2009

What Proportion of Children Stay in the Same Location as Adults, and How Does This Vary Across Location and Groups?

Timothy J. Bartik
W.E. Upjohn Institute, bartik@upjohn.org

Upjohn Institute Working Paper No. 09-145

WHAT PROPORTION OF CHILDREN STAY IN THE SAME LOCATION AS ADULTS, AND HOW DOES THIS VARY ACROSS LOCATION AND GROUPS?

Upjohn Institute Staff Working Paper No. 09-145

Timothy J. Bartik
W.E. Upjohn Institute for Employment Research
Bartik@upjohninstitute.org

February 2009

ABSTRACT

This paper provides new information on what proportion of individuals spend their adult work lives in their childhood metropolitan area or state. I also examine how this proportion varies across different demographic groups, and with the size and growth rate of the metropolitan area. I find that the proportion of individuals who spend most of their adulthood in their childhood metropolitan area is surprisingly high. Furthermore, this proportion does not go down as much as one might think for smaller or slower-growing metropolitan areas, or for college-educated persons. These findings imply that state and local investments in children may pay off for the state or local area that makes these investments. A surprisingly large proportion of the individuals who benefit from these childhood investments will remain in the same state or local area as adults, thereby boosting the local economy.

JEL Classification Codes: R23, J61, R28, R11

I thank Wei-Jang Huang and Claire Black for assistance with this paper.

What Proportion of Children Stay in the Same Location as Adults, and How Does This Vary Across Location and Groups?

INTRODUCTION

This paper examines how many children stay in the same state or metropolitan area as adults. This topic is relevant to state and local policymakers considering investments in children. "Investments in children" includes high-quality early childhood programs, programs to improve K–12 school quality, or other youth programs.

From a local perspective, part of the return to investments in children is only realized if the child stays in the same area as an adult. Suppose these investments in children have positive effects on adult outcomes. These improved adult outcomes could include greater educational attainment, greater earnings, less criminal activity, or less welfare dependence. These positive effects on individuals create benefits for the local community. Community benefits include the following: greater economic activity, a higher tax base, less need for safety net and criminal justice system spending, lower crime, and a higher local quality of life. If the children live in the same community as adults, a state or local government investing in children's programs will reap some community benefits. If they leave the area, some other area will reap these benefits.

Therefore, how many children stay in their home community as adults is important. A benefit-cost evaluation from a state or local perspective must consider mobility. For example, in previous work, I have examined how preschool affects a state's economic development (Bartik 2006, 2008). This previous work made assumptions about how many preschool participants will remain in the same state as adults.

There is no research that provides an answer to this question: what proportion of children stay in the same local economy as adults? And how does this proportion depend upon the characteristics of the local economy or characteristics of the individual?

The size of the local economy might make a difference. Larger economies provide thicker and more diverse employment opportunities. Therefore, perhaps people are more likely to remain in their childhood home in larger communities. But social ties may be stronger in smaller economies. Therefore, perhaps more people would remain in smaller home communities as adults.

The growth of the local economy might make a difference. Faster local growth may be associated with improved local employment options. On the other hand, faster growth is associated with changes in the character of the local community. For persons who preferred the original community character, these changes may increase out-migration.

A person's education might make a difference. Labor markets for more-educated persons are believed to be more national in character than labor markets for less-educated workers. This might increase mobility of more educated workers. On the other hand, more educated persons might be more able to choose the location they want, which could be their home community.

A person's race might make a difference. Racial minorities might have fewer options for making job contacts in distant locations. On the other hand, racial minorities might have to search further to find a better job.

This paper provides descriptive information on how some location characteristics and individual characteristics predict whether a child will live in the same local area as an adult. This descriptive information has some immediate applicability in helping predict the state and local return to early childhood programs. But it also suggests hypotheses that might be explored with more detailed models.

This model estimated is a reduced form model that collapses many decisions into an observed location outcome. Whether an adult lives in his or her childhood area is an amalgam of many decisions that have occurred over time. For example, this will include location decisions by the individual's parents during his or her childhood. This will also include location decisions to move out of the local area as well as decisions to move back in.

The model is descriptive. It is unclear that all of this paper's predictors for "staying in the same place" are fully exogenous. For example, there may be some cultural differences between families choosing larger metropolitan areas versus smaller metropolitan areas. This paper's estimated "effects" of metropolitan area size could reflect these cultural differences as well as true effects of size. As another example, education is chosen by individuals. In some cases, an individual might go to college so that he or she can get away from home—permanently. This paper's estimated effects of education may in part reflect such reverse causation.

However, we currently know little about the correlates of children staying in the same local area as adults. Some descriptive reduced-form information is useful.

DATA

This paper mainly relies on data from the Panel Survey of Income Dynamics (PSID). I also use some data from the U.S. Census Public Use Microdata Samples (PUMS).[1]

[1] Some calculations also rely on population data from the U.S. Census, race and education data from the American Community Survey, and race, gender, and education data from the Current Population Survey–Outgoing Rotation Group. I will explain how these variables are constructed when they are introduced.

The PSID follows the same individuals over time. These data are used to see what proportion of children stay in the same metropolitan area as adults. The PSID is unusual in having a long enough panel that we observe individuals both as children and as adults well into their careers.

An appendix provides some information on how the PSID was processed. I used the Geocode version of the PSID. The Geocode version provides annual information on the county location of each individual in the PSID.[2] Counties were matched to the 2008 definitions of metropolitan areas. The Geocode version of the PSID reliably provides such location information from 1970 to the most recent survey. For this paper, the most recent survey used is for 2005.

The PSID's definition of an individual's "location" has one oddity: dependent children who are in college are considered to be located at their parents' home. A separate location is not assigned until after college.

All my calculations of the probability of staying in the same metropolitan area between two time periods use the PSID's individual weights. The original PSID sample oversampled low-income families. The use of individual weights corrects for this oversampling. It also corrects for other sampling issues with the PSID. This use of individual weights is important so that the estimated staying probabilities will represent the U.S. population. Individuals from low-income families might be less mobile.

Some of the calculations separate individuals by educational attainment. For all calculations doing so, I classify individuals by educational attainment based on their "final"

[2] For recent years, the PSID information is only available every other year.

educational attainment levels observed in the PSID. Individuals who were less than 23 when their final education level was observed are classified as having "missing" educational attainment.

Some examinations of "staying" behavior use data from the Public Use Microdata Series (PUMS) for both 2000 and 1990. Specifically, I use information from PUMS on whether individuals live in their state of birth. All PUMS calculations for percentage living in the state of birth used the appropriate person weights in this data set.[3]

PUMS data obviously does not follow the same individuals over time. However, location patterns by age may suggest "staying" behavior if such behavior has not changed drastically over time.

Finally, for various calculations I used data from other sources. Information on metro area population and population growth, which are used as independent variables, is based on census data on counties aggregated up to the current metropolitan area definitions. The percentage college-educated and percentage black in each metropolitan area is also used as an independent variable. These variables are calculated using data from the 2006 American Community Survey. Calculations for the "typical" percentage of the national metropolitan population that is in different race, gender and education categories are used to calculate average staying probabilities for various "typical" groups. These demographic breakdowns were calculated from the 2007 Current Population Survey–Outgoing Rotation Group.

[3] I would have liked to use PUMS data on how metro-area location varies over time. However, "metro area of birth" is not an available variable in PUMS. In addition, for the five-year migration data in PUMS, only 28 percent of individuals have valid codes for the metro area they lived in five years ago, probably because of suppression of data for confidentiality reasons.

SOME PRELIMINARY DESCRIPTIVE DATA

This section provides some descriptive data from the PSID, and the PUMS, on the proportion of individuals who stay in the same location as adults that they were in as children.

Overall Comparisons

Figure 1 provides a comparison of location in the same state over time using the PSID and PUMS. The PUMS data looks at whether individuals of different ages live in their state of birth. One PUMS line shows results from the 2000 Census, the other from the 1990 Census. This does not look at how the location of the same individual changes over time. Individuals of different ages are in different cohorts. Changes with age in the proportion of individuals living in their birth state reflect trends in how individuals move as they age. But it may also reflect how the tendency to stay in the same state has changed with different cohorts.

The PSID data do enable us to determine whether an individual stays in the state that individual lived in as of some base year. The base years used here are age 4 and age 14. Age 4 is chosen because it is a typical age for preschool. Age 14 is chosen because school reforms to improve high schools might include this age. I want these estimates to be relevant to whether investments in preschool or high school reform will pay off for the local community.

The PSID calculations involve aggregating all individuals in different years whose base year for calculating staying probabilities is age 4 or age 14. For example, the calculated average staying probabilities from age 4 to age 5 would aggregate the staying probabilities of persons staying in the same state from 1970 to 1971, with staying probabilities of other individuals examined from 1971 to 1972, etc.

Because of the nature of the PSID, there are more observations for staying probabilities when the "ending year" is closer to the "base year." For example, consider the calculation of staying probabilities from age 14 to age 15. The PSID provides observations on 8,349 individuals who were age 14 as of some year and for whom we have an observation for the next year. In contrast, consider the calculation of staying probabilities from age 14 to age 45. Over this age range, the PSID provides fewer observations. There are 524 individuals in the PSID for whom we observe their location at age 14 and for whom we also observe their location 31 years later. Therefore, estimates of staying probabilities become more imprecise the greater the difference between the base year and the end year. This imprecision can be seen in the figure. The staying probabilities jump around more as the difference between the base year and the end year increases. I stop the calculations after a 31-year gap because of this increasing imprecision.

One conclusion from Figure 1 is that there has been little change from 1990 to 2000 in staying behavior for different cohorts. The two PUMS lines almost exactly coincide. This suggests some stability in staying behavior.[4]

Another conclusion is that the PSID and PUMS data imply similar patterns of staying propensity with age. There is a moderate decline in staying propensity from birth to age 18. There is then a dramatic decline until age 30. (This decline is somewhat delayed in the PSID data compared to the PUMS data. This delay probably reflects that PSID data record residence

[4]Other research suggests some stability in staying behavior over longer periods. For example, research suggests that the percentage of 50- to 59-year-olds living in their state of birth was around 60 percent for the entire 20th century (Rosenbloom and Sundstrom 2003, Figure 3). I calculate 59.2 percent for 50- to 59-year-olds from the 1990 PUMS and 57.6 percent from the 2000 PUMS. The percentage of 50- to 59-year-olds in their state of birth was around 50 percent in the late nineteenth century. One change that has taken place over time is that people seem to "settle" sooner in the U.S. in the 1950 and 1960 and subsequent censuses. In censuses from 1850 to 1930, each older age cohort was significantly less likely to live in their state of birth for a given census year. Since 1950, there is not as big a difference in the percentage living in the state of birth between 30- to 39-year-olds and 70- to 79-year-olds. See discussion later in this paper for confirmation of this from the 1990 and 2000 PUMS.

for college students as their parent's residence.) After age 30, the staying propensity stabilizes and only declines slightly with age. In the PUMS data, there is roughly the same proportion in the state of birth after age 55.

The staying probabilities appear to be heading towards a similar long-run propensity to stay in the same state. This long-run propensity is somewhere in the low 60 percent range. The PUMS data have a somewhat lower staying propensity than this range. However, the PUMS data need to be adjusted upwards somewhat because of the nature of "state of birth" as a location variable. From birth to age 1, the percentage living in the state of birth drops abruptly. The drop is from 100 percent to 91.6 percent in 2000 and to 92.0 percent in 1990. But the percentage living in the state of birth then only declines by about 2 percent a year for the next several years. The abrupt drop from birth to age 1 reflects recording the state of birth based on the location of the hospital.[5] Therefore, the staying probabilities estimated in PUMS probably need to be adjusted upwards by 6 percent if they are to reflect differences between the "state of birth residence" and current state of residence. The proportion of individuals in the 2000 PUMS who are in their state of birth averages 58.8 percent from ages 46–64 and 57.6 percent over ages 50–80. The 1990 PUMS gives similar percentages of 59.0 percent from ages 36–64 and 58.9 percent from ages 50–80. Adjusted upwards by 6 percent, this would give a long-run propensity to stay in the state of birth residence of 64 or 65 percent.

[5] According to ICPSR (1996, p. 224), prior to 1990, the PUMS tried to get respondents to identify place of birth as the state the individual's mother lived in at the time she gave birth. However, research suggested that this distinction between actual place of birth and where the mother was living was generally ignored or misunderstood. In 1990 and 2000, there was no attempt to separate the actual state of birth from where the mother lived at the time of the birth. Of course, it is possible that some respondents did give the location of the mother at the birth. But the sharp drop from birth to age 1 suggests otherwise.

In the PSID, the staying probabilities in the later years are close to these long-run values. The percentage of individuals who are living in the same state they lived in at age 4 averages 72.2 percent for ages 31 through 35. The percentage of individuals who are living in the same state they lived in as of age 14 averages 69.1 percent for ages 41 to 45. These figures are not too much over a possible long-run value of 64 or 65 percent living in the same state.

These patterns of change over time reflect choices made by parents when the individual is a child, and choices that the individual makes as an adult. The modest declines prior to age 18 may reflect parents' interests. Parents may want to maintain their own job networks and social ties and provide geographic stability for the child. The more dramatic declines after age 18 and until age 30 reflect the child maturing and finding his or her desired location. This desired location may be affected by the child's college location as well as his or her chosen career. The relative stability after age 30 probably reflects some offsetting trends. On the one hand, there will inevitably be some changes of location to reflect changing career opportunities or a desire for different location amenities. On the other hand, some individuals may choose to return from different locations to their original location. These two trends at some point become roughly of the same magnitude. (In contrast, when the child is young, the probability of returning to the child's original state of residence is much lower. A parental return to the parent's state of birth residence is quite likely to be a different state than the child's state of residence as of some base year.)

The PUMS data also suggest some slight differences over cohorts in staying behavior. There is a dip in the staying probability for cohorts born prior to 1945.[6] This may reflect World War II–related change of state of residence.

The PSID allows us to easily compare the propensity to stay at different geographic units of aggregation. Figure 2 compares the propensity to stay in the same location from some base year to some end year for two definitions of location: 1) the same state and 2) the same metropolitan area. As one would suspect, the propensity to stay in the same metropolitan area trends below that of the same state. (This is not inevitable because an individual can stay in the same metropolitan area but in a different state. Some metropolitan areas straddle state boundaries.)

The propensity to stay in the same metropolitan area appears to become roughly stable after age 30. However, the ratio of the percentage staying in the same metropolitan area to the percentage staying in the same state is about 0.78 or 0.79. The percentage of individuals who live in the same metropolitan area they lived in at age 4 averages 56.2 percent from ages 31–35. The percentage of individuals who live in the same metropolitan area they lived in at age 14 averages 54.3 percent from ages 41–45.[7]

These figures suggest that the percentage of individuals staying in their childhood metropolitan area for most of their working career is likely to be in the 45 percent to 55 percent range. This extrapolation relies on considering plausible declines in the percentages observed in

---

[6] An excel spreadsheet available from the author has the actual percentages for each age.

[7] The comparison of these numbers with similar percentages for the same state gives the ratios cited in the text. For example, 56.2 percent in the same MSA as they were in at age 4, divided by 72.2 percent, gives a ratio of 0.779; 54.3 percent in the same MSA as they were in at age 14, divided by 69.1 percent, gives a ratio of 0.786.

the early 30s and early 40s age ranges. These plausible declines are based on trends observed in the PUMS data on percentages in their birth state.

Comparisons by Education

I also examine some descriptive information on how the propensity to stay in the same location varies by whether the individual has a four-year college degree.

Using PUMS data, I can only examine information on the education of the individual as of a particular age. Therefore, it only makes sense to examine differences between college graduates and noncollege graduates for individuals at least 22 years of age.

Figure 3 uses the PUMS to compare the proportion of individuals at different ages that live in their state of birth, for individuals who have a four-year college degree versus those who do not. From ages 22 onward, college graduates have a lower propensity to stay in the same state. The proportion living in the birth state drops especially fast for college graduates during their 20s, before stabilizing.

There also is an abrupt dip in the percentage living in the same state for college graduates born between 1945 and 1950. This may reflect a cohort effect related to World War II and its aftermath. The 1990 PUMS suggests a similar dip for college graduates born between 1925 and 1930.

Figures 4 and 5 use the PSID data to follow the same individuals over time, but classified by their final educational attainment. By “final” educational attainment, I mean the last educational attainment I observe for that individual in the PSID. I required that this final educational attainment be observed at age 23 or later. (For individuals whose final educational attainment is observed before age 23, final educational attainment is classified as missing.)

Figures 4 and 5 look at whether individuals of different educational attainment stay in the same metropolitan area. I also looked at whether these individuals stayed in the same state. These qualitative results were similar, except with generally higher levels staying in the same state.[8]

Both Figure 4 and Figure 5 show similar patterns of how the staying propensity varies over time for college graduates versus nongraduates. The staying propensity is remarkably similar between college graduates and nongraduates until the early 20s. The staying propensity for college graduates then rapidly declines until about age 30, while the staying propensity for non-college graduates only moderately declines. After age 30, staying propensities for both groups stabilize.

The percentage of college graduates staying in the same metropolitan area they lived in at age 4 averages 42.3 percent at ages 31–35. The percentage of college graduates staying in the same metropolitan area they lived in at age 14 averages 40.6 percent for ages 41–45. This suggests that the percentage of college graduates living in their childhood metropolitan area during the majority of their adult working career is in the 35 to 45 percent range. The percentage of college graduates living in their state of birth is probably in the 49 to 59 percent range, judging from the PUMS 2000 data. (This adjusts for the likely differences between state of birth

[8]An Excel spreadsheet showing these state results is available from the author.

residence and state of birth). The upper end of that range seems more plausible in the absence of some special effects, for example a major war.[9]

The long-run propensity of college graduates to live in their childhood location as adults is one-quarter to one-third lower than for noncollege graduates. Consider persons ages 31–35 and whether they live in the same metropolitan area as they did at age 4. The ratio of the percentage of college grads living in the same area to noncollege grads living in the same area is 0.679 (42.3 versus 62.3 percent). Consider persons ages 41–45 and whether they lived in the same metropolitan area as at age 14. The ratio of the percentage of college grads living in the same area to noncollege grads is 0.677 (40.6 versus 59.9 percent). Consider persons ages 30–64 and whether they are living in their state of birth. The ratio of the percentage of college grads in their birth state to noncollege grads is 0.755 (48.5 versus 64.3 percent).

What about noncollege graduates? About 55 to 65 percent of noncollege graduates stay in the same metropolitan area they lived in as a child for most of their working career. Seventy percent of noncollege graduates stay in the same state they lived in as a child for most of their working career. This is important because of the large number of noncollege graduates. Even among younger cohorts, most Americans do not get a four-year college degree. For example, based on 2007 Current Population Survey data, only 34.7 percent of Americans ages 28–32 have at least a four-year college degree.

[9] This is consistent with the PSID when analyzed at the state level. The PSID results show that the percentage of college grads living in the same state as of age 4 averages 59.3 percent for ages 31–35. The percentage of college grads living in the same state as of age 14, averaged over ages 41–45, is 58.7 percent. Allowing for some decline over time, this would be consistent with long-run percentages of college graduates living in the same state that they lived in as children, which would be in the 50 to 60 percent range.

Comparisons by Race

I also analyze how the propensity to stay in the same location varies between blacks and whites.

Figure 6 examines data from the PUMS. In the 2000 PUMS, for ages 54 and below, a higher percentage of blacks than whites live in their birth state. For ages 55 and above, a lower percentage of blacks than whites live in their birth state. The relative birth-state residence of blacks versus whites starts declining at about age 44. At ages 58 and above, the staying percentage of blacks seems to stabilize at about 5 to 7 percent below that of whites.

The 1990 PUMS suggests that this pattern is due to differences in migration by birth cohort. The 1990 PUMS shows a similar pattern, but at ages that are 10 years younger.

These patterns are probably associated with the migration of blacks from the South to the North that occurred between the early 1940s and the late 1960s (Lemann 1991). The PUMS data suggests that this migration significantly reduced the staying probabilities of blacks relative to whites for persons born before 1942. To a lesser extent, this migration reduced the staying probability of blacks relative to whites for persons born between 1942 and 1956.

This pattern shows that the correlates of staying in the same location may vary across cohorts. We should be cautious when using this descriptive data to make predictions.

Figures 7 and 8 use data from the PSID to look at the staying propensity of blacks vs. whites. The PSID is looking at staying behavior of persons who were children in 1970 or later. In this time period, blacks are much more likely than whites to remain in the same metropolitan area. Consider persons ages 31–35 and whether they lived in the same metropolitan area as at age 4. The percentage of blacks in the same area was 80.4 percent, versus 51.4 percent for whites. Consider persons ages 41–45, and whether they lived in the same metropolitan area as at age 14. The percentage of blacks in the same area was 70.6 percent, vs. 51.6 percent for whites.

Suppose these patterns do not change in the future. Under that assumption, two-thirds or more of blacks will stay in their childhood metropolitan area for most of their working career, versus 45 percent or more of whites. Of course, these relative migration patterns may change.

Comparisons by Size of Metropolitan Area

Using the PSID data, I calculate staying probabilities by metropolitan areas divided into population quartiles. These quartiles are population-weighted quartiles. (Therefore, each quartile contains about one-fourth of the metro area population in the United States.)[10]

Figures 9 and 10 show these calculations. Smaller metropolitan areas have somewhat lower probabilities of individuals staying in the same metro area. However, the differences are not huge. Consider individuals ages 31–35 and whether they stayed in the same metro area they lived in at age 4. For the smallest metro areas, the percentage staying is 51.6 percent. The overall average is 56.8 percent. Consider individuals ages 41–45 and whether they stayed in the same metro area they lived in at age 14. For the smallest metro areas, the percentage staying is 47.0 percent. The overall average is 54.4 percent.

Above the smallest metropolitan areas, it is not clear how the staying propensity varies by metro area population size. Consider persons ages 31–35 and the percentage staying in the same metro area they lived in at age 4. The percentage staying varies by metro area size as follows: 51.6 percent for the smallest population quartile, 57.6 for the second quartile, 55.3 for the third

[10] This quartile calculation is based upon the logit estimation sample used below for staying between age 14 and age 30. Specifically, this takes the metro area's size as of age 14 for the metro area that the person lived in at age 14 and calculates the population-weighted quartiles for that metro area size variable.

quartile, and 62.6 for the largest population quartile. This perhaps shows a slight tendency for larger metro areas to have a higher staying propensity. But consider persons ages 41–45 and the percentage staying in the same metro area they lived in at age 14. The percentage staying varies by metro area size as follows: 47.9 percent for the smallest areas, 61.8 percent, 54.1 percent, and 54.6 percent for the largest areas. The second smallest quartile of metro areas has a higher staying percentage than the largest quartile of metro areas.

These descriptive data only control for one variable at a time. I now move to a multivariate description of how characteristics of metro areas and individuals are correlated with staying in the same metro area.

MODELING INDIVIDUAL PROPENSITIES TO STAY IN THE SAME METROPOLITAN AREA

I use PSID data to look at the propensity of individuals to stay in their childhood metropolitan area. I examine the determinants of this "staying propensity" for three age ranges: 1) from age 4 to age 30, 2) from age 14 to age 30, and 3) from age 14 to age 40. The dependent variable measures at the latter age whether the individual has "stayed" in the same metropolitan area that that individual lived in at the earlier age.

This staying propensity is estimated in this section using individual data. The staying propensity is a zero-one dummy variable. Therefore, the effects of its determinants should be estimated with a discrete choice estimation method. I report logit estimates in this paper. However, estimates are similar using probit or linear probability models.

The individual's staying propensity is estimated as a function of both individual characteristics and metro area characteristics. The individual characteristics include gender, education, and race. The metro area characteristics are for the individual's "original metropolitan area." One metro area characteristic is the metro area's population as of the base age. The other metro area characteristic is its average annual population growth rate between the base age and the "final age."[11] I also include dummy variables for the calendar year of the final age. These dummies control for any national year effects on mobility. Finally, I include a dummy variable for whether the individual was part of a family that was in the original low-income sample of the PSID. I will discuss this sampling issue below.

This model can be written as the following equation:

(1) $\Pr(\mathrm{Stay}_{ima0a1})$ = f(Gender, Education, Race, MSA pop, MSA pop growth, Year, Low-income sample family member)

$\Pr(\mathrm{Stay}_{imat0a1})$ is the probability of individual I staying in metro area m from age $a0$ to $a1$. The model is separately estimated for each of the three sample age ranges (ages 4 to 30, 14 to 30, 14 to 40). These three age ranges are chosen as a compromise between two contradictory goals. First, I want to examine staying behavior as far into the adult's prime work career as possible. Second, I want an adequate sample size. The PSID allows a reasonable sample size (see below) for up to a 26-year age range. A 26-year age range pushed the samples with age 4 and age 14 starting ages about as far into the work career as possible with a reasonable accuracy of estimates. The smaller 16-year age range for the ages 14 to 30 sample allows for more precision of estimation.

[11] This is the population growth rate for the original metro area. The population of any other metro area chosen at the final age is irrelevant.

Education is measured by the individual's final educational attainment that is observed in the PSID. I required that this final educational attainment be observed at age 23 or above. For a very few individuals in these samples, educational attainment at age 23 or above is not observed. These individuals are treated as having "missing education." A dummy variable for "missing education" is included.

Educational attainment is defined by a set of dummy variables for the individual's educational attainment. The set of mutually exclusive and exhaustive educational categories include the following: high school dropout, high school graduate but no higher education, some college but no bachelor's or associate degree, college degree but no higher degree, some postcollege degree, and missing final educational attainment variable. In the empirical estimation, "high school dropout" is the omitted dummy variable.

Race is also defined by a set of mutually exclusive and exhaustive categories that each define a dummy variable. These categories include white non-Hispanic, black non-Hispanic, Hispanic, and the "other racial" category. In the empirical estimation, white non-Hispanic is the omitted dummy variable.

Metro area population size is measured as the natural logarithm of the population of the individual's metro area of residence as of the base age. This logarithmic specification allows larger population to have a percentage effect on staying propensity, which seems a reasonable hypothesis. However, I also explore other functional forms, as I will describe further below.

Metro area growth is measured as 100 times the average annual logarithmic population growth of the individual's original metro area of residence between the base age and the ending age. Specifically, this growth rate is equal to the following expression:

$$(2) \qquad GR_{ma0a1} = 100 * [\ln(POP_{ma1}) - \ln(POP_{ma0})]/(a1 - a0).$$

To avoid being misunderstood, both metro area population numbers refer to the population of the individual's metro area of residence as of age $a0$, the baseline age. Even if the individual moves to a different metro area between age $a0$ and $a1$, the population of the new metro area is irrelevant to this calculation.

I also explore alternative functional forms for growth rate affecting the staying propensity, on which more below.

Metro area population is measured using official Census Bureau data on the population of individual counties. Counties are combined according to the 2005 definition of metropolitan areas. These same definitions of metropolitan areas is used in classifying individuals' locations.

Year dummies control for the calendar year in which age $a1$ occurs. This will vary across individuals. For the first and third samples, which comprise 26 years from the base age to the final age, the final age could occur in the following possible years: 1996 and 1997, and on alternate years from 1999 to 2005. The final age cannot occur prior to 1996 because reliable geographic location data are not available in the PSID prior to 1970. After 1997, the final year can only be every other year because the PSID switched from annual to biannual data collection. Similarly, for the second sample, the final age could occur in the following possible years: each year from 1986 to 1997, and every other year from 1999 to 2005.

I include the entire PSID sample in the estimation. However, the PSID is not a simple random sample. Among other things, a little under 40 percent of the original sample families when the PSID started in 1968 were low-income families with a family head less than 60 years old.[12] Among the remaining 60 percent of the original sample, the sample was selected via a stratified random sample. In addition, there has been some differential attrition from the PSID. For all these reasons, any particular sample from the PSID will not necessarily be nationally representative.

In calculations involving sample means, I deal with the nature of the PSID sample by using the PSID's individual sample weights. These weights are meant to account for the probability of any particular individual being selected for the sample. These weights are also adjusted to U.S. population totals to reflect differential attrition.

However, in the logit estimation of the determinants of the staying propensity, I do not use sample weights. (This practice is common in economists' use of data sets such as the PSID.) I assume that the model controls for the nature of the PSID sample by (among other things) the controls for education, race, and gender. I also include a dummy variable for whether the individual was in the original SEO low-income sample. This controls for any other economic or cultural factors that might differentiate individuals who were in low-income families as of 1968 from other families. Using sample weights in the estimation throws away a great deal of the information in the low-income sample, which will yield less precise estimates. I assume that throwing away this information is not necessary as I can control for the separate sample and for other individual factors affecting the staying propensity.

[12]To be more precise, 39.0 percent: 1,872 completed interviews are with families from the low-income Survey of Economic Opportunity Sample, and 2,920 completed interviews are with a sample meant to be nationally representative.

As I will detail below, in making predictions based on the estimates, I do not use the dummy variable for the low-income sample. I also use independently determined population weights in calculating average behavior.

I explored some alterative functional forms in the metro area variables. I start from a specification in which I include all the individual variables, plus the log of population and the growth rate. I then test specifications that add higher-order terms in the log of population. I stop doing so when the Akaike Information Criterion is no longer reduced by adding an additional higher-order term in the log of metro area population. Once I determine the functional form for metro area population, I then add higher order terms in population growth. This process then stops when the Akaike Information Criterion is no longer reduced by adding in higher order terms in the growth rate.

As it turns out, the optimal specification for the first and third sample is linear in both the log of metro population and the logarithmic metro area population growth rate. The second sample includes squared terms as well. The somewhat more complex functional form may in part reflect the larger sample size of the second sample. The larger sample size allows more precision in allowing for different effects of metro area variables on the staying propensity.

Table 1 reports the weighted means and unweighted means of the dependent variables and independent variables in these three samples. The weighted means seem reasonable. The unweighted means, as one would expect, diverge somewhat because of the oversampling of individuals from families that were low-income as of 1968.

Table 2 provides the logit coefficients for each of the three samples. These estimates are for the "optimal specification." This specification includes higher-order terms in metro area population and growth for the second sample.

I first focus on the statistical significance and sign of the estimated coefficients. I then turn to interpreting the size of the effects. This requires some work because of the discrete estimation methods.

Focusing on the statistical significance and sign of the estimated coefficients for individual characteristics, the following seem the most important findings:

1) Higher educational attainment is strongly associated with a lower probability of staying in the same metro area. This is particularly true for individuals with a college degree or above versus those with less than a college degree.

2) African Americans are estimated to be significantly more likely to stay in the same metropolitan area than whites.

3) After controlling for the other variables, whether the individual comes from the 1968 low-income family sample is not significantly associated with whether he or she stays in the same metropolitan area.

What about the effect of metro area characteristics? Individuals are more likely to stay in the same metro area from age $a0$ to $a1$ if the metro area is larger or growing faster. However, these effects are not statistically significant for the first or third samples. This may be due to the relatively modest sample size in these samples.

For the second sample, the effects of metro area population are statistically significant at conventional levels of significance (probability less than 0.001). The effects of metro population growth are not statistically significant at conventional levels (probability of 0.1279).[13]

Because this is a logit estimation, the coefficients do not directly show the effects of marginal or discrete changes in the independent variables on the staying propensity. The usual rule of thumb is that dividing logit coefficients by 4 will give approximate effects of the right-hand side variables on probabilities (Wooldridge 2002). However, the exact effects depend upon the sample probabilities. The figures in brackets represent calculations of the effects of a 1-unit change in each independent variable at the sample mean probabilities. As expected, these bracketed effects on the staying propensity are in the range of 20 to 25 percent of the logit coefficients.

To interpret these bracketed terms, note that most of the variables are discrete dummy variables (gender, education, race, low-income family sample). For such variables, a 1-unit change represents a change from the omitted category to the dummy variable category, evaluated starting out at the sample mean probability of staying. For example, the bracketed term for the college graduate dummy variable is −0.317 for the first sample. This means that for an observation that initially had a logit index value that would generate a sample mean probability of staying of 0.552 and was a high school dropout observation, a change from high school

[13]This conclusion is robust to assuming alternative variance structures. Allowing for clustering at the MSA year level does not much affect the standard error estimation. Allowing for clustering at the metro area level does increase the standard errors on the Metro area variables by a fair amount. Resulting z-statistics on the population terms in the 14–30 logit are 1.55 on ln(pop) and −1.40 for the square of ln(pop). Z-statistics are 1.33 for population growth and −1.31 for the square of population growth. However, the chi-squared for the population level variables has a probability of 0.0155. This is still statistically significant at conventional significance levels. The chi-squared for the population growth variables has a probability of 0.3894.

dropout to college graduate would reduce the staying probability by 0.317, to 0.235 (= 0.552 − 0.317). This effect seems substantial.[14]

For the population size term, a "1 unit" change in the natural log of population size is an increase in metro area population of 171 percent. For the quadratic specification used in the ages 14 to 30 sample, this increase is evaluated at the weighted sample mean probability and at the weighted sample mean of ln(population size). For the growth rate term, a "1 unit" change is an increase in the logarithmic average annual population growth rate of 1 percent. Again, for the quadratic specification used in the ages 14 to 30 sample, this is evaluated at the weighted sample means of the probability of staying and of average population growth rates.

However, these effects are only roughly estimated effects at the sample means. In the quadratic specification, the effects of changes in population size and population growth vary with starting values of population size and growth. In addition, the effects of the different dummy variables will differ at different values of the other variables. Therefore, I do some more precise calculations of the effects of different changes in the independent variables. I do this for the ages 14–30 estimates. I choose these estimates because they are more precise given the larger sample size.

I set up a program that allows for a calculation of the probability of the individual being in the same metro area at ages 14 and 30 as a function of any value of the right hand side variables. I then calculate this probability for all possible combinations of the variables. There are two gender categories, five education categories, and four race categories. Multiplying

[14] Note that this does not imply that the average college graduate has a staying probability of 0.235. Rather, this would be the behavior of college grads whose other variables would yield them an average staying probability even if they were a high school dropout. As high school dropouts would tend to have above-average staying probabilities, this implies that other variables are tending to reduce the staying probability.

together gives 40 possible gender/education/race categories in which the individual might fall. Using the data, I calculate the following percentiles of metro population and population growth: 5 percent, 10 percent, 25 percent, 50 percent, 75 percent, 90 percent, 95 percent. These percentiles are calculated using the PSID weights. Therefore, they represent population-weighted figures for metro area size and growth. These seven categories of metro area population, and seven categories of metro area population growth, result in 49 other categories for which I examined staying probabilities. I calculate all these probabilities using the average of the constant term and the various year dummies as representing "average behavior" over the sample. To make the estimates representative of the general population, I assume that the value of the 1968 low-income family variable is zero.

The resulting calculations are in a spreadsheet. This spreadsheet uses the logit coefficients to calculate the probability of staying in the same metropolitan area for 40 individual characteristic cell categories times 49 possible combinations of metro area population size and growth variables = 1,960 probability calculations.[15] The spreadsheet also has formulas that allow a recalculation of probabilities for any other population or population growth figures.

To present the implications of this spreadsheet requires some simplification. Therefore, I use the spreadsheet to examine the effect of one variable at a time, holding constant other variables at typical values.

First, I use the 2007 Current Population Survey, Outgoing Rotation Group to calculate the proportion of the U.S. population ages 28–32 in each of the 40 demographic groups given

[15] The spreadsheet is available on request from the author.

above. I use ages 28–32 to try to capture "final" educational attainment, while also reflecting recent educational attainment trends.

I use these weights to determine the probability of staying for a population that is representative of current 28- to 32-year-olds in the U.S. To isolate effects with population, I calculate effects with population growth held constant at its population-weighted median of 0.8 percent (an average annual change in logarithm of population of 0.007559). The resulting calculations are shown in Figure 11. Table 3 shows the same numbers.

What does this figure and table show? For larger metro areas (above the 50th percentile), the probability of staying varies little with metro population size. The probability of staying gradually declines as we consider smaller metro areas. Then it declines much more rapidly at the smaller metro area percentiles.

However, even in the smaller metro areas, the probability of staying in the same metro area is not too far below the average. For example, in the median metropolitan area, with a population size of about 1 million, the probability of staying is about 67 percent. This declines for a metro area at the 5th percentile of the metro population distribution, with a population of a little over 100,000. But the decline is only to about 53 percent staying. All of these calculations hold the demographics of the metro area constant. These calculations also hold constant the population growth of the metro area.

To isolate effects of population growth, I do a similar exercise. I again consider a metro area population that represents the U.S. average distribution of population characteristics for 28- to 32-year-olds in 2007. I consider the staying propensity at different percentiles of metro population growth. Metro population is held constant at its median level of 1.083 million.

Figure 12 and accompanying Table 4 show the estimates. As this table shows, metro areas with average or somewhat above average growth have a higher staying probability than slower-growing metro areas. However, the staying probability tails off for the fastest-growing metro areas. Growth may open up opportunities. But rapid growth may disrupt social ties in a metro area.

However, the main point is that metro population growth makes little difference to the probability of staying in the metro area. All these staying probabilities only vary between 64 and 68 percent.

I also used the logit estimates for the 14–30 specification to examine the effects of education and race. How do education and race affect the staying probability, holding other factors constant? To examine effects of education, I did the following: I assumed a metro area with median population size and median population growth. I then compared the staying probabilities for a population that had the same gender/race composition but was placed in each of the five different education categories in turn. The gender/race composition is held constant at the percentages in each gender/race group that were typical of 28- to 32-year-olds in the United States in 2007.

The results of these calculations are shown in Figure 13. The result shows large differences in staying probabilities between those with a college bachelor's degree and above, and those with less education than a college bachelor's degree. For example, the staying probability is 53 percent for someone with a bachelor's degree, versus 80 percent for someone with just a high school diploma. Within these two broad education groups, the differences in staying probabilities are not as marked.

I do a similar calculation for different races. I hold metro area population and population growth constant at their population-weighted median values. Education and gender are held constant by assuming each race has the same overall population composition for education and gender, held constant at the mean values for 28- to 32-year-olds in the 2007 CPS-ORG.

Figure 14 shows the results. There are large differences in the staying probability between races. For example, the estimated staying probability is 80 percent for blacks versus 64 percent for whites.

There are some limitations to these results. One problem is the limited sample size available for estimating the effects of metro area characteristics. The estimated quadratic effects of population and population size yield some interesting results. But these results are based on accepting the quadratic specification. Yet all we can really say is that adding higher-order terms does not improve the fit enough, as judged by the Akaike Information Criterion. But limited sample size may prevent our uncovering more complex relationships between population size, population growth, and the staying probability.

How can we increase the sample size? This leads to the notion of somehow combining all the data for staying probabilities at different ending ages. I explore this possibility in the next section.

Another limitation of these results is that the effects of race and education are only the effects of the individual variables. There may also be effects of the race or education composition of the metro area. The limited size of the age 14-to-30 sample makes it impossible to explore these other metro variables. Again, this requires expanding the sample size, which I explore in the next section.

MODELING AGGREGATE METRO AREA STAYING PROPENSITIES AT DIFFERENT AGES AS A FUNCTION OF METRO AREA CHARACTERISTICS

This section describes and presents estimates for a model that examines the effects of metropolitan area characteristics on the proportion of individuals in the PSID who stay in their childhood metro area. One set of models uses a starting age of 4, the other uses a starting age of 14. For a given starting age, the model uses as a dependent variable sample cell means (using PSID weights) for a given metropolitan area and ending age. The dependent variable is the proportion of individuals who started out in a given metropolitan area who are still in the metropolitan area as of some later ending age. The estimation is based on pooling data over various ending ages and metropolitan areas. This staying propensity is estimated as a function of the ending age, and of various metropolitan area characteristics. Metro area characteristics included are proportion college-educated, proportion black, metro population size, and average annual metro population growth. The model uses tobit estimation to reflect the truncation of the dependent variable at zero or one. The model uses weights for the number of observations generating each metro area/ending age cell mean for the proportion staying in order to correct for the relative precision of the dependent variable estimates. The resulting tobit estimates are then used to see how the staying propensity at different ending ages varies with different metro area characteristics.

Why this additional estimation? The motivation was described in the previous section. There are sample size problems when using data for only one set of starting and ending ages. But it seems likely that data on staying propensities from different ending ages provides additional information. For example, data on staying propensities from ages 14 to 39, or 14 to

41, says something about staying propensities from ages 14 to 40. Therefore, it seems reasonable that some way of combining the data should be tried.

Unfortunately, the usual hazard model of examining some irreversible decision is not applicable. It is not applicable because this decision is not irreversible. People can and do return to the metro area they lived in as a child after leaving the metro area for some time. A model using individual data would need to be quite complex. We need to allow for correlations across different points of time. Instead, I decided to estimate a simpler, descriptive model using aggregate data. This descriptive model adds some simple modeling elements to the descriptive figures used in this paper's section 2.

The dependent variable that is used is aggregate metro area/ending year cell means for the proportion staying in the same metro area from some childhood year to some ending year. One set of estimates is for "beginning age" 4. For each metropolitan area, I calculate cell means for the proportion staying in that same metro area between age 4 and some ending age $a1$, where $a1$ varies from age 5 to age 35. Another set of estimates is for the beginning age of 14. I calculate cell means for the proportion staying in the same metro area between age 14 and some ending age $a1$, where $a1$ varies from age 15 to age 45. These calculated cell means for each metro area/ending year cell aggregate across all individuals in the sample and all years in the sample.

For the beginning age 4, I end up with 6,399 cell means for the proportion staying for a particular metro area/ending age combination. The 6,399 cell means come from 291 distinct metro areas. The 6,399 cell means are based on the observation of 105,107 "staying choices" between age 4 and ending ages from 5 to 35; 1,334 cell means are zero (no one in that cell stayed in that metro area for that ending age), and 1,491 are 1 (everyone in that particular metro area for that ending age stayed in the metro area. The remaining 3,574 cell means are between zero and one.

For the beginning age 14, I end up with 5,176 cell means for the proportion staying in the same metro area for a particular metro area/ending age combination. The 5,176 cell means come from 260 different metro areas. The 5,176 cell means represent 100,051 “staying choices” between ages 14 and ending ages from 15 to 45; 804 cell means are zero, 1,205 are one, and the remaining 3,167 are between zero and one.

These cell-mean-calculated proportions use the PSID individual weights. The individual weights are meant to make the PSID sample representative of the U.S. population. Therefore, these cell means represent the “staying behavior” of the U.S. population.

This particular estimation strategy seeks gains from combining data from multiple “ending ages.” However, there can be too much of a good thing. Preliminary estimation suggested some problems with estimating the probability of staying as some polynomial function of the ending age. The model had a tendency to significantly underestimate or overestimate staying probabilities close to the beginning age, or close to the last ending age examined. Predicted staying probabilities were frequently inaccurate for the older ending ages. This is unfortunate. One of the goals of this paper is to look at long-run tendencies to be located in the metro area of one’s childhood.

To deal with this problem, the estimation for each beginning age was separated into five “ending age” segments. Each of these five ending-age segments was separately estimated, with coefficients allowed to vary freely across the five segments. This allows a better fit of predicted staying probabilities with actual probabilities for each ending-age segment.

I chose the ending-age segments to roughly match what we think of as different life stages. In addition, I tried to match up with what the figures in section 2 suggested about different "shapes" of the probability of staying in the same metro area over different ending ages.

For beginning age 4, the estimation was separately done for five ending-age segments. These five segments are 1) ending age 5–14, 2) ending age 15–17, 3) ending age 18–23, 4) ending age 24–29, and 5) ending age 30–35. For beginning age 14, the estimation was also done for five ending-age segments: 1) ending age 15–17, 2) ending age 18–23, 3) ending age 24–29, 4) ending age 30–35, and 5) ending age 36–45. The segments roughly correspond to "pre–high school" (ages 5–14), "high school" (ages 15–17), "college and early postcollege" (ages 18–23), "career beginning" (ages 24–29), "early 30s" (ages 30–35), and "midcareer" (ages 36–45). These segments logically correspond to periods during which we think behavior might be different. In addition, I chose some segments based on ensuring the segments from the beginning ages of 4 and 14 coincided.

For each of the five ending age segments, I estimated models of the following form:

(3) Proportion staying from beginning age $a0$ to ending age $a1$ in same metro area m = function of ending age $a1$, % college education in the metro area, % black in the metro area, average population size of the metro area, and average population growth rate of the metro area.

I will elaborate more on how these variables were defined below, and how the functional form was determined.

This equation was estimated for each of five segments, with a different range of ending ages. However, the beginning age is always the same.

The estimation was done using tobit. Tobit estimation is used because the dependent variable by its definition cannot be less than zero or more than one. Furthermore, in practice a considerable proportion of the observations were clumped at zero or one.

The estimation was done using weighted tobit. The weights used were the number of individuals used to calculate each cell mean. Obviously the estimated cell means for the propensity to stay are more precisely estimated when the sample size used to estimate this proportion is larger. This estimation procedure puts more weight on the observations that we are more confident are precisely estimated.[16]

Why these independent variables in the estimation? The ending age is used to allow for the staying propensity to change during each of the separately estimated segments. (Obviously this will be more important for segments with a wider range of ending ages, or if staying behavior is changing more over time during that segment.) Metro area population size and metro area population growth are included to control for two variables that plausibly affect the probability of staying in a community. These two metro characteristics were shown to matter empirically in the previous section.

[16]This weight is optimal if the only source of randomness in the model is measurement error in measuring the cell mean proportions. Because there are other sources of randomness, there might be somewhat better weights to optimally correct for the true heteroskedasticity. But these weights should be fairly good weights. In addition, the estimated standard errors are robust standard errors that correct for any remaining heteroskedasticity, of unknown form.

The percent college-educated or black in the metro areas controls in part for two individual determinants of the staying propensity. The previous section estimated that these two individual characteristics were the most important in affecting staying behavior. If aggregate behavior is just the sum of individual behavior, then these two variables should be important determinants of the aggregate average staying propensity for a particular metro area. In addition, percent college-educated or black in a metro area may have effects in and of themselves as determinants of staying behavior. These two variables might be correlated with and proxy for various amenities that may make a metro area more or less attractive.

How are these independent variables measured? The ending age is available from our processed version of the PSID data set. (A data appendix to this paper discusses some processing that was done with PSID-reported ages in this project.) Proportion college-educated and proportion black in the metro area were measured for each metro area using data from the 2005 American Community Survey. Proportion college-educated used all persons age 25 and over, and proportion black used all persons. The ACS was used because its metro area boundaries correspond exactly to the metro area boundaries used here. (Census 2000 metro area boundaries would not so correspond.) Metro area population was measured using average metro area population, from official census data, for each metro area over the entire 1970 to 2005 period of these data. Metro area population growth was measured as 100 times the average annual change in the natural logarithm of metro area population from 1970 to 2005.

There are 10 estimation samples. (Ten samples equal two possible beginning ages times five ending-age segments for each beginning age.) The functional form of each of these 10 estimation samples was built up from a simple to a complex specification.[17] First, for each segment, I looked at alternative specifications in ending age, with no other variables included. I considered alternatives ranging from no ending age included up to a fourth power in the ending age. For each segment, I picked the ending-age specification that minimized the Akaike Information Criterion. This approach ended up choosing specifications with no age terms in four of the ten segments, specifications with a linear age term in five segments, and a specification with a quadratic in age in one segment.

Then, I considered whether percentage college and percentage black should be added by themselves, or interacted with the age terms in the specification. I first did this for the proportion college-educated, then for the proportion black. The AIC criterion never preferred interacting proportion in college with the age terms, so none of these interactions were done. For proportion black, some of the specifications preferred interactions with the age terms. Therefore, I interacted proportion black in all segments with whatever age terms were present in that specification.

I wanted to allow as much flexibility as possible in how the staying propensity was affected by metropolitan area population and population growth. I represented these two variables by two sets of dummy variables. For example, for the metro population variable, the independent variables were five dummy variables for whether the metro area was in various quintiles as ranked by metro area population size. These metro population quintiles were population-weighted quintiles. That is, the top quintile represents the largest metropolitan areas that include 20 percent of the total population that lives in metro areas. Therefore, there are

[17] I considered the alternative to estimating an all-inclusive specification and paring it down. However, this approach, sometimes recommended by econometricians, would have been much too complex and unwieldy.

fewer metro areas in the top quintile than in the bottom quintile, but the same number of people is estimated to be in each quintile.

The metro growth variable was represented by a set of five dummy variables that are based on classifying each metro area into one of five population growth quintiles. These quintiles are also metro population–weighted quintiles. One-fifth of the total metro population lives in each quintile of metro area growth.

It is of course impossible to include all five of any set of quintile variables without a constraint. STATA does not allow tobit to be estimated with linear constraints, so I dropped the middle quintile for the actual estimation. However, the results as reported represent the coefficient on each quintile (including zero for the middle quintile) minus the average of all quintile coefficients. This coefficient represents the effects of each quintile dummy relative to the population-weighted average across all quintiles.

I explored whether the set of quintile population dummies, and growth dummies, should be interacted with the age terms. The Akaike Information criterion suggest such an interaction is optimal in only one case, for the quintile population dummies for one of the segments.

Finally, in the estimation I corrected the estimated standard errors for the correlation of the disturbance terms within the same metropolitan area.[18] This standard error correction is important because most of the independent variables only vary across metro areas. Work by Moulton (1990) and others have shown that misleading inferences are more likely if the clustering of the disturbance terms is not corrected for and happens to coincide with the clustering of an independent variable.

[18]The estimation did not use the correlation to improve upon the usual Tobit estimates, but to make sure the Tobit estimates have the correct standard errors.

Table 5 shows some descriptive statistics for the proportion college-educated and proportion black variables. These statistics are metro population–weighted. As these statistics show, there is a much wider range across metro areas in percentage black than there is in percentage college-educated.

Table 6 shows some statistics on the quintiles used for metro area population size and growth. There is a large amount of variation across quintiles. The lowest population quintile is metro areas with fewer than 333,000 people. The largest population quintile is metro areas with more than 2.6 million people. The lowest population growth quintile is metro areas with annual population growth rates of less than 0.22 percent. The highest population growth quintile is metro areas with population growth rates of more than 1.15 percent annually. In addition, all the quintiles have a fair number of cells for the estimation.

Table 7 presents the tobit estimates. These tobit coefficient estimates show the effects of a one-unit change in the independent variable on the tobit index. This will be close to the effect on the actual proportion of persons who stay in the metro area when the average probability is close to 0.5. This tends to be more true for the older ending ages. For probabilities closer to one, the actual effects on the probability of staying will be attenuated from the tobit coefficients. Later on, I will present simulations of what these tobit coefficients imply for effects on actual staying probabilities.

In analyzing these tobit estimates, I initially focus on the estimates that are significantly different from zero. The age terms, as one would expect, are negative. Based on the AIC, these age terms do not need to be included when the ending-age estimation segment considered encompasses a narrow range of ages or for older ages. This reflects that the staying probability does not vary much with ending age within these segments.

The proportion of the metro area that is college-educated tends to have negative effects. However, these effects are never statistically significant. The percentage black occasionally does have significant effects. These effects appear to be positive (that is, increasing staying) for later ending ages for the beginning age 14 specification.

For the population and growth quintile estimates, all estimates are effects relative to the average effects for all five quintiles. For example, the effects for the lowest population metro areas, quintile 1, are relative to the average effect for all metro areas. The *t*-statistics in parentheses are *t*-statistics for testing whether the quintile's coefficient is significantly different from the average of all five quintiles. In brackets below the quintile 5 estimates, I report the *t*-statistic for the difference between the quintile 5 effect and the quintile 1 effect.

The most consistent effects of metro area population size is that the smallest metro areas, those in quintile 1, have a significantly lower proportion staying in the metro area at later ending ages. This negative effect for the lowest population quintile increases with later ending ages. (In fact, for the beginning age 14 estimation, the lowest population quintiles actually have somewhat greater proportions of the sample staying in the same metro area for ending ages 15–17 and 18–23. The lowest population quintile only begins to significantly lose residents as they reach their later 20s and older ages.)

The remaining population quintile effects do not show a consistent pattern. Effects sometimes go down and sometimes go up as we go to quintiles with larger population. These patterns could reflect a number of unobserved metro area characteristics that are correlated with population size.

In looking at the difference between the largest metro areas (greater than 2.5 million population) and the smallest metro areas (less than 0.333 million population), these also are consistently positive and significant for the older ending ages.

For population growth, effects generally are not as statistically significant. The lowest population growth quintile, quintile 1, with less than 0.2 percent average annual population growth, tends to have a lower proportion of its residents staying in the same metropolitan area for later ending ages. These differential effects for the lowest population growth metro areas are sometimes statistically significant.

The remaining population growth quintiles do not show a consistent pattern. Effects tend to be statistically insignificant. Staying proportions sometimes go up and sometimes go down with larger population. This is particularly true for the beginning age 4 estimation. In the beginning age 14 estimation, the staying proportion goes up fairly regularly for later ending ages as the population size goes up from quintile to quintile.

For beginning year 14 and the last two groups of ending years (ages 30–35, ages 36–45), the fastest growth quintile (more than 2.1 percent annual growth) has a statistically significant higher staying probability than the lowest growth quintile (less than 0.2 percent annual growth).

These tobit coefficients are used in simulations to show how they translate into effects on staying probabilities. For the population and population growth quintiles, I simulate effects at all possible 25 combinations of these quintile effects. For these simulations, proportion college-educated and proportion black are held constant at their population-weighted median level for all metro areas. For proportion black, I simulate effects at five levels of proportion college-educated: the population-weighted 20th percentile, 40th percentile, 50th percentile (median), 60th percentile, and 80th percentile. These simulations assume a metro area in the middle population-size quintile and the middle population-growth quintile and at the median proportion college-educated. (I also did this for the proportion college-educated, but I do not report these results because none of the underlying coefficient estimates are statistically significant.) All these simulations are done at all ending ages from 5 to 35 for beginning age 4, and ending ages 15 to 45 for beginning age 14.

An Excel spreadsheet available from the author includes all these calculated staying probabilities for these various simulations. In this paper, I summarize the key information in figures.

Figures 15 and 16 show how the staying probability varies at different ages for different population quintiles. These figures are for a metro area in the median population growth quintile, and with median values of proportion college-educated and proportion black.

As the figure shows, the clearest pattern is that for older ages, the lowest population quintile (metro population less than 333,000) has a significantly lower staying probability. The pattern for other population-size quintiles is not consistent.

For a beginning age of 4, and an ending age of 30–35, the staying percentage for the lowest population quintile (less than 333,000 in population) is 20 percentage points less than that for the highest population quintile (more than 2.6 million in population). The percentage staying for the smallest metro areas is 38.7 percent versus 59.0 percent for the largest metro areas. For a beginning age of 14 and an ending age of 36 to 45, the staying percentage for the smallest metro areas is 16 percentage points less than that for the biggest metro areas. The percentage staying is 39.5 percent for the smallest metro areas, versus 55.6 percent for the largest metro areas.

These differences by population size are somewhat greater than that suggested by the simple cross-sectional comparisons in section 3. The comparisons in section 3 did not control for other features of metro areas. The comparisons here do.

However, even for the smallest population metro areas, for typical values of growth and other variables, the percentage staying in the same metro areas is probably in the 30 to 45 percent range for most of these persons' working careers. This statement extrapolates the staying percentages at ages in the 30s and 40s and into later ages. This is a plausible inference based on the PUMS data as discussed previously.

Figures 17 and 18 present staying probabilities at different ending ages for different quintiles of metro area population growth. Metro area average population is for the middle population quintile (in the range from 0.9 million to 1.8 million). The metro area proportion college-educated and proportion black are held constant at their median levels.

The most consistent pattern in these figures is the difference in staying proportions at later ending ages between the slowest-growing metro areas and the fastest-growing metro areas. The pattern on the in-between-growth metro areas is not consistent.

For a beginning age of 4 and an ending age of 30–35, the staying percentage for the lowest-growth metro areas (average annual growth less than 0.2 percent) is nine percentage points less than the staying percentage for the highest growth metro areas (average annual growth more than 2.2 percent). The lowest-growth metro areas have a staying percentage of 55.6 percent versus 64.9 percent for the highest-growth metro areas. For a beginning age of 14 and an ending age of 36–45, the staying percentage for the lowest-growth metro areas is 20 percentage points less than that for the highest-growth metro areas. The lowest growth areas have a staying percentage of 51.5 percent versus 71.8 percent for the highest-growth areas.

Suppose we consider metro areas with the lowest population and population growth. These areas would have metro populations less than 333,000 and average annual population growth of less than 0.2 percent. Even for these metro areas, the staying percentage for the beginning age of 4 and ending age of 30–35 is 37.1 percent. For a beginning age of 14 and an ending age of 36–45, the staying percentage for these small and slow-growth metro areas is 36.4 percent. Even if a metro area is small and slow-growing, at least 30 percent of its childhood residents stay in that metro area for most of their working careers.

Figures 19 and 20 show the difference in the staying percentage at different proportions of black population in the metro area. These simulations hold other variables constant at median values. There are little differences with metro race composition up to ending ages in the 20s. For older ending ages, there are some modest effects across the range of proportion black in different metro areas. These modest effects are larger for the age 14 starting age.

At the starting age of 4 and ending age of 30–35, the difference in staying percentage between metro areas at the 20th percentile of proportion black (0.048) and the 80th percentile (0.210) is 1.9 percentage points. The percentage staying for the lowest proportion black metro areas is 56.9 percent, versus 58.8 percent for the highest proportion black metro areas. At an age 14 starting age and ending age of 36–45, the difference in staying percentage between the lowest and highest proportion black metro areas is 5.2 percentage points. In the lowest proportion black metro areas, the percentage staying is 52.7 percent, versus 57.9 percent for the highest proportion black metro areas.

These differences across metro areas are somewhat larger for the age 14 starting age than one would expect based on the results in section 4. Using individual data in section 4, I estimated that holding all else constant, the staying percentage was 64.0 percent for whites and 80.5 percent for blacks. A switch from a metro area population whose proportion black is 0.048 to a metro area whose proportion black is 0.210 would be predicted to reduce the staying percentage by 2.7 percentage points (= [80.5 − 64.0] × [0.210 − 0.048]). The actual difference that is observed for the beginning age of 14 and ending age of 30–35 in this aggregate tobit estimation is 3.5 percentage points. (The simulated staying percentage is 66.3 percent for metro areas at the 20th percentile, versus 62.8 percent for metro areas at the 80th percentile.)

MORE ON THE SMALLEST METRO AREAS

Based on the previous section, the staying percentage is significantly lower in the smallest quintile of metro areas. These are metro areas with populations of less than 333,000. A natural question is whether the staying percentage gets even smaller for the smallest metro areas within this category.

To look at this, I redid all the tobit estimates of section 5, but only for metro areas of less than 333,000. I include dummies for the four "5 percentile" ranges in this lowest population quintile. These are as follows: 1) the smallest 5 percent of all metro areas (less than 123,000), 2) the 5th through 10th percentiles (123,000 to 171,000), 3) the 10th through 15th percentiles (171,000 to 271,000), and 4) the 15th to 20th percentiles (271,000 to 333,000).

An appendix shows all the tobit-estimated effects. The estimation reveals that there are no statistically significant differences across the four population groups within this lowest population quintile. Furthermore, there is no sign of any substantively important tendency for the very smallest metro areas to have a lower staying percentage. However, particularly for the age 14 beginning age, there are statistically significant differences in staying propensity across different growth rates.

Figures 21 and 22 show how the staying percentage varies across these five percentile metro population groups. (A spread sheet available from the author shows the actual numbers.) From this figure, it is apparent that there is not a clear systematic relationship between metro population size and the staying probability within these small-population metro areas. For example, for the starting age of 4, the staying percentage from ages 30–35 is 63.0 percent for the smallest metro areas (smaller than 123,000) versus 60.3 for the largest small-metro areas (271,000 to 333,000). For the starting age of 14, the staying percentage for ages 36–45 is 51.2 percent for metro areas smaller than 123,000 in population, and 53.9 percent for metro areas in the 271,000-to-333,000 range. In fact, metro areas in a "mid-small" range of 171,000 to 271,000 on average have the smaller staying percentage. Perhaps as metro areas get smaller, at some point social capital and other amenities may increase the staying percentage somewhat.

Figures 23 and 24 show how the staying percentage varies across the five growth quintiles for these smaller metro areas. For these smaller metro areas, variation in average metro-area growth continues to show a strong effect. For these small metro areas, for a beginning age of 4 and an ending age of 30–35, the staying percentage for the lowest-growth metro areas (less than 0.2 percent growth) is only 31.8 percent, versus a staying percentage of

49.6 percent for the highest-growth metro areas (greater than 2.2 percent annual growth). For a beginning age of 14 and an ending age of 36–45, the staying percentage for the lowest-growth metro areas is 32.2 percent, versus 59.4 percent for the highest-growth metro areas.

These differentials in staying propensity with growth are greater for the smallest population quintile than they are for the entire sample. For example, the differential between the fastest-growth and slowest-growth metro areas, for a beginning age of 14 and an ending age of 36–45, is 27.2 percent for these smallest metro areas. The differential is 20.3 percent for the sample of all metro areas.

CONCLUSION

The headline from this paper is that a surprisingly high proportion of Americans stay for most of their worklife in their childhood state and metro area. It is commonly believed that Americans are extremely mobile and are becoming increasingly so (Wolf and Longino 2005). It is understandable that social scientists at top-ranked research universities may believe that Americans are extremely mobile: professors at top-ranked research universities are an extremely mobile group. However, most Americans are far less mobile than top professors. Close to two-thirds of all Americans spend most of their careers in the U.S. state of their childhood. About half of all Americans spend most of their careers in their childhood metro area.

Mobility goes up for some demographic groups. For example, Americans with a 4-year college degree or higher are more likely to end up their career outside their childhood metro area or state. However, even among Americans with a college degree, perhaps 40 percent spend the bulk of their career in their childhood metro area.

Adult outmigration from childhood metro areas is greater for metro areas that have smaller populations or are slower growing. However, only the smallest and slowest growing metro areas show large differences in the percentage of children in the metro area who remain there as adults. But even among the smallest and slowest growing metro areas, around 30 percent of persons who grew up in that metro area will spend most of their working career in that metro area.

These findings are relevant to state and local decisions about investing in children. A significant proportion of the children who receive the benefits from state and local investments in early childhood education or K–12 education will live in that state and local area as adults. If these investments in children result in improving their later adult outcomes, some of these benefits will accrue to the state or local area making the investments.

REFERENCES

Bartik, Timothy J. 2008. *The Economic Development Effects of Early Childhood Programs*. Issue Paper No. 6. Washington, DC: Partnership for America's Economic Success. http://www.partnershipforsuccess.org/uploads/20080723_Bartikformatted.pdf (accessed February 23, 2009).

———. 2006. *Taking Preschool Education Seriously as an Economic Development Program: Effects on Jobs and Earnings of State Residents Compared to Traditional Economic Development Programs.* Washington, DC: Committee for Economic Development. http://www.ced.org/images/library/reports/education/early_education/report_prek_bartik_2006technical.pdf (accessed February 23, 2009).

Inter-university Consortium for Political and Social Research. 1996. *Third ICPSR Release of Census of Population and Housing, 1990: Public Use Microdata Sample: 5-Percent Sample*. ICPSR 9952. Ann Arbor, MI: ICPSR.

Lemann, Nicholas. 1991. *The Promised Land: The Great Black Migration and How It Changed America*. New York: Alfred A. Knopf.

Moulton, Brent R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics* 72(2): 334–338.

Rosenbloom, Joshua L., and William A. Sundstrom. 2003. "The Decline and Rise of Interstate Migration in the United States: Evidence from the IPUMS, 1850–1990." NBER Working Paper 9857. Cambridge, MA: National Bureau of Economic Research.

Wolf, Douglas A., and Charles F. Longino Jr. 2005. "Our 'Increasingly Mobile Society'? The Curious Persistence of a False Belief." *Gerontologist* 45(1): 5–11.

Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Table 1 Descriptive Statistics for Variables Used in Three Logit Estimations of Probability of Staying in the Same Metropolitan Area

Variable	Ages 4 to 30 sample		Ages 14 to 30 sample		Ages 14 to 40 sample	
	Weighted mean	Unweighted mean	Weighted mean	Unweighted mean	Weighted mean	Unweighted mean
Stayed in same metro area	0.552	0.627	0.641	0.724	0.555	0.65
Female	0.541	0.582	0.531	0.55	0.541	0.571
High school dropout	0.088	0.104	0.092	0.136	0.084	0.129
High school graduate	0.335	0.38	0.359	0.402	0.366	0.407
Some college	0.266	0.275	0.27	0.268	0.283	0.274
Bachelor's degree	0.182	0.138	0.172	0.12	0.15	0.104
Postcollege degree	0.109	0.079	0.1	0.064	0.118	0.085
Missing education variable	0.02	0.024	0.007	0.009	0	0.001
White	0.804	0.568	0.776	0.495	0.758	0.488
Black	0.128	0.385	0.157	0.455	0.17	0.469
Hispanic	0.056	0.035	0.053	0.038	0.044	0.026
Other race	0.013	0.013	0.014	0.012	0.027	0.017
In initially low income sample	0.114	0.376	0.226	0.546	0.141	0.461
ln(metro area population as of age 4 or 14)	13.665	13.72	13.816	13.913	13.763	13.87
Original metro area average annual population growth between earlier age and later age	1.01	1.08	0.88	1.0	0.88	0.97

NOTE: The three samples refer to the ages over which whether the person stays in the same metropolitan area is measured: from age 4 to age 30; from age 14 to age 30; or from age 14 to age 40. The sample sizes for the three samples are: 710 observations, 2,853 observations, and 862 observations. (The actual estimation for the age 14 to age 40 drops one observation because the "missing education" variable perfectly predicts that one observation for which the educational variable is missing.) Unweighted means are the simple means for these samples of each variable. Weighted means use the PSID person weights. These PSID person weights correct for the oversampling of initially low-income persons because a big part of the initial PSID sample was a special sample of low-income persons, as well as for sample attrition and sample stratification. The education variables are based on the person's final educational attainment as measured in the PSID. If no educational attainment was measured at or after age 23, final educational attainment was treated as missing, which only affects a few observations in these samples. For the face variables, Hispanic overrides other racial categories, so white should be interpreted as "white non-Hispanic." The low-income sample variable measures whether this person comes from a family that originally was in the low-income sample portion of the PSID. Population and population growth are based on the metro area the person lived in as of the "base age" (ages 4 or 14). Metro area population is the population as of the year that person attained that base age in that metro area. Average annual metro area growth rate is 100 times the change in the natural logarithm of metro area population between the base age and the final age, divided by the number of years between the two ages. The standard deviation of most of the variables is derivable from their unweighted means. The standard deviation of ln(metro area base population) for the three samples is 1.241, 1.164, and 1.180. The standard deviation of average metro-area growth rate is 0.99 for the first sample, 1.02 for the second sample, and 0.90 for the third sample. The estimation also includes a constant and year dummies for the "end year." The end years that are possible for the first and third sample are 1996 and 1997, and then alternate years from 1999 to 2005. (The PSID switched from an annual sample to an every-other-year sample in 1997.) The end years possible for the second sample are 1986 through 1997 and alternate years from 1999 to 2005. The possible end years are dictated because geographic location is only reliably coded in a timely way in the PSID starting in 1970, which means that the possible end years start only "*x*" years later, where *x* is 26 years for the first and third samples and 16 years for the second sample.

Table 2 Logit Estimates for the Propensity of Individuals to Stay in the Same Metropolitan Area as Adults that They Lived in as a Child

	Logit Coefficient (*z*-statistic is in parentheses below estimated coefficient)		
	Results from age sample:		
Independent variable	4 to 30	14 to 30	14 to 40
Female	-0.0865 [-0.0215] (-0.51)	0.158 [0.036] (1.74)	0.293 [0.071] (1.89)
High school graduate	-0.460 [-0.114] (-1.41)	0.125 [0.028] (0.80)	0.0199 [0.0049] (0.08)
Some college	-0.653 [-0.161] (-1.94)	-0.343* [-0.082] (-2.13)	-0.366 [-0.091] (-1.39)
Bachelor's degree	-1.388* [-0.317] (-3.80)	-1.140* [-0.278] (-6.27)	-0.946* [-0.229] (-2.92)
Postgraduate degree	-1.195* [-0.280] (-2.97)	-1.295* [-0.313] (-6.28)	-1.204* [-0.283] (-3.65)
Missing education variable	-0.374 [-0.093] (-0.63)	-0.351 [-0.084] (-0.74)	
Black	1.080* [0.232] (3.77)	0.898* [0.173] (7.24)	0.500* [0.118] (2.03)
Hispanic	0.184 [0.045] (0.44)	0.347 [0.075] (1.48)	0.140 [0.034] (0.29)
Other race	-0.742 [-0.182] (-1.01)	-0.141 [-0.033] (-0.39)	-0.291 [-0.073] (-0.60)
1968 low-income family member	-0.170 [-0.042] (-0.59)	-0.0066 [-0.0015] (-0.06)	0.433 [0.103] (1.77)
ln(metro area population)	0.124 [0.030] (1.73)	1.929* [0.025] (incl square effects) (2.47)	0.133 [0.033] (1.94)
Square of ln(metro area population)		-0.0635* [see above] (-2.24)	
Metro area average annual population growth from ages *a*0 to *a*1	0.128 [0.031] (1.47)	0.166 [-0.041] (incl square effects) (1.91)	0.047 [0.012] (0.54)
Square of metro area population growth		-0.0616* [see above] (-1.98)	
Weighted sample mean probability of staying (from Table 1)	0.552	0.641	0.555

Table 2 (continued)

NOTE: These estimates are from three separate logit estimations on three different samples. Dependent variable is zero-one indicator for whether that individual at age $a1$ was living in the metro area he or she lived in at an earlier age $a0$. A "1" indicates the individual stayed, "0" that he or she moved to somewhere different than that metro area. The remaining variables are control variables and are described in the text more fully: dummy variables for gender, educational attainment, race, and whether the individual is part of the sample drawn from families that were low-income in 1968; and continuous variables describing the population and population growth of the original metro area that the individual lived in at age $a0$. The logit estimation also included year dummies and a constant term. The omitted dummy variable categories are three: 1)male, 2) high school dropout, and 3) white non-Hispanic. In brackets below the estimated logit coefficients are the estimated effects of a "1-unit" change in each independent variable. This is computed at the sample-weighted mean probability, and (in the quadratic specification used for the population terms in the ages 14 to 30 sample) at the sample-weighted mean for ln(metro pop.) and metro pop. growth. Given that dummies are omitted, the change of 1 unit for the gender, education, race, and low-income family sample dummies represents a change from the omitted category (white male high school dropout that is not part of the low-income family sample) to a value of 1 for that particular dummy. As is expected from the literature, the effect of a 1-unit change in these variables is about one-fourth of the logit coefficients. Given the units in which metro area population is calculated, the 1-unit change for that variable is a 1-unit change in the natural log, or a 172 percent increase in metro area population. A 1-unit change for the population growth rate term represents a 1-percent change in the log growth rate. The effects of these discrete changes for the quadratic specification are calculated using both the linear and squared terms at the weighted sample means for ln(pop) and the growth rate. *Z*-statistics (ratios to standard errors) are included in parentheses below the estimated coefficients. All standard errors that are the basis for these *z*-statistic calculations are based on robust variance estimation methods which should be robust to heteroskedasticity of unknown form. * indicates statistical significance at the 5% level using a two-tailed asymptotic test. Sample size for estimation is 710 for the ages 4 to 30 sample, 2,853 for the ages 14 to 30 sample, and 861 for the ages 14 to 40 sample. At the probabilities estimated here, the effect of a 1-unit change in any right-hand side variable is equal to between 20 and 25 percent of the estimated coefficient.

Table 3 Probability of Being in the Same Metro Area at Age 14 and Age 30, as a Function of Metro Area Population at Age 14

	Percentile of metro area pop distribution						
	5th	10th	25th	50th	75th	90th	95th
Metro area pop (in millions)	0.118	0.170	0.425	1.083	2.266	6.791	10.100
Probability of staying	0.527	0.561	0.629	0.672	0.689	0.689	0.682

NOTE: See Figure 11 for a graphic depiction of these results. All estimated probabilities are from the logit 14–30 specification with quadratic specification in both metro area population and metro area population growth. Estimated probabilities are weighted average probability for a population whose gender, education, and race composition matches the composition of the U.S. population ages 28–32 in 2007. Calculations assume median metro area population growth of 0.8 percent per year. The average logit year effects are used, and the effects for the non-low income sample are used. Percentiles of the metro area population are calculated using PSID weights for the logit 14–30 sample.

Table 4 Probability of Being in the Same Metro Area at Age 14 and Age 30, as a Function of Original Metro Area's Population Growth Between Ages 14 and 30

	Percentile of metro area pop growth distribution						
	5th	10th	25th	50th	75th	90th	95th
Metro area annual pop growth rate (%)	-0.44	-0.26	0.12	0.76	1.46	2.20	2.63
Probability of staying	0.636	0.644	0.658	0.672	0.676	0.667	0.656

NOTE: All estimated probabilities are from the logit 14–30 specification with quadratic specification in both metro area population and metro area population growth. Estimated probabilities are weighted-average probability for a population whose gender, education, and race composition matches the composition of the U.S. population ages 28–32 in 2007. Calculations assume median metro area population of 1.08 million. The average logit year effects are used, and the effects for the non-low income sample are used. Percentiles of metro area population growth are calculated using PSID weights for the logit 14–30 sample. Population growth is for the metro area lived in at age 14. What happens in any new metro area of residence as of age 30 is irrelevant.

Table 5 Descriptive Statistics for Metro Proportion College-Educated and Metro Proportion Black, All U.S. Metro Areas, Population-Weighted Statistics

	Percentile				
	20th	40	Median	60	80th
Proportion college-educated	0.237	0.273	0.296	0.310	0.338
Proportion black	0.048	0.089	0.118	0.148	0.210

NOTE: These data are calculated over all U.S. metropolitan areas from the American Community Survey for 2005. The calculations use population weights for each metropolitan area based on average metro-area population from 1970 to 2005. Proportion college-educated is proportion with a bachelor's or higher degree of those age 25 and above. Proportion black is based on all persons.

Table 6 Description of Metro Population Quintiles and Metro Population Growth Quintiles for Estimation Samples

	Lowest pop. or growth (≥)	Highest pop. or growth (<)	Cells in quintile: Age 4 beginning age	Cells in quintile: Age 14 beginning age	Metro areas in quintile: Age 4 beginning age	Metro areas in quintile: Age 14 beginning age
Pop. quintile 1 (lowest pop.)	26,000	333,000	3,133	2,120	162	137
Pop. quintile 2	333,000	907,000	1,650	1,459	73	67
Pop. quintile 3	907,000	1,798,000	821	808	30	30
Pop. quintile 4	1,798,000	2,578,000	547	541	18	18
Pop. quintile 5	2,578,000	10,500,000	248	248	8	8
Growth quintile 1 (lowest growth)	-0.83%	0.22%	965	800	41	36
Growth quintile 2	0.22%	0.78%	1,250	1,027	56	51
Growth quintile 3	0.78%	1.27%	1,292	1,067	59	54
Growth quintile 4	1.27%	2.15%	1,550	1,343	72	66
Growth quintile 5	2.15%	7.87%	1,342	939	63	53

NOTE: Quintiles of both population and population growth are defined based on average population of all 381 metro areas in United States, using 2005 definitions, from 1969 to 2005. Quintiles are defined based on population-weighted statistics. Metro areas on dividing line are classified in higher-population quintile. The cells used in estimation represent proportions for combinations of metro area and ending age that are in that estimation sample. Metro areas are distinct metro areas in that quintile for that estimation sample. I aggregate overall estimation samples with the same beginning age.

Table 7 Tobit Estimates of How Metro-Area Staying Probability Between Fixed Beginning Age and Various Ending Ages Is Affected by Various Ending Ages and Metro Characteristics

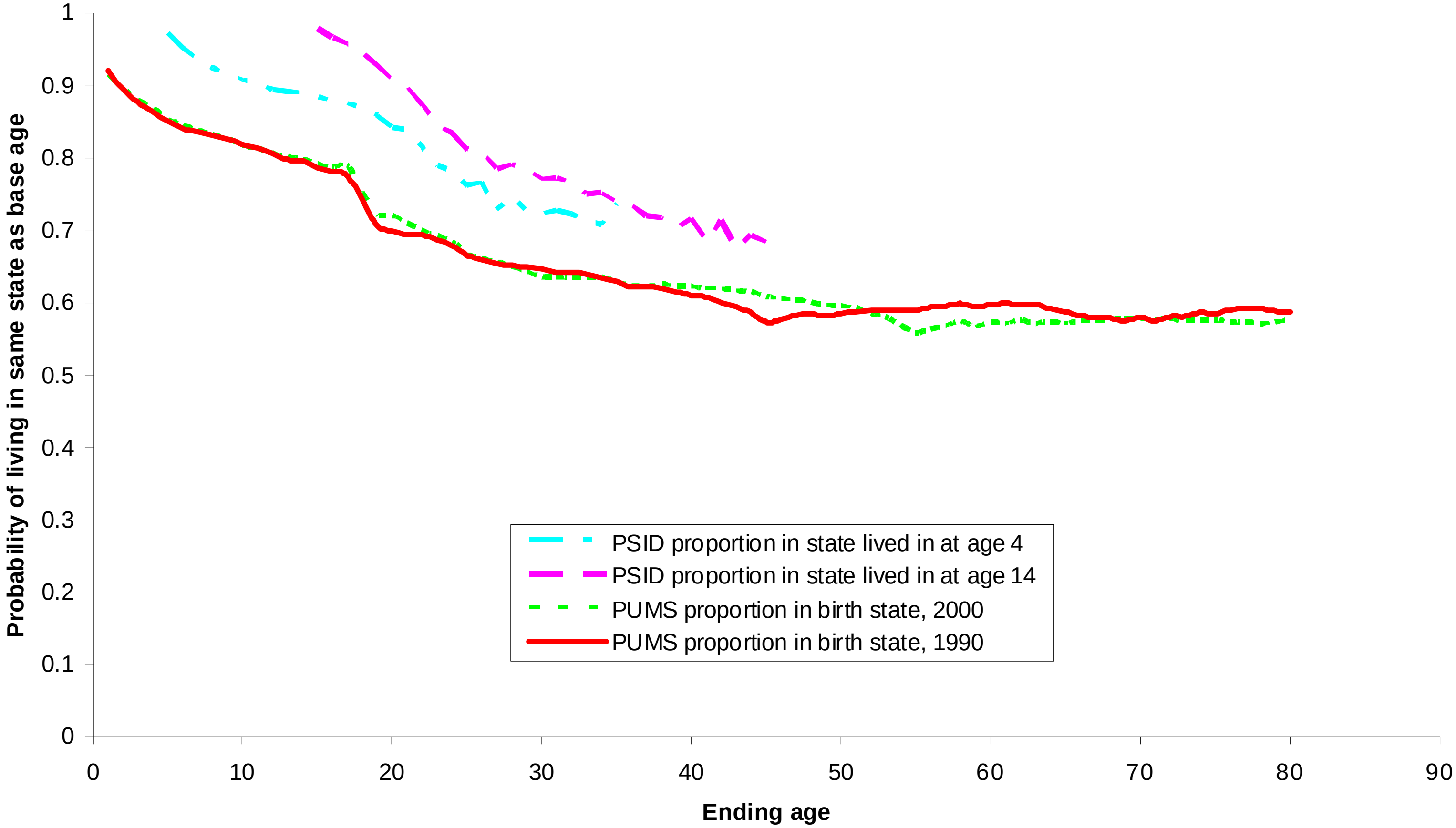
Beginning age	Age 4					Age 14				
Ending ages	5–14	15–17	18–23	24–29	30–35	15–17	18–23	24–29	30–35	36–45
Age	-0.0570*		-0.0266*	-0.0122*		-0.0334*	-0.0345*	-0.0174*		
	(-7.32)		(-5.45)	(-1.98)		(-5.75)	(-6.31)	(-5.01)		
Age^2	0.00195*									
	(5.28)									
Proportion college in metro area	-0.090	-0.158	-0.119	-0.268	-0.723	-0.097	0.019	-0.170	-0.383	-0.340
	(-0.56)	(-0.53)	(-0.38)	(-0.66)	(-1.35)	(-0.69)	(0.13)	(-0.72)	(-1.16)	(-0.81)
Proportion black in metro area	-0.322*	0.0354	-0.241	0.481	0.161	-0.910*	-0.460	-0.218	0.338*	0.366*
	(-2.24)	(0.31)	(-0.64)	(0.57)	(0.81)	(-2.59)	(-1.37)	(-0.51)	(2.39)	(2.10)
Age × proportion black	0.0416		0.0124	-0.0144		0.0553*	0.0258	0.0183		
	(1.29)		(0.77)	(-0.44)		(2.58)	(1.51)	(1.02)		
Age^2 × proportion black	-0.00111									
	(-0.73)									
Metro population in quintile 1 (lowest)	-0.0277	-0.0501	-0.1029*	-0.1786*	-0.1775*	0.0449*	0.3896*	-0.0905*	-0.1287*	-0.1751*
	(-1.28)	(-1.41)	(-2.62)	(-3.24)	(-2.17)	(2.42)	(3.71)	(-3.15)	(-3.22)	(-3.37)
Population quintile 2	0.0237	0.0206	0.0220	0.0038	-0.0186	0.0118	-0.0404	0.0090	0.0653*	0.1436*
	(1.72)	(0.80)	(0.70)	(0.11)	(-0.44)	(0.87)	(-0.48)	(0.34)	(1.99)	(3.80)
Population quintile 3	-0.0211	-0.0154	-0.0267	0.0261	0.0879	-0.0171	-0.1029	0.0143	0.0105	-0.0004
	(-1.48)	(-0.57)	(-0.83)	(0.58)	(1.81)	(-1.49)	(-1.12)	(0.53)	(0.30)	(-0.01)
Population quintile 4	0.0135	0.0215	0.0502*	0.0550	0.0003	-0.0155	-0.1094	0.0348	0.0248	0.0213
	(1.19)	(1.12)	(2.08)	(1.72)	(0.01)	(-1.53)	(-1.38)	(1.64)	(0.77)	(0.55)
Population quintile 5	0.0115	0.0234	0.0575	0.0937*	0.1078	-0.0241	-0.1370	0.0325	0.0281	0.0105
	(0.51)	(0.60)	(1.67)	(2.15)	(1.91)	(-1.69)	(-1.75)	(1.24)	(0.72)	(0.21)
	[1.00]	[1.09]	[2.49]	[3.20]	[2.29]	[-2.39]	[1.23]	[2.72]	[2.31]	[2.07]
Age × Population quintile 1							-0.0204*			
							(-3.89)			
Age × Population quintile 2							0.0025			
							(0.58)			
Age × Population quintile 3							0.0048			
							(1.00)			

Table 7 (continued)

Beginning age	Age 4					Age 14				
Ending ages	5–14	15–17	18–23	24–29	30–35	15–17	18–23	24–29	30–35	36–45
Age × Population quintile 4							0.0062 (1.52)			
Age × Population quintile 5							0.0070 (1.67)			
Metro population growth, quintile 1 (lowest)	0.0131 (1.11)	-0.0012 (-0.05)	-0.0126 (-0.46)	-0.0682* (-1.96)	-0.0945* (-2.15)	0.0061 (0.61)	0.0045 (0.30)	-0.0389 (-1.54)	-0.0617 (-1.78)	-0.0766 (-1.88)
Metro population growth quintile 2	0.0334* (2.22)	0.0497* (2.12)	0.0471 (1.65)	0.0620 (1.59)	0.0871 (1.71)	-0.0004 (-0.04)	0.0083 (0.67)	0.0161 (0.80)	-0.0070 (-0.25)	-0.0409 (-1.08)
Metro population growth quintile 3	-0.0065 (-0.32)	-0.0331 (-0.92)	-0.0013 (-0.04)	0.0051 (0.12)	-0.0709 (-1.45)	0.0163 (1.22)	0.0080 (0.69)	0.0256 (1.06)	0.0132 (0.36)	-0.0396 (-0.77)
Metro population growth quintile 4	0.0050 (0.36)	0.0109 (0.46)	0.0053 (0.19)	0.0060 (0.16)	0.0438 (0.95)	-0.0230* (-2.02)	-0.0159 (-1.09)	-0.0140 (-0.55)	-0.0036 (-0.12)	-0.0111 (-0.35)
Metro population growth quintile 5	-0.0450* (-2.17) [-2.53]	-0.0263 (-0.85) [-0.72]	-0.0386 (-1.28) [-0.61]	-0.0049 (0.12) [1.08]	0.0344 (0.56) [1.57]	0.0011 (0.10) [-0.34]	-0.0049 (-0.29) [-0.36]	0.0112 (0.43) [1.18]	0.0592 (1.53) [2.11]	0.1682* (3.29) [3.48]
F-test prob: Pop terms, Growth terms, Age terms, Black terms	0.1078, 0.0523, 0.0000, 0.0167	0.4287, 0.2506, NA, 0.7566	0.0252, 0.4460, 0.0000, 0.7207	0.0211, 0.2812, 0.0001, 0.7033	0.1001, 0.0605, NA, 0.3787	0.1268, 0.2502, 0.0000, 0.0353	0.0200, 0.7319, 0.0000, 0.2270	0.0322, 0.4145, 0.0000, 0.0306	0.0095, 0.3061, NA, 0.0172	0.0001, 0.0084, NA, 0.0361

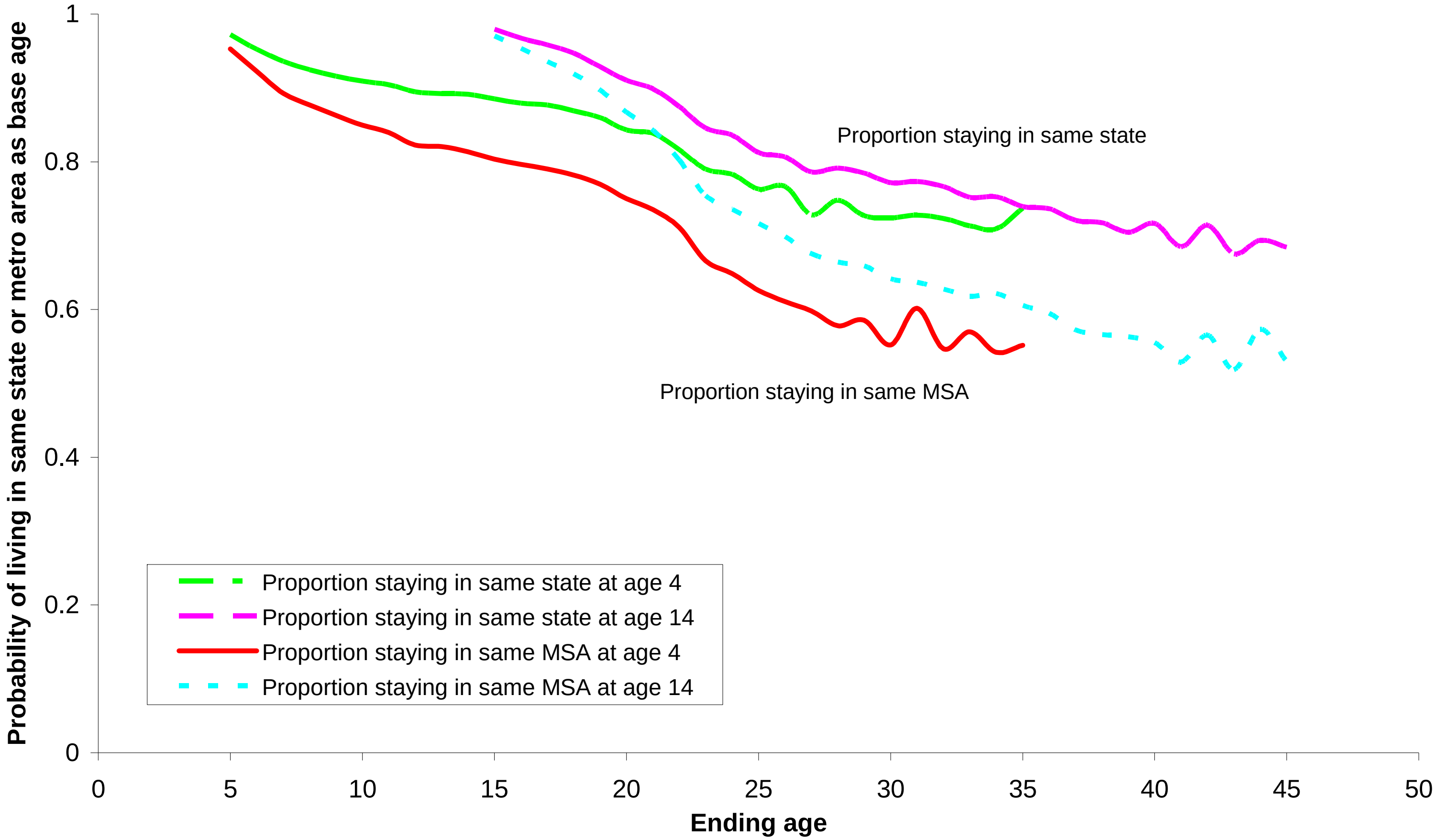
NOTE: Coefficients are Tobit coefficients. Quintile coefficients are effects of quintile vs. average of all quintiles. Numbers in parentheses below are *t*-statistics. Numbers in brackets below 5th quintile are *t*-statistics for difference between 5th quintile and 1st quintile. *F*-tests are probabilities for joint significance of variables in the indicated categories.

Figure 1
How Probability of Staying in Same State Location Varies with Age



Note: Estimated probability is proportion of individuals in same state at an ending age as they were at a base age. PUMS data is from 2000 and 1990 censuses. Estimated probability is proportion of individuals living in same state as birth state. PSID data is from pooled data from Geocode PSID data, 1970 to 2005. One line shows estimated proportion living in same state lived in at age 4. Other line shows estimated proportion living in same state lived in at age 14. Proportions pool data over all year pairs. All estimated proportions use appropriate individual weights in the two data sets.

Figure 2
Probability of Staying in Same Metro Area vs. Same State

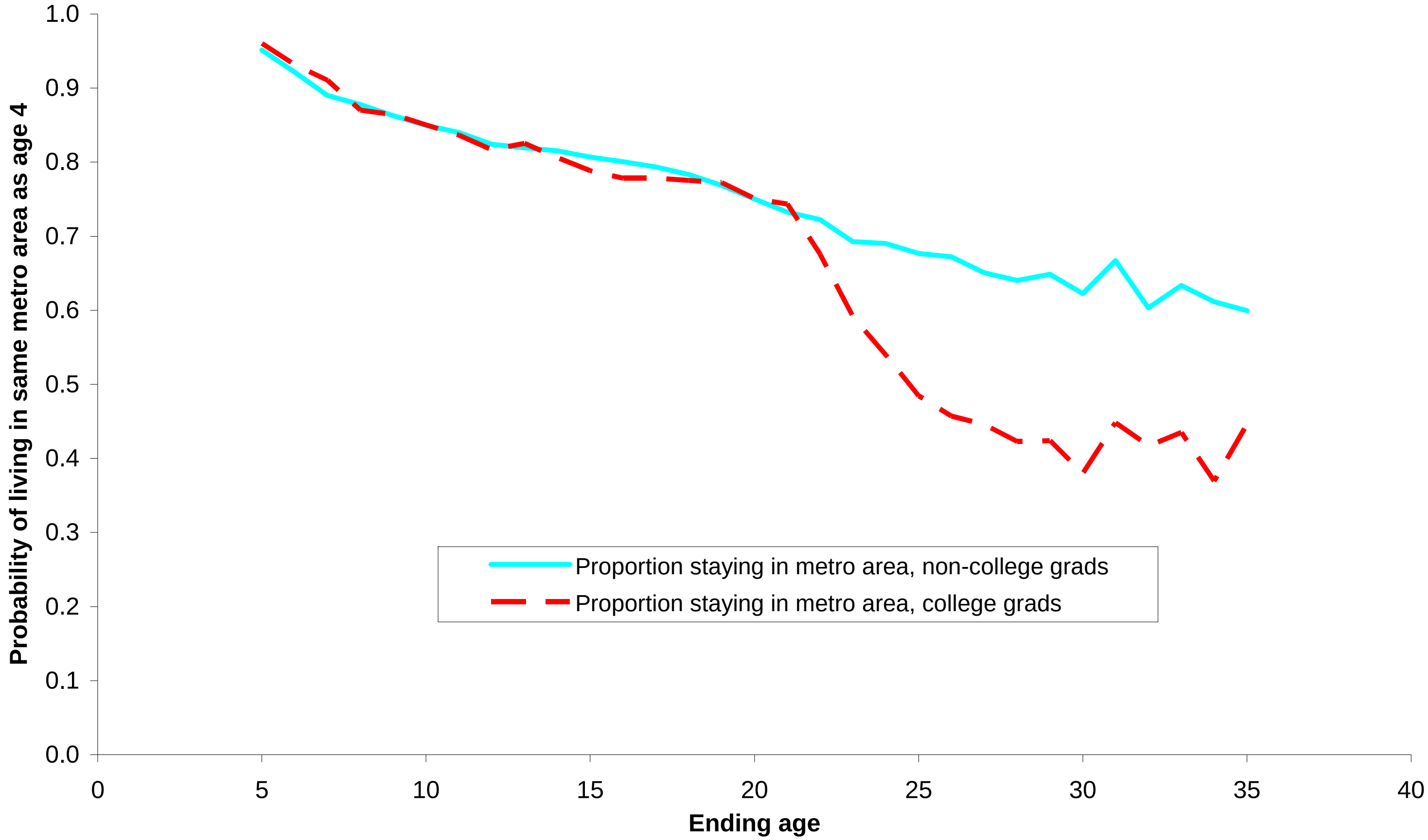


Note: Estimated probability is proportion of individuals in same state or same metro area at an ending age as they were at a base age. PSID data is from pooled data from Geocode PSID data, 1970 to 2005. Two lines show estimated proportions living in same state or same metro area as lived in at age 4. Another two lines shows estimated proportion living in same state or same metro areas as lived in at age 14. Proportions pool data over all year pairs. All estimated proportions use appropriate individual weights in the PSID. All metro area definitions for all year pairs are based on year 2008 metro area definitions.

Figure 3
Probability of Staying in Birth State, College Grads vs. Non-College Grads

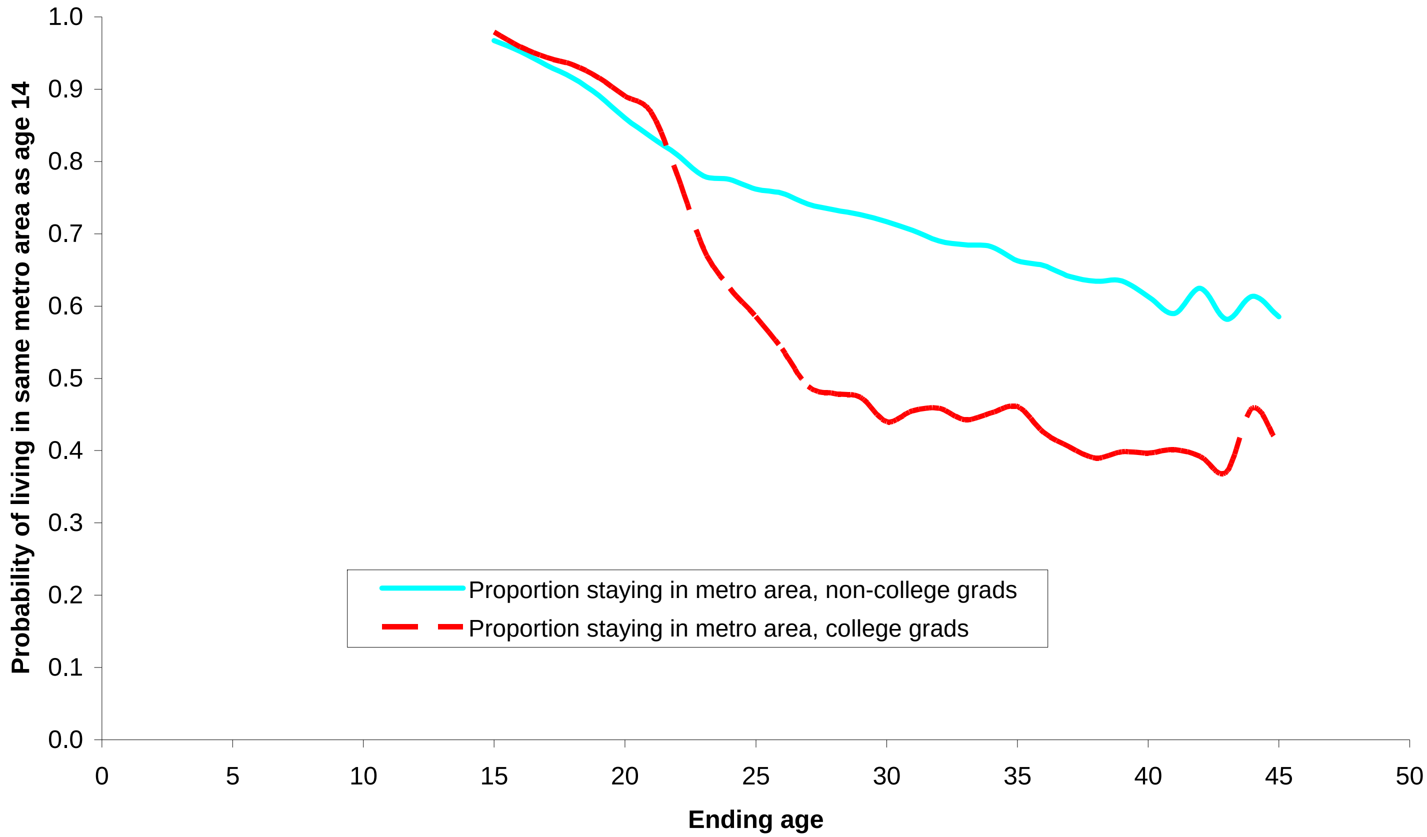
Note: Estimated probability is proportion of individuals living in birth state, for each of two education groups: 4-year-college degree or more vs. less than 4-year-college degree. The two groups together sum to entire population. All data are from 2000 PUMS and 1990 PUMS. All estimated proportions use appropriate individual weights in the PUMS.

Figure 4
Probability of Staying in Same Metro Area As Lived in at Age 4, College Grads vs. Non-College Grads



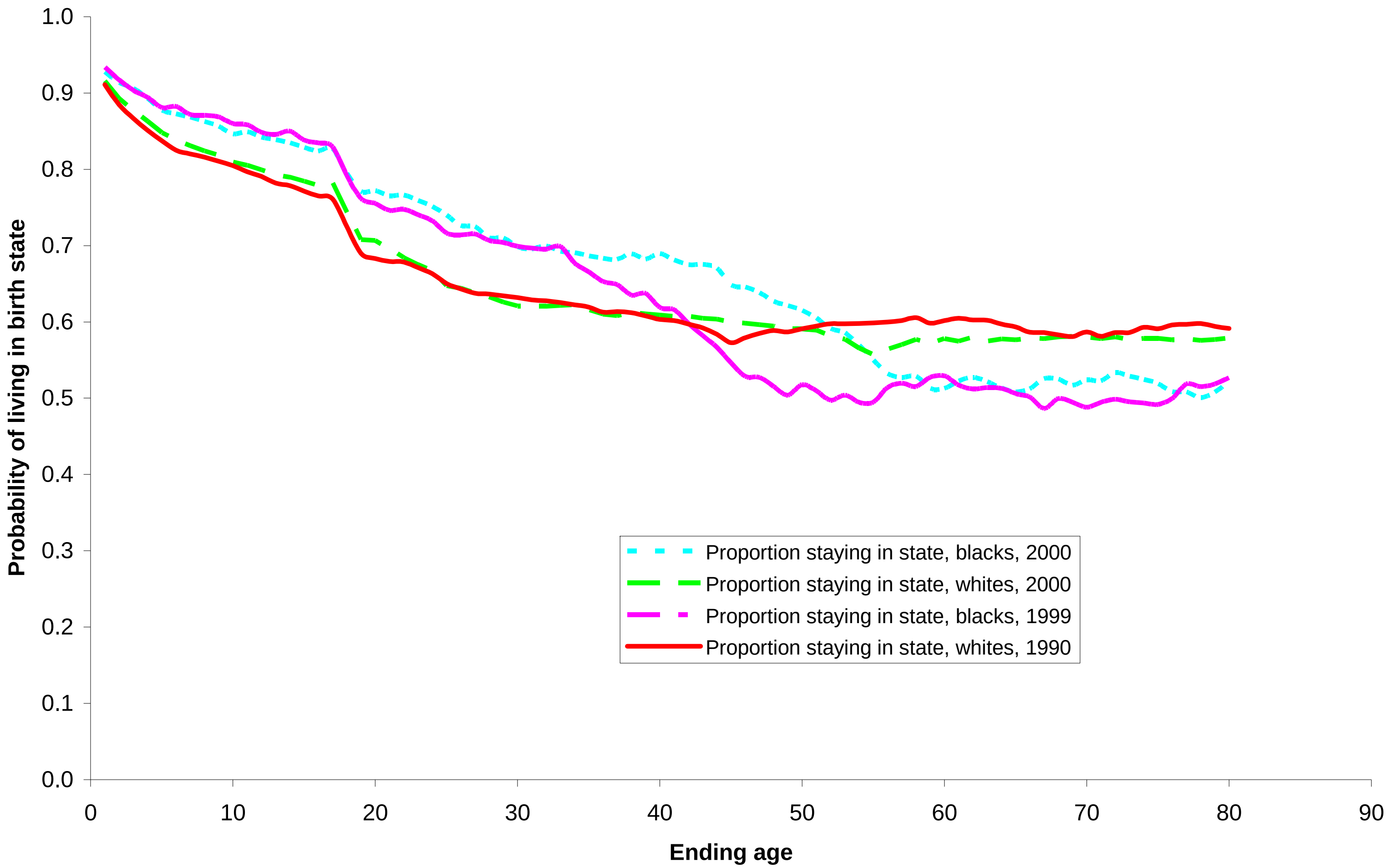
Note: Estimated probability is proportion of individuals living in same metro area as lived in at age 4, for each of two education groups: 4-year-college degree or more versus less than 4-year-college degree. The two groups together sum to entire population with observed final educational status. All data are from PSID. All estimated proportions use appropriate individual weights in the PSID. Education status is measured as of final observation on that individual in PSID, not as of ending age. Final observation on individual must be age 23 or above for final educational status to be treated as non-missing.

Figure 5
Probability of Staying in Same Metro Area As Lived in at Age 14, College Grads vs. Non-College Grads



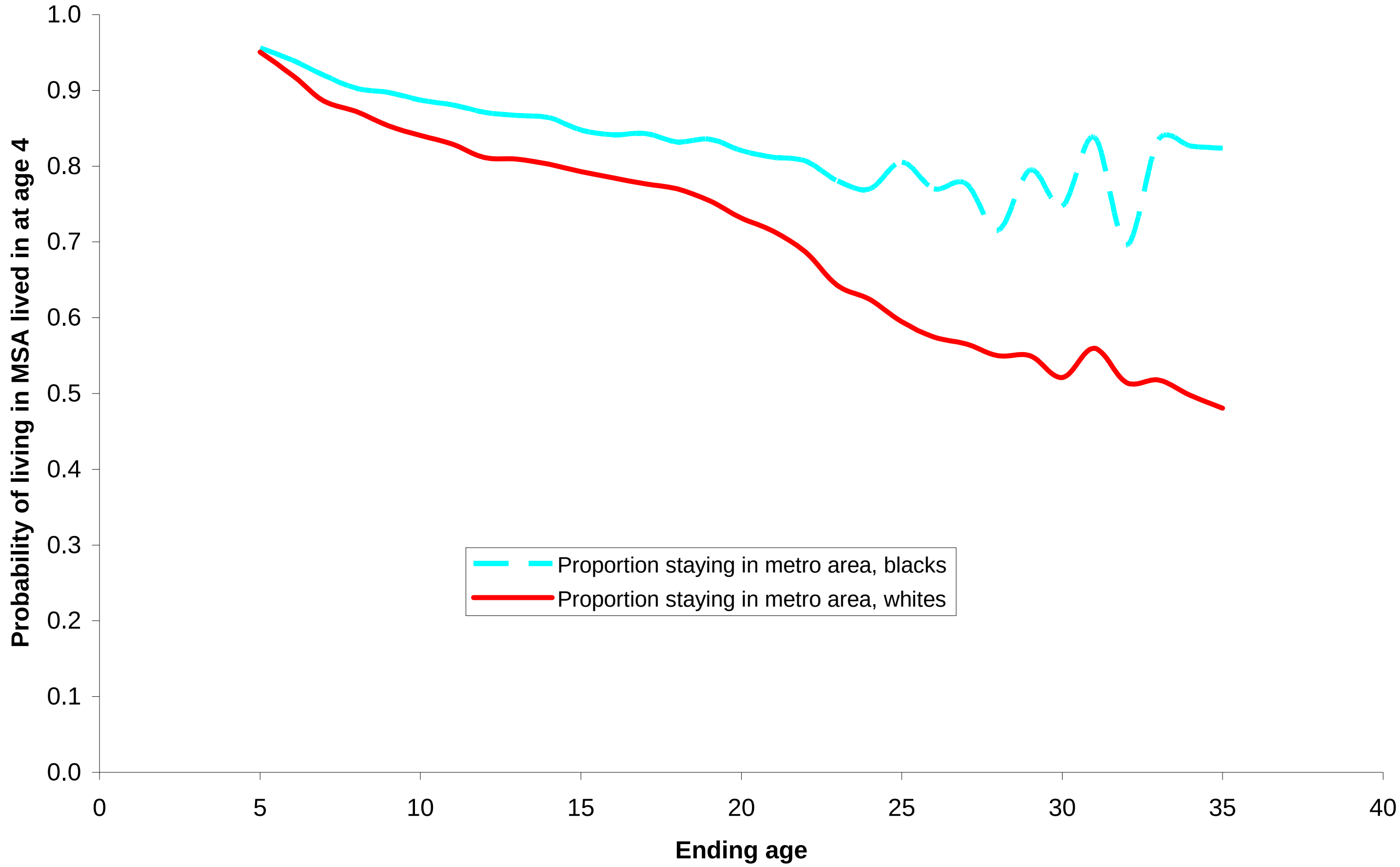
Note: Estimated probability is proportion of individuals living in same metro area as lived in at age 14, for each of two education groups: 4-year-college degree or more vs. less than 4-year-college degree. The two groups together sum to entire population with observed final educational status. All data are from PSID. All estimated proportions use appropriate individual weights in the PSID. Education status is measured as of final observation on that individual in PSID, not as of ending age. Final observation on individual must be age 23 or above for final educational status to be treated as non-missing.

Figure 6
Probability of Staying in Birth State, Blacks vs. Whites



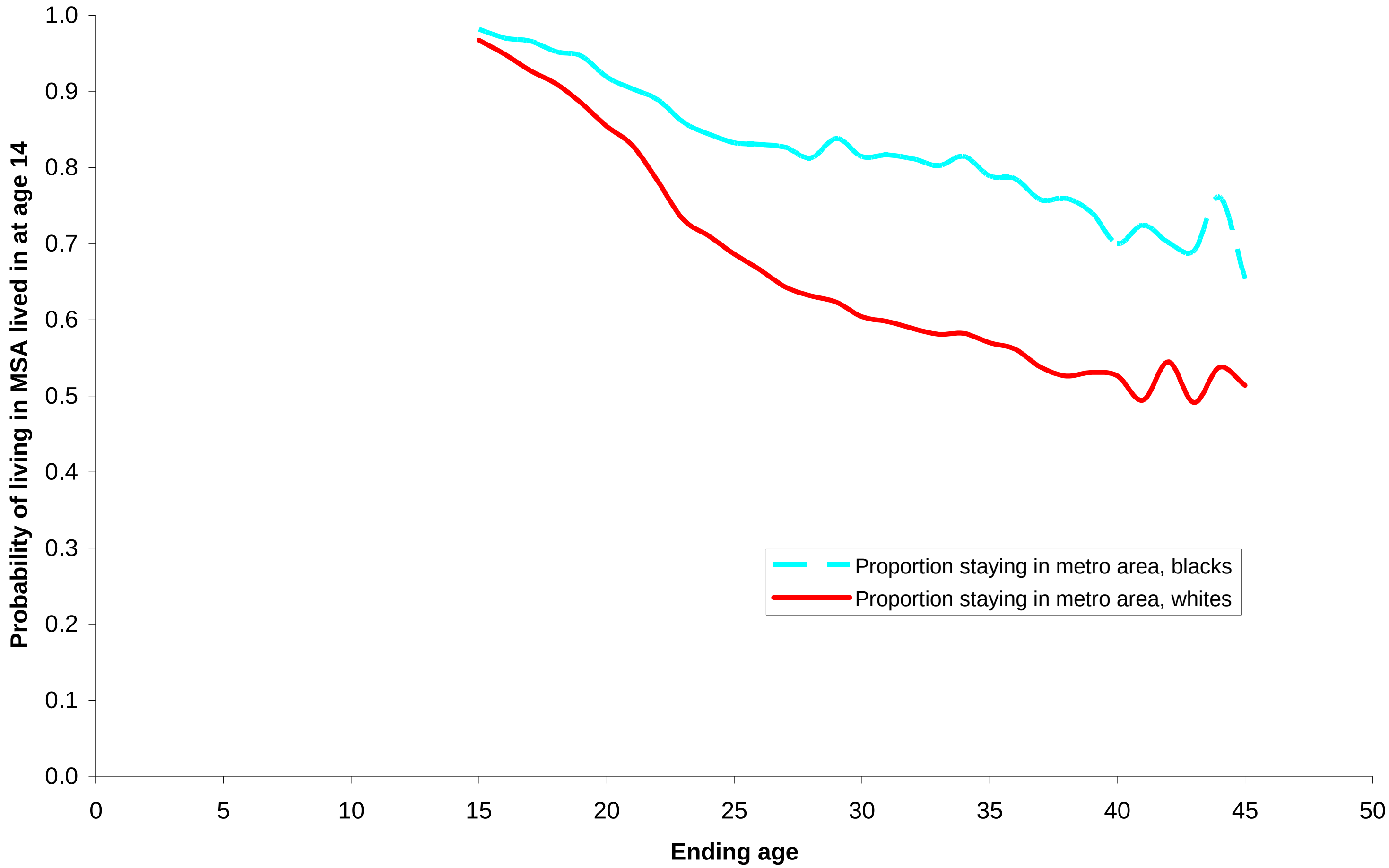
Note: Estimated probability is proportion of individuals living in birth state, for each of two racial groups: blacks and whites. All data are from 2000 PUMS and 1990 PUMS. All estimated proportions use appropriate individual weights in the PUMS.

Figure 7
Probability of Staying in Same Metro Area as Lived In at Age 4,
Blacks vs. Whites



Note: Estimated probability is proportion of individuals living in state lived in as of age 4, for each of two racial groups: blacks and whites. All data are PSID. All estimated proportions use appropriate individual weights in the PSID.

Figure 8
Probability of Staying in Same Metro Area as Lived In at Age 14,
Blacks vs. Whites



Note: Estimated probability is proportion of individuals living in state lived in as of age 14, for each of two racial groups: blacks and whites. All data are from PSID. All estimated proportions use appropriate individual weights in the PSID.

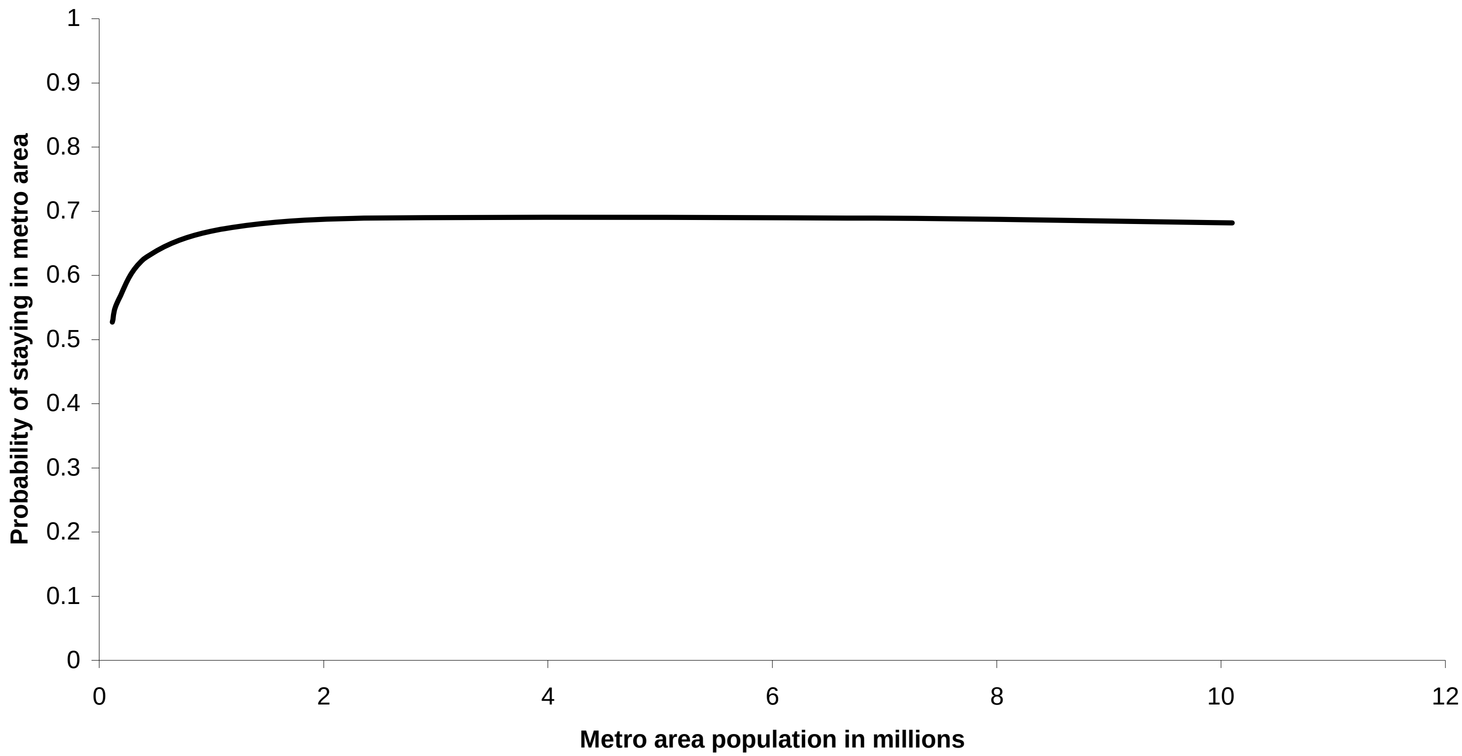
Figure 9
Probability of Staying in Same Metro Area as Lived In at Age 4, Metro Areas of Different Population Sizes

Note: Estimated probability is proportion of individuals living in state lived in as of age 4, for metro areas of different population sizes. Population size breakdowns divide all metro areas into four population-weighted quartiles. All data are from PSID. All estimated proportions use appropriate individual weights in the PSID.

Figure 10
Probability of Staying in Same Metro Area as Lived In at Age 14,
Metro Areas of Different Population Sizes

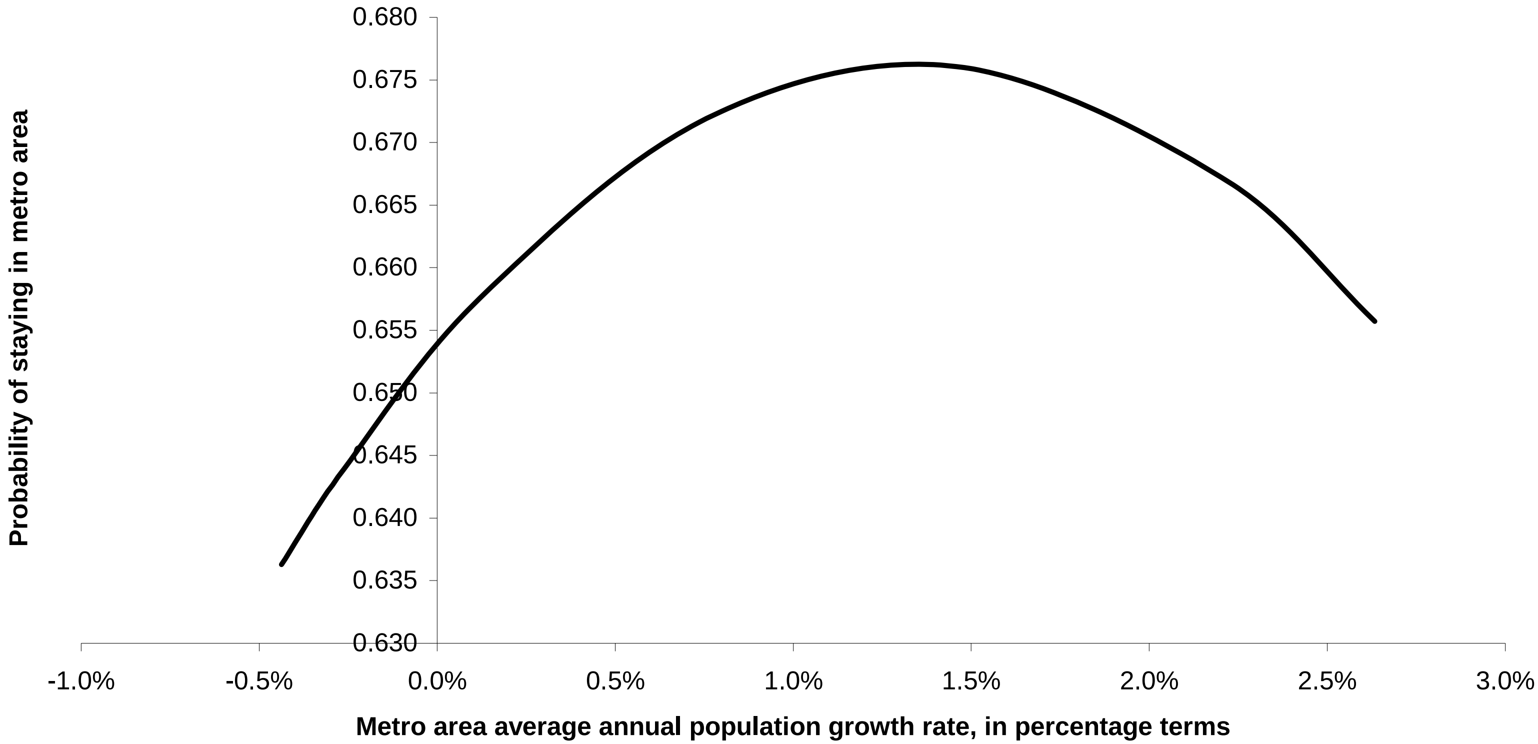
Note: Estimated probability is proportion of individuals living in state lived in as of age 14, for metro areas of different population sizes. Population size breakdowns divide all metro areas into four population-weighted quartiles. All data are from PSID. All estimated proportions use appropriate individual weights in the PSID.

Figure 11
Probability of Being in the Same Metro Area at Age 14 and Age 30, as a Function of Metro Area Population at Age 14



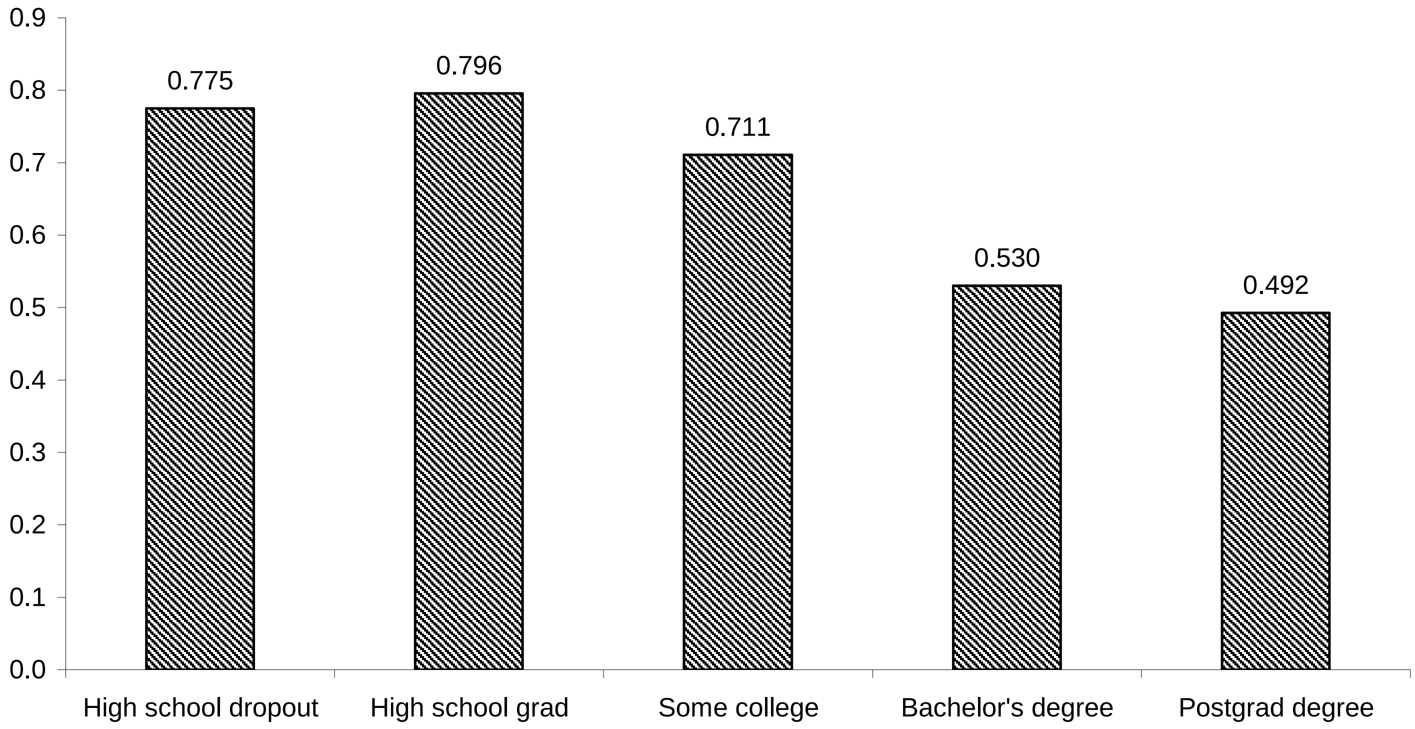
Note: See Table 3 for numerical summary of these results. All estimated probabilities are from the logit 14–30 specification with quadratic specification in both metro area population and metro area population growth. Estimated probabilities are weighted average probability for a population whose gender, education, and race composition matches the composition of the U.S. population ages 28–32 in 2007. Calculations assume median metro area population growth of 0.8% per year. The average logit year effects are used, and the effects for the non-low income sample are used. Percentiles of the metro area population are calculated using PSID weights for the logit 14–30 sample.

Figure 12
Probability of Being in the Same Metro Area at Age 14 and Age 30,
as a Function of Original Metro Area's Population Growth Between Ages 14 and 30



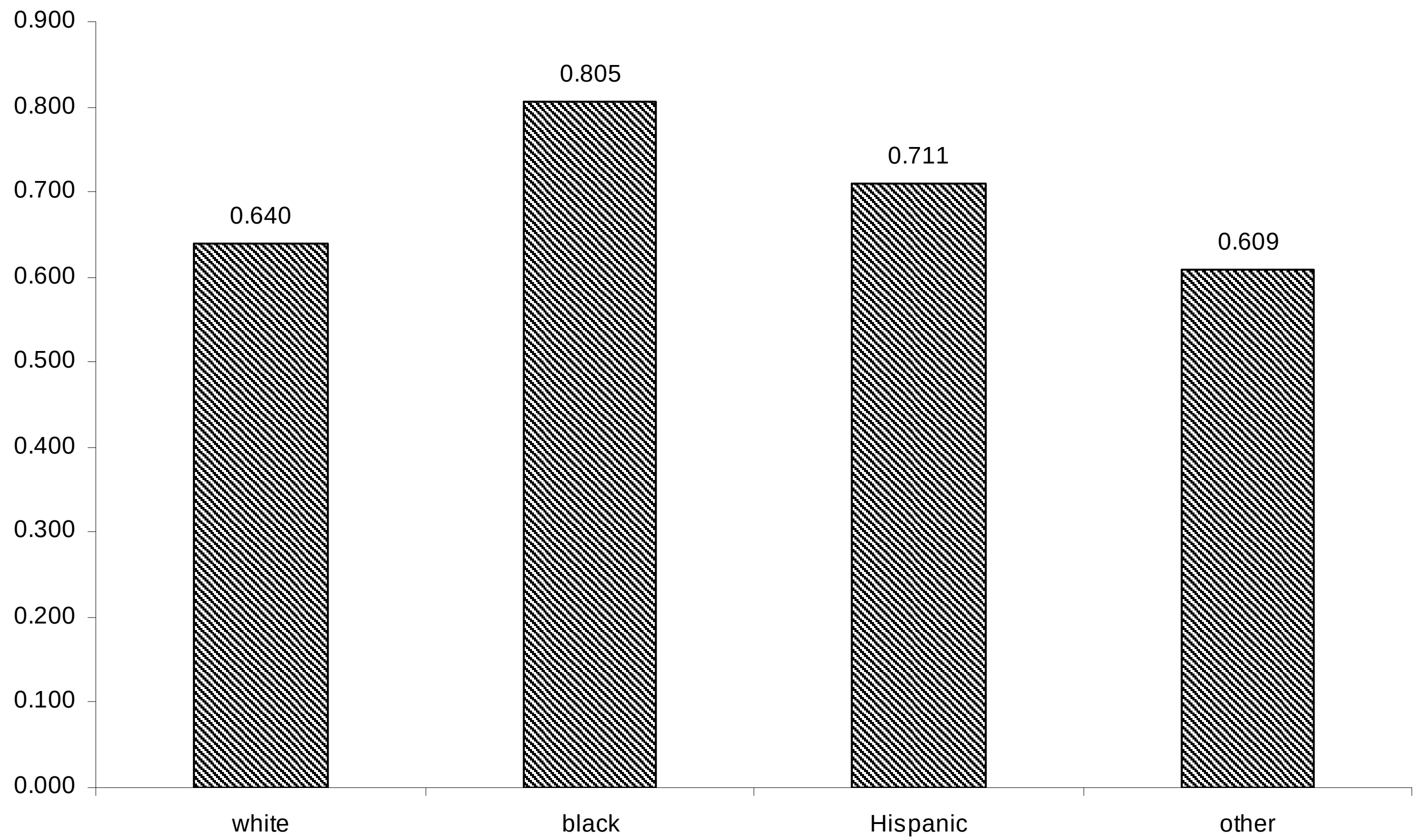
Note: All estimated probabilities are from the logit 14–30 specification with quadratic specification in both metro area population and metro area population growth. Estimated probabilities are weighted average probability for a population whose gender, education, and race composition matches the compositiono f the U.S. population ages 28–32 in 2007. Calculations assume median metro area population of 1.08 million. The average logit year effects are used, and the effects for the non-low income sample are used. Percentiles of metro area population growth are calculated using PSID weights for the logit 14–30 sample. Population growth is for the metro area lived in at age 14. What happens in any new metro area of residence as of age 30 is irrelevant.

Figure 13
How the Probability of Staying in the Same Metro Area Varies by Education Group, Based on Logit Estimates from Ages 14 to 30



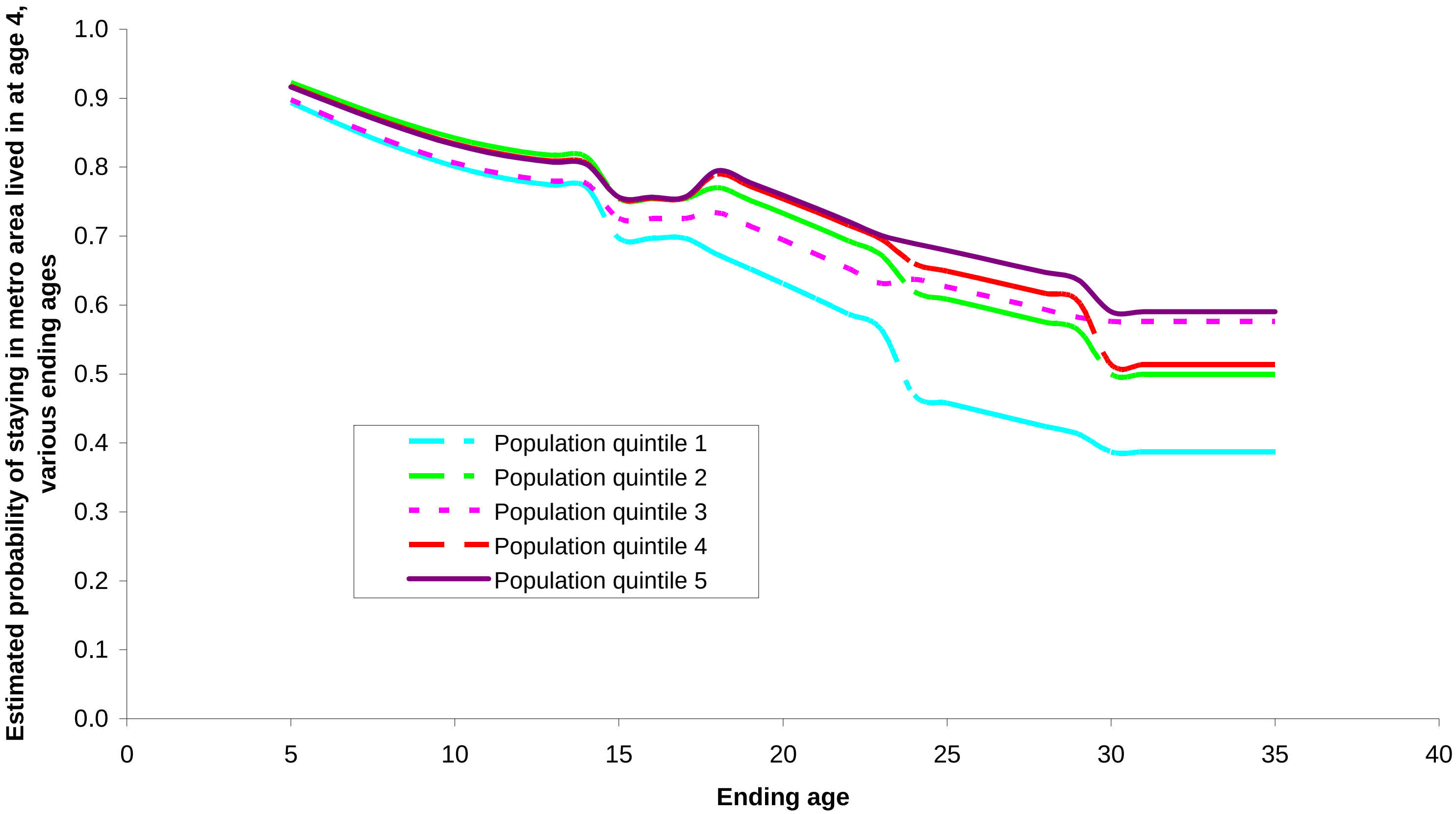
Note: These estimates are based on a metro area with median population of 1.083 million and median population growth of 0.8 percent annually. In addition, this calculation holds race/gender constant by assuming a constant race/gender composition of each education group. The degree categories show highest degree. The categories are mutually exclusive and exhaustive. Educational classifications are based on final educational achievement level observed in PSID for that individual.

Figure 14
How the Probability of Staying in the Same Metro Area Varies with Race, Based on Logit Ages 14 to 30 Estimates



Note: All estimates hold metro area size and growth constant at population-weighted median values. Education and gender are held constant by assuming that all race groups have some education/gender composition, based on mean values for 28–32 year olds in 2007 CPS-ORG.

Figure 15
How Probability of Staying in Metro Area Varies Across Metro Area Population Quintiles, Starting Age of 4 and Various Ending Ages



Note: All calculations hold constant metro growth at median level, and also hold constant metro percentage college-educated and percentage black at median level. Median values and cutoffs for population quintiles are given in previous tables. Simulations are based on five different tobit regressions, as described in Table 7.

Figure 16
How Probability of Staying in Metro Area Varies Across Metro Area Population Quintiles, Starting Age of 14 and Various Ending Ages

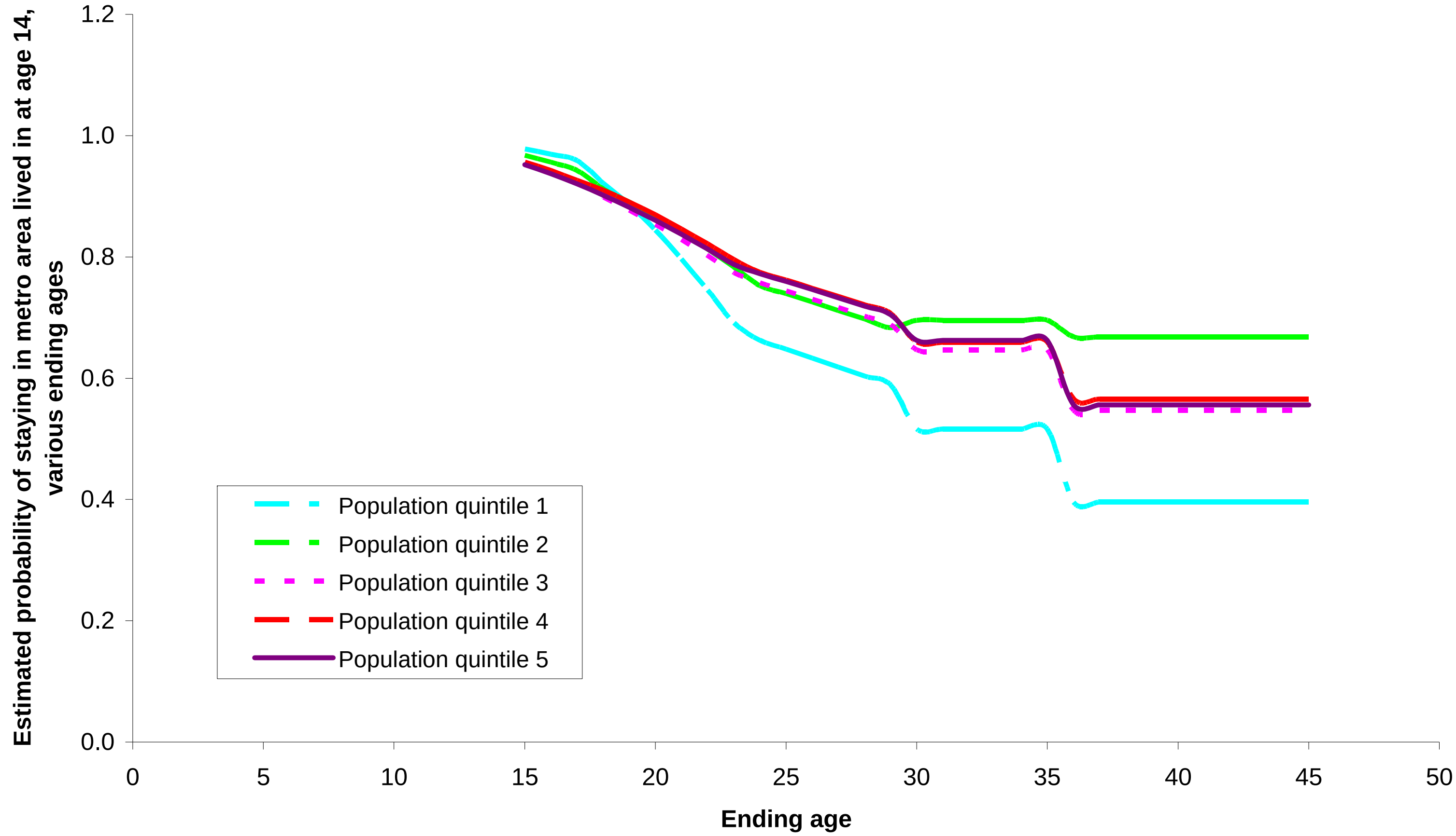
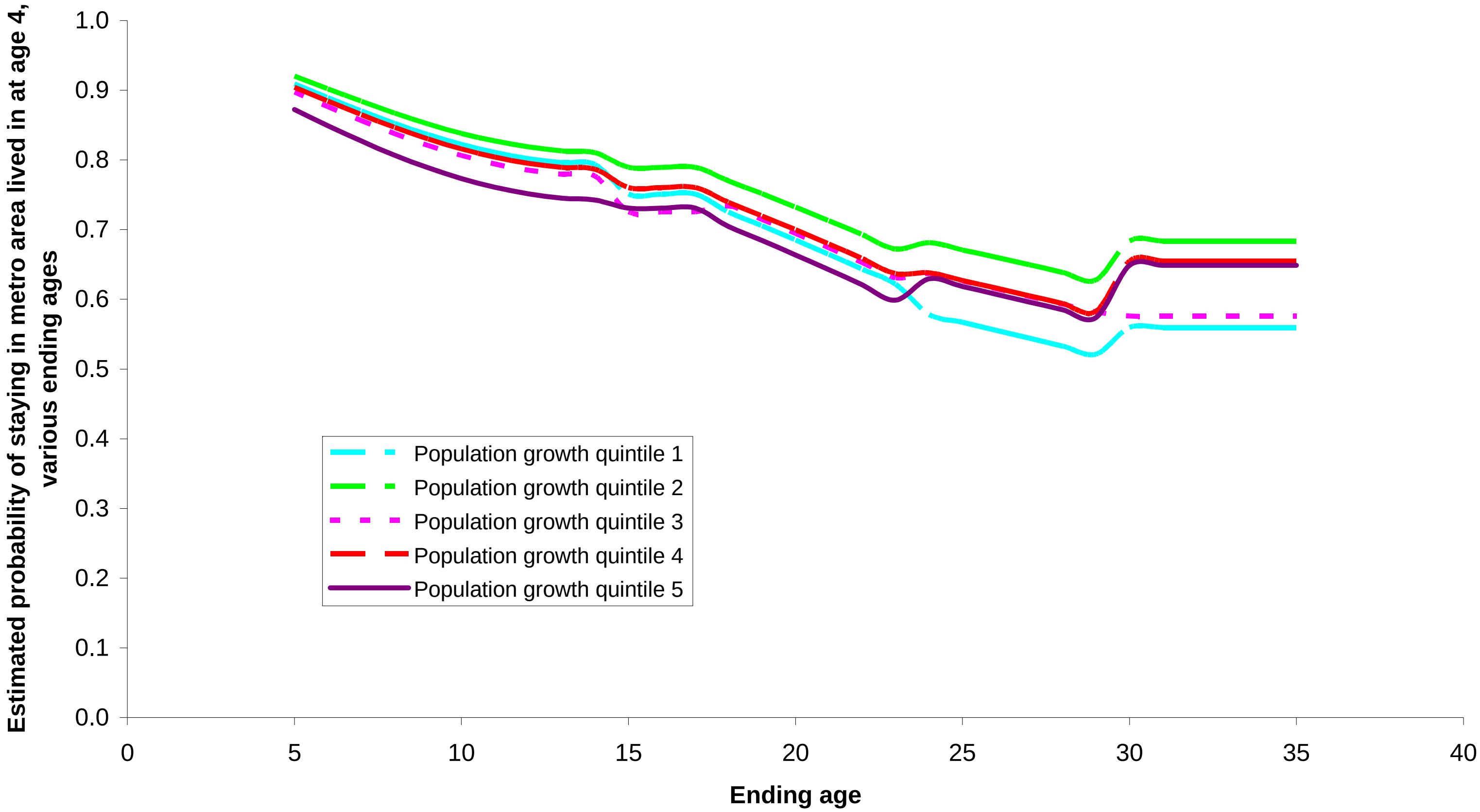
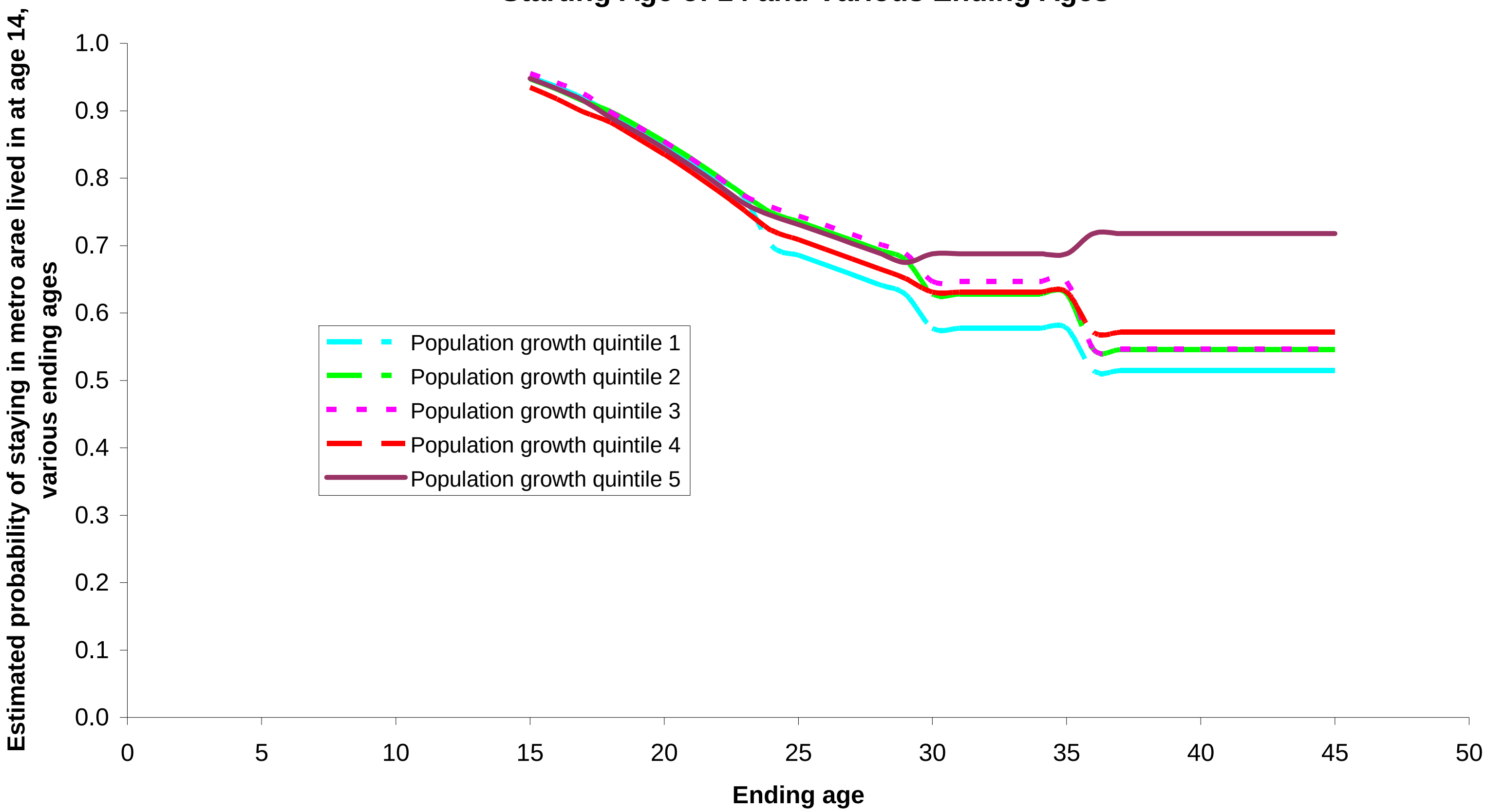


Figure 17
How Probability of Staying in Metro Area Varies
Across Different Metro Area Population Growth Quintiles,
Starting Age of 4 and Various Ending Ages



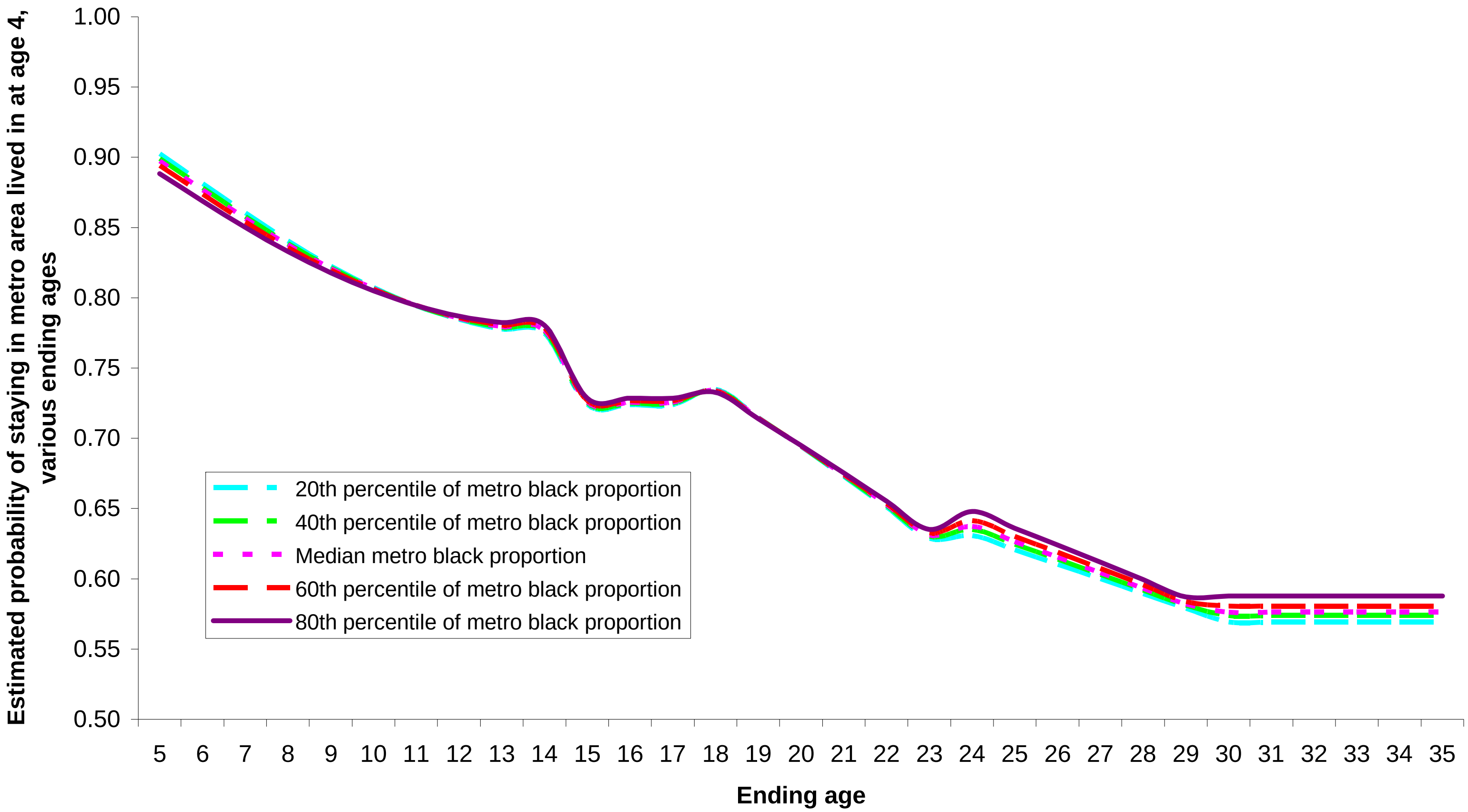
Note: All calculations hold constant metro population, proportion college-educated, and proportion black at median levels. Median values and cutoffs for different growth quintiles are given in previous tables. Simulations are based on five different tobit regressions, as described in Table 7.

Figure 18
How Probability of Staying in Metro Area Varies Across Different Metro Area Population Growth Quintiles, Starting Age of 14 and Various Ending Ages



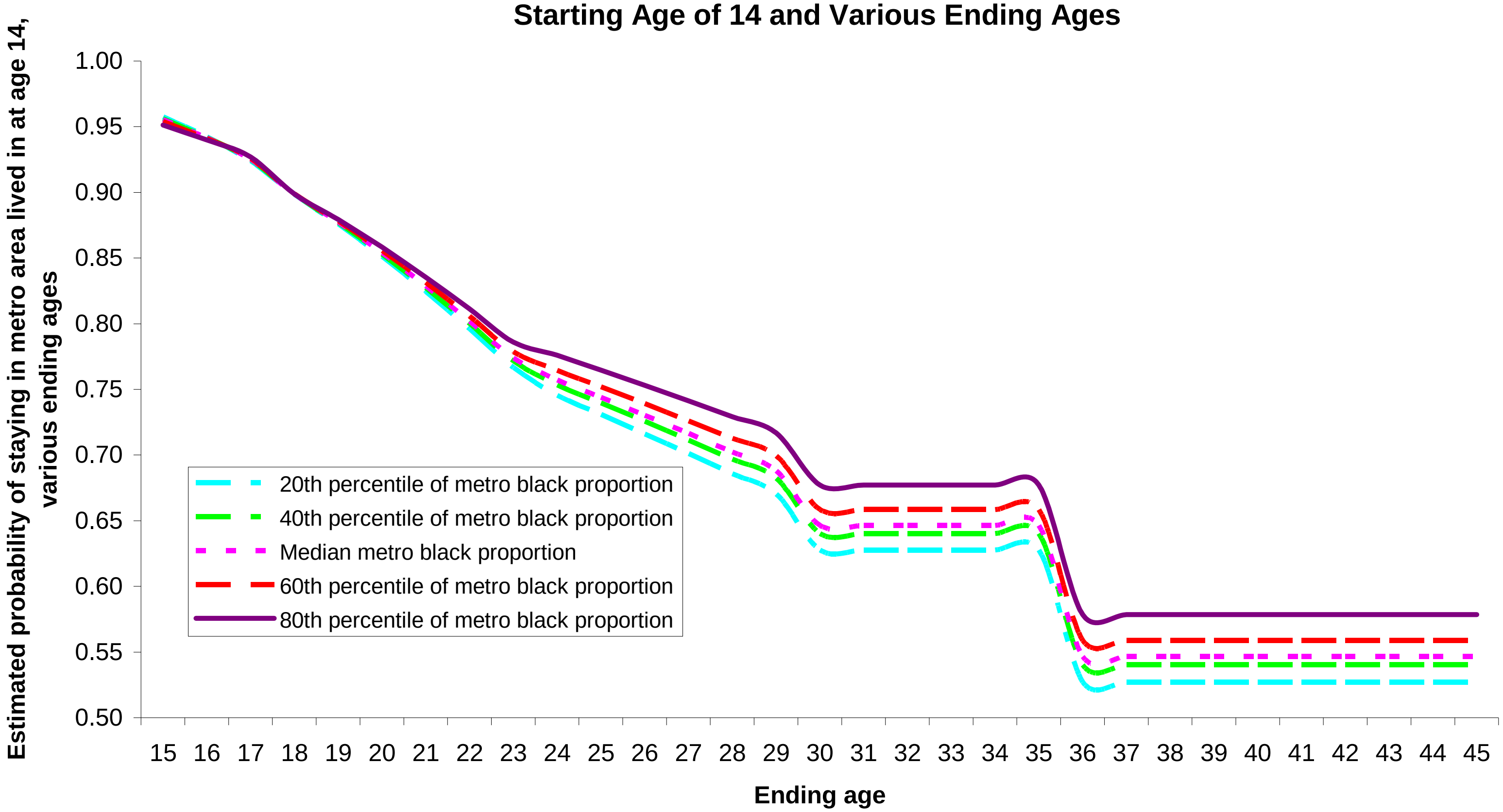
Note: All calculations hold constant metro population, proportion college-educated, and proportion black at median levels. Median values and cutoffs for different growth quintiles are given in previous tables. Simulations are based on five different tobit regressions, as described in Table 7.

Figure 19
How Probability of Staying in Metro Area Varies Across Different Metro Areas, Black Proportions, Starting Age of 4 and Various Ending Ages



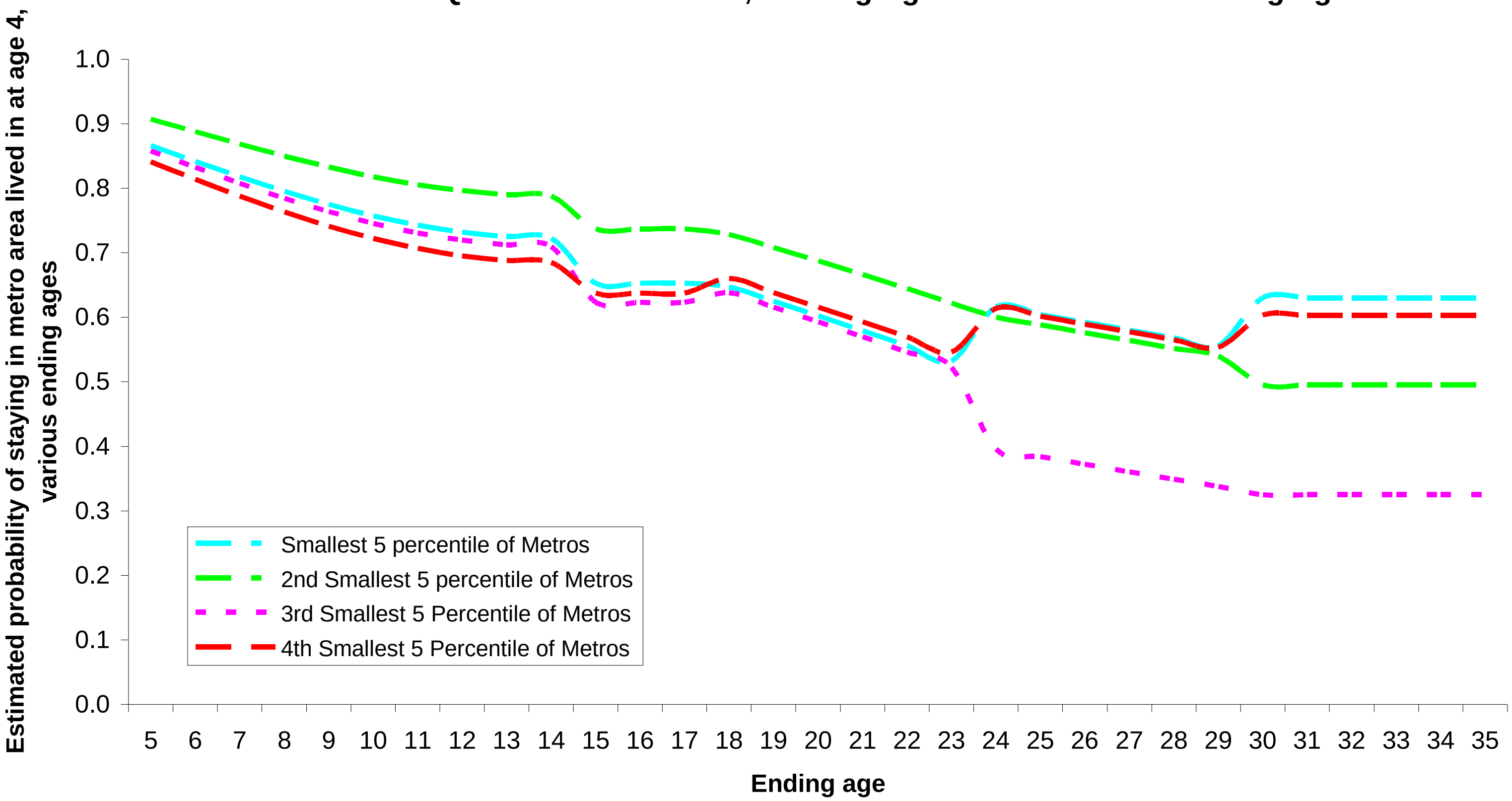
Note: All calculations hold constant metro population, population growth, and proportion college-educated at median levels. Median values and cutoffs for different percentiles for proportion black in metro area are given in previous tables. Simulations are based on five different tobit regressions, as described in Table 7.

Figure 20
How Probability of Staying in Metro Area Varies Across Different Metro Areas, Black Proportions, Starting Age of 14 and Various Ending Ages



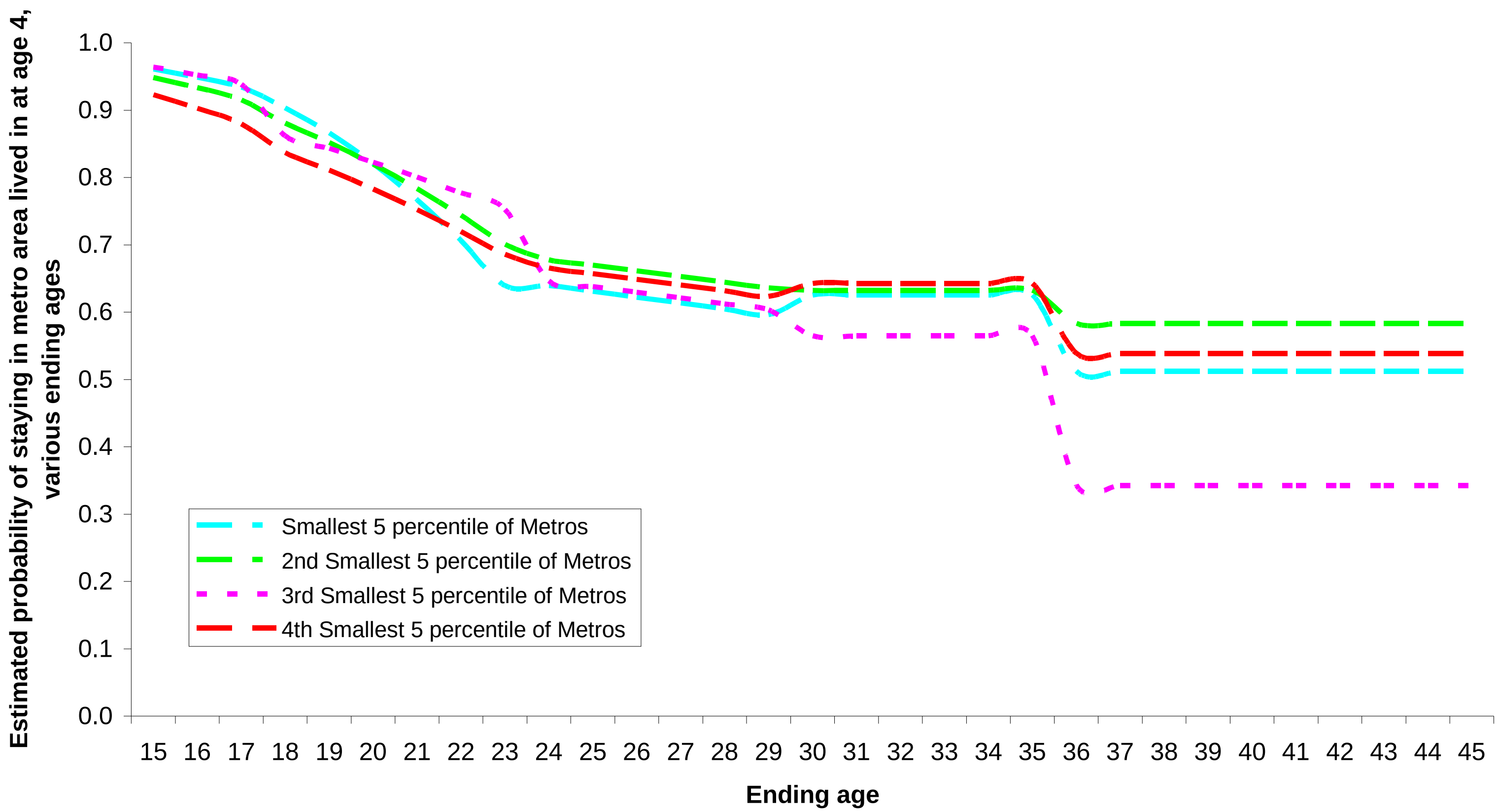
Note: All calculations hold constant metro population, population growth, and proportion college-educated at median levels. Median values and cutoffs for different percentiles for proportion black in metro area are given in previous tables. Simulations are based on five different tobit regressions, as described in Table 7.

Figure 21
How Probability of Staying in Metro Area Varies
Across Different Metro Area Population Sizes,
Within Smallest Quintile of Metro Size, Starting Age of 4 and Various Ending Ages



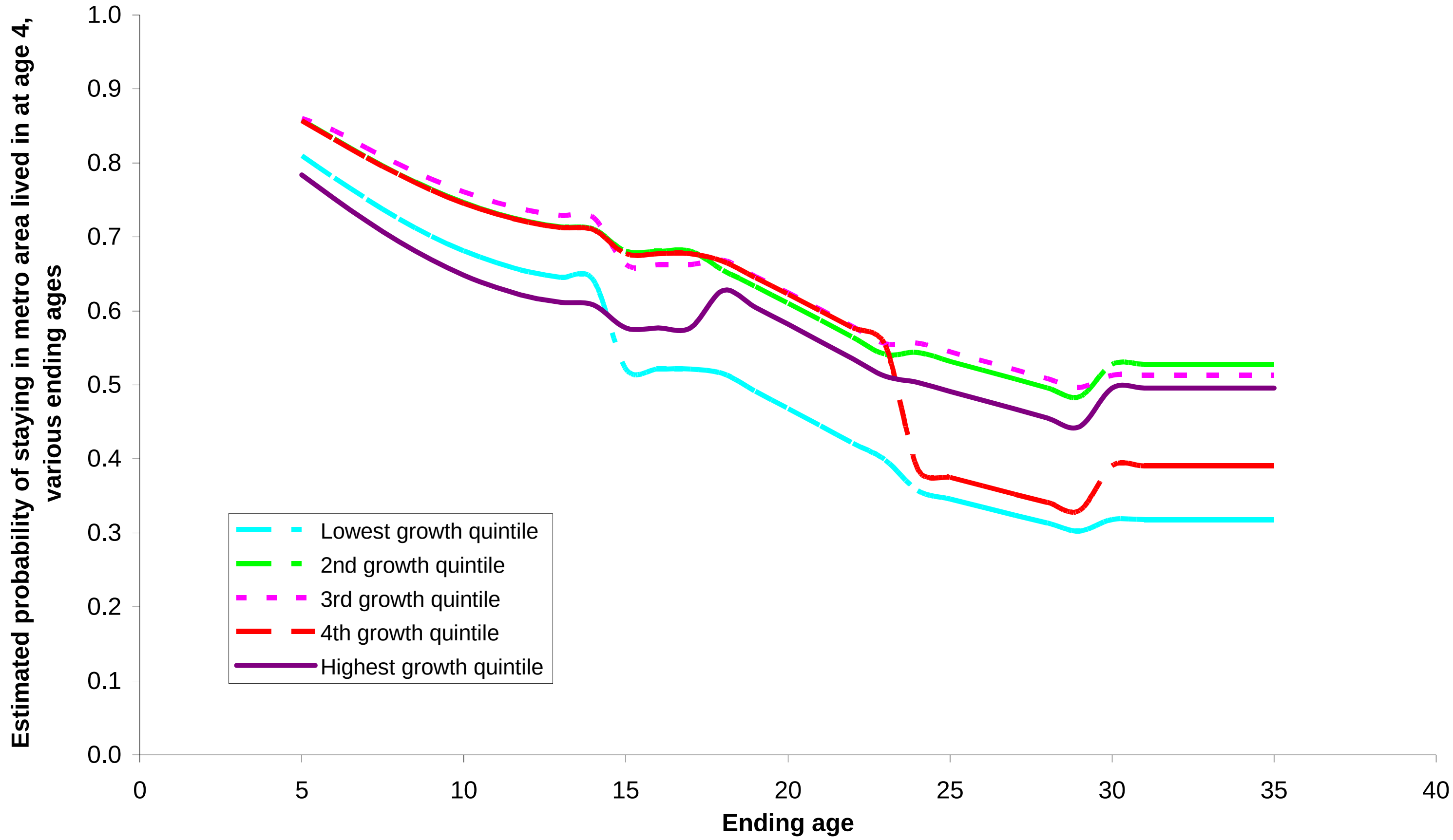
Note: All calculations hold constant metro population growth, proportion college-educated, and proportion black at median levels. Median values and cutoffs for the five percentile population groups is given in text. Simulations are based on five different tobit regressions, as described in Appendix Table B1.

Figure 22
How Probability of Staying in Metro Area Varies
Across Different Metro Area Population Sizes,
Within Smallest Quintile of Metro Size, Starting Age of 14 and Various Ending Ages



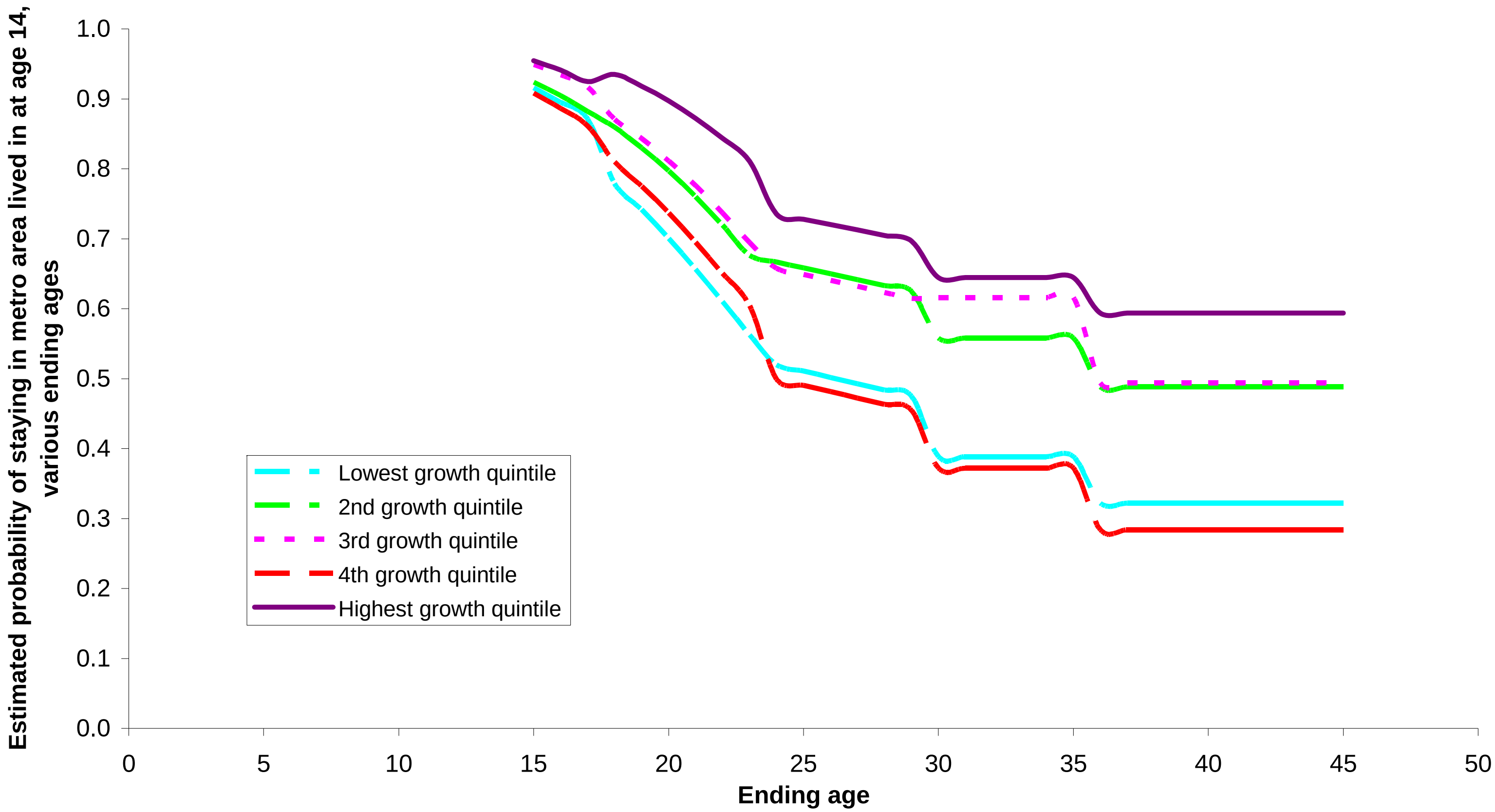
Note: All calculations hold constant metro population growth, proportion college-educated, and proportion black at median levels. Median values and cutoffs for the five percentile population groups is given in text. Simulations are based on five different tobit regressions, as described in Appendix Table B1.

Figure 23
How Probability of Staying in Metro Area Varies Across Different Metro Area Growth Rates, Within Smallest Quintile of Metro Size, Starting Age of 4 and Various Ending Ages



Note: All calculations hold proportion college-educated, and proportion black at median levels. All calculations are averaged over four possible five-percentile groups of population size within smallest population quintile metro areas. Median values and cutoffs for the five-percentile population groups is given in previous tables. Simulations are based on five different tobit regressions, as described in Appendix Table B1.

Figure 24
How Probability of Staying in Metro Area Varies Across Different Metro Area Growth Rates, Within Smallest Quintile of Metro Size, Starting Age of 14 and Various Ending Ages



Note: All calculations hold proportion college-educated, and proportion black at median levels. All calculations are averaged over four possible five-percentile groups of population size within smallest population quintile metro areas. Median values and cutoffs for the five-percentile population groups is given in previous tables. Simulations are based on five different tobit regressions, as described in Appendix Table B1.

Appendix A

Notes on Processing the PSID Data

Some processing of the PSID data was done to calculate the variables used in estimation as accurately as possible. Specifically, there are issues with age, educational attainment, race, and location. These processing procedures were developed by the author and Wei-Jang Huang, senior research analyst at the Upjohn Institute, and implemented by Ms. Huang.

Age

The estimation stresses estimating staying behavior between two ages. Therefore, it is particularly important to make sure that the age variables are as accurate and consistent as possible.

In the PSID, the age variable frequently does not increment by one year for each year. Sometimes it increments by zero years or two years. In this case, the lack of a consistent one-year increment could in theory be due to the timing of birthdays and the data collection. In other cases, an age will stay the same for several years in a row, and then suddenly jump by several years. In still other cases, the age will decrease from one year to the next. Finally, in some cases the age will make a large discrete jump from year *t* to year *t*+1, before returning in year *t*+2 to an age that is consistent with year *t*'s reported age. It is obvious therefore that in some cases that the age variable is mistaken, either due to errors by the respondent or in data entry.

My goal in processing the age data in the CPS was to get a data series that for each individual would increment by exactly one year for each calendar year, yet would be as consistent as possible with the actual observed data. Therefore, when examining whether an individual stays from age $a0$ to age $a1$, we know that those two ages are separated by exactly $a1-a0$ calendar years, no more and no less.

Out of the 67,271 individuals observed in this PSID sample, 213 individuals had cases where the age decreased by more than two years from one year to the next. These records were manually corrected. Thirty-five of these 213 records were so confusing that they were dropped from the sample. This leaves 67,236 individual records.

For these 67,236 records, for each age observed for each calendar year for each individual, we created a separate possible age series that is consistent with each calendar year observation for each individual. Thus, for example, for an individual observed for the entire PSID sample, there will be age records for each of the years 1968 until 1997, and for alternate years from 1999 until 2005. For each of the observed ages, we create a possible age series that is incremented by one year each year. For each of these possible age series, we calculate the sum of absolute deviations of the observed ages from the possible age series. We picked as the "correct" age series the one that minimized this sum of absolute deviations. Where there's a tie between two possible series, we picked the possible series that makes fewer changes to observed ages.

Out of the 67,236 records, about a quarter had no changes in reported ages because of this procedure. The median percentage of reported ages that were changed was 9 percent and the mean percentage changed was 16 percent. Only 5 percent of the records had more than half the reported ages changed.

Some 115 of the 67,236 individuals had zero age in all reported years. These 115 individuals were dropped from the sample.

Educational Attainment

For each individual, we have various educational attainment variables observed for different years. Where possible, we used the most recent observed educational attainment from the individual data files. If the individual education variable was missing (which it almost always was from 1970 to 1974), we used the head's or wife's educational attainment variable when the individual was a head or wife. For each individual, we only retained the final observed educational attainment variable. We also examined the individual's age for the final observed educational attainment variable. If that age was less than 23, we recoded "final educational attainment" as missing. Educational attainment was classified into the following categories: high school dropout; high school graduate; some college education, bachelor's degree, post-grad education.

Race

The PSID reports the household head's race each year from 1968 on, the wife's race from 1985 on, and children's race from 1997 on. The PSID follows the old sexist convention of the head always being the husband in a husband-wife family.

To classify each individual's race, we used the most recent head's/wife's race when this was available. Out of 67,121 individuals, 35,471 were assigned a valid race based on this procedure. We then assigned a race variable from the individual's child file for individuals who were never a head or wife, or whose head or wife race information was missing. This assigned race to an additional 6,295 individuals. In cases where none of this information was non-missing, we assigned the mother's race to an individual, or, if not available, the father's race. This assigned race to an additional 15,593 individuals. There are 13,583 individuals out of the 67,121 individuals for whom there is not a valid race assignment. Given how the data are collected, most of these individuals would be children in the early PSID who became non-response before race was collected from them as children or as a head or wife, wives in the early PSID who became non-response before wives' race information was collected, or non–PSID sample members living in sample member households. It should be noted that in almost all of these cases, we would not observe whether they stayed in the same metro area or state from ages 4 or 14 to adulthood in any event, as we would not have a sufficiently long time series on these individuals. So these individuals would have been dropped from the sample for other reasons. The race variables were categorized as white non-Hispanic, black non-Hispanic, other non-Hispanic, and Hispanic.

Location Information

The state of location for each year is available in the regular PSID files. The county of location was obtained for this project using the PSID Geocode data, which is available only under special request and conditions to safeguard respondent confidentiality. The Geocode data also includes a metro area assignment. However, to ensure consistently with the latest metro area definitions in all cases, we coded each observation ourselves into the official metro area definitions as of 2005, based on the observed state and county of residence.

Appendix B

Tobit Estimates for Staying Propensity, Estimation Restricted to Metro Areas in Lowest Population Quintile

Appendix Table B1 reports some additional tobit estimates. The model used is similar to that used to generate Table 7 to the text. However, the sample is restricted to metro areas in the lowest population quintile, which has a metro area population of less than 333,000. Therefore, the population variable is changed from Table 7. The population variable is defined as the metro area population percentiles from 0 to 5, 5 to 10, 10 to 15 and 15 to 20. The cutoffs for these "5 percentile" groups are given in the notes to the table. As in Table 7, all the population variable's coefficients are the effects of that dummy variable relative to the average for all dummies in that category. The term "first 5 percentiles" in the table refers to the smallest metro areas, and so on with the other terms.

Appendix Table B1 Effects on Proportion Staying in Metro Area, from Various Beginning Age to Various Ending Ages, for Metro Areas in Lowest Population Quintile

Beginning age	Age 4					Age 14				
Ending ages	5–14	15–17	18–23	24–29	30–35	15–17	18–23	24–29	30–35	36–45
Age	-0.0848*		-0.0420*	-0.0165		-0.0567*	-0.0550*	-0.0170*		
	(-2.68)		(-3.28)	(-1.11)		(-2.55)	(-3.62)	(-3.10)		
Age^2	0.00259									
	(1.90)									
Proportion college in metro area	-0.087	-0.317	0.232	0.864	2.098	0.245	0.410	0.100	-0.107	0.197
	(-0.17)	(-0.40)	(0.31)	(0.81)	(1.00)	(0.58)	(1.28)	(0.29)	(-0.18)	(0.23)
Proportion black in metro area	-0.191	-0.167	0.319	2.028	-0.498	-0.665	-1.124	-1.061*	0.243	0.305
	(-0.27)	(-0.52)	(0.24)	(0.98)	(-0.46)	(-0.52)	(-1.24)	(-2.03)	(0.66)	(0.75)
Age × proportion black	-0.0857		-0.0185	-0.0938		0.0340	0.0610	0.0510*		
	(-0.62)		(-0.31)	(-1.02)		(0.44)	(1.36)	(2.64)		
Age^2 × proportion black	0.00662									
	(1.04)									
Population 1st 5 percentiles	-0.0088	-0.0221	-0.0447	0.1310	0.3330	0.0456	0.6282*	-0.0245	0.0126	0.0251
	(-0.18)	(-0.24)	(-0.47)	(0.86)	(1.09)	(0.75)	(2.75)	(-0.51)	(0.14)	(0.25)
Population 2nd 5 percentiles	0.1055	0.1559	0.1259	0.0958	-0.0483	-0.0094	0.0623	0.0281	0.0228	0.1208
	(1.91)	(1.72)	(1.47)	(0.76)	(-0.21)	(-0.19)	(0.23)	(0.82)	(0.37)	(1.62)
Population 3rd 5 percentiles	-0.0291	-0.0807	-0.0627	-0.3522	-0.5388	0.0589	-0.4017	-0.0149	-0.0733	-0.2065*
	(-0.47)	(-0.78)	(-0.54)	(-1.69)	(-1.29)	(0.69)	(-1.34)	(-0.26)	(-0.93)	(-2.21)
Population 4th 5 percentiles	-0.0676	-0.0531	-0.0185	0.1254	0.2542	-0.0952	-0.2887	0.0112	0.0379	0.0605
	(-1.31)	(-0.66)	(-0.24)	(1.03)	(0.99)	(-1.52)	(-1.41)	(0.24)	(0.53)	(0.69)
	[-0.75]	[-0.23]	[0.19]	[-0.03]	[-0.20]	[-1.43]	[-0.67]	[0.45]	[0.18]	[0.22]
Age × population 1st 5 percentiles							-0.0306*			
							(-2.83)			
Age × population 2nd 5 percentiles							-0.0024			
							(0.18)			
Age × population 3rd 5 percentiles							0.0211			
							(1.52)			
Age × population 4th 5 percentiles							0.0119			
							(1.20)			

Appendix Table B1 (continued)

Beginning age	Age 4					Age 14				
Ending ages	5–14	15–17	18–23	24–29	30–35	15–17	18–23	24–29	30–35	36–45
Metro population growth, quintile 1 (lowest)	-0.0582 (-1.13)	-0.1985 (-1.78)	-0.2144 (-1.82)	-0.2548 (-1.79)	-0.3963 (-1.27)	-0.0489 (-0.60)	-.0.1459* (-2.22)	-0.1275* (-2.01)	-0.1791 (-1.76)	-0.1601 (-1.51)
Metro population growth, quintile 2	0.0474 (0.92)	0.1118 (1.34)	0.0545 (0.68)	0.1676 (1.38)	0.2341 (1.12)	-0.0252 (-0.39)	0.0024 (0.05)	0.0648 (1.87)	0.0585 (0.88)	0.0754 (0.82)
Metro population growth, quintile 3	0.0725 (1.18)	0.0750 (0.86)	0.0820 (0.86)	0.1961 (1.49)	0.1913 (0.76)	0.0600 (0.77)	0.0276 (0.78)	0.0518 (0.81)	0.1411 (1.51)	0.0832 (0.68)
Metro population growth, quintile 4	0.0459 (0.85)	0.1050 (1.06)	0.0783 (0.82)	-0.1865 (-1.20)	-0.1701 (-0.74)	-0.0690 (-1.52)	-0.0934* (-2.54)	-0.1534* (-3.81)	-0.2031* (-2.84)	-0.2192* (-2.20)
Metro population growth, quintile 5	-0.1077 (-1.20) [-0.45]	-0.0934 (-0.63) [0.53]	-0.0003 (-0.00) [1.13]	0.0776 (0.41) [1.24]	0.1410 (0.40) [0.99]	0.0831 (0.94) [0.93]	0.2093* (3.09) [3.05]	0.1643* (3.13) [2.95]	0.1826* (2.68) [2.40]	0.2207* (2.31) [2.40]
F-test prob: Population terms, Growth terms, Age terms, Black terms	0.2667, 0.5271, 0.0002, 0.0273	0.4008, 0.3658, NA, 0.6047	0.5350, 0.4651, 0.0018, 0.9175	0.3829, 0.1305, 0.0682, 0.5926	0.5304, 0.4882, NA, 0.6458	0.4750, 0.6030, 0.0326, 0.8035	0.0713, 0.0044, 0.0000, 0.3122	0.7977, 0.0000, 0.0059, 0.0093	0.7948, 0.0008, NA, 0.5123	0.1189, 0.0131, NA, 0.4554

NOTE: These are tobit estimates from estimation that only includes metro areas whose population is in the lowest population quintile, smaller than 333,000 in population. The specification is the same as the estimation for the full sample, except that the population dummies are for the lowest population quintile divided into four percentile groups (zero to 5th, 5th through 10th, 10th to 15th, and 15th to 20th). Percentiles are metro-population-weighted quintiles. Cutoffs for 5th, 10th, and 15th percentiles are 123,000; 171,000; and 271,000. As with the previous specification, all of these population group effects and growth effects are effects relative to average effects over all groups. Numbers in parentheses are all *t*-statistics. Numbers in brackets above largest population group and largest growth quintile are *t*-statistics for largest group effects minus smallest group effects.