PRIVATE ENTERPRISE RESEARCH CENTER
TEXAS A&M UNIVERSITY

PERC Study
March 2016

No. 1601

Statistics of Income County-to-County Migration Data: An Overview

Adam Narkiewicz

Private Enterprise Research Center
Texas A&M University

March 2016
No. 1601

Content

Statistics of Income County-to-County Migration Data: An Overview

Adam Narkiewicz

Statistics of Income as a data source

Statistics of Income (SOI) is a federal statistical organization with an annual budget of about $40 million, operating under the Internal Revenue Service (IRS). It fulfils the goal of processing and publishing statistics pertaining to operation of the IRS as mandated by the Revenue Act of 1916.

One of the products developed by SOI are U.S. Population Migration Data. These data are based on incomes, exemptions, and taxpayers' addresses reported on tax returns. SOI matches tax returns from two consecutive years using the taxpayer's identity and establishes whether the taxpayer resided in the same county in both years or moved to a different county. As a result, the data set contains comprehensive information on movement of the population across the country as well as income characteristics of those who did and did not move.

The data are publicly available for each year pair from 1990/1991 to present with an approximately two-year delay. A year pair indicates the two consecutive years in which the matched tax returns were filed. SOI uses tax returns from a given year pair to identify migrants as well as non-migrants. Migrants are then disaggregated by their county of origins and by their destination county, and non-migrants are disaggregated by their county of residence. As a result, the taxpayer population is partitioned into groups identified by the county of origin and the destination county (which can be the same in case of non-migrants). Each of these groups is characterized by three numbers: (1) the number of tax returns, (2) the total number of exemptions, and (3) total income. Both income and the number of exemptions are taken from tax returns in the second year of the given year pair.

For privacy reasons, a group is identified and its characteristics are made publicly available only if the number of returns in that groups is 10 or higher. The data also contain exact aggregate county-level characteristics of all immigrants and emigrants. The number of returns and exemptions in these aggregate data can exceed the corresponding sums for the reported specific county-to-county migration groups, as some of these groups may be censored.

The data from the year pairs '90/'91 and '91/'92 do not have all three aforementioned variables, and the data from '90/'91 through '94/'95 are in a somewhat different format than the subsequent data. Thus, the data set used in this analysis spans only the year pairs '95/'96 through '13/'14. In addition, SOI introduced methodological changes in the data collection process beginning with the year pair '11/'12. Prior to 2011 the Census Bureau developed the migration files using the IRS data, but beginning with the 2011 data, the SOI at the IRS has produced the

files. The methodological improvements include matching on additional taxpayer identification numbers besides only the primary taxpayer. Also the data collection period has been extended to include an entire year rather than ending in September. The first change increased matches by 5% and the second increased matched income, particularly among more complicated high income returns. With these improvements, the time series between 1995 and 2010 are not directly comparable to the time series beginning in 2011.[1]

Notes on methodology

The aim of this document is to present some basic information and general trends in the data used to create the PERC county-to-county migration overview map. As such, the analysis is limited only to variables used on the map: namely, non-migrant population, immigration and emigration rates, exemptions per return, and average incomes. The document contains possible explanations and interpretations of certain features exhibited by the data. These should be treated as suggestions and potential starting points for further research rather than definitive statements, as they are based on simple correlations and are not cross-verified with other sources or established economic theories.

I analyze counties using four basic dimensions: (1) population, (2) income per capita, (3) population growth, and (4) income per capita growth. For each dimension I investigate how the situation evolved over time and how a given dimension relates to other variables in the dataset. I hope that this data exploration can shed some light on the trends occurring in the US economy and can help understand how the economic situation of particular counties fits in the bigger picture.

I use the following proxy variables: (1) population of a county is approximated by the total non-migrant number of exemptions and (2) county income per capita is defined as the total non-migrant income divided by the total non-migrant number of exemptions. Dynamic variables are based on two consecutive year pairs in the following way: (3) population growth is approximated by the percentage increase in the non-migrant population, and (4) income per capita dynamics is approximated by the percentage increase in the non-migrant income per capita as defined in (2).

Additionally, I consider the following variables: (a) immigration and emigration percentage rates as well as their difference: net immigration rate, (b) immigrant and emigrant income per capita, (c) average number of exemptions per tax return for non-migrants, immigrants, and emigrants, and (d) ratios of average immigrant and emigrant incomes to average non-migrant income. Note that incomes of migrants are not necessarily comparable to the incomes of non-migrants as they may be affected by short spells of unemployment associated with changing jobs upon moving.

[1] See Kevin Pierce, SOI Migration Data: A New Approach, *Statistics of Income Bulletin*, Summer 2015 (https://www.irs.gov/pub/irs-soi/soi-a-inmig-id1509.pdf). Also see the list of sources at the end of the document for additional links to the SOI data and documentation.

For the purpose of this analysis I construct three types of graphs. First, for each given year I group the counties into quintiles according to one of the four basic dimensions (say population, as shown in Figure 1 Panel B). Then, for a given variable of interest (say non-migrant income), I calculate median in each quantile. Finally, I construct a graph that represents how five medians – one for each quintile – evolve over time. Each median is consistently marked by the same color and the quintile identifier across figures. Q1 denotes the quintile with the 20% counties at the bottom of the distribution. Q2 denotes the quintile with the next 20%, and so on. In each year the counties are split into quintiles using this year's data and therefore the subgroups in which medians are calculated vary over time. These graphs are shown in Figures 1-10. The naming convention uses the first year of the year pair to denote this year pair on the graph. For growth variables the situation is more complicated – they need two year pairs to be calculated. For them, I use the year associated with the second year pair.

The second and third type of graphs are heat maps that are easy recognizable by their square shape and blue color. They are shown on every figure except for Figure 4. These heat maps represent density of counties placed on a Q-Q plot. That is, for a given heat map, every county is assigned two characteristics (say non-migrant population in 2013 and non-migrant average income in 2013 for the one in Figure 1 Panel B). Then, counties are sorted according to each of these two characteristics in order to determine what quantile each county is. In other words, each county receives two numbers between 0 and 1, corresponding to its position on the sorted lists. These two numbers are the coordinates used to place a county on the Q-Q plot. Finally, based on local density of counties on the Q-Q plot, a heat map is generated with a dark blue signifying high density and white signifying low density.

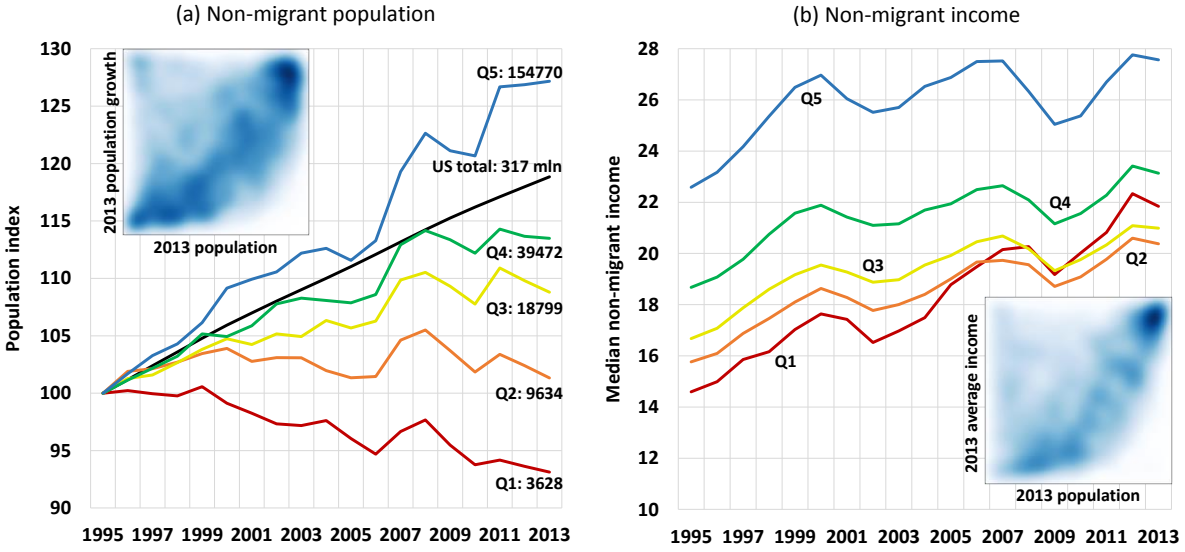
The two types of heat maps are created from two distinct data sets. One is based solely on the data from the year 2013. The other is based on the data from the period between 1995 and 2010 (due to the change in data collection methodology, the data from the period between 1995 and 2010 may not be comparable to the data staring in 2011). The average migration rates are calculated by summing the migrant numbers from each year and summing the non-migrant numbers from each year. The migration ratio averages are thus population weighted. Similarly, averages of incomes and exemptions per return are population weighted.

For each of the heat maps, the corresponding correlation is reported in the text of the document. Sometimes, correlation is also reported for pairs or variables not depicted as heat maps. The correlation is reported in form of a 95% confidence interval. This means, that the correlation is between the two given numbers with a 95% probability. This allows to see both the magnitude of correlation and how precisely was it estimated.

Non-migrant population

The population variability among the 3143 US counties is high. The most populous is Los Angeles County with over 10 million inhabitants and the least populous are Loving County in Texas and Kalawao County in Hawaii, both with a population below 100.[2] The Statistics of Income migration data underestimate true population figures, as it reflects only households for whom tax returns from two consecutive years could be matched. For example, the total number of exemptions claimed in Los Angeles County in each of the years from 2011 to 2013 does not exceed 8 million which indicates that our proxy variable underestimates population of this county by around 20%.

Figure 1. Non-migrant population breakdown: population and income



Note: income in thousands of 2009 US dollars. The US total population time series is from www.bea.gov; the remaining time series are from www.irs.gov.

Let us start by stratifying counties by their non-migrant number of exemptions. For each year, I divide counties into five equally sized groups – population quintiles. Figure 1 depicts how non-migrant population and non-migrant income per capita were changing over time within each quintile. Panel A shows population relative to the year 1995. The least populous counties are marked as Q1 and the most populous counties are market as Q5. The reported numbers are median values within each quintile in the year 2013. That is, in 2013, the 10% least populous counties had a non-migrant population of 3628 or lower, and the top 10% counties had a population of 154,770 or higher. Notice the divergence in terms of population – the 20% least populous counties lose inhabitants and more populous counties grow faster.

The black line on Figure 1 Panel A depicts the relative increase in the total US population. Only the top 20% of counties grew relatively to the entire country and the bottom 80% of counties lost their share in population. Notably, the reported median population is much more

[2] These are the 2014 population estimates from census.gov.

volatile than the total national population. This may be attributed to factors like the business cycle which affects the number of people who file tax returns.

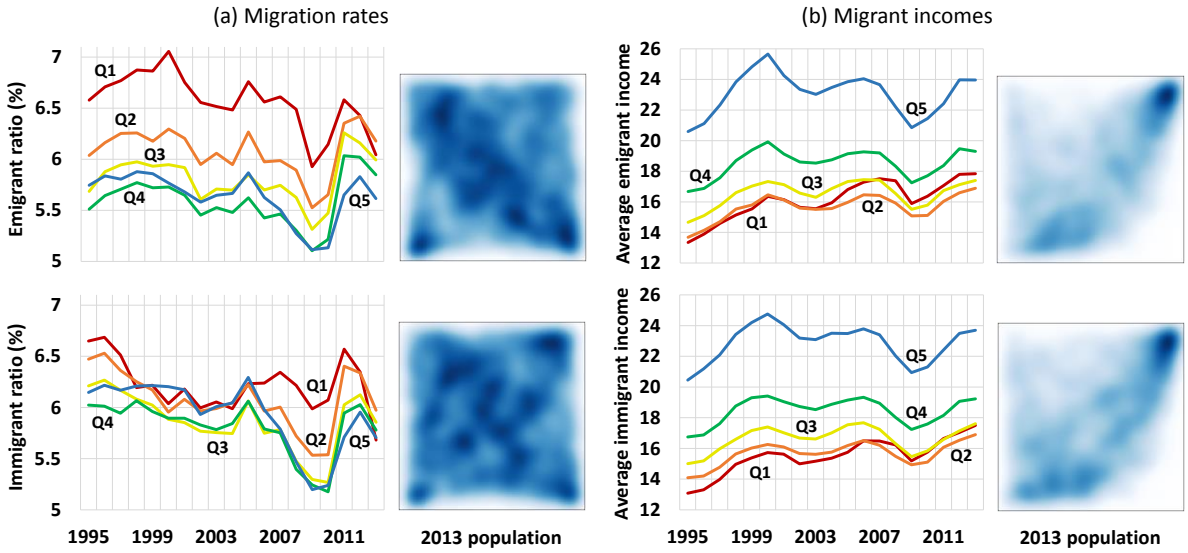
The relationship between population and population growth is confirmed by the heat map in the top left corner of Figure 1 Panel A. The heat map was constructed by assigning to each county two values: its quantile in terms of non-migrant population in 2013 and its quantile in terms of non-migrant population growth between 2012 and 2013. The quantiles were then used as coordinates of counties in the XY plane and the areas with higher density were marked with darker color (see previous section for more details on how graphs are constructed). The correlation between population and population growth in 2013 is positive and its 95% confidence interval is (0.2794, 0.3426). The heat map indicates that there are very few large counties that grew very slowly. Conversely, there are some small counties which grew very quickly.

Figure 1 Panel B depicts changes in non-migrant income in each population-based quintile. Income is higher in more populated areas. The smallest counties constitute an exception. Their economies started to grow faster than economies of the other counties around the year 2004. This may be attributed to economic developments disproportionately affecting counties with small population, like the onset and expansion of the shale oil industry. Another interesting feature of Figure 1 Panel B is cyclicality of incomes. The incomes rise as the economy expands and then they drop when the economy contracts, as for example between years 2007 and 2009. Finally, it is important to keep in mind that the source of data experienced a change in methodology, which may be the reason for the time series jumping up in the year 2011 and staying high from then on, especially for more populous and higher-income counties.

The heat map in the bottom-right of Figure 1 Panel B illustrates the positive correlation between county population and average income in 2013. The 95% confidence interval for the correlation is (0.3165, 0.3780). Although few highly populous counties have low income, many sparsely populated counties have incomes at both ends of the distribution. In general, throughout the entire analysis, high-population counties tend to be more homogeneous and small-population counties tend to be on extremes in nearly every aspect.

The difference between Figure 1 Panel B and Figure 2 Panel B indicates that migrant incomes are seldom higher than non-migrant incomes. Nevertheless, incomes exhibit fairly similar general trends, including the tendency for high population counties to have high-income migrants and for small population counties to be more diverse and more likely to be on both extremes, as shown by the corresponding heat maps on the right side of Figure 2 Panel B. The 95% confidence interval for correlation between 2013 population and 2013 immigrant income is (0.3395, 0.4002) and for emigrant income it is (0.3525, 0.4125).

Figure 2. Non-migrant population breakdown: migration rates and income



Note: income in thousands of 2009 US dollars. All data come from www.irs.gov.

An interesting difference between immigration patterns and emigration patterns can be seen on Figure 2 Panel A. Emigration rates tend to be higher and more dispersed than immigration rates. Both immigration and emigration rates are based on the same group of people moving, therefore having one consistently higher than the other may seem counterintuitive. However, there is a simple explanation for this phenomenon: people from smaller counties with higher emigration rates tend to migrate to larger counties with smaller emigration rates. This explains not only differences in the dispersion but also the general difference in magnitude – higher fractions of people emigrating from small counties get diluted among the smaller fractions (but bigger numbers) of people emigrating from and to large counties, thus making immigration ratios appear smaller than emigration ratios. Also, note that both immigration and emigration appear to be pro-cyclical and that in the year 2005 – the year of hurricane Katrina – there is a visible spike in migration rates for all county groups.

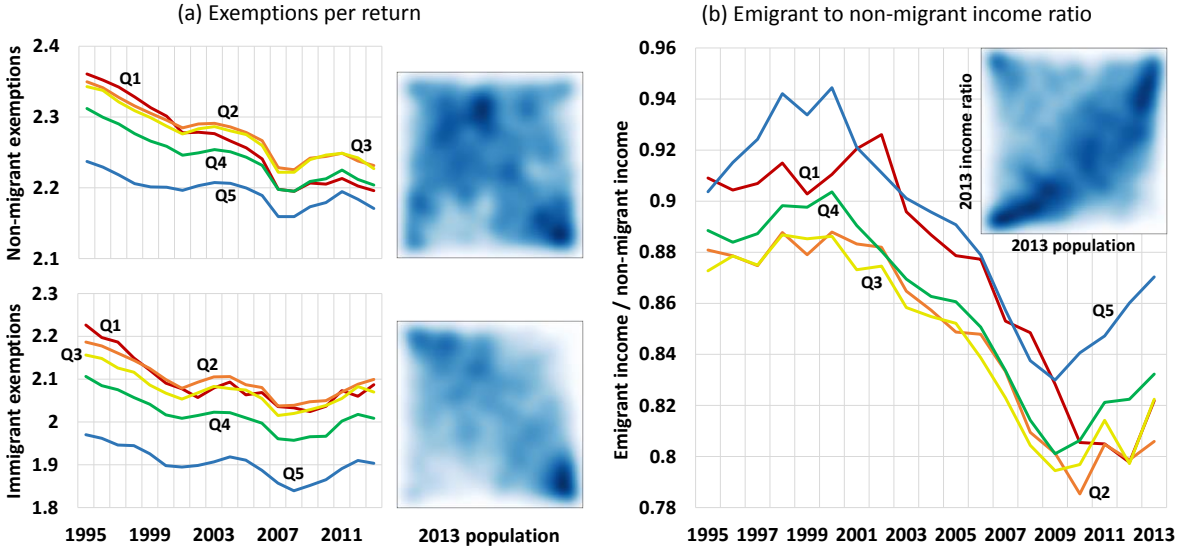
Quintile graphs in Figure 2 Panel A suggest a negative historical correlation between emigration ratios and population (series marked Q1 tend to be above other series). However, the heat maps for 2013 do not confirm this trend. The 95% confidence interval of (-0.1104, -0.0406) for the correlation between population and emigration ratio indicates marginal statistical significance, and the correlation between population and immigration rates is not significant at all at (-0.0514, 0.0190). As a result, net immigration ratio is only slightly positively correlated with population (no heat map depicted) at (0.0435, 0.1134). This is consistent with the idea that bigger counties grow faster.

Migrants have lower incomes per capita on average than non-migrants and they also have smaller households. Figure 3 Panel A shows how number of exemptions per tax return, a proxy for family size, evolves over time for counties segregated by size. For both immigrants and non-

migrants, bigger counties tend to be historically associated with smaller families. Additionally, non-migrant families appear to be bigger than migrant families (the picture of emigrant exemptions per return was omitted due to its similarity with the picture of immigrant exemptions per return).

In 2013 the relationship between population and number of exemptions per return nearly vanishes for the non-migrant population. As the heat maps on Panel A indicate, the correlation is much stronger between population and immigrant exemptions per return at (-0.3230, -0.2586) than between population and non-migrant exemptions per return at (-0.1082,-0.0387). This is caused mostly by a group of very small counties with a very low number of exemptions per return. Again, this might have been induced by recent changes in the economies of these small counties, like migration of oil industry workers. This however cannot be the full explanation, as small counties started to have smaller households around the year 2000, long before onset of the shale oil boom.

Figure 3. Non-migrant population breakdown: exemptions and income ratios



Note: all data come from www.irs.gov.

Figure 3 Panel A also shows a general downward trend in the number of exemptions per return which may be a consequence of declining fertility rates, marriage rates, and, as a result, household sizes. An interesting phenomenon is the difference between non-migrants and immigrants over the period between 2011 and 2013. The former experienced a noticeable drop in the number of exemptions per return while the latter experienced an increase.

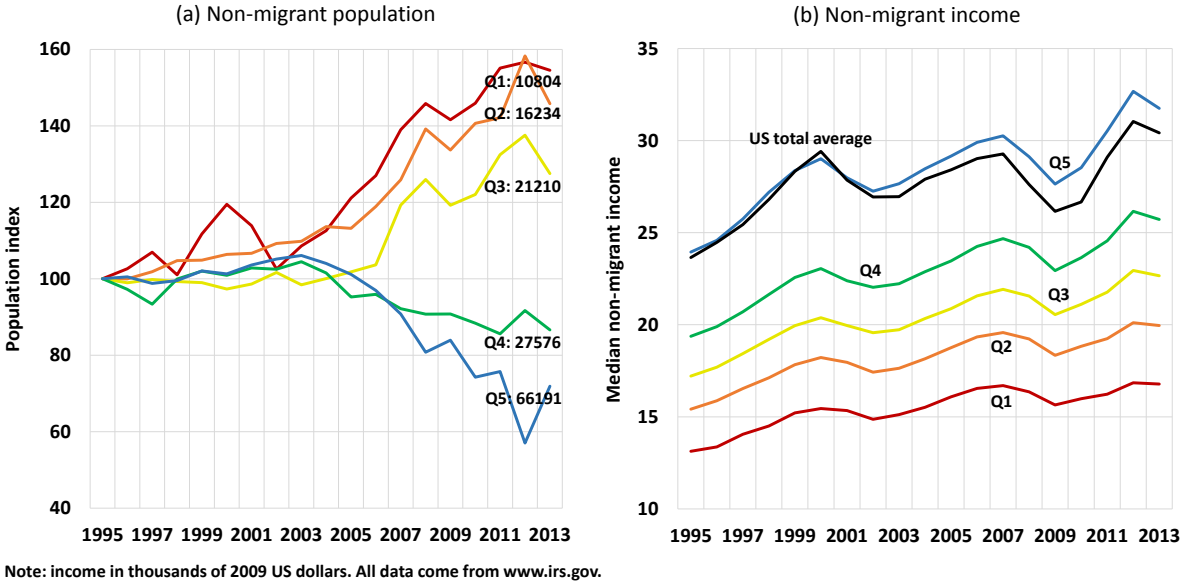
Similarly, Figure 3 Panel B shows a clear downward trend in terms of migrant incomes relative to non-migrant incomes (only the series for emigrants are reported due to their similarity with the series for immigrants). Individuals who move are relatively poorer than people who stay and their incomes are further declining. Since non-migrants have higher average numbers of exemptions per returns, migrant households have even smaller incomes than could be inferred

from the Panel B alone. And, as indicated by the heat map, higher income migrants come from (or go to) highly populated counties, while migrants associated with smaller counties represent both extremes. The 95% confidence interval for the correlation is (0.1382, 0.2063).

Non-migrant income

Figure 1 Panel B indicates that more populous counties are richer and vice versa. This finding is mirrored by Figure 4 Panel A, which stratifies counties according to per capita income and then traces how population changed in such strata over time. Since 1995, the group of high income counties includes less and less populous counties and the group of low income counties starts to include bigger and bigger ones. It may be a sign of convergence in terms of income between poorer and smaller counties on one hand and richer and bigger counties on the other hand. Nevertheless, in 2013 the bottom 20% of counties in terms of income per capita had a median population of 10,804 while the top 20% of counties in terms of income per capita had a median population of 66,191. The correlation between population and income is thus strong, although it decreases over time. Additional research is needed to determine whether it is because of migration of rich families to the suburbs or for any other reason.

Figure 4. Non-migrant income breakdown: population and income



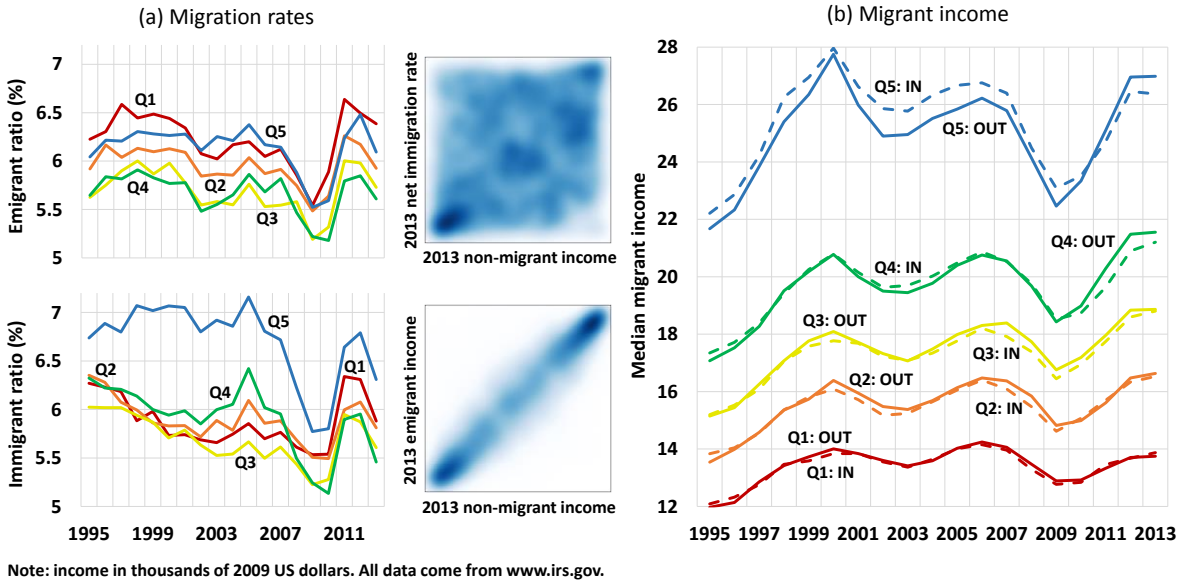
Note: income in thousands of 2009 US dollars. All data come from www.irs.gov.

Figure 4 Panel B displays the 1st, 3rd, 5th, 7th, and 9th decile of the county income per capita distribution (series marked Q1, Q2, Q3, Q4, and Q5 respectively) as well as US average income per capita. Notice the skewness of the distribution: national average is close to the 9th decile of the distribution. This is due to the fact that the top 20% of high income counties also include the most populous counties, most of which are above both the average size and average income. It is also clear that incomes are pro-cyclical, although it is not clear whether incomes in 2012

exceeded those in 2007. Other sources[3] indicate that they did not while SOI data suggests that they did. The increase in the SOI series may be attributed to the change in methodology, making them less credible than other data sources.

Although the immigration rate for the richest counties has been historically higher than the immigration rate of other counties (see Figure 5 Panel A), the correlation between immigration rate and non-migrant income is barely significant at (0.0052, 0.0755). Similarly, the correlation between emigration rate and non-migrant income is barely significant at (-0.0900, -0.0200) and the relationships between the two variables displays no recognizable patterns (heat maps are thus not presented). Nevertheless, non-migrant income in 2013 is correlated with net immigration, suggesting that, on average, people are more attracted to rich places. This is mostly reflected in high number of counties with low income and low net immigration as well as very low number of counties with high income but low net immigration, as depicted by the upper heat map on Figure 5. The correlation between these two variables is within the (0.1272, 0.1958) interval with 95% probability.

Figure 5. Non-migrant income breakdown: migration rates and income



Note: income in thousands of 2009 US dollars. All data come from www.irs.gov.

Figure 5 Panel A indicates that poorest counties (marked Q1) have higher emigration than immigration ratios, and that riches counties (marked Q5) have higher immigration ratios than emigration ratios. The graphs also show pro-cyclicality of migration which supports the idea that its fluctuations are mostly caused by changing job market conditions. Note also a visible spike in 2005, the year of Hurricane Katrina.

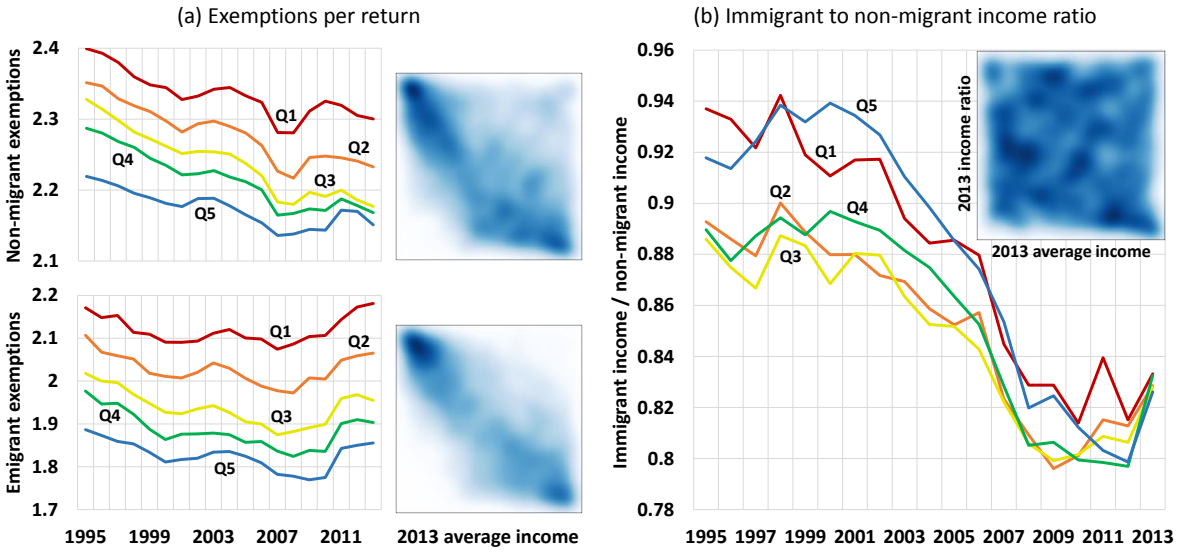
Some of the highest observed correlations in the data are between migrant incomes and non-migrant incomes. The bottom heat map in Figure 5 shows a strong relationship between emigrant

[3] See bea.gov for example.

incomes and non-migrant incomes in 2013. The 95% confidence interval for this correlation is (0.8424, 0.8617) and the 95% confidence interval for the correlation between immigrant incomes and non-migrant incomes (not depicted) is (0.7711, 0.7981). Immigrant incomes and emigrant incomes tend to follow very similar trajectories over time, as illustrated in Panel B of Figure 5. Series marked as "IN" depict median immigrant incomes and series marked "OUT" depict median emigrant incomes. The difference between income of migrants for the 20% poorest counties and the 20% richest counties is almost twofold. However, it is important to keep in mind that income per capita is determined not only by the household earning power or its human capital, but also by the household size. Two families with the same household income can be far apart in terms of per capita income if the number of children varies in each family.

Indeed, the average size of a family, measured by exemptions on a tax return, is strongly negatively correlated with average income per person. The 95% confidence interval for this correlation is (-0.4157, -0.3562) in case of non-migrants (top heat map on Figure 6 Panel A), (-0.5926, -05451) in case of emigrants (bottom heat map on Figure 6 Panel A), and (-0.5651, -0.5153) in case of immigrants (heat map omitted). Interestingly, for non-migrants, there are a number of counties which are at the top of both income per capita and family size distributions.

Figure 6. Non-migrant income breakdown: exemptions and income ratios



Note: all data come from www.irs.gov.

Moreover, non-migrant exemptions per return show a certain degree of pro-cyclicality, especially for low-income counties (Q1, Q2, and Q3 on the top graph in Figure 6 Panel A) around the year 2007. The reasons for the drop in the number of exemptions per return during the economic downturn is a puzzle. One possible explanation is that people with bigger families are losing jobs at a higher rates than people with smaller families.

Finally, the general downward trend in the size of the non-migrant family is not reflected in the group of migrants (graph for immigrants is not shown as it is very similar to the graph for emigrants), whose average exemptions per return in 2013 almost returned to the levels from 1995 after reaching a trough around 2007. This increase may be attributed to a drop in the number of single persons moving, and increase in the number of families moving, or the change in the data collection methodology which might have kept some moving families out of the dataset prior to 2011.

Figure 6 Panel B shows how immigrant income relative to non-migrant income evolved over time in each of the five income groups (analogous picture for the emigrant income ratio is not depicted due to its similarity). The reason for the generally decreasing trend are unknown. One hypothesis may be that the non-migrant families are getting smaller faster than migrant families resulting in the relative decrease in the per capita incomes of the latter. Another hypothesis is that migrant households indeed have relatively declining earning power for some external reason. For example, improvements in the local job markets might allow higher qualified individuals to more easily find jobs without moving.

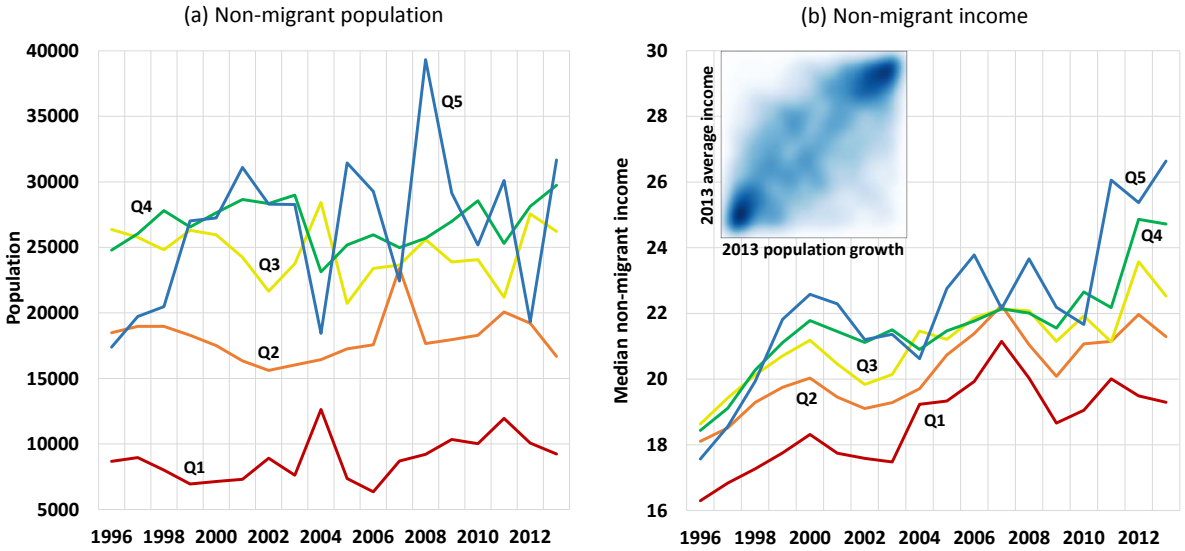
It is also interesting that both the highest income and the lowest income counties tend to have the highest income ratios but the ratios converge in 2013 on the same value of 0.83 both for immigrant income ratios and emigrant income ratio (not depicted). As a result the correlation 95% confidence intervals are respectively (-0.0844, -0.0142) and (-0.0482, 0.0221) and there is no clear pattern on nether heat map (as an example, the heat map for immigrant income ratio is depicted on Figure 6 Panel B).

Non-migrant population growth

Let us turn our attention to the potential correlates of population growth. In particular, let us address the following question: how do fast-growing counties differ from the slow-growing counties? As we have already seen on Figure 1 Panel A, population growth is correlated with population size. Counties diverge in terms of population, as bigger counties grow faster than smaller counties. Similar relationship can be also found in Figure 7 Panel A. The line representing quickly growing counties (Q5) is much above the line representing slowly growing counties (Q1), indicating that the latter are much smaller than the former. Moreover, faster growing counties tend to be richer than slow growing counties (see Panel B). The correlation between these two variables in 2013 was highly significant, within the (0.4252, 0.4807) 95% confidence interval.[4]

[4] The evolution of migrant incomes and the heat maps for the correlation between migrant incomes and population growth are similar to those for non-migrants and were thus skipped.

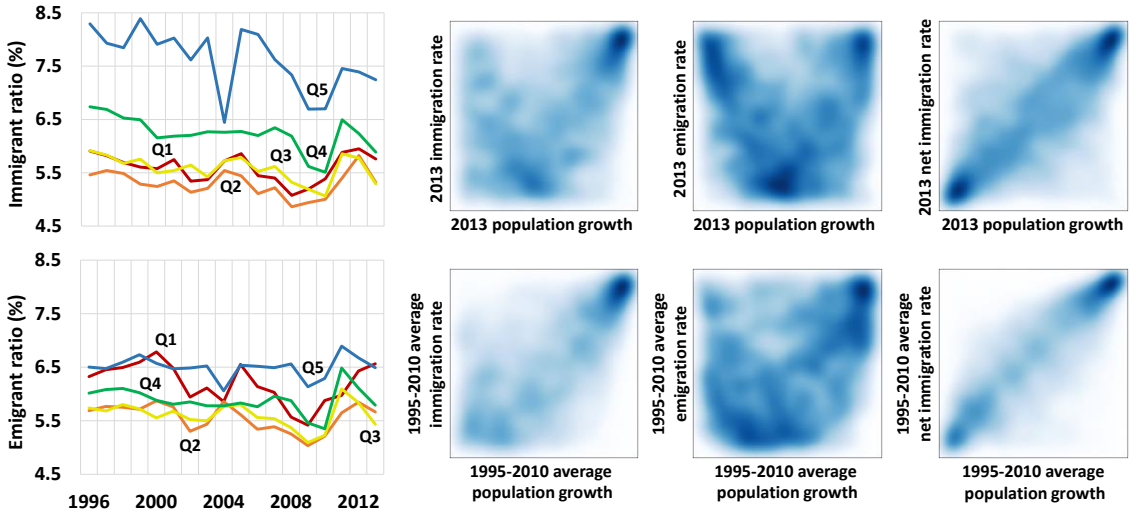
Figure 7. Non-migrant population growth breakdown: population and income



Note: all data come from www.irs.gov.

An interesting question is whether population growth is driven by migration or factors resulting in higher internal population growth. Figure 8 makes it clear that migration plays an important role. Immigration rates tend to be higher than emigration rates in the group of the fastest growing counties (Q5) and tend to be smaller in the group of the slowest-growing counties (Q1). As a result, net immigration rates are highly correlated with population growth, with the 95% confidence interval of (0.3784, 0.4371) for the year 2013 and (0.6045, 0.6472) for the averages over the period from 1995 to 2010 (see the rightmost heat maps on Figure 8).

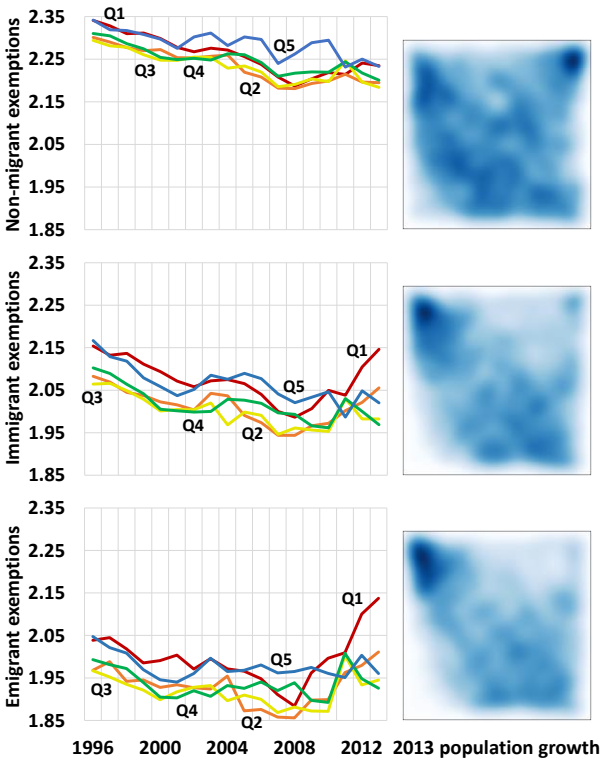
Figure 8. Non-migrant population growth breakdown: migration rates



Note: all data come from www.irs.gov.

Nevertheless, the relationships between the migration rates and population growth are somewhat more complicated. The V-shaped heat maps indicate that counties on both extremes in terms of population growth have higher immigration and emigration ratios than counties with median population growth. Overall, the correlation between immigration rate and population growth in year 2013 had the 95% confidence interval of (0.1947, 0.2615) and the correlation between emigration rates and population growth had the 95% confidence interval of (-0.0353, 0.0349). The lack of correlation in the latter case is caused by the V-shape of the distribution of counties (see the top-middle heat map on Figure 8). On the other hand, the V-shaped distribution of the counties in case of immigration is more asymmetric with a higher concentration of high-growth-high-immigration counties than low-growth-high-immigration counties.

Figure 9. Non-migrant population growth breakdown: exemptions



Note: all data come from www.irs.gov.

This pattern is consistent with what we have already learned about migration across counties. Emigration rates tend to be higher for both the highest and lowest income counties (series Q5 and Q1 on the top picture in Figure 5 Panel A) while the corresponding relationship is less pronounced when it comes to immigration. It is thus possible that the middle income counties are those with small migration rates and median population growth while richest and poorest counties are on the extremes in terms of population growth and have high migration rates both ways. This pattern seems to be somewhat preserved in the long term: both immigration and emigration rates are positively correlated with population growth over the period between 1995 and 2010 but the immigration rate is correlated stronger, at (0.4344, 0.4896) than the emigration rate at (0.1685, 0.2358).

The second potential source of population growth is fertility rate. Under some simplifying assumptions a higher fertility rate translates into larger family size which can be measured by the number of exemptions per return. Figure 9 shows how exemptions per return changed over time for counties grouped by their population growth. As noted before, non-migrant families are bigger than migrant families. Also, both for migrants and non-migrants, counties with moderate annual population growth (Q2, Q3, and Q4) have mostly smaller average family sizes than counties with very low (Q1) and very high (Q5) population growth. The correlation between non-migrant number of exemptions per return and population growth in 2013 is near zero with the 95% confidence interval of (-0.0546, 0.0153). The correlation would be negative and statistically

significant if not for a group of counties with high population growth and high number of non-migrant exemptions per return visible in the upper-right corner of the top heat map of the Figure 9. In addition, the correlation between average non-migrant exemptions per return and average non-migrant population growth between 1995 and 2010 is marginally significant with a (0.0328, 0.1027) 95% confidence interval (heat map not depicted). These data do not indicate that population growth is driven by fertility rates. It is likely that migration contributes more. This hypothesis is further supported by analysis of case studies like Travis County, Texas, where changes in population growth coincided with changes in the sign of net migration rate while the number of exemptions per return remained fairly constant (see our online map).

Unlike the case of non-migrants, the relationships between migrant exemptions per return and non-migrant population growth are more pronounced, and they are negative. For the year 2013, the correlation between immigrant exemptions per return and population growth was in the (-0.2395, -0.1722) 95% confidence interval and correlation between emigrant exemptions per return and population growth was in the (-0.3017, -0.2365) 95% confidence interval (corresponding heat maps are depicted on Figure 9). Apparently, in 2013, slowly growing counties generated and attracted migrants with the biggest families. This relationship is however not supported by the long run analysis – in the period between 1995 and 2010 corresponding correlations are very close to zero (not depicted). This is a counterintuitive phenomenon that deserves further study.

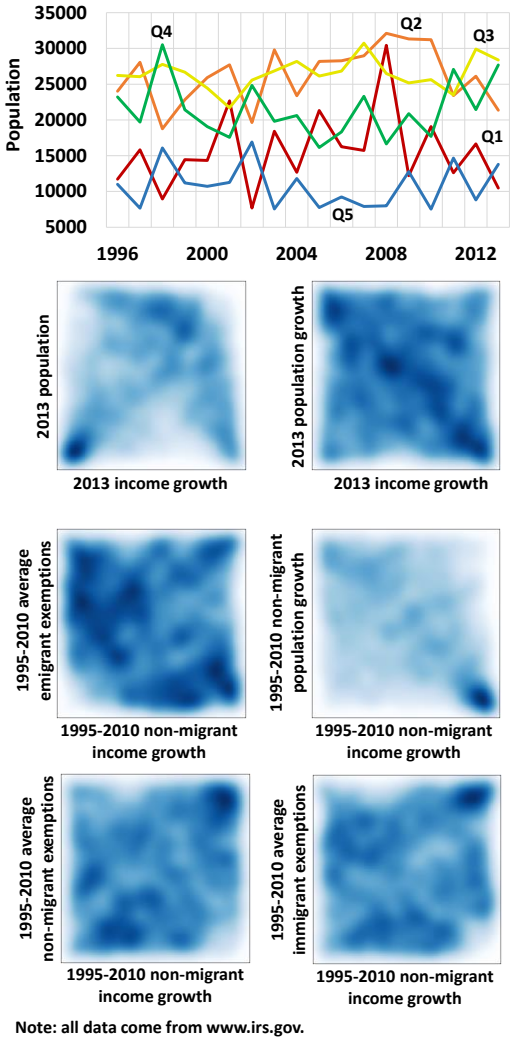
Another unexpected relationship is between migrant income ratios and population growth (not depicted). Between 1995 and 2010, average immigrant income ratio and population growth are positively correlated at (0.1600, 0.2276), while average emigrant income ratio and population growth are negatively correlated at (-0.2014, -0.1360). These findings indicate that counties that grow faster attract people who already have incomes relatively high in comparison to the people living in the place they are moving to, while people moving out of the fast-growing places are relatively poorer than those moving out of slow-growing places. However, these long-run findings are not confirmed by the 2013 data, when both immigrant and emigrant income ratios were correlated with population growth at around (0.04, 0.11). That is, in 2013, effects were generally smaller in magnitude and the relationship for emigrants reversed.

Non-migrant income growth

Counties that have extreme income growth rates are more likely to be small. Intuitively, it is easier to change the economy of a small county than of a large one, for example, by building a factory. This idea is confirmed in Figure 10, where counties with the smallest income growth (Q1) and the largest income growth (Q5) tend to have lower population than the counties in the middle of the distribution. This relationship is further supported by the A-shape on the top-left heat map in Figure 10. The correlation based on this heat map is marginally significant with the 95% confidence interval of (0.0169, 0.0867).

Groups of counties with rapidly growing incomes tend to include counties with slowly growing population. This negative correlation is more pronounced between 1995 and 2010 (the middle-right heat map of the Figure 10) than as a characteristic of the year 2013 (the top-right heat map). The correlation of the former places itself in the (-0.2900, -0.2245) 95% confidence interval while correlation of the latter is smaller, at (-0.1218, -0.0523). This may be a reflection of the general trend towards suburbanization.

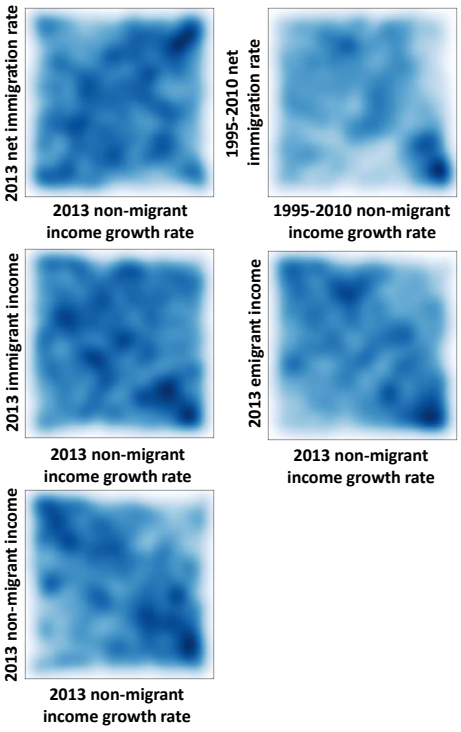
Figure 10. Non-migrant income growth: population and exemptions



Note: all data come from www.irs.gov.

Finally, the 1995 to 2010 average non-migrant income growth rate seems to be weakly positively correlated with the immigrant exemptions per return at (0.0379, 0.1077) and more strongly negatively correlated with the emigrant exemptions per return at (-0.1405, -0.0711). In other words, counties which grow faster attract slightly bigger families and lose slightly smaller ones. In addition, counties with bigger non-migrant families tend to grow rich slightly faster, with the 95% confidence interval for correlation of (0.0755, 0.1468) (see Figure 10 for the corresponding heat maps). In contrast, the data for 2013 (not depicted) tells a slightly different story, in which correlation between immigrant exemptions per return and non-migrant income growth is similar to the long run, but the correlation between emigrant exemptions per return and non-migrant income growth is of a similar magnitude as the long run but of the opposite sign. Moreover correlation between non-migrant exemptions per return and non-migrants income growth is statistically indistinguishable from zero.

Figure 11. Non-migrant income growth: migration and income



Note: all data come from www.irs.gov.

This inconsistency between the period from 1995 to 2010 and the year 2013 reveals itself also in the analysis of net immigration rates (see the two top heat maps of Figure 11). The correlation between net immigration rate in 2013 and non-migrant income growth is between 0.0331 and 0.1032 with probability 95%. Meanwhile, the 95% confidence interval for the correlation between average net immigration rate over the period between 1995 and 2010 and non-migrant income growth rate is (-0.1356, -0.0661). The long-run and short-run correlations are of the opposite sign, although neither of them is strong. This change in correlation, from the negative to the positive sign, may reflect recent reversal in migration or growth patterns across counties.

Despite an apparent divergence in terms of population, counties seem to be converging in terms of income, at least in 2013 (see the bottom heat map of Figure 11). The 95% confidence interval for the correlation between non-migrant average income in 2013 and non-migrant income growth rate is (-0.1744, -0.1059) which implies that lower income counties are growing faster. It also seems that quickly growing counties have relatively lower income emigrants (correlation of (-0.1841, -0.1154)) but this may come from the fact that more rapidly growing counties are lower income and high income individuals do not migrate from low income counties. A weaker, but still negative correlation at (-0.0926, -0.0225) between immigrant income and non-migrant income growth rates supports these findings (see middle row on the Figure 11).

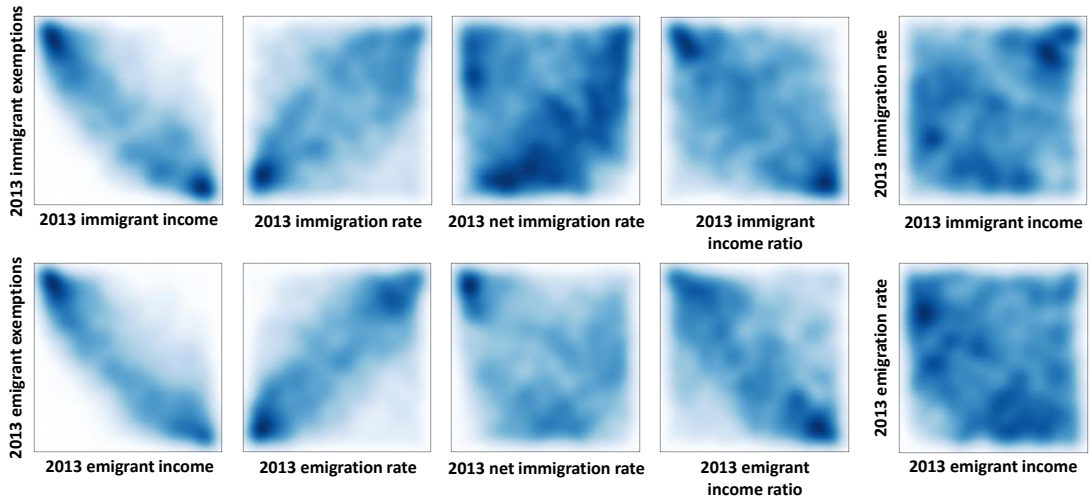
Other relations

Income per person in this analysis is driven not only by the earning power or human capital of individuals but also by the family size. As the top heat map in Figure 6 Panel A indicates, there is a negative correlation between non-migrant incomes per person and non-migrant exemptions per return. This correlation is even stronger in the population of immigrants at the (-0.6384, -0.5949) 95% confidence interval and in the population of emigrants at the (-0.6509, -0.6085). Corresponding heat maps – the leftmost on the Figure 12 – indicate that the relationship between exemptions per return and income exhibits the same pattern in groups of migrants as in groups of non-migrants.

The fact that migration is measured at the level of individuals rather than at the level of households is very important from the point of view of migration analysis. The migration decisions are likely made by a household as a unit and are driven by factors experienced by the

entire household. However, migration rates typically count individuals rather than households. Thus, a county with a high emigration rate may experience this rate not because people have unusually strong reasons to leave but because people in this county have unusually large families. This hypothesis is supported by positive correlation between 2013 immigrant exemptions per return and the 2013 immigration rate in the (0.2431, 0.3081) 95% confidence interval and the analogous correlation for emigrants in the (0.3765, 0.4352) 95% confidence interval. Corresponding heat maps are shown in the second column from the left on Figure 12.

Figure 12. Miscellaneous correlations



Note: all data come from www.irs.gov.

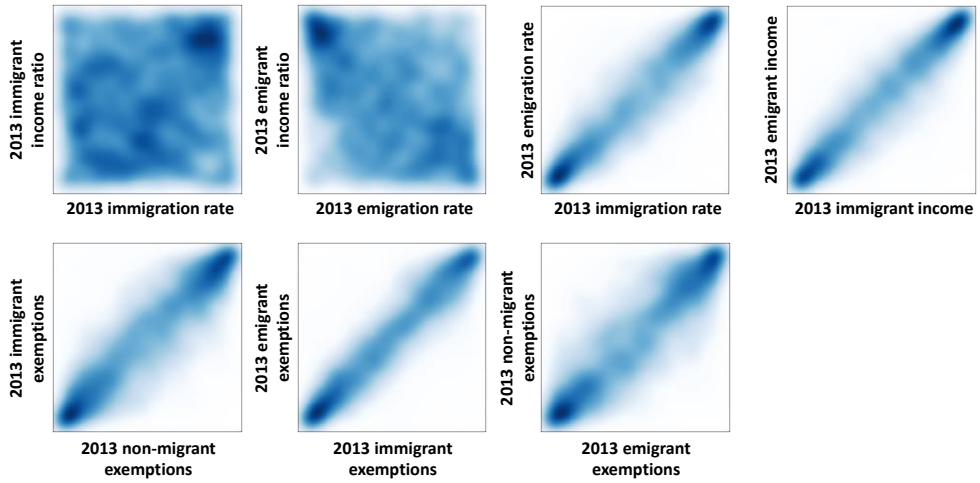
As a result, higher number of immigrant exemptions per return is associated with higher net immigration, and higher number of emigrant exemptions per return is associated with lower net immigration, although the relationship is more pronounced for emigration. The corresponding confidence intervals are (0.0167, 0.0869) and (-0.1858, -0.1170) and the corresponding heat maps can be found in the middle column of Figure 12.

Finally, migrants with bigger families have relatively lower income ratios. This can be seen on the heat maps in the second column from the right on Figure 12. In both cases the correlation is virtually the same: for 2013 emigrant exemptions per return and emigrant income ratio it is at (-0.3038, -0.2387) and for corresponding immigrant variables it is at (-0.3009, -0.2357). One could presume that since bigger families imply both smaller incomes and larger migration rates, the migrant incomes and migration rates should be negatively correlated. However, although they are marginally negatively correlated in the case of emigrants (at (-0.1557, -0.0865)), the correlation between immigrant incomes and immigration rates is marginally positive (at (0.0487, 0.1186)), as depicted in Figure 12. This implies that there are other more important factors at play.

Virtually identical correlation between emigrant and immigrant income ratios and their respective exemptions per return does not translate into similar correlations between income

ratios and corresponding immigration rates. As with the case of migrant incomes the figure for emigration has the expected positive sign – the correlation between emigration rates and emigrant income ratios in 2013 was in the (-0.2348, -0.1674) 95% confidence interval. However, the immigration rates and immigrant income ratios were marginally positively correlated with the 95% confidence interval of (0.0145, 0.0848). This suggests some unknown structural difference between emigration and immigration patterns (see two leftmost heat maps in the top row on the Figure 13).

Figure 13. Miscellaneous correlations continued



Note: all data come from www.irs.gov.

It is worth investigating correlations between certain immigrant, emigrant, and non-migrant characteristics. As many of these characteristics exhibit very high degree of correlation, it seem plausible that some counties are more naturally prone to certain phenomena than others. For example, the second from the right heat map in the top row of the Figure 13 depicts correlation between immigration rates and emigration rates in 2013. The 95% confidence interval for this correlation is (0.7886, 0.8138), very high. Thus, it seems that some counties have inherently smaller migration rates, possible due to bigger labor market offering more opportunities that do not require moving. This hypothesis is mildly supported by the data presented in Figure 2 Panel A.

Despite the fact that both immigration rate and emigration rate have the expected signs when it comes to their correlation with net immigration (that is for immigration rate it is positive and for emigration rate it is negative), they differ in magnitude. The correlation between immigration and net migration rate in 2013 is in the (0.4017, 0.4590) 95% confidence interval, while the correlation between emigration rate and net immigration rate in 2013 is in the (-0.1458, -0.0763) 95% confidence interval (none of them depicted). A potential explanation is that the relationship between net immigration rate and immigration rate is simple, that is low net immigration rate derives from low immigration rate and high net immigration rate derives from

high immigration rate, while for emigration it is more complex, for example counties with high net immigration can also have high emigration rates.

Another highly correlated pair of variables are average immigrant income and average emigrant income in 2013 (see the rightmost heat map in the top row if Figure 13). These two variables are correlated with the 95% confidence interval of (0.7878, 0.8131). This is consistent with an idea that richer counties attract richer immigrants and lose richer emigrants. However, the similar correlation done using emigrant income ratio and immigrant income ratio (not depicted) is much weaker, with the confidence interval of (0.4109, 0.4677).

Finally, let us consider the triplet of non-migrant exemptions per return, immigrant exemptions per return, and emigrant exemptions per return (bottom row of Figure 13). As with other variables (say income), comparison between immigrants and emigrants yields stronger correlations than that between migrants and non-migrants. The 95% confidence intervals of the corresponding three heat maps on the Figure 13 from left to right are (0.7202, 0.7524), (0.8014, 0.8252), and (0.6967, 0.7311). This suggests that there may be some unknown underlying characteristic that makes a county attract and lose smaller families, while its influence over non-migrants is not as strong.

Summary

Using Statistics of Income data I construct an overview of county-to-county migration patterns. The analysis focuses on population, income, and migration flows. It is based on the exemptions and incomes reported on tax returns. Migrants and non-migrants are identified by matching tax returns filed by individuals in two consecutive years and comparing taxpayers' addresses.

I group counties into quintiles based on four main characteristics: population, average income, population growth, and income growth. Then, I analyze what trends can be found in these quintiles between years 1995 and 2013 and what are the interesting correlations between various county characteristics. The main findings include:

- Counties diverge in terms of population: larger counties have faster population growth rates than smaller counties.
- Overall, individuals tend to migrate from smaller counties to larger counties.
- Differences in population growth are driven more by migration and less by internal growth of the non-migrant population.
- Larger counties tend to be more homogeneous and smaller counties tend to be on both extremes in virtually every dimension.
- Larger counties tend to have higher average incomes and have smaller migration rates.
- Migration rates are pro-cyclical: people migrate less during downturns.

- Income of migrants relative to income of non-migrants has decreased over past 20 years.
- Migrants tend to have smaller households than non-migrants.
- Only around top 10% counties are above the national average in terms of income per capita. This inequality is slightly declining.
- There is a general declining trend in the average family size.

To explore further the dynamics of income and migration visit the Private Enterprise Research Center's webpage. There you will find interactive maps depicting county level migration and income data for migrants and non-migrants: http://perc.tamu.edu/perc/maps/

Sources

Statistics of Income Migration Data:
https://www.irs.gov/uac/SOI-Tax-Stats-Migration-Data

Migration Data User Guides:
https://www.irs.gov/uac/SOI-Tax-Stats-Migration-Data-Users-Guide
https://www.irs.gov/pub/irs-soi/1213inpublicmigdoc.pdf

Statistics of Income Migration Data New Approach (change in methodology):
https://www.irs.gov/pub/irs-soi/soi-a-inmig-id1509.pdf

PERC Maps:
http://perc.tamu.edu/perc/maps/