A Dataset for U.S.-Mexico Border Research

Joan B. Anderson
University of San Diego
James Gerber
San Diego State University

This note describes a new dataset for research on the United States-Mexico border regions. The dataset forms the empirical core of the newly published book, *Fifty Years of Change on the U.S.-Mexican Border: Growth, Development, and Quality of Life* by Joan B. Anderson and James Gerber (2007), and is contained in 75 Excel files hosted by the Center for Latin American Studies at San Diego State University. The files can be found at http//latinamericanstudies.sdsu.edu/BorderData.html.

Coverage: Geography and political units

The data are a compilation of Mexican and U.S. census data from the 6 decennial censuses of population and housing taken from 1950 to 2000, along with estimates of a few variables for which there are no census reports (e.g., local income). The political boundaries of the data are counties (U.S.) and municipios (Mexico) while the geographical definition of the border includes all counties and municipios that touch the international border. In total there are 25 U.S. counties in 4 states, and 38 Mexican municipios in 6 states.[1] The Excel data tables are numbered to correspond to the numbering of figures in the book. For example, Table 2.1a presents the Mexican data by nation, border region, border state and municipio and Table 2.1b presents the comparable data from the U.S. side. These two tables correspond to Figure 2.1 in Chapter 2. The 75 files are composed of 27 files with demographic data, 5 files with language and ethnic

[1] As described below, there were changes in political boundaries. Consequently, the number of municipios in our dataset is slightly different from the today's count. We explain this below.

data, 16 files of labor force characteristics, 5 files on income and poverty, 20 files relating to housing and quality of life, including education and infant mortality, and 2 tables of human development indexes, along with the sub-indexes. Most files span the period 1950 to 2000 in the ten-year increments corresponding to the decennial censuses.

Mechanics of the constructed variables

Data for each border region as a whole is the sum of the values for each of the counties and municipios. For example, the U.S. border region population in 2000 was 6,312,252 which is the sum of the population in each of the 25 border counties in 2000. All percentage changes and ratios for border regions were calculated by adding the raw numbers for the counties or municipios to find total for the border and then the percentages or ratios were calculated.

To calculate the annual growth rate over a ten year period we calculated a simple compound rate by using the formula:

$$Y_1 = Y_0*(1+g)^t,$$

where Y_1 is the end value, Y_0 is the starting value, g is the rate of growth, and t is the number of years between the starting and ending values. For more detail see Box 2.2 in Chapter 2 of the book (Anderson and Gerber, 2007). Fertility rates are measured in the Mexican data as the total fertility rate, which is equal to the total number of children ever born divided by the total number of women of childbearing age. We have this data for the U.S. as well, from 1950 through 1990, but the U.S. census did not present total fertility rates for women in 2000. The fertility rate, measured as births per 1000 women between 15 and 44 years of age, is available for the whole period for the U.S. Net migration was calculated as the residual difference between total population growth and

growth due to natural increase. While the population growth was the simple compound annual average for each decade, the natural increase was calculated as births minus deaths of the census year in question. For more detail see Box 2.3 in Chapter 2 of Anderson and Gerber (2007). In both countries internal migration is measured by the census data as "born in other states". Foreign migration is measured by the number of "foreign born".

Calculating the labor force participation rate is a good example of changes over time and space with respect to the definition of population eligible for the labor force. Labor force participation rate is defined as the proportion of "working-age" that is actively in the labor force (economically active). The problem is that the definition of working-age varies with economic development, especially the minimum age for labor force participation. In Mexico it has remained at 12 years of age over the fifty year period. In the U.S., the Census Bureau defined the minimum working age as 14 for 1950, 1960 and 1970, but then raised it to 15 in 1980 and thereafter. We used the census definition which may account for at least some of the lower labor force participation rate in Mexico compared to the U.S. for much of this period. Discussion of the definitions of the labor force participation rate and unemployment rate for each country is contained in Box 6.1 in the book. Age-segmented population also differs for calculations of educational attainment between the U.S. and Mexican censuses. The U.S. measures attainment by the proportion of population 25 years and older while the Mexican census uses the proportion of population 15 years and older for all education levels except for attainment beyond high school. We used the detailed tables on educational attainment by

age contained in the Mexican census to convert the Mexican high school attainment variable to that for population 25 years and older.

Another challenge was to estimate per capita income for the Mexican municipios. The main problem is that Mexico's Instituto Nacional de Estadística, Geografía, e Informática (INEGI) provides estimates of Gross State Product, divided into 9 sectors,[2] but no income estimates at the municipio level. Gross State Product was allocated to the municipios in accordance with employment by sector. In other words, municipio m's share of state income is equal to the sum of the products of state-level sectoral income shares times municipio-level employment shares. This will give a reasonable approximation as long as we can assume that productivity within a sector is the same across all the municipios of a given state. This income is divided by the population of the municipio and then translated into constant 1996 U.S. dollars using purchasing power parity exchange rates based on the Penn World Table, version 6.1 (Heston, Summers, and Aten, 2002). More detail on the income variable can be found in the technical appendix to chapter 7 of the book or in Anderson and Gerber (2004).

Reliability and comparability

The chapter on methodology in the *Oxford Latin American History Database* notes that:

> All comparative analyses, particularly when extended over many countries and many years, encounter methodological problems such as the reliability, consistency, and comparability of the data series. The reliability of statistical data depends to a large extent on the quality of the source material….

[2] The division of economic activity varies overtime, but earlier censuses divisions can be combined into aggregates that match later censuses.

While we believe that the national censuses are the best available data, we are not of the opinion that they are 100 percent error free. The remainder of this note attempts to alert the reader to some of the areas where problems may exist.

First of all, a data set that covers six censuses and fifty years will inevitably have changes in the units of analysis and variables collected, both of which can reduce data comparability through time. Rapid population growth caused changes in political divisions in Mexico, while in the U.S., county lines generally remained fixed once they were established and regardless of how much population expanded. As a result, several new municipios were created over the 50 year period of the study. Between 1950 and 1960 Miguel Aleman broke off from Mier (Tamaulipas), Valle Hermosa from Matamoros (Tamaulipas), and Puerto Peñasco from San Luis Río Colorado (Sonora). During the 1960's, Gustavo Diaz Ordaz and Rio Bravo (both in Tamaulipas) came into being. These new municipios were simply added to the data set when they were created. This causes some distortion in the population growth rates for the municipios that were divided, but the border region aggregates are not affected at either the state level or for the border region as a whole. After 1970, there was much less division and only two more municipios were created. During the 1980s, General Plutarco E. Calles split off from Puerto Peñasco (Sonora) and in the 1990s Playas de Rosarito from Tijuana (Baja California). Given that these last two new municipios have a very short history relative to the data set, we ignored the separation and added the numbers for the newly created municipios back into the original. Therefore, the values for General Plutarco E. Calles are contained within Puerto Peñasco for 1990 and 2000 and Playas de Rosarito numbers are added to Tijuana for 2000.

In addition to changing political boundaries, a second obstacle to the creation of comparability over time is that the variables collected by the respective census bureaus are sensitive to the level of economic development. This causes problems of comparability over time, since a great deal of economic development has occurred on both sides of the border during this fifty year period. It also causes problems of U.S.-Mexico data comparability, given the difference in levels of development of the two nations. For example, in the 1950 U.S. census, the population breakdown between rural and urban appears in the first population table while by 2000 it is buried and difficult to find. In addition, the U.S. ceased to collect data on basic literacy after 1970 and the only housing quality data on plumbing is whether or not the house has full plumbing. On the Mexican side, plumbing is divided into a number of categories, including connection to sewage systems, the type of sewage systems, availability of running water to the property, availability of water to the inside of the house and the bathroom(s) and so forth.

A third and more intractable problem is the accuracy of the census enumeration. Head counts are inevitably political acts since federal financial support is often allocated on the basis of population. Consequently, even a perfectly accurate count would likely generate complaints of undercounting by sub-national units. The mayors of Tijuana, for example are famous for arguing that the "real" population of Tijuana is at least 3 million (1.4 million were counted in the *2005 Conteo de Población*). Add political intrigue, lack of transparency in federal funding, extremely rapid urban sprawl, and tens of thousands of recently arrived squatters in the peri-urban areas of rapidly growing cities, and it is not surprising that there are many voices claiming that the population is undercounted. More

sober and impartial analyses have shown that in some cases at least, there is some truth to the charges.

The United States began to formally evaluate its census accuracy in 1950 and by 1980, there had developed more than 40 separate projects as part of its Research, Evaluation, and Experimentation program. Among the better known are the pos-enumeration program and the content re-interview survey. These and others have demonstrated that ethnic minorities (Blacks and Latinos) tend to be undercounted, as do people living in Western states, and the homeless (U.S. Bureau of the Census, 1989). The Mexican case tends to be more complex, in part because there is not the same published documentation of census checks, and also because checks on the census using demographic estimation techniques tends to be more difficult given the higher frequency for under registering births and deaths.

Rodolfo Corona (1986) provides one of the few evaluations of the Mexican census. Corona looked primarily at the 1980s census, and to a lesser degree, the 1970 census. He also examined data by state. A common technique for checking reliability and consistency is to construct estimates using techniques from demographic theory, and then to compare the estimates to the census enumeration. This is not always successful at uncovering inconsistencies, but it is one of the few ex-post methods available. In Corona's case, he begins with registered births which are acknowledged by everyone to be an undercount. Using life tables, he projects the population for various age cohorts born between 1970 and 1980 and finds the estimated population of 0 to 5 and 5 to 14 to be much larger than the actual count. This is prima facie evidence of a census undercount given that he begins with what everyone agrees is an undercount of births.

Furthermore, investigation of the preliminary results for the 1990 census show a similar pattern according to the same researcher (Corona Vázquez, 1991), the declarations of accuracy by INEGI notwithstanding.

There are two main difficulties for an accurate census count in the Mexican border region: migration and the definition of residency. To some extent, the problem of large migratory flows is akin to the problem of trying to count homeless people, since many migrants may not have a permanent residence and are therefore difficult to count. This is true for the U.S. as well as Mexico. A related problem is the issue of residency and whether a migrant should be counted in their place of origin or their place of residence. The answer depends in part on the definition of residence, as well as the ability of census enumerators to uncover the information. Mexico's census bureau, INEGI (Instituto Nacional de Estadística, Geografía y Informática), applies a definition of residency that requires a person to be in a geographical place for 6 months or more, or to intend to be there for 6 months or more. This definition does not appear to have changed over time, and was applied at least as far back as 1970. The idea is that tourists and temporary visitors to a region should not be counted as residents, but it is often difficult to discover intentions or length of stay, and as everyone knows, temporary migrants often have a way of becoming permanent. Corona (1986) found that at least for the 1980 census, the reported net migration data was inconsistent with the state of birth data.

A common complaint about the Mexican census, the alleged failure to include new squatter settlements that appear over night on the periphery of many border cities, may be less important than many believe. INEGI attempts to conduct its count in the

shortest time possible (for example, the 1980 census was done in one day, more recent ones in a couple of weeks) and census takers are instructed to amend their maps of neighborhoods if they encounter new and unexpected communities while canvassing the outer reaches of an urban area. To the extent possible, squatter settlements are counted.

In sum, it is impossible to provide a standard error for the data but we note that there is a continuing problem of undercounting, particularly in the border region where the characteristics of those less likely to be counted are more frequently observed. Later censuses in both countries appear to have become more sophisticated in their attempts to address this issue, but the problem does not appear to have been resolved completely. Hence, we hope that data users will recognize that the numerical values are suggestive rather than precise measurements. Nevertheless, they are the best available source of information and our hope is that this data will be useful for other border researchers. If you have additional questions about the data in the website you may contact Joan Anderson at joana@sandiego.edu or Jim Gerber at jgerber@mail.sdsu.edu.

References

Anderson, Joan B. and Gerber, James (2007). *Fifty Years of Change on the U.S.-Mexican Border: Growth, Development, and Quality of Life.* Austin, Texas: University of Texas Press.

Anderson, J.B. and Gerber, J. (2004). A human development index for the United States-Mexico border. *Journal of Borderlands Studies*, *19*, 1-26.

Corona, Rodolfo (1986). *Evaluacíon de los Datos Censos de 1980: Población Residente y migración en Baja California.* Tijuana, Baja California: Centro de Estudios Fronterizos del Norte de Mexico.

Corona Vázquez, Rodolfo (1991). Confiabilidad de los resultados preliminares del XI Censo General de Población y Vivienda de 1990. *Estudios Demográficos, 16,* 33-68.

Heston, A., Summers, R., and Aten, B. (2002). *Penn World Table Version 6.1*. Center for International Comparisons at the University of Pennsylvania (CICUP).

Oxford University Press (2007). "Methodology," *Oxford Latin American History Database.* Available: http://oxlad.qeh.ox.ac.uk/sources.php

U.S. Bureau of the Census (1989). *Census of Population and Housing (1980). History.* Washington, DC. U.S. Government Printing Office.