Analysis of Population Change in U.S. Metropolitan Statistical Areas

Wenyi Tang
Department of Resource Analysis, Saint Mary's University of Minnesota, Winona, MN 55987

Keywords: Population Change, Geographic Information Systems (GIS), Metropolitan Statistical Area (MSA), Factors, Correlation, Moran's I, Regression

Abstract

Understanding population change helps when studying local, regional, and even national trends affecting economies, environments, political decision-making, and other social factors. Analysis on why and how population changes is meaningful and important when studying population change and trends. Population change and factors such as employment status, education level, age, and housing value were researched using statistical methods such as Pearson's correlation coefficient, Global Moran's I, and regression. The 30 largest U.S. metropolitan statistical areas according to the 2010 census survey were selected and sampled to analyze population change of highly developed areas between 2013 and 2014 using American Community Survey (ACS) 5-Year Estimates data. Population change may be influenced by employment status, education level, age, and housing value. Of the variables studied, change in the number of people over age 65 appears to affect overall population change most.

Introduction

Background Literature

Understanding population change helps when studying local, regional, and even national trends affecting economies, environments, political decision-making, and other social factors. Ehrlich and Holdren (1971) indicate the problem of population, resources, and environment is often the focus of papers, books and symposia. They also believe population growth has a negative impact on the environment. Zhu and Zeng (2004) state population change affects and interacts with other factors, such as economy, military, environment, and politics, which together determine the strength of the country. DeFries, Rudel, Uriarte, and Hansen (2010) present that forest loss is positively correlated with population growth and agricultural products exported. In addition, Brown and Flavin (1999) report that human population grew fourfold and the world economy became 17 times as large during the 20th century; however, they also state this growth damaged natural systems while it helped people improve living standards.

Helping to generate background information on the topic, Bloom, Canning, and Sevilla (2003) indicate that population growth can be harmful or beneficial to economic development depending on circumstances; they also mention there is a neutralist view that rapid population growth neither

Tang, Wenyi. 2017. Analysis of Population Change in U.S. Metropolitan Statistical Areas. Volume 20, Papers in Resource Analysis. 12 pp. Saint Mary's University of Minnesota University Central Services Press. Winona, MN. Retrieved (date) from http://www.gis.smumn.edu.

promotes nor impedes economic growth.

General Overview

As population change is closely related with humanity's future through its interaction with other factors, it is important to analyze how population changes. Because larger cities are more developed, they are more stable and representative than the small cities when performing a statistical analysis from a more generalized standpoint. As a result, metropolitan areas were selected to sample and study in this paper.

The Statistical and Science Policy Branch, Office of Information and Regulatory Affairs and OMB (Office of Management and Budget) (2003) determined that "Metropolitan Statistical Areas have at least one urbanized area of 50,000 or more population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties."

There are many factors that have an influence on population change in urban areas. Short and Mussman (2014) point out that there is a relationship between political economy and urban population change; they indicate the revalorization of a metropolis and the restructuring of economy in United States make population change more complex. In addition, Moomaw and Shatter (1996) state population migration between urban and rural areas, as well as the level of economic development and earning differences, are primary factors influencing population change. Zhang and Song (2003) also indicate that the level of economic development and earning differences have significant influence on urban population growth.

It is necessary to mention that education plays an important role in the process of urbanization. Chen and Huang (2003) demonstrate that better education, medical treatment, and higher income are important factors that accelerate population migration from rural to urban locations, according to their survey in Zhejiang province in China. Zhu (2003) also confirms this by stating higher education accelerates population migration from rural to urban areas. In Zhu's statistical analysis, education levels of Bachelor degrees or higher significantly affected accelerating urban population growth. This shows that better educational facilities can attract migration. The result is consistent with the conclusion of Glaeser, Scheinkman, and Shleifer (1995).

Additionally, housing conditions are an important index of living quality in urban areas. Zhu (2003) shows that better housing conditions have significant attraction in population migration. In addition, Zhu suggests people who are highly educated can work in better companies. Chen and Huang (2003) indicate housing is an important factor that affects population migration from rural to urban areas.

The age structure of the population, which means how the population is distributed across different age groups, is also important. “Taking account of age structure provides powerful confirmation of the age-old view that, when it comes to the determination of living standards, population does, indeed, matter” (Bloom *et al*., 2003).

Population change also has spatial dependence. Current literature highlights the importance of location when modeling population change. For example, Lewis and Stanley (2016) note population change in one area is dependent on population change in neighboring areas. It means the location of a study area also has influence on changes within the population. Thus, economy, education, age structure, and housing are variables explored in this study.

Data and Study Area

Study Area

In this study, 30 metropolitan statistical areas, which had a population of approximately 2,000,000 or more in the 2010 census survey, were selected for study. These metropolitan statistical areas and their 2010 population are listed in Appendix A.

Data

Data used in this analysis was obtained from American Community Survey (ACS) 5-Year Estimates. The 2010-2014 ACS 5-Year Estimates include statistics from January 1, 2010 to December 31, 2014. The 2010-2014 ACS 5-year data products include estimates of demographic, social, housing, and economic characteristics for people living in housing units and group quarters. For this study, five factors were examined influencing population change: economy, education, age structure, housing, and location. These factors directed the data collection process.

The survey tables contained population change information and data from the study areas. The names of the tables and datasets utilized are (a) 2014 ACS 5-year estimates dataset – ACS Demographic and Housing Estimates, (b) 2014 ACS 5-year estimates dataset – Earnings in the Past 12 Months (in 2014 Inflation-Adjusted Dollars), (c) 2014 ACS 5-year estimates dataset – Educational Attainment, and (d) 2014 ACS 5-year estimates dataset – Financial Characteristics for Housing Units.

All tables and datasets were derived from the U.S. Census Bureau's American FactFinder. American FactFinder provides access to data about the United States, Puerto Rico and Island Areas. Data in American FactFinder come from several censuses and surveys.

From these tables, the following data from 2013 to 2014 was used in this analysis: percent change of total population, percent change of population 16 years and over with earnings, percent change of population which have Bachelor's degree or higher, percent change of population which are 65 years of age and over, and the percent change of the median house value. All these variables help represent a snapshot of time in the study areas to aid in examining changes from 2013 to 2014.

Methods

Several statistical tests were used to examine relationships between the aforementioned variables.

Pearson's Correlation Coefficient

To describe the relationship between two variables, the covariance (Eq.1) was calculated.

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1} \qquad \text{(Eq.1)}$$

Pearson's correlation coefficient (Eq.2) can be considered as a standardized covariance (Rogerson, 2010) where cov is the covariance and s_x is the standard deviation of x, and s_y is the standard deviation of y.

$$r = \frac{cov(x,y)}{s_x s_y} \quad \text{(Eq.2)}$$

This correlation coefficient shows the strength of the linear relationship between two variables which ranges from -1 to 1. When points lie perfectly on a line, the negative slope returns a result of -1 while the positive slope returns a result of 1. However, a result of zero does not necessarily mean the two variables are not related; they may just not be associated with each other in a linear function (Rummel, 1976; Cohen, 1988; Buda and Jarynowski, 2010).

A significance test (Eq.3) for r is usually applied to test if the true value of the correlation coefficient is equal to zero.

$$\boldsymbol{t} = \frac{r\sqrt{\boldsymbol{n-2}}}{\sqrt{\boldsymbol{1-r^2}}} \quad \text{(Eq.3)}$$

Global Moran's I

Moran's *I* is a measure of spatial autocorrelation developed by Patrick Alfred Pierce Moran (Moran, 1950) (Eq.4).

$$I = \frac{n\sum_i^n \sum_j^n w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{(\sum_i^n \sum_j^n w_{ij})\sum_i^n (y_i-\bar{y})^2} \quad \text{(Eq.4)}$$

where *n* is the number of the regions; w_{ij} is a weight of describing the relationship between regions i and j.

Values near 1 indicate a strong spatial pattern, while values near 0 indicate lack of spatial pattern. Spatial autocorrelation is more complex than one-dimensional autocorrelation because spatial correlation is multi-dimensional (ie. 2 or 3 dimensions of space) and multi-directional. In this study, the Moran's *I* was generated in ArcGIS. Because the closer the cites are in space, the more likely they are to interact/influence each other, the inverse distance squared was used in this method as the conceptualized spatial relationship.

Regression

Regression analysis is a statistical process that analyzes relationships between a dependent variable and independent variables. The equation for linear regression is:

$$\hat{y} = a + b_1x_1 + b_2x_2 + \cdots + b_px_p \quad \text{(Eq.5)}$$

where $\hat{y}$ is the predicted value of the population change.

With the observations, the goal of this method is to find *a* and $b_1, b_2, \ldots, b_p$ which minimize the sum of the squared residuals (Seal, 1967).

This procedure was performed in the *IBM SPSS* statistical software. Some assumptions are made in applying a linear regression model:

a. The relationship between *y* and *x* is linear;
b. The errors have a mean of zero and constant variance;
c. The residuals are independent; the value of one error is not affected by the

value of another error.
d. For each value of x, the errors have a normal distribution about the regression line.

Tests for the regression model includes an F-test (Eq.6) using coefficient of determination (Eq.7) for residuals (Glantz and Slinker, 1990), t-test (Eq.8) for b, and a variance inflation factor (VIF) (Eq.9) used to assess multicollinearity (Allison, 1998).

$$F = \frac{r^2(n-2)}{1-r^2} \quad \text{(Eq.6)}$$

where n is the sample size and r is the coefficient of determination.

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \quad \text{(Eq.7)}$$

The coefficient of determination squared measures the proportion of the total variability in y explained by the regression equation. The F-test is used to decide if the proportion of the variability in y explained by x is equal to zero.

$$t = \frac{b}{s_b} \quad \text{(Eq.8)}$$

where s_b is the standard deviation of the slope, which is given by:

$$s_b = \sqrt{\frac{s_e^2}{(n-1)s_x^2}} \quad \text{(Eq.8)}$$

Test for beta examines the true value of the slope relative to zero.

$$VIF = \frac{1}{1-r_i^2} \quad \text{(Eq.9)}$$

where r is the coefficient of determination.

A common rule for VIF is that a value greater than 5 indicates that multicollinearity may exist among the independent variables. In addition, in order to choose the appropriate independent variables which best explain the dependent variable, a backward selection can be used. This uses all possible variables in the first iteration and removes those that contribute least to r^2 until the removal of any variable will result in a significant reduction of r^2.

Results

In this study, the relationship between percent change of population and its factors were examined. The factors were the percent change of population 16 years of age and over with earnings, the percent change of population which have Bachelor's degrees or higher, the percent change of population which are 65 years of age and over, and the percent change of the median house value. For convenience, these variables were titled for brevity in the study: population change, employment, highly educated, age, and housing value, respectively.

The assumption was these factors are related to the percentage of population change. However, considering the possibility of collinearity among these factors was also important. A regression model was applied to the percent of population change and its factors to show how well the percent of population change can be explained by its factors.

Figure 1 shows the percentage change of population of selected metropolitan areas.

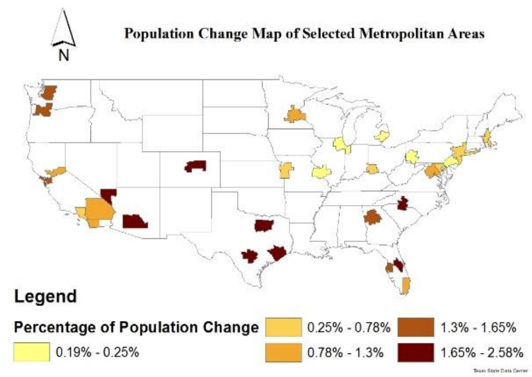

Figure 1. Population change map of selected metropolitan areas. The darker areas have larger population change percentage than the lighter areas. In general, population changes faster in southwest areas than northeast areas.

Pearson's Correlation Coefficient

Percent population change was calculated by dividing population change from 2013 to 2014 by the total population in 2013. Pearson's correlation coefficient was calculated for each factor in comparison to percent population change to determine if a correlation was present (Table 1).

Table 1. Correlation coefficient between population change and each of the four variables displayed (N=30 for all).

Variables	Pearson Correlation	Sig. (2-tailed)
Employment Change	0.727	0.000**
Highly Educated Change	-0.419	0.021*
Elder Change	0.883	0.000**
House Value Change	0.383	0.037*
*Significant at the 0.05 level (2-tailed)		
**Significant at the 0.01 level (2-tailed)		

Table 1 shows Pearson's correlation coefficient of each factor in relationship to percent population change. All of the four Sig. 2-tailed levels are less than 0.05 meaning that there is significant correlation between population change and these four factors. The relationship between population change and the factors is shown by the Pearson correlation value. The positive value means that as one variable increases or decreases so will the other one, while the negative value means that as one variable increases or decreases the other variable will move in the opposite direction. Among the four factors, change in the percent of the population age 65 and older had the strongest correlation with population change.

Global Moran's I

Some variables mentioned have a random spatial distribution, while others have a clustered spatial distribution, as determined from calculation of Global Moran's *I*. Table 2 shows the p-value and z-score of each variable in Global Moran's *I*. Figure 2 shows the graph of Global Moran's *I*.

Table 2. Variables and their values in Global Moran's *I*. Population and housing value were determined to be clustered spatially with the significance level indicated. The remaining variables were found to have a random distribution.

Variables	Significance Level (p-value)	Critical Value (z-score)
Population	0.05	1.96-2.58
Employment	---	-1.65-1.65
Highly Educated	---	-1.65-1.65
Elder (65+)	---	-1.65-1.65
Housing Value	0.01	>2.58

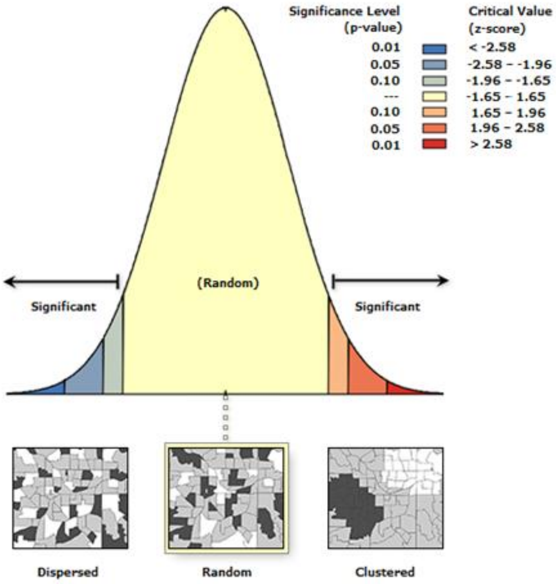

Figure 2. Global Moran's *I* graph. When the p-value is less than 0.10 and the z-score is less than 1.65, it has a dispersed spatial distribution, which means it has negative spatial correlation. When the p-value is greater than 0.10 and the z-score is between -1.65 and 1.65, it has a random spatial distribution, which means it does not have spatial correlation. When the p-value is less than 0.10 and the z-score is greater than 1.65, it has a clustered spatial distribution, which means it has positive spatial correlation.

Table 2 and Figure 2 provide the spatial distribution of each variable. The Global Moran's *I* of employment change, highly educated change, and elder (65+) change have random spatial distribution, meaning they do not exhibit spatial autocorrelation. This can be used in support of the regression analysis assumption that the variables are spatially independent.

The Global Moran's *I* of population change and housing value have clustered spatial distribution with positive spatial correlation. As Lewis and Stanley (2016) stated, population change in one area is dependent on population change in neighboring areas. Spatial dependence also affected the housing value for the same reason. This may have an influence on the regression model and is discussed later.

Regression

Explanatory variables analyzed in this paper were employment, college education, age (65+), and housing value. Since they were all correlated to the percent population change, according to Pearson's correlation coefficient, multiple regression analysis was conducted for the dependent variable (population change) and these factors. Table 3 provides the R and R^2 values. The R value represented the simple correlation and was 0.919 (the "R" Column), which indicated a high degree of correlation. The R^2 value (the "R Square" column) indicated how much of the total variation in the dependent variable can be explained by the independent variable. In this case, 84.5% could be explained, which was very large.

Table 3. Model summary table of linear regression. The dependent variable is population change, while the independent variables are employment, highly educated, age 65+, and housing value.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.919a	0.845	0.820	0.32196%

a. Predictors: (Constant), 2013-2014 House Value change, 2013-2014 Employment change, 2013-2014 Highly educated change, 2013-2014 Elder change

The ANOVA table reported how well the regression equation fit the data. In Table 4, results indicated the regression model predicted the dependent variable significantly well. The significance (the "Sig" column) indicated the statistical significance of the regression model that was run. This value, which was less than 0.05, indicated the regression model statistically predicted the outcome variable.

The coefficients table, Table 5, provides predictive values of population change from factors, as well as determines whether each factor contributes statistically to the model (the "Sig." column).

Table 4. ANOVA table of linear regression.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14.092	4	3.523	33.987	.000[b]
	Residual	2.591	25	0.104		
	Total	16.683	29			

ANOVA[a]

a. Dependent Variable: 2013-2014 population change
b. Predictors: (Constant), 2013-2014 House Value change, 2013-2014 Employment change, 2013-2014 Highly educated change, 2013-2014 Elder change

Table 5. Coefficients table of linear regression.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	-0.888	0.387		-2.295	0.030		
	2013-2014 Employment change	0.367	0.121	0.308	3.032	0.006	0.602	1.661
	2013-2014 Highly educated change	0.044	0.391	0.012	0.111	0.912	0.542	1.846
	2013-2014 Elder change	0.429	0.074	0.666	5.772	0.000	0.467	2.140
	2013-2014 House Value change	0.016	0.016	0.099	1.001	0.326	0.637	1.569

a. Dependent Variable: 2013-2014 population change

Factors' influence on population change can be ranked by using the values in the "Beta" column under the "Standardized Coefficients" column. Elder (age 65+) affected population change most. After that was employment. The third was housing value. Highly educated affected population change the least.
All the *VIF* values were less than 5, meaning that multicollinearity did not exist among the independent variables. The significance of each factor is listed under the "Sig" column. Only the elder (age 65+) had a significance less than 0.05, which means this model needed to be rebuilt. Backward selection was used in this step. The coefficients table, Table 6, provides the model iterations.

Table 6. Coefficients table of linear regression iterations using backward selection method.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	-0.888	0.387		-2.295	0.030		
	2013-2014 Employment change	0.367	0.121	0.308	3.032	0.006	0.602	1.661
	2013-2014 Highly educated change	0.044	0.391	0.012	0.111	0.912	0.542	1.846
	2013-2014 Elder change	0.429	0.074	0.666	5.772	0.000	0.467	2.140
	2013-2014 House Value change	0.016	0.016	0.099	1.001	0.326	0.637	1.569
2	(Constant)	-0.852	0.206		-4.129	0.000		
	2013-2014 Employment change	0.370	0.117	0.310	3.168	0.004	0.624	1.603
	2013-2014 Elder change	0.426	0.067	0.661	6.354	0.000	0.553	1.808
	2013-2014 House Value change	0.016	0.014	0.093	1.118	0.274	0.858	1.165
3	(Constant)	-0.836	0.207		-4.042	0.000		
	2013-2014 Employment change	0.358	0.117	0.300	3.062	0.005	0.629	1.589
	2013-2014 Elder change	0.452	0.063	0.701	7.159	0.000	0.629	1.589

a. Dependent Variable: 2013-2014 population change

In the third model of Table 6, both significance values of the factors were less than or equal to 0.005. This model was used as the final model. The R-squared value for it was 0.84, which means 84% of population change can be explained by employment change and elder change. Population change can be predicted by using the values in the "B" column under the "Unstandardized Coefficients" column in model 3 (Eq.10).

$p = -0.836 + 0.358 * em + 0.452 * el$
(Eq.10)

p = population
em = employment
el = elder (age 65+)

Discussion

Methods used helped to explore general relationships between variables and the percent change of population in several large cities. Results of correlation coefficient calculations show all factors examined could be used in linear regression.

Results of Global Moran's *I* show, except for the population change and the housing value, variables were spatially independent. Spatial autocorrelation affected population change and housing value. Results of linear regression illustrate how all of the factors impact population change. The formula (Eq.10) was derived by using values in the B column in the third model of the coefficients table.

Findings suggested the factors studied are related to the percent of population change and statistical and spatial statistics can help to infer some level of predictive analysis to understand population change based on demographic factors.

For future work, one may consider adding additional variables or studying a longitudinal impact by including data from additional years. By using new methods and more data, greater trends and factors may be identified.

Conclusion

The relationship between the percent change of population in metropolitan statistical areas and its factors were examined. The four factors were (a) the percentage change of population 16 years and over with earnings, (b) the percentage change of population who have Bachelor's degrees or higher, (c) the percentage change of population 65 years of age and over, and (d) the percentage change of the median house value.

Pearson's correlation coefficient, Global Moran's *I*, and regression were used to help analyze relationships between variables. As a result, employment and elder age (65+) significantly related to the percent of population change, of which the "elder" variable (age 65+) had the most significant effect.

Acknowledgements

It has been a long time since this project started; I appreciate everyone who gave suggestions and help to me during this period. I would like to express my gratitude to my advisor, Mr. John Ebert. He gave me a lot of advice on this project and helped me a lot through my entire study life on campus.

I would also like to say thank you to Mrs. Greta Poser. She taught me a lot of useful knowledge and technology, which was used in this project and will help a lot in my future life. She also answered and helped me a lot in my daily life.

At last, I want to thank all my friends and my parents. Thank all of you for supporting me in my student life in Saint Mary's University of Minnesota.

Above all, thank you.

References

Allison, P. D. 1998. *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge Press.

Bloom, D. E., Canning, D., and Sevilla, J. 2003. *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change.* Santa Monica, CA: Rand.

Brown, L. R., and Flavin, C. 1999. A new economy for a new century. *State of the World 1999. A Worldwatch Institute Report on Progress Toward a Sustainable Society*, 3-21.

Buda, A., and Jarynowski, A. 2010. *Life time of correlations and its applications*. Wydawnictwo Niezależne.

Chen, X., and Huang, Z. 2003. Migration of Surplus Agricultural Labor in the Process of Economic Transition. *Chinese Rural Economy, 5,* 33-37.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

DeFries, R. S., Rudel, T., Uriarte, M., and Hansen, M. 2010. Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience, 3,* 178-181.

Ehrlich, P. R., and Holdren, J. P. 1971. Impact of Population Growth. *Science, 171(3977),* 1212-1217.

Glaeser, E. L., Scheinkman, J. A., and Shleifer, A. 1995. Economic Growth in a Cross-Section of Cities, *Journal of Monetary Economics, 36*, 117-144.

Glantz, S. A., and Slinker, B. K. 1990. *Primer of Applied Regression & Analysis of Variance.* McGraw-Hill.

Lewis, W., and Stanley, B. 2016. An Integrated Spatial Model of Population Change in South Carolina Counties. *The Review of Regional Studies, 45(2),* 127-142.

Moomaw, R. L., and Shatter, A. M. 1996. Urbanization and Economic Development: A Bias toward Large Cities? *Journal of Urban Economics, 40(1),* 13-37.

Moran, P. A. P. 1950. Notes on Continuous Stochastic Phenomena. *Biometrika, 37,* 17-23.

Rogerson, P. A. 2010. *Statistical Methods for Geography: A Student's Guide*. Los Angeles, CA: Sage.

Rummel, R. J. 1976. *Understanding Correlation*. Retrieved from https://www.hawaii.edu/powerkills/UC.HTM.

Seal, H. L. 1967. *The historical development of the Gauss linear model.* Yale University.

Short, J. R., and Mussman, M. 2014. Population Change in U.S. Cities: Estimating and Explaining the Extent of Decline and Level of Resurgence. *The Professional Geographer, 66 (1),* 112-123.

Statistical and Science Policy Branch, Office of Information and Regulatory Affairs and Office of Management and Budget. 2003. *Metropolitan Statistical Areas, Micropolitan Statistical Areas, Combined Statistical Areas, New England City and Town Areas, and Combined New England City and Town Areas*. Retrieved from https://www.whitehouse.gov/sites/default/files/omb/bulletins/b03-04_attach.pdf/.

Zhang, K. H., and Song, S. 2003. Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses. *China Economic Review, 14(4),* 386-400.

Zhu, J. 2003. Effects of Expansion of Higher Education in China on Urbanization and Economic Growth. *Chinese Journal of Population Science, 1,* 32-38.

Zhu, N., and Zeng, Z. 2004. The Determinants of Urban Population Growth in China. *Chinese Journal of Population Science, 5*, 9-18.

Appendix A

Table A1.
Thirty metropolitan statistical areas which have more than or about 2,000,000 total population in the 2010 census survey in the United States by rank level and their 2010 census values. Higher ranks listed equate to greater population.

Rank	**Metropolitan Statistical Area**	**2010 Census**
1	New York-Newark-Jersey City, NY-NJ-PA Metro Area	19,567,410
2	Los Angeles-Long Beach-Anaheim, CA Metro Area	12,828,837
3	Chicago-Naperville-Elgin, IL-IN-WI Metro Area	9,461,105
4	Dallas-Fort Worth-Arlington, TX Metro Area	6,426,214
5	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area	5,965,343
6	Houston-The Woodlands-Sugar Land, TX Metro Area	5,920,416
7	Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area	5,636,232
8	Miami-Fort Lauderdale-West Palm Beach, FL Metro Area	5,564,635
9	Atlanta-Sandy Springs-Roswell, GA Metro Area	5,286,728
10	Boston-Cambridge-Newton, MA-NH Metro Area	4,552,402
11	San Francisco -Oakland-Hayward, CA Metro Area	4,335,391
12	Detroit-Warren-Dearborn, MI Metro Area	4,296,250
13	Riverside-San Bernardino-Ontario, CA Metro Area	4,224,851
14	Phoenix-Mesa-Scottsdale, AZ Metro Area	4,192,887
15	Seattle-Tacoma-Bellevue, WA Metro Area	3,439,809
16	Minneapolis-St. Paul-Bloomington, MN-WI Metro Area	3,348,859
17	San Diego-Carlsbad, CA Metro Area	3,095,313
18	St. Louis, MO-IL Metro Area	2,787,701
19	Tampa-St. Petersburg-Clearwater, FL Metro Area	2,783,243
20	Baltimore-Columbia-Towson, MD Metro Area	2,710,489
21	Denver-Aurora-Lakewood, CO Metro Area	2,543,482
22	Pittsburgh, PA Metro Area	2,356,285
23	Portland-Vancouver-Hillsboro, OR-WA Metro Area	2,226,009
24	Charlotte-Concord-Gastonia, NC-SC Metro Area	2,217,012
25	Sacramento-Roseville-Arden-Arcade, CA Metro Area	2,149,127
26	San Antonio-New Braunfels, TX Metro Area	2,142,508
27	Orlando-Kissimmee-Sanford, FL Metro Area	2,134,411
28	Cincinnati, OH-KY-IN Metro Area	2,114,580
29	Kansas City, MO-KS Metro Area	2,009,342
30	Las Vegas-Henderson-Paradise, NV Metro Area	1,951,269