Data Appendix for

“Assessing the Importance of Tiebout Sorting: Local Heterogeneity from 1850 to 1990.”
Rhode and Strumpf, American Economic Review.

Introduction.

Reasons for Moving

American/Annual Housing Survey (AHS)

Source: AHS national surveys (see references for a full list of sources). Availability: 1974-1981 (annual), 1983-1995 (biennial). There are two panels in the AHS, 1974-1983 and 1985-1995 (there is also data for 1973 but it cannot be linked to either panel). All AHS values are calculated separately for 1974-1983 (annual frequency) and 1985-1995 (biennial frequency), since there is a new sample in 1985. Only homes present in all sample years are included. The non-Tiebout reasons for a recent move include specific characteristics of the house structure in 1985-1995. The local services figure is based on a broad definition including schools, crime, sanitation, and roads.

Current Population Survey (CPS)


Section II.

Mobility Costs: Discussion in Text and Sources of Figures 1-2


For average real labor returns, see Lebergott data on annual earnings -- Series D 726 (1860-1900) and D 736 (1900-60)-- in Historical Statistics of the United States, and average real weekly earnings in the private sector, 1960-2000, from Table B-47, p. 330 of the 2001 Economic Report of the President.

For automobile costs discussed in text, see Motor Vehicle Manufacturers Association, Motor Vehicle Facts and Figures: 1999, p. 51; Hiram Percy Maxim, "Some Data on the Cost of Operating Automobiles for Commercial Purposes" Scientific American Supplement No. 1479, 7 May 1904, pp. 23694-95. (Also see sources for Figure 1.) For road miles, see US Federal Highway Administration, Highway Statistics, various years.


The U-Haul company information for 1945 is from http://www.uhaul.com/corporate/company_history.html, downloaded 16 June 2000; current rental rates are based on a 19 June 2000 phone conversation with I-5 Auto Sales, Woodland WA.


Sources for Figure 1: Transportation Costs

The passenger car mileage cost series are built up from estimates of the fixed and variable costs of owning and operating an automobile (excluding the value of the driver’s
time) and assume 10,000 miles of use annually. The 1904 figure is from Hiram Percy Maxim, "Some Data on the Cost of Operating Automobiles for Commercial Purposes" Scientific American Supplement No. 1479, 7 May 1904, pp. 23694-95. This number is likely low because Maxim does not include the expenses of licenses, fees, taxes, garage/parking, or the special clothing many early automobilists purchased. The figures from 1924 to 1940 were based on authoritative studies conducted at the Iowa State Engineering Experiment Station. The sources are 1924: T. R. Agg and Harold S. Carter, "Highway Transportation Costs," Iowa State College Engineering Experiment Station Bulletin 69 (Ames, IA, 1924) p. 25; 1928: T. R. Agg and Harold S. Carter, "Operating Cost Statistics of Automobiles and Trucks," Iowa State College Engineering Experiment Station Bulletin 91 (Ames, IA, 1928) p. 19; 1931: Robley Winfrey, "Automobile Operating Cost and Mileage Studies," Iowa State College Engineering Experiment Station Bulletin 106 (Ames, IA, 1931) p. 18; 1934: Robley Winfrey, "Motor Vehicle Operating Costs as Affected by Roadway Surfaces," in National Research Council, Proceedings of the Fourteenth Annual Meeting of the Highway Research Board, 1934, (Washington, DC, 1935) p. 42; and 1939 average from R. A. Moyer and Robley Winfrey, "Cost of Operating Rural-Mail-Carrier Motor Vehicles on Pavement, Gravel, and Earth," Iowa State College Engineering Experiment Station Bulletin 106 (Ames, IA, 1939) p. 34. The figures for 1950 are from the AAA "Your Driving Cost" series are reported in Motor Vehicle Manufacturers Association, Motor Vehicle Facts and Figures: 1999, p. 51 and Motor Vehicle Facts and Figures: 1983, p. 43. These series are not completely consistent because among other issues, the vehicle model shifts from a full-size to intermediate car between 1973 and 1974. Data from the Federal Highway Administration Cost of Owning and Operating Automobiles and Vans publications for 1982 and 1984 indicate that the costs per mile for intermediate cars averaged about 90 percent of those for large cars. Thus the costs, consistently measured, would be higher than displayed in the post-1973 period. Data on finance charges are not available in the AAA series over the 1950-79 period and are excluded throughout. Including such costs does not meaningfully change the overall pattern.

Sources for Figure 2: Real Cost of Communications

Section IV.A.
Minor Civil Division (MCD)

Sample
Because data was entered from hardcopies and there are so many MCDs (over 30,000 in the current period), a sub-sample was necessary. Because MCD names vary over time, the sub-sample is based on a random, 1-in-10 sample of counties (N=311). The MCD sub-sample includes all MCDs (or “Civil Divisions” in the 19th Century) present within
this sample of counties. The same approach was used to generate a sub-sample of school districts.

**Data**
The number in parentheses following each data source indicates the years for which that source is used.

**Taxes/Spending**

Variable definitions: Taxes include all levied taxes including alcohol, public utility taxes and license taxes. School district taxes in 1880 and 1890 include school taxes levied by municipalities, and the data are at the MCD-level. Total spending includes all current operations expenditure excluding utilities and liquor stores. Protection Spending is current operations expenditure on fire plus police services. Education spending is current operations spending on education.

Notes: In all cases, MCDs or school districts with taxes/spending were dropped from the calculations (this does not noticeably alter the heterogeneity indices).


**Black population share**

**Young population share**

**Old population share**

**Owner-Occupied Homes**
Years: 1980, 1990. This variable is defined as the proportion of all homes which are owner occupied.

**Education**
Years: 1970, 1980, 1990. This is based on the population shares of the following five groups: 0-8 years education, 9-11 years education, 12 years education, 13-15 years education, 16+ years education. The data is for those who are at least 25 years old (the size of this group serves as the population weight in the heterogeneity measures).
Foreign-born population share
Years: 1870, 1930, 1970, 1980, 1990. The data for 1930 are just for white foreign-born (these values are normalized to the total white population.

Income Distribution
Years: 1969, 1979, 1989. Observation unit: families. Data note: the distribution is reported as the number of families/households whose income falls in a given income interval. There are 15 listed income groups for 1969, 17 income groups in 1979 and 25 groups in 1989.

Within-MCD (Tract Level) Data
The 1980 racial composition data comes from the tract-level version of ICPSR study 9693 (1980 data is used since the Census tract extract file for 1990 does not contain MCD identifiers). The sample is all tracts/block numbering areas and Enumeration Districts. All areas in the country in 1980 were tracted and/or in an Enumeration District (the ICPSR study 9693 documentation points out that no geographic area appears twice in the dataset). We use Census Enumeration Districts for areas where there are no tracts. Census documentation suggests these are comparable units, and excluding the enumeration districts does not noticeably alter any of our results. The term “tract” in the text refers to both tracts and Enumeration Districts.

In computing the within-MCD component, we use ICPSR study 9693’s assignment of each tract and Enumeration District to exactly one MCD or Census County Division (the study’s documentation states the latter can be treated as equivalent to a MCD).

Section IV.B.
Boston SMSA Data
All data are for the 92 municipalities in the 1980 definition of the Boston SMSA.

Party vote shares in presidential elections
Frequency: every 4 years. Availability: 1868, 1884-1988. Source: Massachusetts General Court (various years).

Current expenditure

Demographics (population, race, nativity, age)

Income

Section V.
County Data
The variables below are considered at the county level. All counties are included though data for Alaska is omitted because of inconsistencies in its county codes. The number in parentheses following each data source indicates the years for which that source is used.

Policy outcomes
Education Spending

Notes: In all cases the values sum up education spending by all governments within each county and include capital expenditures. The exact definitions of the spending categories are: for 1890, “ordinary expenditures on public common schools;” for 1932 “[school] government-cost payments operation and maintenance;” for 1957-1992 “direct general expenditures for education” which include spending on all education institutions within the county. See also the note in the taxes and revenues section below regarding New York area counties.

Taxes and Revenues
Frequency: Approximately every 10 years (1870-1942); every five years (1957-1992). Sources: the Census of Government. ICPSR study 3 (1870-1880), US Census Office,
Notes: (i) Data for the five counties in the greater-New York area (Bronx, Kings, New York, Queens, Richmond) are aggregated into one county. This is necessary since the Census puts all city spending in just one county, but shifts the identity of this county over time. All results are qualitatively similar if instead these five counties are excluded. (ii) Rhode Island is excluded because its counties have no fiscal responsibilities. (iii) Only certain variables are available each year. The four possible categories are: taxes collected by counties, Tax1; taxes collected by all local governments—county, minor civil divisions, school districts, etc.—within the county, Tax2; revenues—excluding intergovernmental grants but excluding inter-local government grants—collected by counties, Rev1; and revenues collected by all local governments within the county, Rev2.


Preference Proxies

Party vote shares in presidential elections
Frequency: every 4 years. Sources: ICPSR studies 8611 (1850-1972) and 13 (1976-88). Each election includes vote shares for Democrats and Republicans (except 1852 when Whigs are included) and significant minor parties (Socialists are included in 1912-32). In addition, each election included an “all others” category to ensure the vote shares sum to 1. For certain observations in ICPSR study 8611, the sum of the party vote shares is less than or greater than 1 (these discrepancies were confirmed by Erik Austin, the data archive director at ICPSR). In such cases the vote share are re-allocated in proportion to their stated values with the resulting shares now summing to 1.

Black population share
Frequency: every 10 years. Sources: ICPSR studies 3 (1850-70), 9693 (1980) and U.S. Census Bureau, 1994b (1990). The data for 1850-60 are the sum of the population categories “free colored” and “slave.” The data for 1870-80 are for the population category “colored.” The data for 1890-1970 are for the population category “negro.” The data for 1980-90 are for the population category “black.”

Religion shares
Frequency: every 10 years for 1850-70, 1906-36, 1980-90. Additional observations in 1890, 1952, 1971. Sources: ICPSR studies 3 (1850-70, 1890), 8 (1906-52), and Roper Center (1952, 1971, 1980, 1990). The data for 1850-70 is for the number of accommodations or sittings; the data for 1890 is for both the number of seatings and members; the data for the remaining years is for the number of members. No data was collected in 1880. The data for 1952 is from both ICPSR study 8 and Roper Center. Because the particular groups included in the files change over time (due largely to schisms and mergers), the data was aggregated into the following 27 denominational families: Adventist, Baptist, Black Baptist, Catholic, Communal, Congregationalist, Disciples of Christ, Eastern Liturgical, Episcopalian, Friends, Fundamentalist, Holiness,
Jewish, Lutheran, Mennonite, Metaphysical, Methodist, Black Methodist, Moravian, Mormon, New Age, Pentecostal, Presbyterian, Reformed, Unitarian-Universalist, Members of other bodies, and Non-classified. This classification scheme is based on the list in Melton (1978). Some denominations are not present in all years. All values are normalized by the county population. In one of our classification schemes we also generate an “unaffiliated” or non-sampled group to ensure the groups sum to 1. In certain years the sum of the group shares exceed 1 due to the manner in which the data is collected (these cases were confirmed by consulting the hard-copy versions of the various studies). When this happens, the groups are re-allocated in proportion to their stated values with the resulting shares now summing to 1. We uncovered some errors in the data which we corrected by consulting hard-copy versions of the statistics. A full list of these discrepancies as well as details of our classification scheme are available upon request.

**Young population share**
Frequency: every 10 years. Sources: ICPSR studies 3 (1850-1950, 1970), 7736 (1960), 9693 (1980), U.S. Bureau of the Census, 1994b (1990). Generally, the young are defined as those 5-20 years old. Because of data coding, this had to be modified in several years. 1850-60: young = 5-19 years old; 1870: young = 5-18 years old; 1880: young = 5-17 years old; 1910: young = 6-20 years old; 1920-30: young = 7-20 years old; 1980: young = 6-17 years old + 0.5×18-24 years old.

**Old population share**
Frequency: every 10 years (with gaps). Sources: ICPSR studies 3 (1850-60), U.S. Bureau of the Census, 1932 (1930), 7736 (1950-70), U.S. Bureau of the Census, 1998b (1980-90). The old are defined as those at least 65 years old. In 1850 and 1860, half of the individuals in the age category 60-69 year olds were counted as old as well as all individuals 70 years and older. No electronic data exists for this age group between 1870 and 1940.

**Marriage rate**
Frequency: every 10 years (with gaps). Sources: ICPSR study 3 (1930, 1950), U.S. Bureau of the Census, 1964 (1960), U.S. Bureau of the Census, 1973 (1970), U.S. Bureau of the Census, 1998b (1980-90). No marriage data are available for 1940. The Census defines these groups for those who are at least 14 years old (in 1950-70) and for those who are at least 15 years old (in 1930, 1980-90). These age categories are used to normalize the marriage rate and also serve as the population weight in the heterogeneity measures. The data for 1970 only include males.

**Owner-Occupied Homes**
Frequency: every 10 years. Sources: US Census Office, 1896 (1890), US Census Office, 1902 (1900), U.S. Bureau of the Census, 1913 (1910), U.S. Bureau of the Census, 1921 (1920), U.S. Bureau of the Census, 1933 (1930), ICPSR studies 3 (1940), 7736 (1950-70), U.S. Bureau of the Census, 1998b (1980-90). This variable is defined as the proportion of all homes which are owner occupied. Housing data for 1930 is unavailable in tape form. Instead we use values for the variable “families” listed in the hardcopies
(the prefatory remarks of the 1930 Census reads: “Since a home is defined as the living quarters occupied by a family, the number of homes is always the same as the number of families”).

**Education**
Frequency: every 10 years. Sources: ICPSR study 3 (1940-50), U.S. Bureau of the Census, 1964 (1960), ICPSR study 7736 (1970), U.S. Bureau of the Census, 1998b (1980-90). Our categories are those with only a high school degree and those with a college degree or more. The Census defines these groups for those who are at least 25 years old (this age category is used to normalize the education groups and also serves as the population weight in the heterogeneity measures). No education data are available prior to 1940.

**Foreign-born population share**
Frequency: every 10 years. Sources: ICPSR studies 3 (1850-1950), 7736 (1960), 9694 (1970), 9693 (1980), U.S. Bureau of the Census, 1994b (1990). The data for 1850-60 include both free and slave population. The data for 1910-30 are just for white foreign-born (these values are normalized to the total white population); in 1900 and 1940 the white and overall foreign-born shares are comparable. For 1950 the data is just for foreign-born aged 21 years or older (these values are normalized to the total population aged 21 years or older).

**Income Distribution**
Frequency: 1949, 1969, 1979, 1989. Sources: U.S. Bureau of the Census, 1952 (1949), ICPSR study 9694 (1969), U.S. Bureau of the Census, 1998 (1979, 1989). Observation unit: we separately consider both families and households (“families and unrelated individuals” in the 1949 Census). Household data for 1969 is not available in electronic form. Data note: the distribution is reported as the number of families/households whose income falls in a given income interval. There are 14 listed income groups for 1949, 15 listed income groups for 1969, 17 income groups in 1979 and 25 groups in 1989. The Census did not begin collecting income data until 1940 and detailed, county-level income distribution data is not available in electronic form prior to 1970 (records in the County and City Data Book generally include only 3 income groups). We collected a random, 1-in-10 sample (N=311) for 1949 from hardcopies in U.S. Bureau of the Census, 1952. To maintain comparability, the same 1-in-10 sample counties are used in the later years.

**Results Omitted from Text (Sections III, V, VI).**

**Local Heterogeneity (County and Contiguous Neighbors)**
Sources: The basic county contiguity data comes from ICPSR study 9835. The county boundary change data is based on ICPSR study 6576 (there are some obvious formatting errors in this study which we corrected). We followed the latter and formed counties into clusters which have “common historical origins” and can be thought of as the smallest units for which longitudinal analysis can be conducted. We form clusters based on the earliest year in our sample (typically 1850) and use the same set cluster definition in all later years. One problem with this is that areas without counties in the beginning of the
sample (typically western and southwestern states) are omitted in all years. To investigate the robustness of our results, we also formed year-specific clusters using ICPSR study 6576 (this means all counties with available data in a year are included). All of the our qualitative results still hold under this alternative approach.

**Net Migration by County**
Frequency: every ten years. Sources: Net migration rates and population shares for blacks/non-whites and the elderly from ICPSR studies 20 (1930-1950), 8493 (1950-1970) and 8697 (1970-1980). For each group and each decade, we regressed the group’s county-level net migration rate on its initial population share including controls for the county’s overall population and (in the black/non-white regressions) whether the county was in the South.

**Metropolitan Area Racial Heterogeneity**
Frequency: 1930, 1990. Sources: For 1930 we use U.S. Bureau of the Census, 1932, *Fifteenth Census, Metropolitan Districts* and for 1990, U.S. Bureau of the Census, 2000. In both years heterogeneity is calculated based on comparing the aggregate black population share across metropolitan areas. In 1930, we use the 96 metropolitan districts listed in the Census. Each district had an aggregate population of 100,000 or more and contained one or more central cities of 50,000 or more inhabitants (p.6). A district included “in addition to the central city or cities, all adjacent and contiguous civil divisions having a density of not less than 150 inhabitants per square mile, and also, as a rule, those civil divisions of less density that are directly contiguous to the central cities, or are entirely or nearly surrounded by minor civil divisions that have the required density.” (Italics in the original, pp. 5-6). In 1990, we use the 284 largest metropolitan statistical areas (MSA).
Data References
All ICPSR files are available at http://www.icpsr.umich.edu/. Census and Massachusetts documents with no on-line or CD-ROM reference were used to manually input data.


ICPSR studies 7736, 8256, 9251 (various years). *County and City Data Book*. Investigator: U.S. Bureau of the Census.


Massachusetts, General Court (various years). *Manual for the General Court.* Commonwealth of Massachusetts, Boston.


Roper Center (various years). *Churches and church membership in the United States.* Unpublished data tape at the Social Science Data Center of the University of Connecticut, Storrs, CN.


U.S. Census Office (1880). *Valuation, Taxation and Public Indebtedness, 1880*. Washington DC, GPO.


