I. Introduction

The Urban Institute's Under Class Data Base (UDB) contains social, demographic, economic, and housing data on census tracts in the United States for 1970, 1980 and 1990. Census tracts are locally-determined geographic units, typically including between 2,500 and 8,000 persons. Tracts are meant to approximate "neighborhoods" by capturing a group of residents with similar social characteristics, economic status, and housing conditions. For 1970, the UDB contains 904 variables for 34,586 census tracts; in 1980, there are 1,098 variables for 43,221 tracts; and in 1990, there are 1,062 variables for 61,258 tracts. In all 3 years, variables range from very general characteristics (e.g., total population and number of housing units in a tract) to detailed cross-tabulations of sub-groups (e.g., race/ethnicity by educational attainment by employment status).

Data in the UDB is based on information gathered by the Census Bureau in its decennial census. The Bureau makes census tract data available to the public in both printed and machine-readable formats, but these products generally force users to focus on one tract and one characteristic at a time. By compiling data on all census tracts into one data file, the UDB allows users to simultaneously analyze numerous (or all) tracts on a host of dimensions.

Another powerful feature of the UDB is its ability to match tracts across all three Census years, enabling users to examine changing tract characteristics between 1970, 1980 and 1990. For most tracts, a unique tract identification code applies to the same physical area in all three years. But because the boundaries of some tracts change between the decennial censuses (either splitting into several tracts, merging with other tracts, or appearing for the first time), a methodology has been developed to link such tracts and their associated data back to their original geographic dimensions.

While census tracts are the UDB's basic unit of observation, information on a tract's metropolitan area, county and state affiliation are also included in the data base. This allows users to sum data from census tracts to broader levels of geography (for selected variables, this has already been done at the metropolitan level). Conversely, users can append information to the UDB from other sources that supply their data at the state, county or metropolitan area level. Finally, other tract-level data bases can be merged to the UDB. For 1980 and 1990, place codes are also provided for each census tract.

A. History and Uses of the UDB

In 1989, the UDB was developed at the Urban Institute by Isabel Sawhill and Erol Ricketts. Initially, the UDB only contained data for 1980. It was later expanded under the supervision of Ronald Mincy and Susan Wiener to include 1970 and 1990 data. The creation of the UDB and distribution of the data have been funded by the Rockefeller Foundation.

Urban Institute researchers are using the UDB to analyze trends in the growth and composition of under class and concentrated poverty areas (see Mincy and Wiener, 1993; Mincy, 1993; Galster, Mincy, and Tobin, 1993). Under class areas, as measured by the Urban Institute, are census tracts that simultaneously score high on four indicators: the fraction of female-headed

families; high school dropouts; males not attached to the labor force; and welfare recipients. Concentrated poverty areas (sometimes called extreme poverty areas) are tracts in which 40 percent or more of the residents live below the official poverty threshold. Urban Institute researchers are also utilizing the UDB to examine changes in the geographic concentration of child poverty, regional patterns of economic distress, and the fortunes of newly arrived immigrants (see Tobin, 1993 and Tobin, Fix and Zimmermann, forthcoming).

In addition to being a rich data source for social science research, the UDB has provided valuable information to state and local officials, helping them identify distressed communities within their jurisdictions. For example, health officials from the State of Virginia have used the UDB to target neighborhoods in need of public health facilities. The Urban Institute also makes data from the UDB available to community-based organizations, assisting them in identifying potential clients and site locations for service delivery. Recently, the One-to-One Partnership used information from the UDB to target their mentoring programs to disadvantaged neighborhoods.

B. Limitations of the UDB

While the UDB has powerful potential as a social science tool and information source, it does have its limitations and drawbacks.

As discussed in detail below, the UDB is based on information provided by the Census Bureau. This agency's decisions about what information to gather in the decennial Census and how to tabulate the data plays a critical role in determining the UDB's format. For obvious reasons, not every piece of information supplied by the Census Bureau has been included--while the UDB was designed for a broad audience of users, it will not satisfy everyone's data needs. The Census Bureau's selection of data itself tends to be stronger in certain areas (e.g., employment) than others (e.g., mobility patterns), and some topics are simply not covered (e.g., workers' wages). The 1990 Census, however, represents a marked improvement from previous years in both the scope of its coverage and the complexity of its cross-tabulations.

Using data from different years creates some difficulties because of changes in Census Bureau methodology. For many variables, we have attempted to find equivalent measures in all three years; but due to changes in the questions asked in the Census and the Census Bureau's tabulations, it has not always been possible to perfectly match variables across years. Notice is provided when such changes have occurred. Likewise, the Census Bureau has revamped major policies (detailed in Sections III and IV) concerning geography and racial identification, which creates further challenges for comparing data between 1970, 1980 and 1990.

It is also important to recognize that the UDB does not provide information on individuals directly--all data is aggregated to the census tract level. This is done to preserve the confidentiality of respondents that is guaranteed by Federal law (see Section IV for more on suppression of data). Thus, while the UDB can be used to measure the percentage of a tract's population occupied as professionals and be used to calculate the tract's median income, the UDB cannot reveal the average income of professionals. Similarly, changes in a tract's characteristics over time indicate little about the fortunes of the individuals initially residing in

the tract since large numbers of people migrate into and out of tracts between the decennial censuses.   For certain variables, cross tabulations by characteristics such as race, family structure and education provide some information about relationships between variables, but this is done at a general level.   In short, the UDB is best thought of as a source of data on neighborhoods and areas.

A final caveat concerns the nature of the UDB's source: a decennial census.   Nearly all data in the UDB are from 3 points in time: 1970, 1980, and 1990.   Although data from these three dates is useful in performing certain dynamic analyses, doing so reveals little about the intervening periods, during which many important and interesting changes may have taken place. For some variables which move relatively slowly and steadily over time (such as racial composition), this may not be a significant problem; but for highly variable or cyclical data (such as employment), the data from 3 points probably hide as much as they uncover--if unemployment rates were low at the end of decades, but high in the interim, simply examining data from 1970, 1980 and 1990 would be misleading.