A New Sample of Males Linked from the Public Use Micro Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules

JOSEPH P. FERRIE*

May 16, 1996

Abstract: Though the geographic, occupational, and financial mobility of average Americans were important aspects of nineteenth century U.S. economic development, the extent and correlates of this economic mobility have remained open to debate in the absence of individual-level longitudinal data. This essay describes a new sample of 4,938 individuals linked from the 1850 Public Use Micro Sample of the federal census of population to the 1860 federal census manuscript schedules, using the new national 1860 federal census index. The linked sample provides information on occupation, wealth, family structure, and location in both 1850 and 1860. The construction of the sample is described in detail, along with tests of its representativeness, and examples of potential uses.

---

* Assistant Professor, Department of Economics and Fellow, Center for Urban Affairs and Policy Research, Northwestern University, Evanston, IL 60208, and Faculty Research Fellow, National Bureau of Economic Research, Cambridge, MA 02138. E-Mail: ferrie@nwu.edu.
Financial support was provided by the National Science Foundation (Grant No. SES-9309689) and a faculty research fellowship from the John M. Olin Foundation. Extremely conscientious research assistance was provided by Katie Holsinger, Taylor Holsinger, Shannon Holsinger, Angela Kehl, Brian Kehl, Jefferson Call, Amber Starr, and Mark Woltuis. I am grateful to Steven Ruggles and the University of Minnesota's Social History Research Laboratory for providing a pre-release version of the 1850 Public Use Micro Sample, Clayne Pope and the Brigham Young University Economics Department for their help in the collation of the 1860 census index that made this project possible, and David Weir for helpful comments. Previous versions of this essay were presented at the National Bureau of Economic Research/ Development of the American Economy Program's Summer Institute in 1994 and 1995. The sample described here will be available in September, 1996 on the World Wide Web (WWW) site maintained by the Northwestern University Economics Department in the directory http://www.econ.nwu.edu/faculty/ferrie/data.

I. INTRODUCTION

The economic mobility of Americans has played a major role in accounts of the development of the nineteenth century U.S. economy. More than a century ago, for example, Frederick Jackson Turner's "safety valve thesis" suggested the salutary effects on the U.S. economy that movement to the American frontier by common laborers could have had: allowing dissatisfied urban wage earners to improve their lot through migration to the less-settled west, entry into agriculture, and the accumulation of land (Turner 1920). Little evidence has been available, however, with which to assess how much geographic, occupational, and financial mobility were actually experienced by the average American. Published data from the decennial censuses of population allow examination of only broad aggregates: net migration flows, the distribution of occupations at a particular place or point in time, and the average level of wealth in various communities (Galloway and Vedder 1971; McClelland and Zeckhauser 1982; Schaefer 1994). Where specific people went, how they changed jobs over the course of their careers, and how successful they were in monetary terms as those changes occurred have remained a mystery.

This essay describes a new sample of 4,938 males linked from the new Public Use Micro Sample (PUMS) of the 1850 federal census of population to the manuscript schedules of the 1860 federal census of population that will help fill this gap in our knowledge. The sample was created by seeking males age 10 or older in 1850 from the PUMS in the new national index to the 1860 federal population census and then retrieving the individuals from

the 1860 manuscript census schedules.[1] Of the 55,852 males in the PUMS, only the 25,586 whose combination of surname and given name appeared ten or fewer times in the index were sought in the 1860 manuscripts. The linkage success rate was 19 percent among these individuals with uncommon names, and 9 percent overall. Both non-migrants and migrants who moved anywhere in the U.S. are included, which provides a picture of antebellum economic mobility more complete than studies that have focused only on individuals who remained in a particular location for a decade or more. This new linked sample more adequately represents the experiences of American males who survived the 1850s than previous samples that have attempted to follow individuals over this decade (Steckel 1988a; Schaefer 1985).

After reviewing the existing work on individuals linked across the 1850s in Section II, I describe the collection of the new sample in detail in Section III. In Section IV, I use these data to examine the geographic mobility of the population (in particular, movement to the western frontier). An appendix contains new life tables for the 1850s—based on manuscript data from the mortality schedules of the 1850 census—that were used to estimate how many survivors could be expected between 1850 and 1860 in the linkage process.

[1] The 1850 PUMS was created by the Social History Research Laboratory at the University of Minnesota. It is part of that group's Integrated Public Use Micro Sample (IPUMS) project. Additional information, as well as copies of the data files, can be obtained from the group's WWW site: http://www.hist.umn.edu/. Though the final sample is about 30 percent larger than the preliminary sample used here, the preliminary sample is also nationally representative. The national index to the 1860 census appears in Jackson (1992). It is described in greater detail below.

II. PREVIOUS RESEARCH

The lack until now of longitudinal evidence on nineteenth century economic mobility has been unfortunate in two respects. First, research by economists working with contemporary data suggests the strength of the relationship between geographic mobility and upward occupational and financial mobility, and how some of the most important effects—the selectivity of migration and the extent and timing of the adjustment migrants make to their new environment—are best observed at the individual rather than the aggregate level, by following specific people over several years.[2] Second, because we lack such evidence for the nineteenth century, we are unable to examine how economic mobility along these dimensions has changed over long periods of time. Excellent longitudinal data now exist for the contemporary U.S. population from sources such as the National Longitudinal Surveys (NLS) or the Panel Study of Income Dynamics (PSID), but we lack an adequate historical benchmark against which to measure the mobility they document.[3]

In the absence of adequate national-level data for the nineteenth century, researchers have turned instead to studies of particular places, generally cities or counties, and tried to

[2] For studies of how international migrants fare in the U.S. labor market, see Chiswick (1978); Chiswick (1994); and Borjas (1987). For internal migrants within the U.S., see Polachek and Horvath (1977); Borjas, Bronars, and Trejo (1992a and 1992b); Bartel (1979); and Schlottmann and Herzog (1984).

[3] Steckel and Krishnan (1993) compare wealth mobility in the NLS and in the sample created by Steckel for the 1850-60 period (Steckel 1988a). Because Steckel's sample consisted of mature men, however, the comparisons they could make across the nineteenth and twentieth centuries were limited to this group. Though the NLS contains similar data for younger contemporary males as well—individuals at a stage in the life-cycle when changes in economic circumstances were even more likely—there has been no data comparable for the antebellum period. Data from the PSID are summarized in Duncan (1984).

trace the experiences of individuals who remained in those communities for a decade or more. But because most individuals did not remain in these communities and could not be observed more than once, such work merely whets the appetites of those interested in making generalizations about economic mobility at the national level. What became of the 60 percent or more of the individuals in these studies who, for one reason or another, were not observed again?[4]

In the 1980s, two studies were undertaken to answer the questions these community-based studies left open, by taking advantage of the creation of indexes to the 1850 federal manuscript census schedules. Richard Steckel and Donald Schaefer both drew samples from the 1860 census manuscript schedules and successfully linked several thousand of these individuals back into the 1850 census manuscript schedules (Steckel 1988a; Schaefer 1985). The research I present here complements their work, linking individuals forward from 1850 to 1860 rather than backward, producing longitudinal data more representative of the antebellum U.S. economy than samples linked backward, and capturing the experiences of younger, more footloose, less established individuals than those samples contained.[5]

---

[4] The average persistence rate in 63 community studies surveyed by Parkerson (1982, 102) was 38.3%.

[5] The new sample is compared with Steckel's and Schaefer's (as well as Soltow's national cross-sectional samples from the 1850 and 1860 federal censuses) below. The sample described here is closer in its construction to that devised by Stephenson (1978), who used the Soundex indexes for the 1880 federal census of population to locate a sample of households drawn from the New York State Census of 1875. The National Panel Study (NPS) linked individuals from the 1880 census manuscripts to the 1900 manuscripts, again using the state-by-state Soundex indexes (Guest 1987). The use of state indexes and the inability to search all state indexes for those who changed states, however, probably leave inter-state migrants under-represented in this sample. Guest reports that “only 160 out of 4,041 cases were found

III. THE SAMPLE

When Steckel and Schaefer created their linked samples, the 1850 census was the only antebellum federal census that met two important criteria: it contained information on place of birth (which would be essential for successful matching) and it had been indexed. The indexes were created at the state level, however, and had not yet been collated into a national index (Jackson 1982). This dictated the sampling strategy used in both studies: since it was necessary to trace individuals *into* the 1850 census (since it was indexed) and since the nearest census with information on place of birth that could be used to verify matches was 1860 (such information was not included in 1840), individuals had to be traced backward from 1860 to 1850 rather than forward. For Steckel, who created a national sample, the appropriate state index had to be chosen out of the collection of more than thirty indexes, so the birthplaces of children age 10 and older in 1860 were used to point to the most likely state index to search. This limited the base (1860) population to families with at least one child who was ten years or older in 1860. Though Schaefer was interested in a smaller set of states (he used the Parker-Gallman sample as his base population but concentrated on those located in 1860 in the "New South"—Alabama, Louisiana, Mississippi, and Tennessee—and the southern frontier— Arkansas and Texas), and did not need to use the birthplaces of children to suggest the state in which to seek an individual in 1850, the exclusion from the 1850

---

out-of-state" (1987, 65). These data are re-examined in Simkovich (1993). Studies by Davenport (1984) and Knights (1991) also involved forward linkage, but both were done with samples originally drawn from two specific communities—Schoharie County in upstate New York, and Boston—so the usefulness of these studies for determining the fate of non-persisters generally is limited by the narrow scope of the locations from which the original samples of individuals to follow were drawn.

census index of members other than the household head and unrelated individuals means that his sample, too, has a larger share of older males who had established households by 1850 than would a sample of all males present in 1850 who survived to 1860.

The 1860 federal census has now been indexed, and the index is national in scope.[6] This means that if a base population is chosen from the 1850 census, that population can be traced forward to 1860 without the need to focus on particular states or the need to limit the base sample to families with children of particular ages. Instead, all individuals who could have set up independent households by 1860 could enter the 1850 base population.[7] For this study, the new Public Use Micro Sample (PUMS) of the 1850 federal census of population, a nationally representative sample of 153,188 individuals in 27,095 families, was used as the base population. The linkage was done in seven steps, which will be illustrated in Table 1 by the case of Sidney Briggs of Calhoun County, Michigan, who appears in the 1850 PUMS along with his wife and four children.

STEPS IN THE LINKAGE PROCESS

1. *Restrict the base sample to males who were 10 years or older in 1850* . This restriction was necessary because the 1860 index includes only household heads and apparently non-related

[6] The index is available both on microfiche and in machine-readable form from Accelerated Genealogical Endexing Services (AGES) of Salt Lake City, Utah.

[7] The 1860 index contains the location in the census of every household head and every individual living in a household whose surname was different from that of the head. Thus, any individual observed in 1850 and sought in 1860 could be found if the individual had either set up a separate household or moved into the household of someone with a different surname.

individuals (those with surnames different from that of the household head). As Table 1 shows, Briggs meets this qualification, though his wife, daughters, and son do not. Females like Briggs's wife and daughters were not sought, as it was likely they would remain listed with their husbands or father in the index or change their name upon marriage. Younger individuals like his son James were unlikely to have left their parents' homes by 1860. This restriction resulted in a population of 55,852 males to be located in 1860.

2. *Code each individual's name phonetically and truncate his given name at the fourth letter.* This restriction was imposed to account for minor variations in the spelling of surnames and given names between 1850 and 1860, as well as transcription errors in the creation of the 1850 PUMS and the 1860 census index. Surnames were coded according to the New York State Identification and Intelligence System (NYSIIS) described by Lynch and Arends (1977). This is similar to the Soundex system in that names are coded identically if they sound roughly the same, though the NYSIIS is more likely than the Soundex system to code misspellings and consonant blends identically.[8] Using the NYSIIS system, the surnames "Briggs," "Bragg," "Bruge," and "Brugge" would all be assigned the same code: "BRAG." In the case of Sidney Briggs, then, instead of searching the 1860 index for exact matches to the name "Sidney Briggs," matches were sought for NYSIIS code "BRAG" and the given name "SIDN."

3. *Sort both (a) the list of individuals being sought and (b) the 1860 census index by NYSIIS code and truncated given name and identify all potential matches in (b) for each individual in*

[8] The algorithm used to code names according to the NYSIIS is described in detail in Atack and Bateman (1992, Appendix B).

(a). For Sidney Briggs, there were four such potential matches in the index. At this point, it is not yet possible to determine which, if any, is the Sidney Briggs being sought, since the index contains only an individual's name and location (state, county, subdivision, page) in the 1860 census manuscripts.

4. Discard individuals with more than ten potential matches in the index, and retrieve all potential matches from the 1860 manuscript schedules for the remaining individuals. The rationale for the elimination of those with more than ten potential matches, and its consequences for the representativeness of the sample, are discussed below. For Sidney Briggs, who had only four potential matches in the index, this restriction was not binding. All four potential matches were retrieved from the 1860 census manuscripts.

5. Compare the 1860 characteristics of all the potential matches from the 1860 manuscripts to the 1850 characteristics of the individual being sought. To save space in Table 1, these comparisons are not shown for Sidney Briggs.

6. Discard a potential match if at least one of the following conditions is true:

> *a. the difference between the age reported in 1850 and the age reported in 1860 is more than fifteen years or less than five years;*
>
> *b. the state of birth (or country of birth for the foreign-born) reported in 1860 does not match that reported in 1850; or*
>
> *c. if the individual was a household head with a family in 1850 and there is no family present in 1860, or the names of the family members age 10 or older in 1860 do not match those of the family members present in 1850.*

For Sidney Briggs, condition (a) eliminated the first potential match, while condition (b) eliminated the second and third. Only the fourth could not be eliminated by these criteria. It was not necessary to use condition (c) for the first three potential matches—they were eliminated by the first two conditions.

7. If two or more potential matches remain after step 6, choose the individual for whom the difference between reported age in 1860 and the anticipated age in 1860 (based on the reported 1850 age) is smallest. Since only one potential match was left after step 6 in the case of Sidney Briggs, this step was unnecessary. There were no cases in which multiple potential matches remained after step 6 that could not be reduced to one match by minimizing the age difference.

RATIONALE FOR THE ELIMINATION OF COMMON NAMES

The most time-consuming step in the linkage process was retrieving the information on each potential match from the 1860 census manuscripts. Since no individual characteristic other than surname, given name, and location were provided, it was necessary to locate each potential match in the 1860 manuscripts and compare his characteristics with the corresponding characteristics of the individual from the 1850 PUMS being sought in order to determine whether the right person had been located. For example, there were four potential matches in the 1860 census index for Sidney Briggs, as shown in the middle panel of Table 1. For individuals with common combinations of surname and given name, this meant that several hundred reels of microfilm might have to be searched to verify a single match.

To overcome this difficulty, only those individuals whose combinations of surname and given name appeared ten or fewer times in the 1860 census index were sought in the 1860 manuscript census schedules. An exception was made for individuals whose combination of surname and given name appeared more than once in the base population (the 1850 PUMS): these individuals were sought if there were ten or fewer matches per individual. Thus, if a particular surname/given name appeared three times, those three individuals were sought if there were thirty or fewer matches in the index for that surname/given name combination. This exception was made because five percent of the individuals in the 1850 PUMS had the same combinations of NYSIIS code and truncated given name.

The cut-off point of ten potential matches was chosen on the basis of previous experience in matching using the 1860 index (Ferrie 1992), some experimentation, and the availability of funding for research assistants. The choice of this cut-off generated 98,451 potential matches to be retrieved from the 1860 census manuscripts, which exhausted my budget for research assistants. It is possible to see *ex post* whether the chosen cut-off was a good one. The relationship between the number of potential matches and the "price" of obtaining a successful match can be determined over a fairly wide range of potential matches since many individuals had more than 10 potential matches searched because their surname/given name combination appeared more than once in the 1850 PUMS. By regressing the ratio of potential matches to successful matches against the number of potential matches, the following equation was obtained:

$$\frac{potential\ \ matches}{successful\ \ matches} = e^{1.8817 + (0.2030)(potential\ \ matches) - (0.0022)(potential\ \ matches)^2}$$

The relationship is non-linear, the coefficients on the linear and quadratic terms are statistically significant, and the fit (as measured by the adjusted R^2 of 0.9679) is quite good. This relationship suggests that if the cut-off were doubled from 10 to 20, the "price" for obtaining a match among those whose number of matches is right at the cut-off—the number of individuals whose records would have to be retrieved in order to obtain one successful match—would increase by a factor of four (from 40 to 157 potential matches per successful match). Though raising the cut-off would yield more matches, then, the cost of each match (in terms of the number of potential matches that must be searched to obtain that match) would rise more rapidly than the number of matches.

THE REPRESENTATIVENESS OF THOSE WITH UNCOMMON NAMES

The exclusion of common surname/given name combinations is a shortcut that dates back to Thernstrom's (1964) first study. There has been, to the best of my knowledge, only one attempt to assess the extent of the bias introduced by this procedure: Steckel (1988a, 54) estimated a logistic regression in which the dependent variable was whether a surname/given name combination appeared more than ten times in the state index being searched.[9] He found

[9] Note that this definition of a "common" name is more restrictive (i.e. its use will categorize fewer people as having a common name) than the standard used here, where the ten name limit was imposed at the national level. An individual who had only five matches in

that the only characteristics associated with whether an individual had a common name were location (region and size of location) and nativity. There was no statistically significant association with migration, occupation, or wealth.

The second column of Table 2 shows the characteristics of individuals who had common names by the definition used in this study. Along most dimensions, those with common names appear similar to the general population: where there are differences, they are generally only a few tenths of a percentage point. The most striking differences appear by occupation and location: the fraction of high white collar workers was 0.7 percentage points lower among those with common names, while the fraction of those located in the Pacific region was 0.4 percentage points lower among those with common names. To assess the independent impact of an individual's characteristics on the representativeness of the sample and the magnitude of any bias introduced by the exclusion of those with common names, I estimated a logistic regression in which the dependent variable is equal to one if an individual's combination of surname and given name appeared more than ten times in the 1860 census index and equal to zero otherwise. The results are presented in the first column of Table 3.

The results show that there are some statistically significant relationships between an individual's characteristics and the probability that he had a common name. The sizes of the marginal effects, however, are generally quite small relative to the predicted mean

---

the Georgia index, for example, and was classified as having an uncommon name using Steckel's standard, will be classified as having a common name here if he had just six matches in the rest of the country.

probability.[10] For example, though the coefficients on the linear and quadratic age terms are both statistically significant at the 99% level, the impact of age on the predicted probability is small: as Figure 1 shows, the probability declines from 54% at age 20 to 52% at age 38, and then rises to 57% at age 65. The range of five percentage points represents just 9% of the total predicted probability of 54.2% that the average individual had a common name. The same is true for the other continuous variables: family size and real estate wealth.[11] Among the categorical variables, only residence in the Pacific region (20 percentage points, or 37

[10] The marginal effect for each independent variable (the impact of a one unit change in the variable on the probability that the dependent variable equals one) was calculated by multiplying the estimated logistic regression coefficients by $(\overline{P})(1-\overline{P})$ where $\overline{P}$ is the probability that the dependent variable is equal to one, evaluated at the sample means for each independent variable:

$$\frac{\partial P}{\partial X_j} = \beta_j(\overline{P})(1-\overline{P}) = \beta_j \frac{e^{\overline{X}\beta}}{(1+e^{\overline{X}\beta})^2}$$

where $\overline{X}$ is a vector of average values for the independent variables, β is a vector of logistic regression coefficients, X_j is a particular variable j, and β_j is its coefficient. Marginal effects were also calculated as the average of the marginal effects for all individuals in the sample. For each individual i, the marginal effect for variable j was calculated as $\beta_j(P_i)(1-P_i)$ where P_i is the probability that the dependent variable is equal to one for individual i, evaluated at the values of each independent variable for that individual:

$$\frac{\partial P}{\partial X_j} = \sum_{i=1}^{N}\left(\frac{\partial P_i}{\partial X_j}\right)\left(\frac{1}{N}\right) = \sum_{i=1}^{N}\beta_j(P_i)(1-P_i) = \sum_{i=1}^{N}\beta_j\frac{e^{X_i\beta}}{(1+e^{X_i\beta})^2}\left(\frac{1}{N}\right)$$

where X_i is a vector of values for the independent variables for individual i, β is a vector of logistic regression coefficients, X_j is a particular variable j, β_j is its coefficient, and N is the number of observations in the sample. The results were identical to the third decimal place.

[11] Someone with a wife and four children was 2.7 percentage points more likely to have a common name than an unmarried individual, a difference of 5 percent. Though the coefficient on real wealth was statistically insignificant, the point estimate suggests that someone with $10,000 of real estate wealth in 1850 was 0.02 percentage points less likely to have a common name than someone with no in real estate in 1850, a difference of 0.03 percent.

percent, less likely to have a common name) and a high white collar occupation (7 percentage points, or 13 percent, less likely to have a common name) had a substantial impact. None of the other categorical variables had an impact greater than 4.5 percentage points or 7.7 percent of the predicted probability.

The effects of residence on the west coast and high white collar occupation are puzzling. An examination of the names recorded in the PUMS suggests that individuals in this region and this occupational category were significantly more likely to have their first and middle initials reported by the census marshal rather than their given name.[12] Since only 1.2 percent of those in the 1850 PUMS were living on the west coast, this probably causes only a slight reduction in the representativeness of the overall sample. In high white collar occupation, the reporting of initials rather than full names was most likely for clergymen, lawyers, and doctors, and less common among proprietors (the largest group in this category) and teachers. Since clergymen, lawyers, and doctors comprised only 0.5 percent of the observations in the 1850 PUMS, it is again unlikely that the bias in the overall sample resulting from the concentration on uncommon surname/given name combinations is great.

EXPLAINING THE LINKAGE RATE

After eliminating 54% of the 1850 PUMS from further consideration because of common names, the remaining 25,586 individuals were sought in the 1860 census index. Of

[12] Individuals in the Pacific region were twice as likely to report initials rather than a given first name as individuals in any other region. Those in high white collar occupations were more than twice as likely to do so as those in any other occupation.

these, 11,274 (44%) had no potential matches in the 1860 census index and could not be pursued further. The remaining 14,312 individuals were sought in the 1860 manuscript schedules. This generated 98,451 potential matches, which yielded 4,938 definite matches. Since 19 percent of those with uncommon names were successfully matched, we must account for the individuals who were not matched and say how they differ from the matched population. The potential sources of linkage failure fall into two categories: mortality and problems in the 1850 census, the 1850 PUMS, the 1860 census, or the 1860 census index.

We know relatively little about the mortality experienced by antebellum Americans, and we know even less about the mortality of particular groups (native-born or immigrant) or those living in particular places (different regions, urban places, or rural places). Life tables for this period have been created using data from Massachusetts and Maryland (Jacobson 1957; Vinovskis 1972). National life tables for the entire white male population have also been constructed from information contained in the mortality schedules of the federal census of population and from information contained in family histories (Haines 1979; Pope 1992). These life tables suggest that perhaps as many as 12 percent of those in the 1850 PUMS may have died between 1850 and 1860.[13] Estimates of mortality based on census survivorship methods suggest that mortality rates may have been as high as 21 percent (McClelland and Zeckhauser 1984).

Though mortality differences by region, urban/rural location, and nativity might result in differences in linkage failure rates, existing antebellum life tables do not provide mortality

[13] These calculations are shown in Table A-1 in the Appendix.

information disaggregated along these dimensions. I constructed new life table estimates using a sample of 30,000 decedents from the manuscript mortality schedules of the 1850 federal census. These calculations are described in detail in the Appendix. These new life tables were applied to the distribution of ages, regions, urban/rural locations, and nativities for the males in the 1850 PUMS at risk to be linked to the 1860 census. The results (shown in Table A-1 in the Appendix) suggest that survival rates in the 1850s may have been as low at 76 percent (for immigrants in urban places) and as high as 89% (for the native-born is the rural Northwest). The overall survival rate was 85 percent. For the present purpose, I will assume that no more than 80% of the males age 10 and over in the 1850 PUMS survived to 1860.

For an individual who survived to 1860, the chances of being successfully linked depended next on the accurate transcription of the three characteristics used to verify matches in each of three sources: the 1850 manuscripts, the 1850 PUMS, and the 1860 manuscripts. If a name, age, or birthplace was mis-reported in any of these three sources, the probability of linkage would be reduced by an amount corresponding to the severity of the error. The linkage procedure was designed to eliminate some of these of errors. The use of NYSIIS codes for surnames and the use of only the first four letters of the given names should reduce the linkage failures from inaccurate name reporting and transcription. To reduce the errors resulting from inaccurate age information, a tolerance of 10 years was used in comparing the expected age in 1860 of an individual from the 1850 PUMS with the age of a potential match in the 1860 manuscripts (i.e. an absolute difference between expected and actual ages of 10 or fewer years was acceptable for a match along this dimension). Some leeway was also allowed

in the comparison of birthplaces, particularly for immigrants who may have reported a very specific birthplace in one source ("County Mayo, Ireland"), but only a very broad description in another ("U.K.").

These procedures, however, cannot account for more serious errors. Individuals who were omitted entirely from the 1860 census, for example, could never be linked. Steckel (1991) reports a range of estimates for the under-enumeration of adult males in the federal censuses of 1850 and 1860, from 9 to 19 percent (1991, 588). It seems reasonable to assume that at least 15 percent of the population may have been missed by these censuses.[14] Knights (1971) notes that 11 percent of the respondents in Boston in 1850 and 1860 gave age differences of five or more years. Steckel (1988a) reports that literacy information was inconsistently reported by 7.3 percent of respondents in another sample. It seems reasonable as a first approximation to use 85% as the enumeration rate for the 1860 census and 95% as the rate of accurate reporting for name, age, or birthplace in each of the 3 sources.

The next source of linkage failure is the 1860 census index. My previous calculations suggest that the accuracy rate was 95 percent for household heads.[15] But not all individuals

[14] The 19 percent rate was for males traced through family genealogies and sought in the 1850 census. The 9 percent figure was for state legislators in the Upper South who were sought in the 1860 census. Other studies examined 1850 voters in Clinton Township, Ohio (12 percent), state legislators in the Lower South in 1850 (15 percent), state legislators in the Upper South in 1850 (14 percent), Schoharie County in New York in 1850 (15 percent), Boston in 1850 (12 to 15 percent), Alexandria, Virginia (12 percent), Washington County, Oregon in 1860 (16 percent), state legislators in the Lower South in 1860 (15 percent), and Boston (14 percent). For citations, see Steckel (1991, 588).

[15] Ferrie (1992, 42-43). Random samples of 100 individuals were drawn from the 1860 manuscript census schedules of eight counties. Of these, 98.5% were subsequently located in the census index.

from the 1850 PUMS sought in the 1860 census were household heads and therefore eligible to appear in the census index in 1860. Since the individuals sought were as young as 10 years of age in 1850, and may not have set up independent households by 1860, and since some were old enough to have retired and perhaps moved into the homes of siblings or children by 1860, perhaps only 70 percent of those who survived to 1860 and had their characteristics successfully recorded would have appeared in the census index.[16]

Table 4 shows the cumulative effect of these sources of linkage failure on the overall linkage rate. To simplify the calculation, I assume that the sources of error are independent, so the overall probability is simply the product of the individual probabilities. Two scenarios are presented: an optimistic scenario using the somewhat conservative assumptions regarding the magnitude of the errors at each stage (which projects a linkage rate of 28 percent) and a pessimistic scenario which uses more liberal assumptions regarding the errors and reduces the accuracy rates at each stage by five percentage points (and projects a linkage rate of 15 percent). The actual linkage rate (19 percent) is close to the mid-point of this range (21 percent), and thus seems reasonable.

THE REPRESENTATIVENESS OF THE LINKED SAMPLE

[16] In the 1850 PUMS, 9 percent of 20 year old males were living outside the homes of their parents, as either independent household heads or residents of group quarters. This percentage increased steadily with age, reaching 48 percent at age 25 and 70 percent at age 30. By applying the fraction living outside their parents' homes at each age from the 1850 PUMS to the age distribution that would prevail in 1860 if everyone survived to that date, it is possible to estimate the fraction of the males age 10 and over in 1850 who were at risk to appear in the 1860 census index, in the absence of a mortality effect. This calculation reveals that only 70% could have appeared in the 1860 index.

Though it is possible that the sources of linkage failure are uncorrelated with an individual's characteristics, other studies suggest that some groups (immigrants, the unskilled, the poor, illiterates) were less likely to be enumerated at all or less likely to be enumerated accurately. Age must have also had an effect on linkage: the probability of survival decreased with age, and the probability of appearing in the census index increased until middle age and then may have decreased. One way to see whether individual characteristics are associated with successful linkage is to compare the 1850 characteristics of those who were linked and those who were not yet linked. Table 2 shows these comparisons for the full 1850 PUMS (column 1) and the those linked to 1860 (column 4). The second column of Table 3 presents a logistic regression on the sample of individuals from the 1850 PUMS with uncommon names; the dependent variable is equal to 1 if a successful match was made and equal to zero otherwise.

As expected, the probability of a successful link falls with age in 1850. Figure 1 shows this effect. It no doubt reflects mortality and perhaps also individuals moving into other households at older ages. Occupation has a negligible impact on linkage: farmers were the most likely to be linked, followed by low white collar workers, craftsmen, unskilled workers, and high white collar workers, though only the effect for farmers is statistically significant. The high rate of linkage among farmers probably reflects their prominence in the community's geography and their low migration rates: it was more difficult for a census enumerator to miss someone living for a lifetime on a hundred acre farm than it was to miss a transient clerk or laborer in a boardinghouse.

The impact of 1850 region of residence on the probability of a successful match was both large and statistically significant. Within the North, linkage rates were high among those who began the 1850s in New England and the Middle Atlantic states, lower among those in the Midwest, and lowest for those in the West and on the Pacific coast. Within the South, rates were higher in the Southeast than in the South Central states. These differences may reflect the ease of enumeration in longer-settled, more densely populated places. Regional differences in linkage rates may also reflect out-migration and mortality.[17] Linkage was less likely among those who began the 1850s in places with populations greater than 2,500 persons than among those in rural places.

The probability of successful linkage was also related to nativity, marital status, literacy, and previous migration. The impact of nativity may reflect levels of immigrant mortality higher than those experienced by the native born (like those seen in Table A-1), or lower census enumeration rates for immigrants than for the native-born. Literate individuals were probably more likely to report their characteristics accurately to census marshals. Individuals who were married had, for the most part, left the homes of their parents; they were therefore more likely to appear in the 1860 census index than unmarried individuals. Those who were living outside their state of birth in 1850 may have suffered higher mortality than those who remained within their state of birth, or may have been more likely to migrate between 1850 and 1860; if those who migrated between 1850 and 1860 were more difficult to

[17] For example, if migrants were more difficult to locate than non-migrants, regardless of the region in which they lived in 1850, and out-migration rates differed across regions, linkage rates would be lower in places with high rates of out-migration.

link than those who remained in their 1850 community, and those who migrated prior to 1850 were more likely to relocate subsequent to 1850, we would expect a lower linkage success rate among this group. There were no substantively or statistically significant effects of family size or real estate wealth on the probability of linkage.

COMPARISONS WITH OTHER SAMPLES

Another way to assess the success of the linkage procedure is to compare the characteristics of those in the linked sample with other samples from the same period. Table 5 compares the 1850-60 Linked PUMS with the national sample of males age 20 and over in 1850 and 1860 collected by Soltow (1975), the national linked sample created by Steckel (1988a, 1988b), and the linked sample of men located in 1860 in the New South and on the southern frontier created by Schaefer (1985). All observations in the Linked PUMS are used for the comparisons with Soltow and Steckel; only those who were located in the New South and on the southern frontier in 1860 are used for the comparison with Schaefer.

Three aspects of Table 5 are particularly noteworthy: (1) the sharp differences between the 1850 characteristics in the Linked PUMS and those in Soltow's sample; (2) the close correspondence between the 1860 characteristics in the Linked PUMS and those in Soltow's sample; and (3) the sharp differences between the Linked PUMS and either Steckel's sample (comparing 1860 characteristics) or Schaefer's sample (comparing 1850 characteristics). The individuals in the Linked PUMS in 1850 are younger and less likely to be farmers than males in Soltow's sample for 1850; by 1860, though, the two samples look quite similar in these

respects.[18] By contrast, individuals in the Linked PUMS are younger and less likely to be farmers than individuals in either Steckel's sample (using 1860 values) or Schaefer's sample (using 1850 values). This suggests the extent of the bias introduced in the latter samples by the need to construct them through backward-linkage.

A more subtle sense of the bias can be seen in comparing the rates of migration over the 1850s in these linked sample with the corresponding rates in the Linked PUMS. In Steckel's national sample, 19 percent made an interstate move between 1850 and 1860, while 28 percent did so in the Linked PUMS. In Schaefer's sample, 28 percent made an intercounty move during the 1850s, while 55 percent in the Linked PUMS did so. This suggests that both Steckel's and Schaefer's samples, in concentrating on those who had already established households by 1850, miss a significant amount of migration, which occurs disproportionately among the young. The fraction of those who had made a previous interstate move by 1850 supports this view: the fraction is higher in Schaefer's sample than in the Linked PUMS, since more of the males in Schaefer's sample are older and past the years when their geographic mobility was great; many of the males in the Linked PUMS are just entering the years of greatest geographic mobility.

Since Soltow's sample was a random sample of males age 20 and over, while Steckel's and Schaefer's samples were composed of men who had established their own households by 1850, these comparisons are not surprising. They indicate that the 1850-60 Linked PUMS

[18] The fraction native-born in 1860 is, however, somewhat larger in the Linked PUMS than in Soltow's 1860 sample. This no doubt reflects the heavy immigration that occurred in the early 1850s. Since the linked PUMS includes only those present in the U.S. in 1850, it does not reflect this post-1850 immigration.

looks quite similar to a random sample of the nation's adult males who were present in the U.S. in 1850 and survived the 1850s. By 1860, the Linked PUMS looks wealthier and more egalitarian than a cross-section of adult males would look, but this no doubt reflects the fact that only males who had established independent households by 1860 were successfully linked from 1850 to 1860. As a result, the lower tail of the distribution of wealth for males age 20 and over in 1860 is smaller in the Linked PUMS than in Soltow's 1860 sample. The Linked PUMS nonetheless looks more like Soltow's nationally representative sample in terms of average wealth, wealth inequality, age, occupation, and nativity than it does like Steckel's or Schaefer's, the best views we have had until now of the experiences of American males over the 1850s.

IV. AN EXAMPLE: GEOGRAPHIC PERSISTENCE

The sample described here provides a new perspective on several aspects of nineteenth century U.S economic development. One question it seems uniquely suited to address is internal migration in the antebellum U.S. Since the pioneering work of Malin (1935), Curti (1959), and Bogue (1963), much attention has been focused on this topic. These early studies focused on the farming frontier, but later work by Thernstrom (1964) focused on urban places as well.[19] In each case, two central questions were addressed: how many people moved and who was most likely to move? The answers have varied considerably from one study to the next.

[19] More recent studies of rural communities in this tradition include Barron (1984), Winkle (1988), and Faragher (1986). Other studies of urban places include Thernstrom (1973), Blumin (1968), Katz (1975), and Griffen and Griffen (1978).

In some places, fewer than half and sometimes fewer than a third of those examined remained in the same city or county for a decade or more (Curti 1959; Blumin 1968; Faraghar 1986; Galenson 1991). Such places included both eastern (Philadelphia) and western (Chicago) cities, and frontier communities (Trempeleau County, Wisconsin and Sugar Creek, Illinois). In other places, more than half of the population remained over a decade or more. This was the case in places as diverse as Hamilton, Iowa (Bogue 1963), East Central Kansas (Malin 1935), Holland, Michigan (Kirk and Kirk 1974), Poughkeepsie, New York (Griffen and Griffen 1978), and Indianapolis (Barrows 1977).

This line of research has run into three difficulties. The first is the inability to generalize from the experience in one city or county. The only clear conclusion to emerge from these studies is that "persistence varied with the nature of social and economic development . . . [and] it is critical to describe the context—the nature of the place—in which any study is set" (Katz et al. 1978, 672). It is difficult to say anything more general about the mobility of nineteenth century Americans. Though the research methodology was similar in all these studies (generally, creating a sample of individuals from census manuscript schedules and seeking the same individuals in the manuscripts of a subsequent enumeration for the same location), each emphasized circumstances in the particular place examined rather than more universal patterns. Information on a wider range of locations might make such general patterns more apparent.

The second difficulty is in the interpretation of non-persistence. Researchers have recognized that non-persistence in these studies (failure to re-appear in a subsequent enumeration for the same location) is not the same as migration. Galenson and Levy (1986)

have pointed out how mortality and census under-enumeration can cause a divergence between non-persistence rates and migration rates. Since mortality rates and enumeration rates can vary by community, even a large number of studies conducted using the methodology of studies like Thernstrom's might not say much about national patterns of internal migration: variation across communities in mortality and enumeration can obscure underlying common migration patterns.

Finally, this line of research has been unable to address two questions that are arguably more interesting than how many people moved and who was most likely to move: where did people go when they moved, and how successful were they at their new locations? Studies based on particular communities have provided only conjectures here. For example, Thernstrom has suggested that internal migrants in the nineteenth century comprised a "*floating proletariat* . . . of men ever on the move spatially but rarely winning economic gains as a result of spatial mobility" (1973, 42 [italics added]). Thernstrom's conjecture seems implausible in light of contemporary evidence on the link between geographic mobility and economic improvement, through either upward occupational mobility or improved incomes.[20]

The Linked PUMS provides information on the migration of people from a wide range of locations, overcoming some of the parochialism of community-based studies. It includes only individuals who were alive in both 1850 and 1860, so it provides a measure of migration

[20] See note 2 above for studies showing the improvements in occupation and income enjoyed by contemporary migrants. The link between geographic mobility and occupational and financial mobility among antebellum European immigrants after their arrival in the U.S. is examined in Ferrie (1995a).

that is more accurate than persistence rates calculated for particular locations.[21] Since the sample follows individuals wherever they moved over the 1850s, it also tells us where people were moving, and what kinds of people moved to different locations. Finally, since it provides information on wealth and occupation in 1860 for both non-migrants and migrants, it also allows us to say how the successes of migrants compared to those of non-migrants over the decade.

Table 6 presents some preliminary results on internal migration using the Linked PUMS. The table shows "conditional" persistence rates: the ratio of individuals observed in the same county in 1850 and 1860 to the total number of individuals who were located in both 1850 and 1860. This rate is thus "conditional" on survival to 1860 and enumeration in the 1860 census. The overall persistence rate of 53 percent is somewhat higher than the average persistence rate of 40 percent calculated by Parkerson (1982) from community-based studies. This suggests the potential magnitude of the discrepancy between non-persistence rates and migration rates resulting from mortality and census under-enumeration in previous studies.

The table provides separate persistence rates by occupation. Clearly, males who entered the labor force between 1850 and 1860 and unskilled laborers were the most geographically mobile: fewer than half remained in the same county over the 1850s. At the other extreme, nearly two thirds of farmers were persisters. White collar and craft workers

[21] The Linked PUMS is not entirely free from the influence of mortality, however. If mortality rates varied across locations, the sample would be more likely to catch migrants from high mortality places and to low mortality places than non-migrants or migrants from low mortality places to high mortality places. In the calculations which follow, the reader should bear in mind that migration between 1850 and 1860 refers to the population *at risk to migrate*: individuals who survived the 1850s.

were somewhere in the middle. This pattern of persistence by occupation can be seen in every region except the Southeast, where laborers were less likely to persist than farmers, but white collar and craft workers were even less likely than laborers to persist.

These patterns are consistent with several explanations. One recognizes that some investments are more location-specific than others. A farmer who has purchased or inherited land in a particular community and learned how to raise crops under those conditions may be reluctant to relocate where his existing skills are less applicable.[22] In the same way, a craftsman or shopkeeper might have built up a clientele in a particular place, an asset that would be impossible to transport to a new location. Because they had made fewer such location-specific investments, new labor market entrants and unskilled workers may have been more willing to relocate in response to relatively small expected improvements in lifetime income.[23] A less sanguine view of the operation of the antebellum economy suggests an alternative explanation: the unskilled were forced to relocate because they had not acquired the skills that ensured success or because they did not have the good fortune to inherit a farm. Their movement reflected desperation more than optimism. These explanations are currently under investigation (Ferrie 1995b).

---

[22] This is the explanation Steckel (1983) offers for the rigid patterns of east-to-west migration he observes between 1850 and 1860.

[23] A variant of this explanation relies on differences in age by occupation: craftsmen, white collar workers, and farmers were generally older on average than new labor market entrants and unskilled laborers. Differences in migration rates may simply reflect an individual's stage in the life cycle. Laborers, in this view, are more mobile than farmers because they are younger; in time, they will make location-specific investments as well, and their mobility rate will fall as a result.

Table 6 also reveals substantial differences in persistence rates by size of location and region. Persistence rates were lowest in the largest cities: places of 50,000 or more had persistence rates below 40 percent in every region except the Southeast. For new labor market entrants, the persistence rate in large cities was even lower: fewer than one in five new entrants remained in large cities between 1850 and 1860. Places with smaller population had higher persistence rates, though there is some evidence that this relationship was not monotonic: it appears that rural places may have had slightly lower persistence rates than places with populations over 2,500 and under 10,000. Across regions, persistence rates generally fall moving from east to west: within the North, they are highest in New England and the Middle Atlantic states, and lower in the Midwest; within the South, they are higher in the Southeast than in the South Central states. These patterns appear to hold across occupations as well, with one exception: in the South, skilled and white collar workers were less likely to persist in the Southeast than in the South Central states.

Some of the earliest research on internal mobility in the U.S. was directed at answering a specific question: how prominent were unskilled urban laborers in the movement from East to West? The view of the frontier advanced by Turner suggested that the westward migration of urban laborers was an important outlet for dissatisfaction with conditions in eastern cities. Goodrich and Davidson (1935, 1936) were among the first to subject Turner's view to empirical scrutiny. They examined newspaper accounts and local records for Fall River, Massachusetts, and other New England urban centers, seeking evidence of the exodus of

urban laborers to the west that Turner predicted.[24] They found no support for Turner's view: "This cumulation of evidence thus points to the conclusion that the movement of eastern wage-earners to western lands was surprisingly small. Too few industrial workers reached the frontier to attract notice in the accounts of settlement" (1936, 114). The Linked PUMS can address this question, since it reveals the specific locations reached by internal migrants over the 1850s.

Figure 2 provides some insight into this issue. It shows the 1860 location of individuals who left their 1850 county of residence. The sample is restricted to those who lived in New England, the Middle Atlantic states, and the Southeast in 1850. To assess the importance of internal migration to urban laborers, separate tabulations are shown by size of 1850 location and occupation. The most striking finding is the large fraction of urban laborers who moved to the frontier.[25] In small towns and rural places, roughly 3 to 4 percent of those in each occupational class (white collar and craft workers, farmers, and unskilled laborers) moved to the frontier during the 1850s. Among unskilled workers located in places larger

[24] They adopted their largely anecdotal approach after noting the then insurmountable difficulties inherent in using the federal census manuscripts for the purpose of determining who was moving to the frontier: ". . . the Census and the records of the General Land Office throw little light upon the precise question at issue . . . at least short of the task of taking the names of individual wage-earners in a given census and then searching for them in the haystacks of succeeding enumerations" (1936, 62). Thernstrom (1964, 86) was also aware of the usefulness of data linking individuals across censuses, but, like Goodrich and Davidson, lacked the resources to create such data.

[25] The frontier was defined for this exercise as any location west of 90º longitude: Minnesota, the Dakota Territory, western Wisconsin, western Illinois, Iowa, Kansas, Nebraska, Colorado, the western two thirds of Missouri, Arkansas, Louisiana, Texas, Arizona, New Mexico, and the Pacific coast (California, Oregon, and Washington).

then 10,000 on the eastern seaboard, however, 12 percent had moved to the frontier by 1860. Of the white collar and craft workers in such cities, only 4 percent had made such a move by 1860.[26] Though this finding is not sufficient to overturn 60 years of scholarship purporting to show the irrelevance of the frontier for urban laborers, it does suggest that previous studies may have missed a significant stream of migration to the west.

V. CONCLUSION

Despite the numerous steps in the linkage process and the low overall linkage rate, the Linked PUMS provides our best view yet of the economic mobility of average Americans during the 1850s. The ability to follow large numbers of individuals wherever they moved over a decade is particularly important for the 1850s: during this decade, the nation's center of population moved a greater distance west than during any other decade in the nineteenth century.[27] The range of locations to which people could have moved thus expanded enormously over this decade, making generalizations regarding the fate of non-persisters even more suspect for the 1850s than they are for other decades.

The last decade has seen a marked decline in the volume of community-based studies of persistence and social mobility. Some of the disfavor into which such work has fallen may be the result of a nagging suspicion on the part of many that a significant part of the story of

[26] In small towns and rural places (population under 10,000), the cell sizes were: white collar and craft workers, 556; farmers, 1069, laborers, 1291. In places of 10,000 or more people, the cell sizes were: white collar and craft workers, 206; laborers, 206.

[27] The center moved from 23 miles southeast of Parkersburg in present-day West Virginia to 20 miles south of Chillicothe, Ohio (U.S. Census Office 1901, 2).

mobility was being missed in the exclusive concentration on those who remained in a particular city or county. Perhaps the ability to say more about the fate of non-persisters using samples like the one described here will help revitalize an area that has provided important insights into the lives of average Americans in the nineteenth century.

APPENDIX: ESTIMATES OF WHITE MALE LIFE EXPECTATION IN THE U.S. FOR 1850 BY REGION, URBAN/RURAL, AND NATIVITY

As part of the 1850 federal census of population, census marshals were instructed to ask at each household they visited whether any member of that household had died in the preceding twelve months. The published totals based on this question have been examined by Condran and Crimmins (1979) who concluded that these data need to be used with care. The enumeration of deaths was probably better in areas that had been settled longer, better after infancy and before old age, and more accurate for deaths that occurred closer in time to the date of the census.

Since the published mortality totals did not provide breakdowns of life expectancy by region, size of location, and nativity, and these were necessary to estimate the fraction of the 1850 PUMS that should have survived to 1860, I used a sample of the data from the manuscript mortality schedules to provide new estimates of life expectation along these dimensions.[28] In order to reduce the errors resulting from the retrospective process by which the data were collected, the life expectations were based only on those who died in the 6 months immediately before the census.[29]

---

[28] The sample was created by Accelerated Indexing Systems, and for 1850 consists of 30,713 decedents in 593 counties in 17 states. The sample contains 18 counties (7 urban) in the Northeast, 303 counties (2 urban) in the South. Urban counties had an 1850 population of 35,000 or more. The 17 states were: Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, Tennessee, Texas, and Virginia (South); Connecticut, Delaware, and Vermont (Northwest).

[29] Each death in these months was multiplied by two to return the mortality rate to a full-year basis. A second adjustment to the mortality-by-age totals in each county was made necessary by the absence of records for the deaths of some individuals from the sample. The shortfall of deaths (the difference between the published total deaths for the white population and the total in the sample) was apportioned to each age, sex, and nativity group in the county according to that group's share in the region's population. The population shares were derived

The life expectations estimated for prime-age males are shown in Table A-1. Rates for infants and the elderly (not shown) are certainly too high. For the present purpose, the first of these is not a problem, since we are interested in the survival of males age 10 and older in 1850. The understatement of the mortality rate for the elderly will be overcome by assuming a 100 percent mortality rate by 1860 for those age 65 and over in 1850 when estimating the number of survivors from the 1850 PUMS. The life expectancies shown in Table A-1 are probably too high for the urban Northeast (since the sample's urban places in the Northeast do not include the largest cities, such as New York, Philadelphia, or Boston) and too low for the urban South (since one of the ten urban counties in the South is Orleans Parish, which contains the City of New Orleans, an unusually unhealthy place). The differences between the native-born and immigrants seen reasonable.

The differences between these new rates and those previously calculated for the antebellum period may reflect shortcomings of the sample of mortality records used here, but those differences might also reflect differences in coverage. For example, the figures for the Northeast presented here are identical to Jacobson's (1957) estimates from Massachusetts and Maryland when the urban and rural populations are combined using their population shares as weights. The weighted life expectations for natives and immigrants are only a half year greater at age 20 than the expectation from Pope's (1992) sample.

REFERENCES

Atack, J., and F. Bateman. 1992. “Matchmaker, matchmaker, make me a match”: a general personal computer-based matching program for historical research. *Historical Methods* 25: 53-65.

Barron, H.S. 1984. Those who stayed behind: Rural society in nineteenth-century New England. New York: Cambridge University Press.

Barrows, R.G. 1977. *A demographic analysis of Indianapolis, 1870-1920*. Ph.D diss., Indiana University.

Bartel, A. 1979. The migration decision: What role does job mobility play? *American Economic Review* 69: 775-786.

Blumin, S.M. 1968. *Mobility in a nineteenth-century American city: Philadelphia, 1820-1860*. Ph.D. diss., University of Pennsylvania.

Borjas, G.J. 1987. Self-selection and the earnings of immigrants. *American Economic Review* 77: 531-553.

Borjas, G.J., S.G. Bronars, and S.J Trejo. 1992a. Self-selection and internal migration in the United States. *Journal of Urban Economics* 32: 159-185.

--------. 1992b. Assimilation and the earnings of young internal migrants. *Review of Economics and Statistics* 74: 170-175.

---

from the 1850 PUMS. The missing records resulted from illegibility and missing pages in the manuscripts.

Bogue, A. 1963. *From Prairie to Cornbelt*. Chicago: University of Chicago Press.

Chiswick, B.R. 1978. The effects of Americanization on the earnings of foreign-born men. *Journal of Political Economy* 86: 897-921.

--------. 1994. The performance of immigrants in the United States labor market. In *Economic Aspects of International Migration*, edited by Herbert Giersch. Berlin: Springer-Verlag.

Coale, A., and P. Demeny. 1966. *Regional model life tables and stable populations.* New York: Academic Press.

Condran, G.A., and E. Crimmins. 1979. A description and evaluation of mortality data in the federal census: 1850-1900. *Historical Methods* 12: 1-23.

Curti, M. 1959. *The making of an American community: A case study of democracy in a frontier county* . Stanford: Stanford University Press.

Davenport, D. 1984. Tracing rural New York's out-migrants, 1855-1860. *Historical Methods* 17: 59-67.

Duncan, G.J. 1984. *Years of poverty, years of plenty: The changing economic fortunes of American workers and families*. Ann Arbor: Institute for Social Research.

Faragher, J.M. 1986. *Sugar Creek: Life on the Illinois prairie*. New Haven: Yale University Press.

Ferrie, J.P. (1992), *"We are yankeys now": The economic mobility of antebellum immigrants to the U.S.* Ph.D. diss., University of Chicago.

--------. 1995a. Up and out or down and out: Immigrant mobility in the antebellum United States. *Journal of Interdisciplinary History* 26: 33-55.

--------. 1995b. Westward who? Internal migration and economic opportunity in antebellum America. Unpublished working paper.

Galenson, D.W. 1991. Economic opportunity on the urban frontier: Nativity, work, and wealth in early Chicago. *Journal of Economic History* 51: 581-603.

Galenson, D.W., and D.S. Levy. 1986. A note on biases in the measurement of geographic persistence rates. *Historical Methods* 19: 171-79.

Galloway, L.E., and R.K. Vedder. 1971. Mobility of native Americans. *Journal of Economic History* 31: 613-649.

Goodrich, C., and S. Davidson. 1935. The wage-earner in the westward movement: I. *Political Science Quarterly* 50: 161-185.

--------. 1936. The wage-earner in the westward movement: II. *Political Science Quarterly* 51: 61-116.

Griffen, C., and S. Griffen. 1978. *Natives and newcomers: The ordering of opportunity in mid-nineteenth century Poughkeepsie*. Cambridge: Harvard University Press.

Guest, A. 1987. Notes from the National Panel Study. *Historical Methods* 20: 63-77.

Haines, M.R. 1979. The use of model life tables to estimate mortality for the United States in the late nineteenth century. *Demography* 16: 289-312.

Jackson, R.V. (1982). *Index to the seventh census of the United States*. Salt Lake City: Accelerated Indexing Systems International.

--------. 1992. *Index to the eighth census of the United States*. Salt Lake City: Accelerated Indexing Systems International.

Jacobson, P.H. 1957. An estimate of the expectation of life in the United States in 1850. *Milbank Memorial Fund Quarterly* 35: 197-201.

Katz, M.B. 1975. *The people of Hamilton West, Canada: Family and class in a mid-nineteenth century city*. Cambridge: Harvard University Press.

Katz, M.B., M.J. Doucet, and M.J. Stern. 1978. Population persistence and early industrialization in a Canadian city: Hamilton, Ontario, 1851-1871. *Social Science History* 2: 208-229.

Kirk G., and C.T. Kirk. 1974. Migration, mobility, and the transformation of the occupational structure in an immigrant community: Holland, Michigan, 1850-80. *Journal of Social History* 7: 142-164.

Knights, P. 1971. *The plain people of Boston, 1830-1860: A study in city growth*. New York: Oxford University Press.

--------. 1991. *Yankee Destinies*. Chapel Hill: University of North Carolina Press.

Lynch, B.T., and W.L Arends, W.L. 1977. Selection of a surname coding scheme for the SRS record linkage system. Sample Survey Research Branch, Research Division Statistical Reporting Service, U.S. Department of Agriculture. Unpublished, Washington, D.C.

Malin, J. 1935. The turnover of farm population in Kansas. *Kansas Historical Quarterly* 20: 339-372.

McClelland, P., and R. Zeckhauser. 1983. *Demographic dimensions of the new republic: American interregional migration, vital statistics, and manumissions, 1800-1860*. Cambridge: Cambridge University Press.

Meech, L. 1898. *Systems and Tables of Life Insurance*. New York.

Parkerson, D. 1982. How mobile were nineteenth-century Americans? *Historical Methods* 15: 99-109.

Polachek, S.W., and F.W. Horvath. 1977. A life-cycle approach to migration: analysis of the perspicacious peregrinator. *Research in Labor Economics* 1: 103-149.

Pope, C.L. 1992. Adult mortality in America before 1900: A view from family histories. In *Strategic Factors in Nineteenth Century American Economic History*, edited by C.D. Goldin and H. Rockoff. Chicago: University of Chicago Press.

Schaefer, D. 1985. A statistical profile of frontier and new south migration, 1850-1860. *Agricultural History* 59: 563-578.

--------. 1994. U.S Migration, 1850-50. In *American economic development in historical perspective*, edited by T.Weiss and D. Schaefer. Stanford: Stanford University Press.

Schlottmann, A.M., and H.W. Herzog 1984. Career and geographic mobility interactions: Implications for the age selectivity of migration. *Journal of Human Resources* 19: 72-86.

Simkovich, B. 1993. Long-term trends in American inter-generational occupational mobility. Unpublished paper.

Soltow, L. 1975. *Men and Wealth in the United States, 1850-1870*. New Haven: Yale University Press.

Steckel, R. 1983. The economic foundations of east-west migration during the 19th century. *Explorations in Economic History* 20: 14-36.

--------. 1988a. Census matching and migration: A research strategy. *Historical Methods* 21: 52-60.

--------. 1988b. Poverty and prosperity. Unpublished, Columbus, Ohio.

--------. 1991. The quality of census data for historical inquiry: A research agenda," *Social Science History* 15: 579-599.

Steckel, R., and J. Krishnan. 1993. Wealth mobility in America: A view from the National Longitudinal Survey. Unpublished, Columbus, Ohio.

Stephenson, C. 1978. *Migration and mobility in late nineteenth- and early-twentieth century America* . Ph.D. diss., University of Wisconsin.

Thernstrom, S. 1964. *Poverty and progress: Social mobility in a nineteenth century city.* Cambridge: Harvard University Press.

--------. 1973. *The other Bostonians: Poverty and progress in the American metropolis, 1880-1970* . Cambridge: Harvard University Press.

Turner, F.J. 1920. *The frontier in American history*. New York: H. Holt.

U.S. Census Office. 1901. The center of population and its median point. *Twelfth Census of the United States, Census Bulletin No. 62*. Washington: GPO.

Vinovskis, M. 1972. Mortality rates and trends in Massachusetts before 1860. *Journal of Economic History* 32: 184-213.

Winkle, K. 1988. *The politics of community: Migration and politics in antebellum Ohio* . New York: Oxford University Press.

TABLE 1
An Example of Linkage from the 1850 PUMS to
the 1860 Manuscript Census Schedules
Using the 1860 National Census Index

1850 PUMS: Calhoun County, Michigan, The Town of Emmett, Page 90

Name	*Sex*	*Age*	*Occupation*	*Origin*	*Real Wealth*	*Pers Wealth*	*M*	*S*	*I*
Briggs, Sidney	M	27	Farmer	Ohio	$0	n.a.	N	N	N
Briggs, Louisa	F	24		New York			N	N	N
Briggs, Joan	F	6		Ohio			N	N	N
Briggs, James	M	4		Ohio			N	N	N
Briggs, Arzina	F	2		Ohio			N	N	N
Briggs, Ellen	F	1		Ohio			N	N	N

1860 NATIONAL CENSUS INDEX:

Name	*State*	*County*	*Page*	*Subdivision*
Bragg, Sidney	CT	Hartford	998	South Windsor
Briggs, Sidney	WI	Green Lake	858	Berlin
Briggs, Sidney	NY	Steuben	782	Cameron
Briggs, Sidney D.	OH	Hancock	117	Blanchard Twp

1860 MANUSCRIPT CENSUS SCHEDULES: Hancock County, Ohio, Blanchard Twp, Page 117

Name	*Sex*	*Age*	*Occupation*	*Origin*	*Real Wealth*	*Pers Wealth*	*M*	*S*	*I*
Briggs, Sidney	M	37	Farmer	Ohio	$0	$250	N	N	N
Briggs, Louisa	F	36		New York			N	N	N
Briggs, Joan	F	17		Ohio			N	N	N
Briggs, James	M	15	Farm Laborer	Ohio			N	Y	N
Briggs, Ellen	F	10		Ohio			N	Y	N
Briggs, John	M	8		Ohio			N	N	N
Briggs, David	M	5		Ohio			N	N	N
Briggs, Amirs	F	3		Ohio			N	N	N

Note: *M* "married during the year"
S "attended school during the year"
I "unable to read or write"

TABLE 2
Characteristics of Males 10 Years and Older
In 1850 Public Use Micro Sample (PUMS)

			Uncommon Names)))))))	
	All Observations	Common Names	Not Linked to 1860	Linked to 1860
Age	29.7	29.8	29.8	28.9
Birthplace (%)				
U.S.	81.5	81.0	81.6	84.0
Foreign	15.3	15.8	15.1	12.9
North & South America	1.1	0.9	1.4	0.6
Europe	14.2	14.9	13.7	12.3
Britain	2.8	3.3	2.3	1.7
Ireland	6.2	7.5	4.3	5.8
Germany	4.3	3.5	5.5	4.1
other	0.9	0.6	1.6	0.7
Unknown	3.3	3.2	3.5	3.1
Family Size	2.7	2.7	2.6	2.8
Married (%)	37.7	38.0	35.9	43.3
Illiterate (%)	5.7	5.9	5.8	4.1
Occupation (%)				
High White Collar[a]	5.4	4.7	6.5	5.1
Low White Collar[b]	1.8	1.7	1.8	2.0
Skilled Blue Collar[c]	12.1	12.1	11.9	13.0
Unskilled Blue Collar[d]	20.7	21.0	20.3	21.1
Farmer[e]	32.6	32.6	31.8	36.6
None[f]	27.3	27.8	27.7	22.2
Region (%)				
New England[g]	14.3	14.5	12.9	18.8
Middle Atlantic[h]	30.0	30.1	28.1	37.2
Midwest[i]	25.8	26.4	25.9	22.3
Southeast[j]	14.7	14.7	15.1	12.9
South Central[k]	13.7	13.1	15.7	8.1
West[l]	0.3	0.3	0.4	0.1
Pacific[m]	1.2	0.8	2.0	0.5
Size of Location (%)				
Rural (under 2,500)	79.4	79.4	79.1	80.8
Urban (2,500 and over)	20.6	20.6	20.9	19.2
2,500-9,999	7.1	7.0	6.9	8.6
10,000-49,999	5.9	6.3	5.6	5.2
50,000-99,999	2.1	1.9	2.5	1.5
100,000 & over	5.5	5.4	5.8	3.9
Migrant[n]	28.2	28.3	29.8	21.1
Real Estate Wealth ($)	641.33	633.95	636.36	708.91
Observations	55,852	30,266	20,749	4,938

(Continued)

TABLE 2
Characteristics of Males 10 Years and Older
In 1850 Public Use Micro Sample (PUMS)
(Continued)

(Continued)

Notes:
[a] Professional, Technical, and Kindred Workers; Managers, Officials, and Proprietors, Except Farm
[b] Clerical and Kindred Workers; Sales Workers
[c] Craftsmen, Foremen, and Kindred Workers
[d] Operatives and Kindred Workers; Private Household Workers; Service Workers, Except Private Household; Farm Laborers and Foremen; Laborers, Except Farm and Mine
[e] Farmers and Farm Managers (including Farm Tenants)
[f] No Occupation Given (most are under age 15 in 1850)
[g] Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
[g] New Jersey, New York, Pennsylvania
[i] Iowa, Illinois, Indiana, Kansas, Nebraska, Michigan, Minnesota, Missouri, Ohio, Wisconsin, Dakota Territory
[j] District of Columbia, Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia
[k] Alabama, Arkansas, Kentucky, Louisiana, Mississippi, Tennessee, Texas
[l] Arizona Territory, Colorado, New Mexico Territory, Utah
[m] California, Oregon, Washington
[n] Living in 1850 in a state different from state of birth

Source: 1850 PUMS linked to the 1860 federal census manuscripts.

TABLE 3
Logistic Regressions on
Whether an Individual Had a Common Name and Was Linkage to the 1860 Census

	Common Name (1=yes, 0=no) $\partial P/\partial X_j$	Linked to 1860 (1=yes, 0=no) $\partial P/\partial X_j$
Age (1850)	-0.004**	0.002*
Age^2 x 10^{-2}	0.005**	-0.005**
Foreign-Born	0.026**	-0.035**
Family Size	0.005**	0.002
Married	0.006	0.051**
Illiterate	0.020*	-0.023*
Occupation (1850)		
High White Collar	-0.072**	-0.018
Low White Collar	-0.019	0.032
Craftsman	-0.010	0.015
Farmer	-0.015**	0.035**
Location (1850)		
Middle Atlantic	-0.010	-0.029**
Midwest	-0.009	-0.084**
Southeast	-0.012	-0.100**
South Central	-0.038**	-0.166**
West	-0.045	-0.221**
Pacific	-0.200**	-0.226**
Urban	-0.003	-0.037**
Migrant	0.019**	-0.050**
Real Wealth x 10^{-5}	-0.002	0.051
Constant	0.105	-0.155
Log-Likelihood	-33,384	-11,951
χ^2	266.073**	907.792**
Mean Probability	0.542	0.177
Observations	55,852	25,586

Notes: * Significant at the 95% level; ** Significant at the 99% level.
The excluded group is native-born, unmarried, literate, laborers living in the rural Northeast in their state of birth. The marginal effect for each independent variable (the impact of a one unit change in the variable on the probability that the dependent variable equals one) was calculated by multiplying the estimated logistic regression coefficient (not shown) by $(\bar{P})(1-\bar{P})$ where $\bar{P}$ is the probability that the dependent variable is equal to one, evaluated at the sample means for each independent variable:

$$\frac{\partial P}{\partial X_j} = \beta_j(\bar{P})(1-\bar{P}) = \beta_j \frac{e^{\bar{X}\beta}}{(1+e^{\bar{X}\beta})^2}$$

where $\bar{X}$ is a vector of average values for the independent variables, β is a vector of logistic regression coefficients, X_j is a particular variable j, and β_j is its coefficient.

Source: For definition of variables, see Table 2.

TABLE 4
Possible Sources of Linkage Failure

	Optimistic Scenario	Pessimistic Scenario
Survived to 1860[a]	0.80	0.75
Enumerated in 1860 Census[b]	0.85	0.80
Correct Name[c]	$(0.95)^3$	$(0.90)^3$
Correct Age[c,d]	$(0.95)^3$	$(0.90)^3$
Correct Birthplace[c]	$(0.95)^3$	$(0.90)^3$
Household Head in 1860[e]	0.70	0.65
Correct Name in 1860 Index[f]	0.95	0.90
	))))	))))
Total Probability	0.28	0.14
Mid-Point of Optimistic and Pessimistic	0.21	

Note: [a] Based on calculations in Table A-1
[b] Based on Steckel (1991)
[c] In three sources: 1850 manuscripts, 1850 PUMS, and 1860 manuscripts
[d] Based on Knights (1971)
[e] Based on age distribution in 1850 PUMS
[f] Based on Ferrie (1992)

TABLE 5
Comparison of 1850 & 1860 Characteristics in
1850-60 Linked Public Use Micro Sample
With Other 1850 & 1860 Samples

	Linked PUMS		Soltow		Steckel	Schaefer
	1850	1860	1850	1860	1860	1850
U.S.						
Age	29	39	37	38	44	
% Farmer	36	45	44	49	63	
% Native-Born	86	86	82	74	92	
Real Estate ($)	1,016	2,004	1,001	1,492	3,739	
Personal Estate ($)		1,382		1,088	3,398	
Gini: Real Estate	.83	.81	.86	.85	.77	
Number of Children		2			5	
% Interstate Movers 1850-60	28			19		
Observations	4,938	4,938	10,393	13,696	1,581	
New South and Frontier						
% Age 40 and Under	82					60
% Farmer	49					81
% Real Estate > 0	29					73
% Previous Interstate Move	49					78
% Intercounty Movers 1850-60	55					28
Observations	239					1,307

Source: Linked PUMS: see text; Soltow: Soltow (1975); Steckel: Steckel (1988a and 1988a); Schaefer: Schaefer (1985).

TABLE 6
"Conditional" Persistence Rates
By 1850 Location and Occupation

1850 Location	Total	None	Unskilled Blue Collar	White Collar and Skilled Blue Collar	Farmers
Total	53.0	41.4	44.1	53.2	65.0
By Size of Location					
Rural	54.7	42.0	45.6	56.8	64.6
2,500-9,999	57.0	54.2	43.4	59.4	74.3
10-49,000	42.7	33.8	44.2	46.6	—
50,000+	31.7	17.1	29.7	36.8	—
By Region and Size of Location					
New England	60.2	46.6	48.5	59.9	77.9
Rural	64.4				
2,500-9,999	60.4				
10-49,000	47.8				
50,000+	—				
Middle Atl	51.4	37.3	44.7	54.3	65.4
Rural	53.7				
2,500-9,999	48.8				
10-49,000	45.3				
50,000+	30.1				
Midwest	45.4	33.6	30.9	41.7	57.5
Rural	47.5				
2,500-9,999	—				
10-49,000	—				
50,000+	25.0				
Southeast	62.2	60.0	59.1	50.5	69.7
Rural	63.7				
2,500-9,999	—				
10-49,000	—				
50,000+	51.4				
South Central	49.8	38.6	34.1	53.9	58.9

Notes: The "conditional" persistence rate is the fraction located in both 1850 and 1860 who remained in the same county between 1850 and 1860. Cells with fewer than 40 observations are not reported. For regions and occupations, see Table 2.

Source: 1850 PUMS linked to the 1860 federal census manuscripts. See text.

TABLE A-1
1850 Male Life Expectancies and Projected Survival Rates 1850-60
By Location and Nativity

	Life Expectancies)))))))			Baseline Population From 1850 PUMS	Estimated Survivors	
	e_{20}	e_{30}	e_{40}	(N)	(%)	(N)
U.S. 1850: NATIVE-BORN						
weighted	44.4	37.7	30.9			
unweighted	45.4	38.8	32.1			
Rural						
unweighted	47.2	40.7	33.9			
weighted	47.6	41.1	34.2			
Northeast	43.4	36.0	28.3	13,613	85	11,571
Northwest	47.6	41.2	34.4	11,466	89	10,205
South	49.5	43.4	36.9	12,999	88	11,439
Urban						
weighted	38.0	30.9	24.3			
unweighted	38.2	31.3	25.2			
Northeast	38.2	31.0	24.1	5,420	81	4,390
Northwest	44.0	37.0	31.4	827	86	711
South	29.8	23.7	19.5	1,106	80	885
U.S. 1850: FOREIGN-BORN						
weighted	35.7	31.2	27.3			
unweighted	38.7	33.8	29.1			
Rural	45.3	38.1	31.6	6,272	86	5,394
Urban	30.6	27.5	24.9	4,149	76	3,153
Total				55,852	85	47,474
OTHER LIFE TABLES						
U.S. 1830-60[a]	40.9	34.5				
U.S. 1830-60[b]	42.8	36.0				
Mass. & Maryland 1850[c]	40.1	33.6	27.1	55,852	88	49,150
U.S. 1850[d]	37.6	31.0				
Mass. towns 1860[e]	44.0	37.4				
Model West, Level 13[f]	41.5	34.0	26.6	55,852	90	50,267

Notes: e_x are the life expectancies at age x. Entries in the "Survivors" column were calculated by applying the male age distribution from the 1850 PUMS to the ${}_nq_x$ values used to generate the life expectancies. This was done by nativity and rural/urban, and for the native-born also by region. "Urban" is a location with a population of 2,500 or more in 1850. "Northeast" includes New England and Middle Atlantic; "Northwest" includes Midwest, West, and Pacific; "South" includes Southeast and South Central.

Source: U.S. 1850: author's calculations from 1850 federal mortality manuscript schedules; (a): Meech (1898); (b): Pope (1992); (c): Jacobson (1957); (d): Haines (1979); (e): Vinovskis (1972); (f): Coale and Demeny (1983).

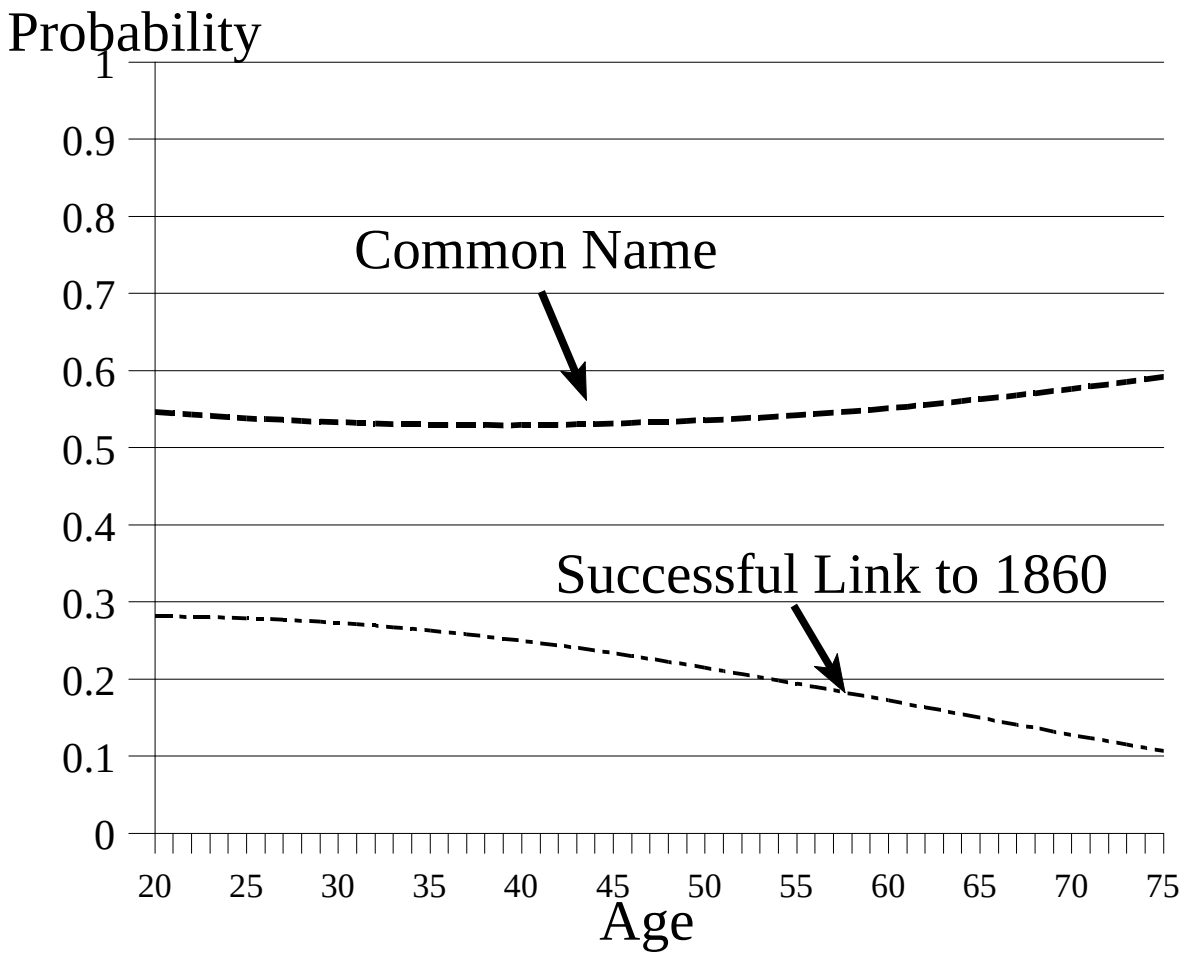

FIGURE 1: Probability of Having a Common Name and Successful Linkage to 1860, By Age

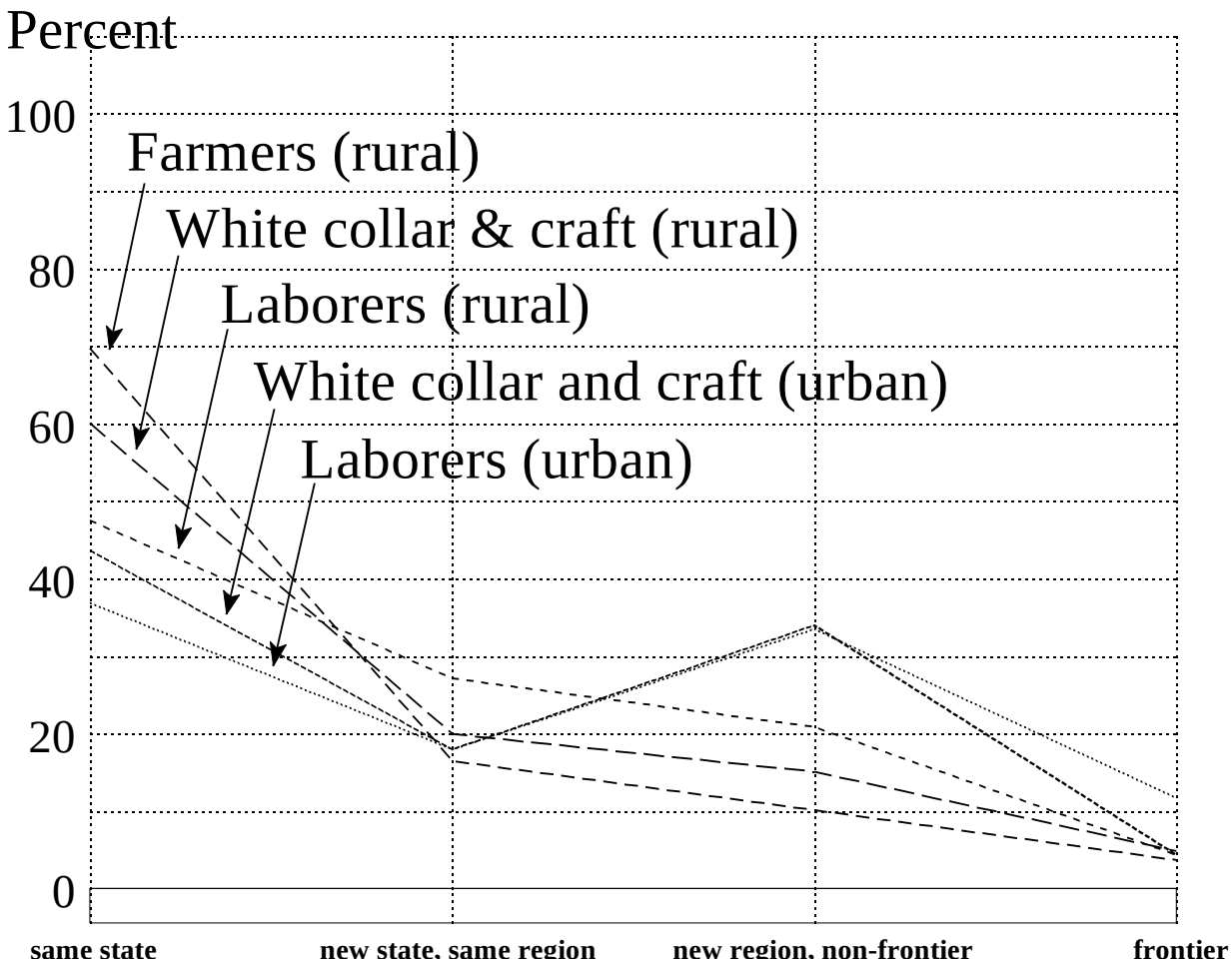

FIGURE 2: 1860 Location of Intercounty Migrants from New England, Middle Atlantic, and Southeast States, By Size of 1850 Location (Urban is 10,000 and over)