Historians and the Census: the Historiography of Census Research

Robert P. Swierenga

ISSN 0003-4827
Copyright © 1990 State Historical Society of Iowa. This article is posted here for personal use, not for redistribution.

Recommended Citation
Available at: https://doi.org/10.17077/0003-4827.9478

Hosted by Iowa Research Online
"CERTAINLY it would be difficult to write history either social or economic or indeed political without statistics," declared Joseph A. Hill of the U.S. Census Bureau in 1908, "and it would be a very defective economic or social history of the United States," he continued, "that ignored the statistics compiled by the United States census." This bicentennial year of the federal census is an appropriate moment to assess how historians have responded to Hill's advice: to survey the development of historical research based on the manuscript census schedules, to describe the current state of the art, and to note future directions.

As the "first new nation," the United States took the unparalleled step of mandating decennial population censuses in its very constitution. Article I, Section 2, declared that "representatives and direct taxes shall be apportioned among the several States . . . according to their respective numbers." To determine those numbers the framers required that an enumeration be made "within three years after the meeting of the Congress of the United States [that is, by 1790], and within every subsequent term of ten years." Accordingly, the first Congress in its second session provided for "taking a census" by enacting the


650
Historians and the Census

THE PIONEERS, 1930–1960

Until the 1930s historians largely ignored Joseph Hill’s invitation to mine the census records. The manuscript schedules were so voluminous that, given the lack of unit-record technology, scholars rightfully shied away from the lists. Also, the first six censuses, from 1790 to 1840, were collected only for households, which limited their usefulness for research. The 1850 census was the first to provide individual-level information, and the schedule was expanded to include questions on occupation, wealth, and nativity, which are crucial behavioral variables.

Frederick Jackson Turner did not use census records for his own research, but in 1922 he encouraged others to use them to test his grand theories. Four historians—James C. Malin of the University of Kansas, Oscar Handlin of Harvard University, Frank L. Owsley of Vanderbilt University, and Merle Curti of the University of Wisconsin—took up his challenge and led the way in census research in the pioneer period, 1930 to 1960. Their questions revolved around the issues of population mobility, immigration, and community development on the midwestern frontier and in the Old South.

Malin inaugurated individual-level census research in 1932. Over a ten-year period he gradually developed a considerable data base that led him to challenge the geographic deter-

---


minism of Frederick Jackson Turner’s frontier theory. Malin began, innocently enough, trying to determine the states or countries of origin of the pioneers in his home community of Lewis in eastern Kansas. He presented his findings in a high school commencement address in Lewis. Since the experience of this community did not jibe with Turnerian views, Malin selected a second sample township in western Kansas. The second study confirmed the first, and, as Malin later recalled, this “led to the addition of new samples and finally to the generalized study of 1935 on population turnover.” In this piecemeal fashion, as the scope of his research expanded and early findings were confirmed, Malin came to realize the larger significance of his work for revising Turnerian historiography.

Malin’s research involved sampling from the federal censuses of 1850 and 1860 and also from the Kansas state population schedules from 1875 to 1925. Many state governments in the mid-nineteenth century followed the lead of the federal government and enumerated their populations at ten-year intervals at the midpoint between federal censuses. These state censuses, which were discontinued after 1925, were not closed to researchers by confidentiality rules as were the federal censuses after 1880. Hence Malin could trace the settlement history of Kansas from 1850 until 1925.

Malin’s innovative research technique was nominal record linkage, which enabled him to follow thousands of specific farmers from one census to the next, virtually throughout their lifetimes. He determined rates of direct and indirect (multiple) migration and also, in the case of large families, charted migration routes based on the state of birth of successive children.


The Kansas historian was thus the first to employ birth-state data and construct residential mobility indexes. Malin’s wide-ranging research into federal and state censuses and his sampling and nominal linkage techniques were so innovative that scholars did not surpass him for thirty years and more.

While Malin carried forward his critique of Turnerian theory, Oscar Handlin applied methods from the emerging field of social science in the late 1930s to the problem of immigrant adjustment in a northeastern city. Handlin chose for his dissertation research the Irish Catholic immigrants who overwhelmed Protestant Boston in the 1840s. He used the 1850 census manuscripts to reconstruct the occupational distribution for the major ethnic groups. Handlin also used the Massachusetts state census of 1855 and the unique Boston city censuses of 1845 and 1850. Handlin’s seminar at Harvard in the 1950s on the history of immigration introduced a generation of students to the possibilities of the census for urban social history. Stephan Thernstrom’s *Poverty and Progress* (1964), a path-breaking study of social mobility in Newburyport, Massachusetts, originated in Handlin’s seminar.6

While Handlin, with the active help of his wife Mary Flug Handlin, completed his initial census research in 1940, Frank Owsley, with the assistance of his wife Harriet Owsley and several graduate students, began an extensive study of the southern states. They based their work on the federal population and agricultural censuses of 1850 and 1860, which they linked to property tax assessment rolls.7 Owsley’s purpose was to determine statistically if nonslaveholders (the “plain folk”) in the Old South owned significant amounts of farmland, in comparison

6. Oscar Handlin, *Boston’s Immigrants, 1790–1880: A Study in Acculturation* (1941; reprint, New York, 1968), 233–37, 270–71, vii–viii; Stephan Thernstrom, *Poverty and Progress: Social Mobility in a Nineteenth Century City* (Cambridge, MA, 1964). Handlin completed his dissertation at Harvard University in 1940. The 1845 Boston census that he used was created by Lemuel Shattuck. It was the first of its kind and served as a model for the director of the 1850 federal census, J. D. B. DeBow, to expand the questionnaire into a socially useful instrument.

to the dominant planter class. His revisionist findings proved that land ownership was widespread even in the lower South and that "the door of economic opportunity swung open easily" for the rank-and-file yeomen. Owsley viewed his work as a contribution to Civil War historiography, but his conclusions also supported Turner's frontier thesis. The plain folk were living proof that the southern frontier produced democratic landowning patterns.

Like Malin and Handlin, Owsley undertook nominal record linkage and areal sampling. But Owsley and his wife were innovators in the application of statistical and social science methods. They used newly developed punch cards to contain their individual-level data, sorted and tabulated variables with electronic sorting machines, and computed statistics with electric calculators. The Owsleys were also the first to use the agricultural census manuscripts. Since their purpose was to compare land tenure between farmers and tenants with and without slaves, they abstracted the names of all heads of farm families in selected counties across the South, arranged the names alphabetically on a master list, and linked the censuses name by name. Finally, they transcribed the linked file to punch cards for sorting and tabulating sums and percentages for the several farm operators. Owsley described his procedure as "exceedingly tedious and long-drawn out," but his new methods provided signposts and guides for social science historians.8

Merle Curti and his graduate assistants embarked in the early 1950s on a major quantitative analysis of the occupational structure of a frontier county in southwestern Wisconsin. Curti's purpose was "to study microscopically" all gainfully employed persons in a developing rural community in order to test Turner's thesis that the frontier fostered equality of economic opportunity. Like Owsley, Curti chose the manuscript population censuses as his primary source and used nominal record linkage, punch card processing, and calculating machines. Curti, however, analyzed one county "as a whole" and included four censuses (1850, 1860, 1870, and 1880) to provide intergenerational data. His codebooks included more

8. Owsley, Plain Folk, 152.
than fifty variables, far surpassing all previous census studies in complexity and scope.

Curti’s census research confirmed the free land hypothesis—that land was readily available and ownership widespread in frontier counties. Owsley had found that 80 percent or more of the southern free farm operators owned land in the antebellum years, and that the “majority of slaveholders and non-slaveholders . . . were farmers falling in the same general economic groups.” Curti similarly calculated that in Trempealeau County, Wisconsin, 85 percent of farm operators and agricultural laborers owned land in 1860 (in 1870, 73 percent did). “Turner’s poetical vision of free land and of relatively equal opportunity,” Curti concluded, “was for a great many people being realized in Trempealeau County. The story of the making of this community is a story of progress toward democracy.”

THE BEHAVIORAL TRIUMVIRATE, 1960–1975

In the decades since Curti and his predecessors popularized census research, interest in the censuses has expanded greatly, fueled by computer technology and the growing scholarly interest in the new social history and its subfields of immigration, ethnic, urban, labor, and family history. Curti’s empirical approach to frontier community development also coincided with the emergence of the “behavioral school” in American historiography.

Samuel P. Hays, a professor at the University of Iowa in the 1950s, issued the call for a new perspective in his 1959 address to the Iowa State Education Association, “History as Human Behavior.” Hays urged history teachers to revitalize their courses by focusing on types of human behavior rather than on formal institutional and “presidential history.” Only by “systematically studying human experience and behavior” at the grass roots, Hays declared, “will solid and concrete generalizations

10. Ibid., 3, 152 (table 18), 448.
emerge.”\(^{11}\) Hays soon had his graduate students at Iowa and after 1960 at the University of Pittsburgh poring over census microfilm in order to study political behavior and ethnicity at the local level throughout the Midwest and on the East Coast.\(^ {12}\)

Meanwhile, Hays’s colleague at Iowa, Allan G. Bogue, who then specialized in midwestern agricultural history and had studied with Malin as a postdoctoral fellow, made extensive use of the Illinois and Iowa population and agricultural censuses to understand farmers’ behavior. His book *From Prairie to Corn Belt* stressed the decisions farmers made, and the text was sprinkled liberally with tabular data drawn from the manuscript censuses. Bogue also was the first to identify tenant farmers in the censuses before 1880 (when the agricultural schedule first included a question on tenancy) by linking farmers listed in the population schedule with farm operators in the agricultural schedule. Those farmers who reported no real property in the population census were classed as tenants. Although subsequent scholars have refined Bogue’s technique, his method remains fundamental.\(^ {13}\) Bogue’s graduate students joined those of Hays in exploiting the census microfilm; a few even read the actual files of the Iowa state censuses (1855–1925) at the State Historical Society of Iowa in Des Moines which had not yet been microfilmed.

The third person in the behavioral triumvirate was Lee Benson, a close associate of Bogue. Benson and other new polit-

---


Historians and the Census

Historians led by Jerome Clubb and Warren Miller created the vast historical data archive at the Inter-University Consortium for Political and Social Research in Ann Arbor, which includes a computer file of all published census data from 1790 through 1980. \(^\text{14}\) Countless scholars have taken advantage of these county-level data in their research.

Moreover, Benson's research agenda changed the direction of American political history from an institutional to a behavioral focus. His research centered on the pithy question "Who voted for whom, when?" \(^\text{15}\) This simple but powerful question spawned dozens of dissertations and books, and eventually produced the now standard ethnocultural interpretation of popular voting behavior. \(^\text{16}\)

For these ethnocultural historians, the crucial independent variables are census data on nativity, occupation, and wealth. By aggregating individual-level population census data by township and ward, they seek to correlate those independent variables with voting patterns in the smallest political subdivisions. Early studies in the 1960s isolated relatively homogeneous communities, such as those with a high percentage of German Lutherans, and generalized from their voting patterns how all German Lutherans voted. By the 1970s, however, multivariate correlation analysis of ecological variables had become the norm. The variables were compiled from the manuscript censuses for entire cities or sample counties in a state. This method

---

largely obviated the charge that homogeneous communities might be atypical. But skeptics may not be convinced until the method is tested nationwide or at least regionally with a random census sample of townships and city wards.17

SOCIAL MOBILITY AND IMMIGRATION STUDIES

Following Curti's example, the new social historians in the 1960s and 1970s made single community studies of geographic and social mobility the cornerstone of their work. Stephan Thernstrom's Poverty and Progress (1964) launched a flurry of research by demonstrating that the manuscript censuses beginning in 1850 provided a primitive social survey of the entire population of the country. Thernstrom measured the extent of upward mobility among unskilled and skilled laborers in the city of Newburyport, Massachusetts, from 1850 through 1880.18 His startling findings of overall socioeconomic improvement prompted other urban historians to test his results elsewhere.19

In 1969 the history department at the University of Pittsburgh began the Pittsburgh Survey, a project to computerize the entire manuscript censuses of the city between 1850 and 1880. The purpose was to provide a data base for student and faculty research projects. Although the project was soon completed, there are few published results. Laurence Glasco, a student of Samuel Hays at Pittsburgh, computerized the Buffalo censuses of 1850 and 1860 to study the effects of ethnicity on social structure among Germans, Irish, and native-born.20 Simi-

18. Thernstrom, Poverty and Progress.
lar projects were undertaken in the other New York cities of Poughkeepsie and Kingston and in Detroit, Atlanta, and Philadelphia. With the Pittsburgh Survey in mind, William Crozier, director of the Minnesota Social History Project, created a computer file that linked all of the residents of Winona listed in the federal censuses from 1850 through 1880 and the 1905 Minnesota state census. Crozier and his associates then prepared a teaching "data package" for the state's secondary and college history teachers to bring census research into their classrooms. This project is typical of many efforts of the 1970s generation of social historians to encourage the systematic use of routinely generated census records to study the history of communities.

Theodore Hershberg's Philadelphia Social History Project (PSHP) quickly overshadowed all other projects. This entrepreneurial scholar raised a million dollars from private and government granting agencies and assembled an interdisciplinary team of scholars and computer experts and some three hundred data entry typists. The PSHP compiled an unprecedented computer file of 2.5 million Philadelphians in the censuses of 1850, 1860, 1870, and 1880. The data was so specific that the residence of any person or family could be located on a block-by-block grid, and for many the journey to work could be


22. The project, funded by the National Endowment for the Humanities, is described in the issues of MISHAP: Newsletter of the Minnesota Social History Project, 1978–1979. A more ambitious 1980s teaching project by Myron P. Gutmann of the University of Texas at Austin is described in Gutmann, "Teaching Historical Research Skills to Undergraduates: Thoughts on Microcomputers and the Classroom," Historical Methods 21 (1988), 112-20.
charted. Such cross-sectional and longitudinal information for a nineteenth-century population on this vast scale was obviously not possible before the availability of powerful computers.

Given his vast resources, Hershberg took a leadership role among social historians in developing standard classification schemes, linkage methodologies, and special-purpose computer programs. To encourage comparability among mobility studies, he assembled a collaborative team of urban census specialists (Stuart Blumin, Laurence Glasco, Clyde Griffen, and Michael Katz), and together they developed a common occupational classification scheme that has become standard. They also compared their findings across five cities to reach broader conclusions about occupational mobility.

During the 1970s and early 1980s Stephan Thernstrom’s cohort of census mobility scholars reaped the first harvest in the quest to understand the effects of urbanization on economic development. Peter Knights’s model study, *The Plain People of Boston*, and Ulf Beijbom’s innovative *Swedes in Chicago* led off in 1971. Next came Howard Chudacoff’s study of Omaha, *Mobile Americans*, which was the first to use the 1900 manuscript census. Thernstrom’s longitudinal study, *The Other Bostonians*, and Laurence Glasco’s study of Buffalo appeared in 1973, and Dean R. Esslinger’s work on South Bend, Indiana, in 1975. Charles Stephenson, meanwhile, demonstrated how the state-by-state Soundex name index system devised by the Census Bureau beginning with the 1880 census could be used to trace people who moved. The next year was a vintage one,


with Stuart Blumin’s book on Kingston, New York, and Kathleen Conzen’s *Immigrant Milwaukee*, which quickly became a classic. In 1978 Gordon Kirk, Jr., published his work on Holland, Michigan, which also drew on the 1894 Michigan state census. Roberta Balsted Miller’s 1979 study of Syracuse, New York, made use of the 1855 New York state census as well. In 1982 Richard L. Doyle completed a massive socioeconomic mobility study of the rural community of Pella, Iowa, 1850 to 1925, which employed the federal censuses for 1850 through 1910 and the six Iowa state censuses from 1856 through 1925, plus county tax lists.26

Olivier Zunz’s book on Detroit, *The Changing Face of Inequality* (1982), is the premier exemplar of the mobility studies. Zunz wove together the story of urbanization, immigration, and industrialization in Detroit from 1880 to 1920. While scholars still eagerly await the opening of the 1920 census manuscripts in 1992, Zunz gained indirect access more than a decade earlier under a special agreement with the U.S. Bureau of the Census, by which the bureau coded a sample of six thousand heads of households in selected geographic areas and provided Zunz with aggregate statistics at the block level.27

Many of these mobility studies of the 1970s and 1980s focused on the experience of immigrants. At the same time, other scholars of immigration and ethnicity turned to the federal population censuses, which in 1850 began to specify place

---


27. Zunz, *Changing Face of Inequality*, 413.
of nativity, to analyze the processes of transplanting and accultur- tation. Ulf Beijbom and Kathleen Conzen had described the process of accommodation and adjustment in the host society, but because they did not trace the story to the Old Country places of origin, the research was truncated.

Since the mid-1970s, immigration historians have routinely traced their groups from sending to receiving societies by linking emigration records of European nations with census and other nominal records in the United States. Historical geographer Robert C. Ostergren demonstrated the new approach in his 1976 dissertation on a Swedish colony in Minnesota, published as *A Community Transplanted* (1988), as did Walter D. Kamphoefner in his 1978 dissertation on Missouri Germans, published in 1987 under the title *Transplanted Westfalians*. Harald Runblom and Hans Norman’s collection of research by the Swedish Uppsala Project, entitled *From Sweden to America* (1976), likewise linked Swedish and U.S. records on migrants. David Vander Stel’s 1983 dissertation on the Dutch of Grand Rapids, Michigan, from 1848 through 1900 also took this transatlantic approach. Jon Gjerde’s 1985 book on Norwegian emigrants from the village of Balestrand thoroughly explored the social and cultural conditions in the homeland and then followed the immigrants throughout the upper Middle West, locating them in the federal and state censuses. My research on Dutch immigration to the United States in the nineteenth century likewise rests on a nationwide census data base of persons of Dutch birth or parentage totaling 116,000 names compiled from the federal censuses of 1850, 1860, and 1870. When linked with Netherlands emigration lists, for example, occupational changes caused by overseas migration can be determined precisely for the entire population. Thus, immigration historians have traced migrants from rural Europe to both rural and urban America, proving that “Americanization” was but part of a transatlantic process of cultural change.28

FERTILITY AND FAMILY

Prior to the early twentieth century, the systematic registration of vital events such as fertility and mortality was unreliable, so demographic historians turned to the manuscript population censuses and developed sophisticated techniques of constructing life tables and other vital statistics. Fertility and mortality patterns have especially drawn the interest of demographic and economic historians since the publication in 1961 of the controversial work of Yasukichi Yasuba, *Birth Rates of the White Population in the United States, 1800–1860*. Richard A. Easterlin, Maris Vinovskis, and Richard H. Steckel, among others, have relied on the manuscript population censuses to estimate directly (or indirectly, if necessary) birth and death rates, health indices, and general population shifts over time. Newer research by Steckel and David Davenport involves tracing families and individuals over several federal and state censuses in order to have longitudinal data to estimate rates of migration.


Other scholars, such as Richard Bernard, have measured the melting pot process by charting intermarriage rates among different nationalities.  

Frank Owsley’s early research on the southern plain folk showed that federal censuses were the best source to recover information about ordinary southerners and their communities. But once the “Vanderbilt School” ended, there was little follow-up research apart from that in the 1940s by Barnes Lathrop and James Bonner. This changed in the 1970s with the book by Roger Ranson and Richard Sutch on black sharecropping during Reconstruction, Morgan Kousser’s ecological regresional analysis of voting patterns in the New South, and Elizabeth Pleck’s research on black family structure in Boston.  

While urban social mobility and migration studies had become the norm, rural life was neglected in the 1970s, although Michael Conzen’s *Frontier Farming in an Urban Shadow* (1971) demonstrated the road to follow in the new rural history. While few in number, several rural historians wrote masterful studies of tenancy and household wealth buildup. Seddie Cogswell and Donald Winters explored Iowa land tenure; Richard Doyle traced wealth accumulation over three generations in Iowa; and J. R. Kearl, Clayne Pope, and Larry Wimmer measured wealth in frontier Utah. In the 1980s, Hal Barron, Vernon Burton, and

34. Michael P. Conzen, *Frontier Farming in an Urban Shadow: The Influence of Madison’s Proximity on the Agricultural Development of Blooming Grove, Wisconsin* (Madison, 1971); Seddie Cogswell, Jr., *Tenure, Nativity, and Age as
Frederick Bode and Donald Ginter published major works. Barron followed the stagnant fortunes of Chelsea, Vermont, farmers who were left behind by the westward migration. Burton studied the black family before and after slavery in a South Carolina county, and Bode and Ginter used an extensive sample of Georgia population and agricultural censuses to determine rates and causes of farm tenancy.35

Burton’s massive study of Edgefield, South Carolina, from 1850 to 1880 is a superb example of a southern community history that will set the standard by which future scholarship in social history will be measured. His sources included the federal population, agricultural, mortality, and manufacturing censuses, plus tax lists, probate and marriage records, and church membership lists. The Edgefield Data Base provides a systematic merged data record combining all available information on every family in the county. Not only is Burton’s data base intimidating, but his striking conclusion that black and white family patterns were similarly patriarchal and that the matriarchal black household is a myth will certainly provide grist for the scholarly mill for many years.


PUBLIC USE SAMPLES

While interest in census research expanded exponentially in the 1970s, fueled by innovations in computer technology and the growing interest in socioeconomic mobility and immigration history, a small band of economic historians, led by William Parker and Robert Gallman, conceived of the ultimate census-based research file—the national sample. Instead of selected city and county census files, as the interests of scholars dictated, the cliometricians envisioned national public use samples of sufficient size to be statistically valid. These samples would be available for all scholars to use for their own purposes.

In the initial southern farm sample, Parker and Gallman selected 5,229 individual farms from the 1860 census of 382 cotton-producing counties in eleven states. The method of preparing the data file was as follows: names of farm household heads in the population census (schedule no. 1) were matched with the same individual in the slaveowners census (schedule no. 2) and the agricultural census (schedule no. 4). The representativeness of the samples was determined by comparing them statistically with the universe of all farms.

Subsequently, various scholars created six additional southern farm samples, using the Parker-Gallman methodology. James Foust selected 897 farms in 1850 in eleven southern states; Mark Schmitz (with Gallman) sampled 1,856 farms in Louisiana sugar counties in 1850 and 1860; Dale Swan added 671 farms in the rice-growing counties of Georgia and South Carolina in 1860; Gallman created a large file of 3,745 farms in Kentucky and Tennessee in 1850 and 1860; and Randolph Campbell and Richard Lowe compiled a sample of 500 heads of households in eastern Texas in 1850 and 1860. Finally, the Southern Economic History Project, under the guidance of

Roger Ranson and Richard Sutch, constructed a massive sample of 11,202 farms from the 1880 manuscript census of agriculture for seventy-three "representative" counties throughout the South. This was supplemented with a subsample of 4,695 farms in twenty-seven cotton counties in 1880, and 1,039 farms in eight cotton counties in 1870, in order to study the economic consequences of emancipation in the postwar South. These southern census files are a primary source of quantitative data on American rural population that holds great promise but yet remains underutilized. Only the slavery cliometricians have fully exploited the files to measure economic growth rates, labor efficiency, and landholding patterns.

When Robert W. Fogel and Stanley Engerman embarked on their econometric analysis of slavery, they realized that a companion northern farm sample would provide a benchmark against which to compare their findings about slave agriculture. They encouraged James Foust and Fred Bateman to seek funding from the National Science Foundation to prepare a census use sample of the rural north. The resulting sample included

37. The major census files compiled by cliometricians are listed in Robert William Fogel and Stanley L. Engerman, Time on the Cross: Evidence and Methods, 2 vols. (Boston, 1974), 2:21-23. Published results are in Gallman, "Self-Sufficiency in the Cotton Economy," 5-23; Foust and Swan, "Productivity and Profitability," 39-62; Randolph B. Campbell and Richard G. Lowe, Wealth and Power in Antebellum Texas (College Station, TX, 1977), 13-31; and Ranson and Sutch, One Kind of Freedom, esp. 283-96. Only the Parker-Gallman sample has yet been subjected to thorough reliability tests, and it failed. See Donald F. Ginter, "Critique of Landholding Variables in the 1860 Census and the Parker-Gallman Sample," paper presented at the 1980 annual meeting of the American Historical Association; Bode and Ginter, Farm Tenancy and the Census in Antebellum Georgia.

21,118 rural households (108,000 persons) selected from the 1860 population and agricultural census in 102 counties in the northeast and central regions. This was a cluster sample, however, rather than an individual sample; it included every household, farm and nonfarm, in one entire nonurban township, selected randomly within each of 102 counties (also chosen randomly) in sixteen states. By collecting data for every household within a township, the northern sample made it possible to estimate total factor productivity in agriculture and to identify the entire agricultural work force, including farm laborers not residing on the twelve thousand farms included in the sample. It also included demographic and economic information. As in the southern samples, names of farm operators listed in the agricultural schedule in the sample townships were linked with the population census. Statistical tests of representativeness reveal the sample to be trustworthy. This northern census sample was little used until the 1980s when Bateman and Jeremy Atack fully exploited it in their prize-winning book, *To Their Own Soil: Agriculture in the Antebellum North* (1987), the first detailed quantitative description and analysis of northern agriculture as a whole that rests on a solid empirical base and employs robust economic theory.

Another noteworthy census sample file for the nineteenth century that includes both rural and urban population is Lee Soltow’s national study of property holding. Soltow compiled random nationwide samples of the free adult male population in 1850 (N=10,393), 1860 (N=13,696), and 1870 (N=9,828). The data are cross-sectional; individuals are not linked among the three censuses. The results thus show structural change, but not longitudinal trends for specific cohorts. Soltow’s sampling method was a self-designed “spin sample” technique, which involved turning a microfilm roll on a reading machine by half-turns of the feeder arm and randomly selecting one line (or


individual) per "spin" on the left-hand pages until the end of the film and then reeling the film backwards by half-turns and randomly selecting a line on the right-hand page. Soltow compared his spin samples to published aggregate census tables and found them to be in "excellent accord." For each individual selected, the author transcribed to magnetic tape the census information on the value of real and personal property, age, sex, color, nativity, occupation, and place of residence. This data base enabled Soltow to determine which population subgroups achieved the American dream of economic improvement.41

The northern and southern farm samples were deliberately biased toward rural counties. To provide a truly national sample for the 1850 through 1880 censuses, a random sample that includes urban places is required. In 1981 the Inter-University Consortium for Political and Social Research (ICPSR) completed the first such data source—the 1900 Public Use Sample, which included one hundred thousand persons and their corresponding household information. In 1989 the University of Pennsylvania created another file funded by the National Science Foundation, the 1910 Public Use Sample, which included 366,000 persons and 89,000 households, or a 1-in-250 sampling fraction. In 1987 the University of Washington created the National Panel Study, which included ten thousand males (aged 5-14 and 25-34) sampled from the 1880 census and then linked forward to the 1900 census. The linkage success rate was a remarkable 39 percent despite the intervening twenty years.42

Genealogists and state and local historians have different agenda; they need complete populations rather than samples.

This has prompted the Mormon Historical Demography Project to computerize the 1880 census of Utah territory (143,000 persons). The Indiana Historical Society did the same for the entire 1860 census of Indiana. This pivotal census numbered 1.3 million residents, and the file incorporates some 18 million responses to census questions; it required 127 million keystrokes to transcribe all of the data. The U.S. Census Bureau has also prepared random national use samples of recent censuses (with names omitted). All of these census use files except the Indiana 1860 census are available from the ICPSR.

Despite the remarkable advances of the past decade, national public use samples are still needed for the federal censuses of 1850 through 1880 and for the 1920 and 1930 censuses. When the 1920 manuscript census opens in 1992, it will receive the highest priority, coming as it does at the end of the great century of immigration.

CENSUS CHALLENGES

As early as the 1960s, social science scholars had begun to cope with challenges to the accuracy and completeness of the censuses. Peter Knights conservatively estimated an undercount in Boston of 8 percent in 1860. Roger Ranson and Richard Sutch calculated an undercount of 6.6 percent in 1870 for the southern black population. Probably 5 to 10 percent of the population was missed in every census, with urban lower classes over-


44. "Society and IUPUI Begin Computerized Census Project," *Indiana Historical Society Newsletter*, No. 77 (June–July 1985), 1, 4. In the early 1970s the Ohio Historical Society sponsored a Federal Census Project to computerize manuscript census files for 1860. Franklin, Wood, and parts of Hamilton County were completed before the project was abandoned.

looked more often than rural middle classes. This is not an insurmountable problem, however, provided that the “missing” population is identified and the bias in the census is reckoned with.

Scholars have also identified the phenomenon of “age-heaping” (reporting age to the nearest year ending in 0 or 5), the problem of identifying household relationships prior to the 1880 census (when the question was added), and the imprecision in classifying specific reported occupations into broadly analytic categories. The dollar values reported for real estate and personal property have long been suspect, and only recently have several researchers compared census wealth with county tax assessments compiled by official assessors under oath. Initial results are positive and give assurances that the census data on wealth are valid.

Another difficulty is the frequent changes in the census questionnaires. The bureau added and deleted questions in response to pressure groups and to clarify answers and reduce the chances for misinterpretation by enumerators. Fortunately, computer technology is sufficiently advanced and powerful to allow end users to modify, reclassify, and reorder data almost at will. This largely obviates the lack of standardization.

Since the thousands of census marshals were variously conscientious or gifted with handwriting skills, and they did not always use the required india ink, legibility problems plague many census pages. Ethnicity specialists and knowledgeable local historians and genealogists are often the best persons to decipher the names of their group or community. Names that are legible but obviously misspelled due to phonetic or other problems pose a different issue. The actual name and also the “corrected” name should be transcribed into the computer file. This upholds the archivist’s duty of preservation

Explanations in Economic History 12 (1975), 8–10; and idem, One Kind of Freedom, 53–54, 224.


47. See comments of Galenson and Pope, “Economic and Geographic Mobility,” 645n.
of the original record, warts and all, but it also satisfies scholars' utilitarian interest in matching names with other records.

In the end all historical sources have failings, but the manuscript census returns are the most comprehensive survey of the American population at given intervals. Only continuous population registers, such as exist in western European countries, would be superior. The census tells who the people are, what they did for a living, and how they lived.

Valuable as the individual censuses are, however, their usefulness increases geometrically when they are linked name by name with other records, such as tax lists, city directories, probate inventories, birth and marriage records, cemetery ledgers, plat maps, immigration and ship passenger lists, school attendance files, and especially church records. Given current computer technology and software, it is now possible to link names automatically, to enter data with optical scanners (which are increasingly able to "read" handwriting), to sort and classify with flexible database management programs, and to compute frequency distributions and other social statistical measures in a wink. 48 One can envision very powerful data files with dozens and even hundreds of variables of entire populations or samples of populations that would have boggled the mind a generation ago.

One of the most important behavioral questions, that of religious affiliation or identity, which the Census Bureau has never asked, may only be uncovered by linkage to church records of membership, pew rentals, and baptism, marriage, and death registers. 49 Religion might also serve as a proxy for ethnicity, which can only be determined for first- and second-generation immigrants for whom the country or state of birth is recorded. Third and later generations, however, are classed as "native born of native parents." Since the end of mass immigra-

48. See, for example, Myron P. Gutmann et al., "Keeping Track of Our Treasures: Managing Historical Data with Relational Database Software," Historical Methods 22 (1989), 128-43.

tion in the 1920s, these native-born people constitute more than two-thirds of the population, and the proportion is rising. Only the 1980 census contained an ancestry query, and that was at the expense of the standard question on parents' nativity. The ancestry question was again deleted from the 1990 census, so the 1980 data may be unique.\textsuperscript{50}

It is impossible to imagine the writing and study of American history without the federal and state censuses. They are foundational sources for uncovering the lives of the inarticulate masses. If one is sensitive to their nuances and uses them in conjunction with other nominal sources, they yield invaluable information. The full potential of these incomparable manuscripts is only now being realized, as computer technology catches up with the dreams of historians for completely representative samples of past populations. The destruction of the 1890 census by a fire, before the manuscripts could be microfilmed, and the detrimental effects on scholarship from this loss, is demonstration enough of the vital nature of this resource. Little could the nation's founders have imagined their contribution to historical understanding that resulted from their decision to mandate "taking a census" every decade.
