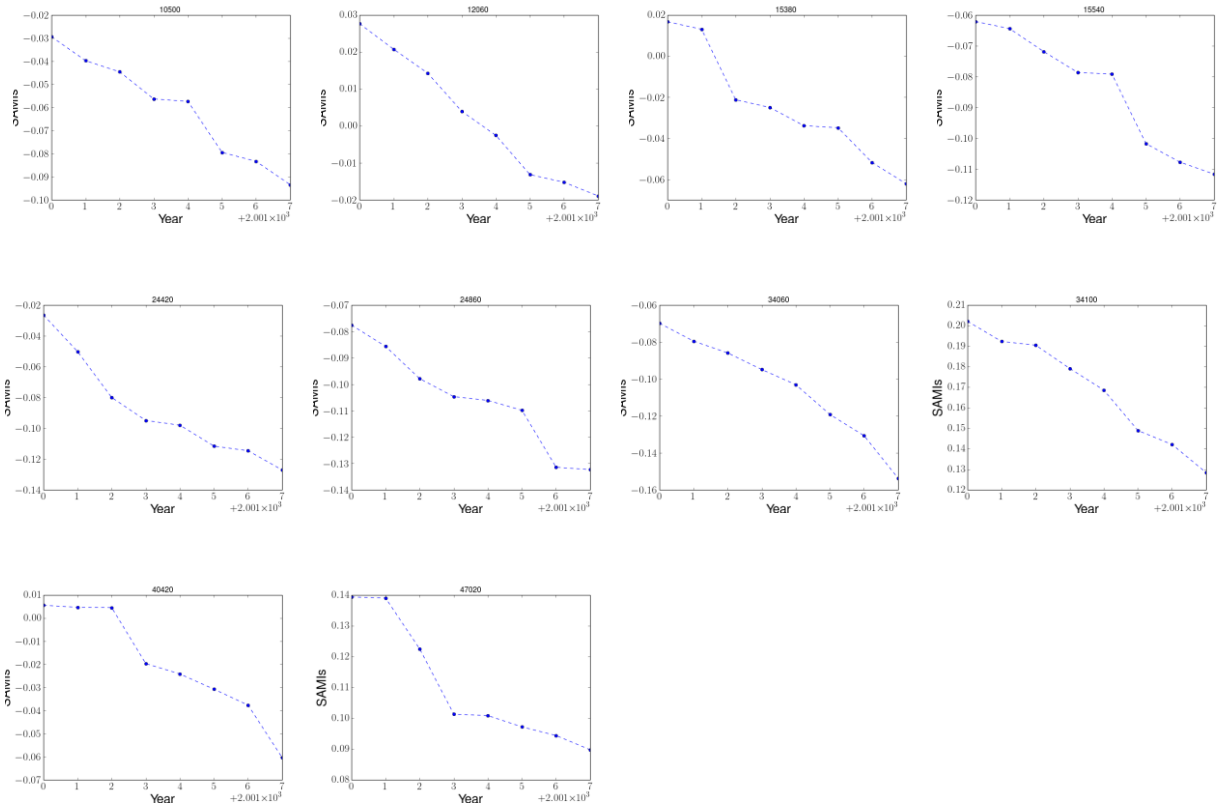
Wages:

No cities moved in the similar way (no lines have the same slope directions between every two points from 1969 to 2008).

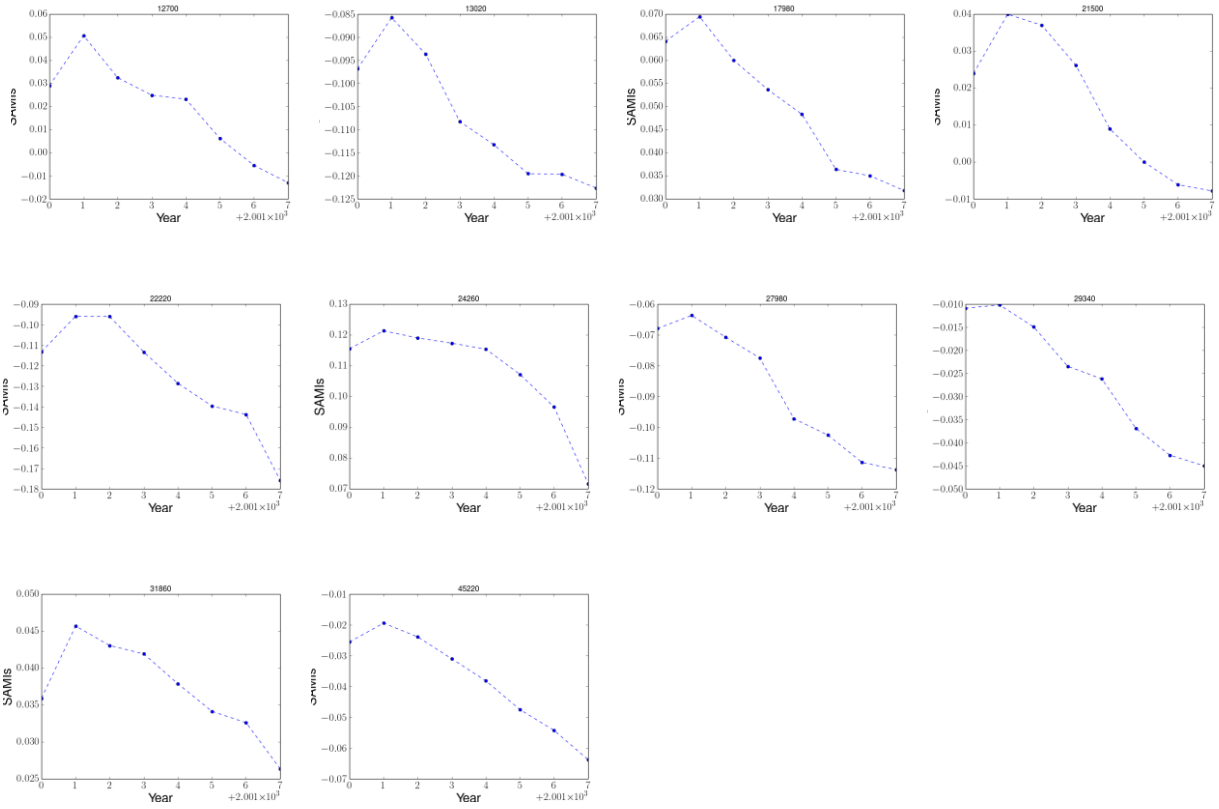
GDP:

Two large clusters (each has ten cities) presented, small clusters not included.

First cluster: Albany GA, Atlanta-Sandy Springs-Roswell GA, Buffalo-Cheektowaga-Niagara Falls NY, Burlington-South Burlington VT, Grants Pass OR, Greenville-Anderson-Mauldin SC, Morgantown WV, Morristown TN, Rockford IL, Victoria TX, and line charts are below (titled in their FIPS):



Second cluster: Barnstable Town MA, Bay City MI, Columbus GA-AL, Erie PA, Fayetteville-Springdale-Rogers AR-MO, Grand Island NE, Kahului-Wailuku-Lahaina HI, Lake Charles LA, Mankato-North Mankato MN, Tallahassee FL, and line charts are below (titled in their FIPS):



From the charts, we can see that cities in the each cluster obviously moved in the similar way from 2001 to 2008, this might be caused by same economic or social characteristics shared by these cities, further detailed investigations into each cluster is needed to accurately answer the question why these cities in the cluster all moved in this way.

[1] Bettencourt (2013). The Origins of Scaling in Cities. *Science* 340.

[2] Bettencourt (2010). Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities. *PLoS ONE*.