The Impact of Ethnic Diversity on Socioeconomic Outcomes in the Metropolitan Area

Dongsuk Jeong
Economics Honors Thesis
Spring 2017

Dr. Sukjin Han
Department of Economics
Supervising Professor

Dr. Stephen J. Trejo
Department of Economics
Second Reader

Abstract

A goal of this research is to see how much changes of residential composition in ethnic diversity affect socioeconomic outcomes such as income per capita and crime rate for the period 1990 and 2000. In this paper, ethnic diversity is defined as a heterogeneous residential pattern within each metropolitan statistical area. Among various indices measuring the level of ethnic diversity, we use the multi-group entropy index based on Census tract regarding the goal of this study. To capture the impact of ethnic diversity on socioeconomic outcomes, we use a panel data model with fixed effect, specifically within estimation, in the metropolitan area since the fixed effect model eliminates unobservable individual effects such as social cognition effects, natural shock effects, and potential growth. Our result suggests that an increase in ethnic diversity helps to decrease total crime rate and increase income per capita.

Content

1. Introduction

With the world's trends of population movement and development of transportation technologies, it gives a better chance for people all over the world to move to the United States and encourage ethnic diversity in society. The increase of ethnic diversity leads to demographical changes that affect economical facts such as labor forces, consumption patterns, and productivity. A generally accepted fact is that ethnic diversity functions positively in our society by improving productivity and generating social cohesion. However, increased ethnic diversity also brings concerns to U.S. residents about increase in crime and related problems in neighborhoods. In addition, along with the trends of anti-immigrant laws, this seems to reflect the concern of not only people, but also the U.S. government. Racial and cultural conflicts between minority and majority groups through history are still the center of social issues in the United States.

This reoccurring concern is not restricted to the U.S. Refugee policies that are heading in the opposite direction in Europe are one of example. According to Rick Lyman—in Warsaw, the capital of Poland—residents are resistant against taking even a tiny fraction of Muslim refugees because of different religious beliefs, while German and other nations have opened the door for them. Even for countries deciding to take refugees, they show contrasting opinions toward how to embrace the refugees. For example, the German and British insist on the “multicultural” model, while the French hold an “assimilationist” approach. We can tell ethnic diversity is an unavoidable issue in today’s society, both socially and economically. Ethnic diversity has been debated for over a decade and there are still no clear results indicated from the various approaches because of these different viewpoints of ethnic diversity.

From Alesina and Ferrara's literature (2005, p. 794) find "rich democratic societies work well with diversity, in the case of the United States very well in terms of growth and productivity". Even within the developing world, similar levels of ethnic diversity are associated with very different degrees of conflict and interethnic cooperation". We can find more results regarding ethnic diversity in other existing literature. According to Alesina and La Ferrara (2005), there is a negative relationship between ethnic diversity and the quality of public goods. Luttmer (2001) finds that ethics diversity negatively influence welfare spending. Fearson and Laitin (2000) concluded that more diversity decrease the civil conflict and trust. Moreover, Alesina, Baqir, and Easterly (1999) think that economic growth slower if there are more races. Montalvo and Reynal-Querol (2005) use the index of polarization to measure religious and ethnic diversity and find a direct negative effect on economic growth.

Unlike the results from previous literatures, we believe ethnic diversity bring the positive effect into today's society. This leads us to find out the impact of ethnic diversity on socioeconomic outcomes such as income per capita and crime rates within metropolitan statistical area. We use the multi-groups entropy index to measure ethnic diversity. Our model is panel data model with fixed effect to eliminate unobserved individual effect. We find that the ethnic diversity is associated with income per capita positively and criminal rate negatively. This implies that ethnic diversity encourages productivity or social cooperation between native residents and immigrants.

This paper is organized as follows. Section II present definition of the ethnic diversity, population density, and dependent variable. Section II also discuss about how we measure the ethnic diversity. Section III describes the panel data models, assumptions, and data collection. Section IV explain the results and possible implications and section V concludes.

2.1 Ethnic diversity and density of population

Ethnic diversity could be interpreted in many different ways. It would be interpreted by primary spoken language, by the ratio of population between racial or ethnic groups, or by different cultural groups. According to the Census Bureau, whites made up 61.6 percent of total population in 2015, while blacks composed 13.3 percent, Hispanics 17.6 percent, Asian 5.6 percent, and other minor ethnic groups composed the remainder. At first thought, this statistic data seems to show the level of ethnic diversity in the United States. However, this data only tells us the percent of the population separated into different groups, not by composition or residence patterns. Since this paper tries to find out the effects of ethnic diversity on socioeconomic outcomes, a simple number of ethnic groups in a community's population is not enough to determine ethnic diversity.

According to Lee, Iceland and Sharp (2012), they define ethnic diversity as, “Population consisting of many groups of equal size would be highly diverse”. In this paper, we consider the definition of highest ethnic diversity as the equal size population of ethnic groups with heterogeneous residential composition. To measure the appropriate ethnic diversity that indicates the heterogeneous ethnic composition in a community, we consider two ethnic diversity indices: interaction index and multi-group entropy index. From Iceland’s literature (2004, p. 7-8), formulation of indices are described below:

Interaction index: $\sum_{i=1}^{n} (\frac{x_i}{X})(\frac{y_i}{t_i})$

Entropy index: $\sum_{r=1}^{n} (p_r) ln(1/p_r), p_r = r_i / t_i$

$Multiple\ entropy\ index$: $\sum_{i=1}^{n} [\frac{t_i(E-E_i)}{ET}]$

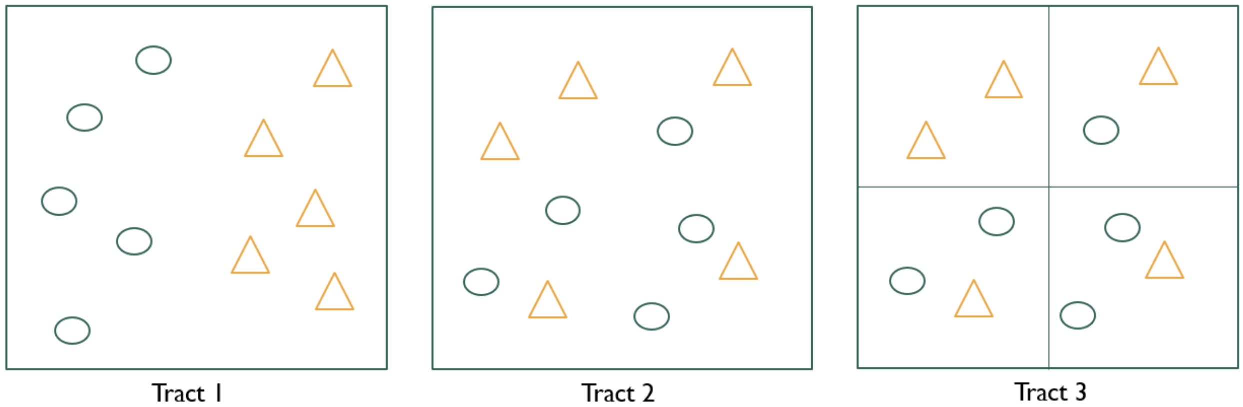
where x_i indicates the minority population of area i, y_i is the majority population of area i, X is the sum of all x_i, and t_i is sum of tracts population. p_r refers to a particular racial/ethnic group's proportion of the whole metropolitan population. t_i refers to the total population of tract i, T is the metropolitan area population, n is the number of tracts, E_i is the entropy score with census tract unit, and E represents tract i 's diversity (entropy) and metropolitan area diversity, respectively.

The interaction index is one of the racial and ethnic residential segregation indices from the U.S Census Bureau's report by Iceland, Weinberg, and Steinmetz (2002, p. 120). According to report, the interaction index is the "measurement of the degree of potential contact, or possibility of interaction, between minority and majority group members". This simply indicates the composition of one racial group in proportion to other groups in the community. The entropy index itself is not considered a measurement of residential patterns of ethnic diversity because it does not measure the distribution of groups within a metropolitan area. For both the interaction and entropy indexes, a high score means general high diversity by population. Multi-group entropy index is the measurement of residential patterns—either homogeneous or heterogeneous.

According to Iceland (2004, p. 6), multi-group entropy index is a measure of evenness—the extent to which groups are evenly distributed among organizational units. The maximum value of multi-group entropy would occur when all ethnic groups have equal proportion within total population. Unlike the interaction and entropy index, a lower score indicates “a population consisting of more equal-sized by all ethnic groups” (Lee, Iceland & Sharp, 2012). In addition, the multi-group entropy index is not influenced by the size of each groups so that it is able to measure how evenly the groups are distributed, while interaction index and entropy index are dependent on the relative size of each group (Iceland, 2004).

The choice of income per capita and crime rate as dependent variables suggests a better idea that which index is best suited regarding to the purpose of our study. Since the measurement of income per capita and crime rate is likely influenced by population size, there is a high chance that the relationship between ethnic diversity and dependent variables can be leaded to wrong ways if we use the interaction index. Because interaction index is highly influenced by the size of each group, it will show the inconsistent level of the ethnic diversity from different sizes of metropolitan areas. This would indicate a wrong relationship to dependent variables. However, the multi-group entropy index will prevent this problem because multi-group entropy index is able to show better measurement of ethnic diversity and suggests the proper relationship to dependent variables. Moreover, entropy index is aimed to measure the evenness of each groups’ distributions and not influenced by size of each groups.

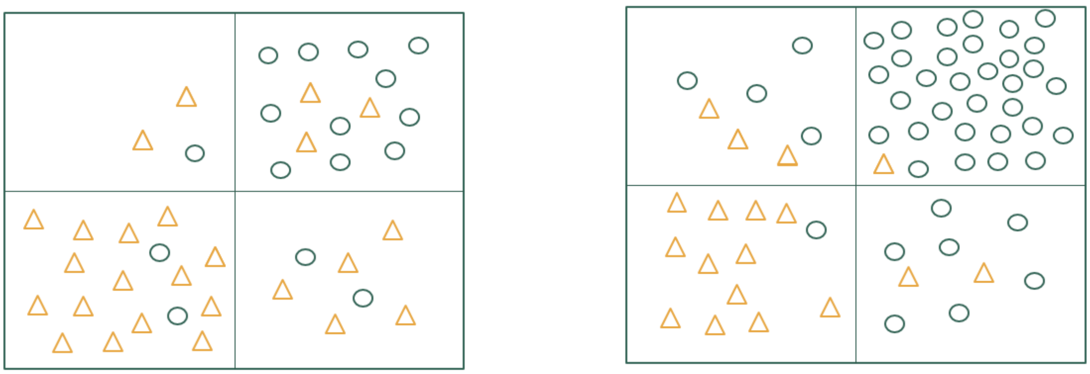
To verify which index is proper to measure ethnic diversity for the purpose of this paper, we conducted a simple experiment by setting up three hypothetical Census tract units.



Each community is composed by two racial groups, X and Y. In tract 1, 5 member of group X live in the west and 5 member of group Y live in the east. In tract 2, 5 members of group X and 5 members of Y live close each other. Considering the format of ethnic composition in tract 1 and 2, tract 1 represents the segregated residential pattern while tract 2 represents the heterogeneous residential pattern. Even though both tracts have the same population, we intuitively consider that tract 2 has a higher level of ethnic diversity compared to tract 1because there would be little interaction between the two groups in tract 1. This indicates that the diversity level in a given area can vary if all minority groups are present, but also can be very highly "segregated" if all groups live exclusively in their own neighborhoods. Therefore, how can the different geographical units be studied?

In tract 3, it has the same population and same residential patterns as tract 2, but is composed by smaller geographical unit. Tract 3 is divided by North West, North East, South West, and South East. If we measure the level of ethnic diversity, tract 3 will show a different score than tract 2 because each small unit will have a different diversity level. This implies that, without proper geographical units, it is more likely that statistics do not reflect the resident patterns in society. If the boundary is too narrow, there is a possibility that it only catches segregated residential compositions. On the other hand, if it is too broad, it is possible it does not reflect the ethnic diversity composed by each ethnic group. Depending on residential pattern and geographical units, the measurement of ethnic diversity can be changed.

Based on these intuitive approaches, we design our experiment to reflect more realistic by setting up hypothetical cities as composed by different population and residential patterns.



City A　　　　City B

Interaction index (City A): $(\frac{2}{25})(\frac{1}{3}) + (\frac{5}{25})(\frac{2}{7}) + (\frac{15}{25})(\frac{2}{17}) + (\frac{3}{25})(\frac{11}{14})$ = 0.25

Interaction index (City B): $(\frac{4}{22})(\frac{4}{8}) + (\frac{1}{22})(\frac{32}{33}) + (\frac{2}{22})(\frac{7}{9}) + (\frac{15}{22})(\frac{1}{16})$ = 0.25

Multi-group entropy index (City A):

$- \{[(\frac{1}{3})ln(\frac{1}{3}) + (\frac{2}{3})ln(\frac{2}{3})](\frac{3}{41}) + [(\frac{2}{7})ln(\frac{2}{7}) + (\frac{5}{7})ln(\frac{5}{7})](\frac{7}{41}) + [(\frac{2}{17})ln(\frac{2}{17}) + (\frac{15}{17})ln(\frac{15}{17})](\frac{17}{41}$

= 0.77

Multi-group entropy index (City B):

$- \{[(\frac{4}{8})ln(\frac{4}{8}) + (\frac{4}{8})ln(\frac{4}{8})](\frac{8}{66}) + [(\frac{32}{33})ln(\frac{32}{33}) + (\frac{1}{33})ln(\frac{1}{33})](\frac{33}{66}) + [(\frac{7}{9})ln(\frac{7}{9}) + (\frac{2}{9})ln(\frac{2}{9})](\frac{9}{66}$

= 0.82

Through computation by each formula, the results indicate that the interaction index score is 0.25 for both cities. Multi-group entropy index score is 0.77 for city A and 0.82 for city B. From the results, this implies that the multi-group entropy index is able to catch more types of residential patterns while the interaction index is not. The interaction index might be good to measure the segregation between two groups, but if we consider multiple groups, then it would not suit the study the best in terms of this paper since we consider heterogeneous residential patterns by multiple groups as high levels of ethnic diversity.

Moreover, the score of the interaction index is resulted by exposure level X to Y and Y to X. If the population of X and Y is different, the score of X to Y and score of Y to X is different. Intuitively, it would be hard to get a proper weighted value if there is more ethnic groups and there would be higher chance of losing accuracy. Since the exposure index is not able to reflect the various residential patterns in today's society, this is not proper to measure the composition of heterogeneous ethnic diversity. As we found out from the above experiment, the multi-group entropy index is an appropriate alternative to measuring ethnic diversity. Multi-group entropy index still has some issues because it is aimed at measuring the evenness of ethnic compositions, it doesn't reflect the kind of ethnic groups. However, regarding the purpose of this paper, the heterogeneous ethnic group itself is what matters rather than the kind of ethnic group. As discussed above, by measuring evenness, it would solve the problem from the exposure index. Moreover, by weighting each entropy score on census tract units within the metropolitan-wide entropy, this problem could be solved by the boundary problem. The more detailed information about census tract will be explained later in this paper. Thus, multi group entropy index is reasonable to use as a measurement of the ethnic diversity for the purpose of this paper.

Another important variable in this paper is population density. Intuitively, higher concentration rates indicate more interchange between different ethnic groups within common geographical areas. Higher interchange among neighbors might cause more conflict because of different cultural factors. Or, it might encourage social cohesion among neighbors, and then generate more ideas and different approaches to the problem in various ways so it brings more productivity into society. Since it is not clear whether the high interchange has a positive or negative effect, the population density has been a subject of debate for a long time. According to Keith (2006, p. 24), high density offers chances for property crimes because of a surrogate for distribution of private property which offers attractive targets to thieves. On the other hand, it prevents violent crimes because there are more chances for people to witness the crime so they are more likely to report them to police. Keith (2006, p. 30) also argues that depending on the definition of high density, type of crimes, or scale or society, the effect would be different and complicated.

2.2. The Dependent Variables: Income per Capita and Crime rate

With a recent change of political trends, an economic recession, and social issues relating to racial conflicts, we select income per capita and crime rate as dependent variables to see the impact of ethnic diversity. Moreover, the reason we select these variables was because both of these variables have a high connection to ethnic diversity, population density, and other covariates in this paper. According to Kathieen (2012), ethnic diversity increases home value and lowers crime in Southern Carolina. Moreover, Mckinsey (2015) found that firms in the top quantile for racial and ethnic diversity are 35 percent more likely to financial return compared to median in their industry.

We can approach the relationship between income per capita and ethnic diversity intuitively. If there are organizations or neighborhoods with well composed various ethnicities, it is likely to improve productivity because a working environment with well-developed diversity generates new approaches of thinking based on diverse cultures and knowledges. If ethnic diversity has a positive impact on productivity, it allows lowering manufacturing costs, reworking, and scrapping costs so it will bring more profits to companies. The increased profits encourage companies to hire more human capital until human capital productivities reaches a maximized point of function, and then it would likely help to decrease the unemployment rate and increase the income per capita. Ankita (2014) shows that employing a diversified work force is very necessary today and definitely leads to improved productivity. However, poor management and cultural differences could be barriers to this. Although there has been much literature research, the relationship between ethnic diversity and productivity still is under controversy since there are many unobserved facts to consider and hardships to measure clear results of productivity. Since this paper is focused on the impact of diverse residential patterns within metropolitan areas, it might be hard to directly connect the relationship between ethnic diversity and productivity. However, this paper will contribute to understanding the relationship between crime rate, income per capita, and ethnic diversity.

3.1 Models

Since the relationship between ethnic diversity and dependent variables, crime rates, and income per capita are likely to have unobserved individual effects, we use the panel data model with the fixed effect model. We denote the fixed effect model as FE model in this paper. We tried to find exogenous variables to eliminate the individual effects, but because of limitations of accessing data and many complicated correlations between dependent and independent variables, the panel data model with the fixed effect fit best for this paper. Panel data model is described as below.

Panel data model: $y_{it} = \beta X_{it} + a_i + u_{it}$

where t denotes a time period 1990 to 2000 and i denotes the metropolitan area. y_{it} indicates income per capita and total crime rate. X_{it} are the multi-group entropy index, population density and other covariates. a_i indicates metropolitan effect in this paper and u_{it} indicates the unobserved error terms.

Since a_i is most likely correlated with other covariates, in this paper, we use the fixed effect model with within estimation. In the FE model, the unobserved individual effects will be eliminated by demeaning variable through the within estimation. Within estimation is described as below.

Within estimation: $y_{it} - \overline{y}_i = \beta(X_{it} - \overline{X}_i) + (u_{it} - \overline{u}_i)$

where $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\overline{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$ and $\overline{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$

There is an alternative to this estimation, first-difference estimation, denoted the FD model. Theoretically, The FE model will be better when the error terms are serially uncorrelated with covariates over FD model. Moreover, the FE model usually gives a better result in many cases in sense of giving precise parameter estimates and small standard errors. Since we are interested in controlling for average differences across metropolitan area, the within-group action is proper rather than first-difference estimation. The effect of ethnic diversity at the group level indicates the better result in terms of our paper. In addition, FE model also loosen the concern about omitted factors that might correlated with predictors at group level. Therefore, FE model is best suit for our study.

In this paper, we set up the two models to see the impacts of ethnic diversity on income per capita and crime rate. These models are aimed to keep multiple entropy index and population density at a significant level. Using both income per capita and total crime rate, we set up independent variables as the multiple entropy index, total population density, each racial population within the metropolitan area, total female population, migration from same county to different county, total poverty level, and total foreign population. For income per capita, we added family structure and education attainments—high school with no diploma, high school with diploma, college with diploma, graduate school and professor and unemployment rate. The model is described below:

Table 1. Description of the Dependent & Independent Variables

Dependent	Independent	
Total Crime Rate & Income per Capita	Multi-Groups Index (entropy)	Total foreign population (totalforeignp)
	Population density (l_density)	Migration (diffcountyp)
	White population (whitep)	Total poverty population (totalpovp)
	Hispanic population (hispanicp)	Family structure (familyp)
	Black population (blackp)	High school with diploma (highdipp)
	Asian population (asianp)	College level (collegep)
	Other group population (otherp)	Graduate & Professor level (grad_profp)
	Total female population (totalfemalep)	Unemployment rate (unemprate)

Note: notations used in model are described inside parenthesis

From the above model, even though the sample size for the income per capita model is larger than the total crime rate models, the reason we set up the same independent variables is that independent variables have a strong connection to both dependent variables. It is a generally known fact in today's society that higher education attainment gives a higher chance of earning high income compared to lower education attainment. At the same time, higher educated individuals are likely to lower the crime rate. Family structure also needs to be considered in this model. Intuitively, a change in family structure would influence children's educational assessment or working conditions, and affects either child's income in the future or labor force's current income in the present. Daron (2001) found that a 10 percent increase in family income is associated with a 1.4 percent increase in attending four-year colleges. Family structure also influences the crime rate in a similar way. Changed educational environment by family structure affects children's behavior and increases the possibility of committing a crime if they get less education.

Moreover, According to a survey conducted by the Census Bureau (2003), it indicates that foreign-born workers are more likely to work in service occupation than native workers. It also shows that 30.5 percent of foreign-born workers have full time occupations and 7.5 percent of foreign-born were unemployed. As known facts, gender and age distributions are likely to have connections to each dependent variable. Ching-Chi (2016), poverty level and income inequality are associated with violent crime. With the intuitive approach, since violent crime and unemployment have similar characteristics, poverty level is also associated with unemployment. Spoken language at home might have less of a relationship with dependent variables, but it affects them at least indirectly or directly. Julian and Robert (2003), for every four immigrant students who spoke languages at home other than English and enroll in public school, one native student switches to a private school. Wolbers (2000) indicates the least educated employees have a higher risk of becoming unemployed than well-educated employees. This suggests that unemployment has a relationship to education which also strongly connects to crime rate and income per capita.

3.2 Assumptions

In our paper, we hold several assumptions relating to the panel data model with fixed effect. The first assumption (Independency) is that collected data is independently and identically distributed, usually denoted as i.i.d. Here, i.i.d would mean that observation for two metropolitan area are independent of each other.

The second assumption (time-invariance) is that unobserved individual effect denoted as a_{it} is time-invariant. In this paper, a_{it} is categorized by three aspects; the natural shock effect, social cognition effect, and potential growth. The natural shock effect would include the effect of catastrophes or weather on society. We assume that damage or influence by weather or temperature does not change rapidly between given time periods. The natural shock effect also generally includes effects generated by geographical conditions such as latitude, altitude, desert, and coast etc., on human nature. Social cognition effect would indicate natural human characteristics such as level of violence and humanity. Potential growths indicate the abilities of human capital such as individual productivity and attitudes toward work, and unexpected economic shocks such as dramatic oil price changes that last long enough to change the macro-economic system.

The third assumption (strict exogeneity) is that covariate x_{it} is uncorrelated with not only current u_{it}, unobserved effect, but also is uncorrelated with all past and future u_i. This assumption is also generally referred to as “strict exogeneity”, which is the main assumption to run the regression in this paper. Even though a_{it} is correlated with the other explanatory variables, the FE model is still valid because $u_{i2} - u_{i1}$ is uncorrelated with x_{i2} and x_{i1} in two time periods. If strict exogeneity holds, an endogeneity problem will not occur, and means the estimation is still consistent.

However, this is a very strong assumption to hold in real life because it is likely the "feedback effect" will occur. This means that today's covariate depends on the past of either u_{it} or y_{it}. To check whether the strict exogeneity holds or not, a comparison between first difference and within estimations is the proper method. If strict exogeneity holds, then the results from both models would be similar, otherwise, the strict exogeneity fails. In the case that strict exogeneity fails, additional instrument variables will be needed. In this paper, since third assumption is such a strong assumption, we loosen this assumption.

3.3 Data collection

In this paper, we used the multi group entropy index constructed by John Iceland (2004). His index is based on the 1980, 1990, and 2000 decennial census data. In his paper, he discussed the main data issue regarding calculating racial and ethnic residential patterns. The first issue is about racial classification. Since data existing from 1980, 1990, and 2000 have different ethnic classifications, Iceland constructed six categories to facilitate the comparison across time: Non-Hispanic Whites, Non-Hispanic African Americans, Non-Hispanic Asians and Pacific Islanders, Non-Hispanic American Indians and Alaska Natives, Non-Hispanics of other races, and Hispanics. According to him, "Having mutually exclusive and exhaustive categories is essential for constructing a single multiracial index. For Census 2000, this involved combining the Asian and Native Hawaiian or other Pacific Islander groups. In addition, non-Hispanic people who identified themselves as being of two or more races in 2000 were also categorized as "Other" since people could not mark more than one race in 1980 or 1990".

The second issue is about geographic area. Since his index is aimed to measure the residential patterns within a larger area, he defined both an appropriate larger area and its component parts. A larger area is represented by an independent metropolitan statistical area as it is generally considered a reasonable approximation of the housing market and component parts indicate the census tract unit. He used the data from Census 2000 to compare a difference over time. According to his additional explanation, "In 2000, there were 331 MAs in the U.S. For this analysis, six MAs were omitted, Barnstable-Yarmouth, MA, Flagstaff, AZ-UT, Greenville, NC, Jonesboro, AR, Myrtle Beach, SC, and Punta Gorda, FL, because they had fewer than 9 census tracts and populations of less than 41,000 in 1980". All other MAs used had populations of at least 50,000 in 1980, which is typically one of the criteria for defining an area an MA. Regarding the components of the geographic area, he used the analysis of census tract unit. Since the census tract is designed to represent neighborhoods, it is reasonable to use this unit from a heuristic perspective. As we discussed earlier in this paper, population density is defined by total population divided by metropolitan land area. Since population density has integer number in data, we use the natural log function to match with other variables.

In terms of collecting data for covariates, we used the data set from the Census Scope. The data from Census Scope is analyzed by the Census and Social Science Data Analysis Network (SSDAN). The boundary of these data sets is also followed the metropolitan area from Census 2000, it is reasonable to use it for comparison over time. Some adjustments were made in a process of analysis in this paper because, compared to numbers of predictors, sample sizes were small. For population estimate, Hawaiians and Pacific Islanders were included in non-Hispanic Asians in 1980 and 1990 in a given data set from Census Scope. Relating to this paper, we combined the minor ethnicities, Hawaiians and Pacific Islanders and others, into "othersp". In family structure, family means more than one person living together in the same place, either married or of the same bloodline. Based on a given data set, we combined married couples, female householders, and male householders into family. The rest of the proportion in family structure is non-family. For age and gender distribution, we used the proportion of total female population. This will indicate the comparison between female and male proportions in the model. For migration, we divided data into the unit of county levels. Since multiple entropy index is based on census tract units, sorting mobility by county will reflect a better result regarding residential patterns. For total foreign population and total poverty population, we used simply a given data set since it was already sorted by total population. The remainder of the total foreign population is total native population and the rest of total poverty population is non-poverty population. Since all of the data set from Census Scope is formed as percentages, there will be no issue regarding combining the data sets into one.

For total crime rate, we collected data from the report by Elizabeth Kneebone and Steven Raphael. They use data from 1990, 2000, and 2008 uniform crime report, denoted as UCR offenses and cleared by data sets compiled by FBI. UCR data provide counts of each crime rates reported to the police. They selected the locations within the country's 100 largest metropolitan areas, as defined by the U.S. Office of Management and Budget in 2008. To determine the location of reporting unit in metropolitan areas, they identified the unit of geography with police agency is associated. For demographic data, they matched community-level crime data from the decennial census and the American Community Survey.

For unemployment rate data, we used the table for civilian labor force and unemployment by metropolitan area seasonally adjusted from Bureau of Labor Statistics. In this data set, unemployment rate is defined as the number of unemployment divided by labor force on each metropolitan area. According to a description of data, this files contain estimates from January 1990 forward. The statistical technique used to adjust estimates is called SEATS, Signal Extraction in Auto regressive integrated moving average Time Series. For each area and division, separate SEATS specifications are used to seasonally adjust employment and unemployment levels. The SEATS specifications also vary by to minimize problems from decennial discontinuities in the not-seasonally-adjusted input data. On November 2, 2016, data for metropolitan areas and metropolitan divisions in Wisconsin or with parts in Wisconsin and data for metropolitan areas in Montana were subject to revision from January 1990 through December 1999 to incorporate special corrections to the not-seasonally-adjusted input data for selected months and subsequent re-fitting of the seasonal adjustment models for that decade.

For income per capita, we use the table of personal income summary within metropolitan statistical area from Bureau of Economic Analysis, denoted as BEA. In this table, Per capita personal income is calculated as the total personal income of the residents of a given area divided by the population of the area. In the computation of per capita income, BEA uses Census Bureau mid-year population estimates. Personal income is measured before the deduction of personal income taxes and other personal taxes and is reported in current dollars, not adjusted for inflation. In this table, several adjustments made. The estimate for the District of Columbia was separated from Valencia in June 1981 because it differs slightly from the estimate of personal income in the national income and product accounts because of differences in coverage. La Paz

County, AZ was also separated from Yuma County on January 1, 1983. The Yuma, AZ MSA contains the area that became La Paz County, AZ through 1982 and excludes it beginning with 1983. Virginia combination areas consist of one or two independent cities with 1980 populations of less than 100,000 combined with an adjacent county. Since the individual unit of this paper is metropolitan areas, which have different standard of living, this table is useful to compare the wealth of one population with those to others.

4.1 Estimation & Result

The column (1) from Table 1 indicates that income per capita has a negative relationship with multiple entropy index. The entropy index keeps the significant level at 1 percent. The relationship with each ethnic group indicates that White and Hispanic have a negative relationship with income per capita at 1 percent and 10 percent significant levels, respectively. Other minor ethnic groups hold 1 percent significant level with a positive relationship. We added female and migration variables in column (2) from (1) to see a change of relationship. Coefficients did not change significantly, but p-value did. Entropy index, white, and other ethnic group keep significant levels with 1 percent and population density becomes 10 percent. Total female population and total foreign population have a negative relationship with 10 percent and 1 percent significant levels, respectively.

In column (3), we added migration, total poverty population, and family structure. The entropy index still holds a 1 percent significant level and density became more insignificant again. Only Hispanic showed a 10 percent significant level with a positive relationship. Foreign population becomes less significant from 1 to 10 percent. In column (4), we added education attainments and unemployment rate. Now, the entropy index shows a strong significant level with a positive connection to income per capita. Population density became strongly significant with a negative relationship. Hispanic also became 5 percent significant. With education variables, all educational attainments have a positive relationship with income per capita. High school with diploma indicates strong association with 1 percent significant level while college and graduate school indicate 5 and 10 significant levels, respectively. Unemployment rate showed a statistically insignificant level in this model. An interesting point in this model is that the higher educational attainments are associated with income per capita with higher coefficients, but become less significant.

The relationship between entropy index and total crime rate was positive only with ethnic group population variables on column (1) in table 2. However, after adding more variables, the coefficient of entropy index showed a negative relationship to total crime rate, but showed insignificant levels on column (2) and (3). The population density remained with a strong significant level on column (1) and (2), and then became less significant once more variables were taken into account in the model. For ethnic groups, only Asian showed a strong significant level with a positive relationship. Since there were smaller sample sizes in table 2, more independent variables indicated significant as we added more attainments. Both entropy index and population density indicate a negative relationship with total crime rate with 5 percent significant level in column (4). Female population, migration, family structure, poverty population, and all of the educational variables indicated insignificant results in this model. However, unlike the previous model, unemployment rate, now has strong impacts in the model. Unemployment rate showed a positive relationship with 1 percent significant level. This suggests that because unemployment has more impact than other variables, other independent variables indicate an insignificant level even though it would be considered important factors in relationship with total crime rate.

Table 2. Fixed Effects Estimates of the Income Per Capita

IPC	**(1)**	**(2)**	**(3)**	**(4)**
Entropy	-.185*** (.045)	-.138*** (.52)	.218*** (.044)	.295*** (.046)
l_density	.002 (.017)	.032* (.018)	-.018 (.015)	-.048*** (.018)
Whitep	.-372*** (.127)	-.593*** (.192)	-.109 (.088)	-.017 (.056)
Hispanicp	-.272* (.164)	-.141 (.203)	.169* (.097)	.226** (.1002)
Blackp	.074 (.182)	-.183 (.214)	-.024 (.148)	-.076 (.126)
Asianp	-.212 (.247)	-.014 (.119)	-.039 (.105)	-.092 (.109)
Othersp	3.99*** (1.35)	3.86 (1.32)***	1.376* (.766)	1.620** (.789)
Totalfemalep		-.819* (.477)	-.396 (.264)	-.137 (.264)
Totalforeignp		-.828*** (.202)	-.251* (.128)	-.223 (.137)
Diffcountyp			-.051 (.059)	-.0101 (.056)
Totalpovp			-.055* (.0327)	-.028 (.027)
Familyp			-1.263*** (.191)	-.741*** (.139)
Highdipp				.269*** (.077)
Collegep				.388** (.175)
Grad_profp				.468* (.239)
Unemprate				-.001 (.23)
Observations	587	584	584	579
Groups	296	295	295	293

Table 3. Fixed Effects Estimates of the Total Crime Rate

TCR	(1)	(2)	(3)	(4)
Entropy	.0008 (.006)	-.00001 (.007)	.-014 (.012)	-.017** (.008)
l_density	-.012*** (.003)	-.015*** (.004)	-.009* (.004)	-.009** (.003)
Whitep	.001 (.005)	.015 (.011)	.006 (.0102)	.004 (.008)
Hispanicp	-.017 (.014)	-.021 (.017)	-.029* (.017)	-.021 (.015)
Blackp	-.028 (.027)	-.005 (.031)	.009 (.025)	.008 (.027)
Asianp	.016* (.008)	.007 (.009)	.007 (.008)	.022*** (.008)
Othersp	-.034 (.091)	-.029 (.088)	.012 (.084)	-.024 (.084)
Totalfemalep		.033 (.078)	.063 (.096)	.069 (.082)
Totalforeignp		.048** (.023)	.035* (.021)	.042* (.023)
Diffcountyp			.027 (.018)	.015 (.017)
Totalpovp			.052** (.021)	.024 (.021)
Familyp			.036 (.028)	-.011 (.029)
Highdipp				-.003 (.016)
Collegep				-.023 (.033)
Grad_profp				-.029 (.045)
Unemprate				.0005*** (.0002)
Observations	191	189	189	187
Groups	96	95	95	94

4.2 Implication

One possible implication for the result of total crime rate model is that total crime rate model only has 187 in the sample size. If we consider the number of predictors in this model, it is a relatively small sample size. Insufficient sample size would be one of the reasons for these outcomes. Overall, most coefficients from table 1 and 2 indicate the same directions and plausibility, so structure appears valid by the robustness check. Regarding to the significant level of group variables, such as racial and educational variable, each variables would likely to influence each other and result less significant level on each variables so we conducted the group testing to see the overall effects of group variables. From table A2 and A3 in appendix, ethnic group variables indicate the 10 percent significant level on both income per capita and total crime rate model. Educational variables showed 1 percent significant level on income per capita and insignificant level on total crime rate. Even though some of ethnic variables indicated the insignificant level individually, we can tell each ethnic variables are still valid with 10 percent level from the result of group testing.

4.3 Interpretation

Our results indicate that ethnic diversity has a strong connection to the income per capita positively with a consideration of education. Population density also is associated with income per capita negatively. This implies that people share more information relating to jobs in society with high population density. This leads to more people available in the labor market, which generates more competition between labor forces. Since there are many labor forces available, a company does not need to pay higher wages to find employees. Another finding in this model indicates that the higher educational attainments are likely to increase income per capita, which make sense in the real world. Higher education levels indicate strong magnitude on income per capita, but become less significant. This result suggests that educational attainments have a strong impact in income per capita compared to other covariates in this model.

The second models indicate that both ethnic diversity and income per capita have a strongly negative connection to total crime rate. This suggests that higher levels of ethnic diversity are likely to lower the crime rate. One possible explanation for the relationship between population density and total crime rate is that a higher population density would generate social cohesion or cooperation in society. More communication or interchange among residents leads society to experience positive results, rather than conflicts. An interesting point in table 1 is that the increase in foreign population actually has a positive relationship with crime rate. This suggests that new cultural facts might cause conflicts between people. Language barriers or different cultural interpretation would explain this phenomenon. Female and family structure showed insignificant levels in this model, but unemployment showed a strong impact on total crime, suggesting that crime rate is likely to increase during economic recession.

5. Conclusion

Throughout the history of the United States, many social events and recent political policies place ethnic diversity in the center of debate regarding whether diversity is good or bad in society. Since an increase in the number of immigration highly contributes to demographical change in society, it affects economic factors such as labor forces and market trends. Ethnic diversity is not only about social aspects, but also economic issues. In our paper, we used the multi-group entropy index to measure the heterogeneous residential patterns in metropolitan areas. We used the panel data with fixed effect, specifically within estimation, to find out the causal effect between ethnic diversity and socioeconomic outcomes; income per capita and total crime rate. Since the unobserved individual effect, metropolitan effect, is eliminated in the model, this loosens concerns about endogeneity problems in the model.

Our results indicate that ethnic diversity is associated with income per capita positively and total crime rate negatively. Higher population density indicates a negative relationship to crime rate and income per capita. Unemployment rate take more account with total crime rate, on the other hand, educational attainments have more impact on income per capita. Since our results indicate the strongly positive relationships between ethnic diversity and income per capita, this carefully implies the positive relationship between ethnic diversity and productivity. For total crime rate model, because of limitation for collecting data of total crime rates within metropolitan area , our results indicate insignificant relationship between total crime rate and educational variables even though we believe educational variables have a strong impact on total crime rate. Based on this study, we can improve our model to find out the relationship between productivity and ethnic diversity in the future.

Acknowledgements

I would like to greatly thank to Dr. Han and Dr. Trejo for their dedicated advices and supports in overcoming numerous obstacles I have been facing through research. I am also grateful to consultants of department of Statistical and Data Science to get results of better quality in this paper.

References

Alesina, A. and E. La Ferrara, 2005; Ethnic Diversity and Economic Performance, Journal of Economic Literature Vol. XLIII, pp. 762–800

Alesina, A. and E. La Ferrara, 2000: "Participation in Heterogeneous Communities", Quarterly Journal of Economics, Vol. 115(3), pp. 847-904

Alesina A., R. Baqir and W. Easterly, 1999: "Public Goods and Ethnic Divisions", Quarterly Journal of Economics, Vol. 114(4), pp. 1243-1284

"Census Trend Chart." Census Scope . Social Science Data Analysis Network (SSDAN)., n.d. Web. 04 May 2017.

Ching-Chi Hsieh, 2016: "Poverty, Income Inequality, and Violent Crime: A Meta-Analysis of Recent Aggregate Data Studies", Criminal Justice Review, Vol 18, Issue 2, pp. 182 – 202

Daron Acemoglu, J.-S. Pischke, 2001: "Changes in the wage structure, family income, and children's education", European Economic Review 45: 890-904

Fearon, J. D. and D. Laitin, 2000: "Violence and the Social Construction of Ethnic Identity", International Organization, Vol. 54(4), pp. 845-877

Graif, C., and R. J. Sampson, 2009: "Spatial Heterogeneity in the Effects of Immigration and Diversity on Neighborhood Homicide Rates." Homicide Studies 13.3: 242-60

Iceland, John, Daniel H. Weinberg, and Erika Steinmetz, 2002: Racial and ethnic residential segregation in the United States, 1980-2000. Washington D.C.: U.S. Census Bureau.

Iceland John, 2004: "The Multi group Entropy Index (Also Known as Theil's H or the Information Theory Index)." Unpublished manuscript, University of Maryland

Julian R. Betts, Robert W. Fairlie, 2003: “Does immigration induce ‘native flight’ from public schools into private schools?” Journal of Public Economics 87: 987–1012

Keith Harries, 2006: “Property Crimes and Violence in United States: An Analysis of the influence of Population density.” University of Maryland Baltimore County, USA, International Journal of Criminal Justice Sciences Vol 1 Issue.2 July

Kelvin M. Pollard, William P. O’Hare, 1999: “America’s Racial and Ethnic Minorities.” Population Reference Bureau, Vol. 54, No. 3

Kneebone, Elizabeth and Steven Raphael, 2012: City and Suburban Crime Trends in Metropolitan America, with Elizabeth Kneebone, The Brooking Institution, Washington D.C.

Lee Barrett A., Iceland John, Sharp Gregory, 2012: “Racial and Ethnic Diversity Goes Local: Charting Change in American Communities Over Three Decades”, Department of Sociology and Population Research Institute

Luke J. Larsen, 2003, “The Foreign-Born Population in the United States: 2003”, U.S. Census Bureau

Luttmer, E., 2001: "Group Loyalty and the Taste for Redistribution", Journal of Political Economy, Vol. 109(3), pp. 500-528

Lyman, Rick. "Eastern Bloc's Resistance to Refugees Highlights Europe's Cultural and Political Divisions." The New York Times. N.p., 12 Sept. 2015. Web. 04 May 2017

Miles, Kathleen. "Ethnic Diversity Increases Home Value And Lowers Crime In Southern California, Study Says." The Huffington Post, 07 June 2012. Web. 04 May 2017

Saxenaa Ankita, 2014: "Workforce Diversity: A Key to Improve Productivity", Procedia Economics and Finance 11, pp.76 – 85

"Seasonally Adjusted Metropolitan Area Estimates." U.S. Bureau of Labor Statistics. U.S. Bureau of Labor Statistics, n.d. Web. 04 May 2017

Vivian Hunt, Dennis Layton, and Sara Prince. "Why diversity matters." McKinsey & Company. N.p., Jan. 2015. Web

Wolbers, M. H. J., 2000: "The Effects of Level of Education on Mobility between Employment and Unemployment in the Netherlands." European Sociological Review 16.2: 185-200

Appendix

Table A1. Descriptive Statistics

	N	Mean	Std. Dev.	Min	Max
IPC	588	0.0943	0.0739	0.0001	0.4811
TCR	191	0.0074	0.0053	0.0003	0.0261
entropy	975	0.2131	0.1170	0.0210	0.5734
l_density	936	5.3979	1.0747	2.4394	12.9738
whitep	936	0.7868	0.1605	0.0492	0.9910
hispanicp	936	0.0781	0.1315	0.0025	0.9428
blackp	936	0.1044	0.1089	0.0003	0.8206
asianp	936	0.0184	0.0381	0.0009	0.5980
othersp	934	0.0089	0.0132	0.0008	0.1125
totalfemalp	622	0.5108	0.0111	0.4019	0.5330
totalforeignp	624	0.0613	0.0674	0.0039	0.5094
diffcountyp	626	0.2016	0.0675	0.0702	0.5054
totalpovp	626	0.1314	0.0568	0.0424	0.9095
familyp	622	0.6928	0.0460	0.5275	0.8561
highdipp	630	0.5088	0.0547	0.3146	0.6407
collegep	630	0.2034	0.0476	0.0877	0.3689
grad_profp	630	0.0776	0.0334	0.0262	0.2145
unemprate	603	0.0525	0.0236	0.0186	0.2238

Table A2. Group Testing on IPC model

Dependent: IPC	
(1)	whitep = 0
(2)	hispanicp = 0
(3)	blackp = 0
(4)	asianp = 0
(5)	othersp = 0
	F(5, 292) = 1.99
	Prob > F = 0.0796
(1)	highdipp = 0
(2)	collegep = 0
(3)	grad_profp = 0
	F(3, 292) = 5.39
	Prob > F = 0.0013

Table A3. Group Testing on TCR model

Dependent: TCR	
(1)	whitep = 0
(2)	hispanicp = 0
(3)	blackp = 0
(4)	asianp = 0
(5)	othersp = 0
	F(5, 93) = 2.15
	Prob > F = 0.0661
(1)	highdipp = 0
(2)	collegep = 0
(3)	grad_profp = 0
	F(3, 93) = 0.48
	Prob > F = 0.6991