

**DATA APPENDIX TO**  
**“UNDERSTANDING THE IMPACT OF IMMIGRATION ON CRIME”**

*A. Crime Data*

All measures of crime are based on agency level data on the number of crimes reported to the police, as compiled by the Federal Bureau of Investigations (FBI) in its *Uniform Crime Reporting* program and distributed by the Inter-University Consortium for Political and Social Research (ICPSR). Data for 1980 and 1990 are contained in Study No. 9028, and information on reported crimes in 2000 is distributed as part of Study No. 3447.

The number of reported offenses is available for the seven Index I crimes: murder, rape, aggravated assault, robbery, burglary, larceny, and motor vehicle theft. The FBI classifies the first four of these as violent crimes; the latter three are denoted property crimes. A single reported incident involving different crimes is scored only once. It is counted only under the most serious crime involved. E.g. two offenders breaking into a car dealership killing the night guard in the process would be counted as one homicide, not as burglary. See U.S. Department of Justice (2004) for a detailed guide on scoring and classifying offenses.

Data on reported crimes are available at the national, state, county, and agency level. Due to changes in the imputation procedures in 1994 county level data before and after 1994 are not comparable. This necessitates the use of agency level data, which has to be appropriately adjusted for non-reporting by police agencies and aggregated to the county level. That is, if a police agency submits reports for at least 1 month, but less than 12, in a

given year, the total number of crimes it reports for this year is inflated by a factor of  $\frac{12}{\#reports}$ . Agencies reporting 0 months are not considered.

An alternative way of adjusting for non-reporting suggested by the National Archive of Criminal Justice Data (NACJD) is to inflate the number of reported crimes for agencies submitting reports for at least 3 months by a factor of  $\frac{12}{\#reports}$ , and to assign an imputed value to agencies reporting 2 or fewer months. NACDJ uses the mean value of agencies reporting 12 months in the same geographic stratum for its imputations.<sup>1</sup> This method has the downside that it introduces substantial correlation in the error terms across counties. The results of the paper remain qualitatively unchanged when using this alternative adjustment mechanism.

Aggregating agency level crime data to the county level is done by adding the adjusted number of crimes of all agencies in a given county. ICPSR Study No. 4634 provides a crosswalk between agency identifiers (ORI) and county codes (FIPS).

To avoid losing approximately one third of the sample, data for 1970 (distributed in ICPSR Study No. 4198) has been used in estimating the first differences models in Tables 4 and 5.

For the years 1970, 1980, 1990, and 2000 the following variables are used in the analysis:

**Violent Crimes** is the sum of all murders, rapes, aggravated assaults, and robberies known to police in a given county during a particular year.

---

<sup>1</sup> See the description at <http://www.icpsr.umich.edu/NACJD/ucr.html> for details.

**Murder** refers to the crime of murder and non-negligent manslaughter and is defined as the willful non-negligent killing of a human being by another one.

**Rape** refers to the crime of rape by force. Rape by force defined as the carnal knowledge of a female forcibly against her will, where carnal knowledge is the act of a man having sexual intercourse with a woman.

**Aggravated Assault** is defined as an unlawful attack by one person upon another one for the purpose of inflicting severe bodily injury. Aggravated assaults are often accompanied by the use of a weapon.

**Robbery** is defined as the taking or attempt to take anything valuable from its owner or custodian by force, threat of force, or intimidation. Both armed and unarmed robberies are subsumed in this category.

**Property Crimes** is the sum of all burglaries, larcenies, and motor vehicle thefts known to police in a given county during a particular year.

**Burglary** is defined as the unlawful entry into a structure with the intent to commit a felony or theft. Forcible entry, unlawful entry without the use of force, and attempted forcible entry are subsumed in this category.

**Larceny**, or theft, is defined as the unlawful taking away of property from the possession of its owner or custodian. Pocket-picking, purse-snatching, shoplifting, thefts from motor vehicles, thefts of motor vehicle parts and accessories, theft of bicycles, theft from buildings, theft from coin-operated devices or machines, etc. are included in this category.

**Motor Vehicle Theft** is defined as the theft or attempted theft of a motor vehicle. A motor vehicle is defined as a self-propelled vehicle running on land surface and not on rails.

### *B. County Level Covariates*

In its *City and County Data Books* and *USA Counties* the US Census Bureau publishes information on county characteristics in various years. Both publications contain a collection of data from the Census Bureau itself and other government agencies based on, for instance, *Decennial Censuses of Population and Housing*, *Censuses of Government*, *Economic Censuses*, etc.

As information in *USA Counties* for years prior to 1977 is very sparse, data for preceding years, which is used in estimating the first differences models, has been taken from the *City and County Data Books* for 1967, 1972 and 1977. *USA Counties* has been obtained from the US Census Bureau website;<sup>2</sup> and ICPSR Study No. 2896 contains the data from the *City and County Data Books* used in this paper.

All dollar values have been converted to 2000 dollars using the Consumer Price Index for All Urban Consumers, and control variables, for which no information is available in a particular year, have been linearly interpolated based on the two closest available years. For instance, in the raw data Police Expenditure per Capita is not available for 1980, but only for 1977 and 1982. A value for 1980 is then imputed based on linear interpolation from 1977 to 1982.

---

<sup>2</sup> See <<http://censtats.census.gov/usa/usa.shtml>>.

County level information for different years has been merged on FIPS codes. Counties that could not be matched consistently over time have been dropped from the analysis. The final sample includes 3,117 counties.

Below follows a description of each variable used in the analysis, and its original source.

**Immigrants** is defined as the number of foreign-born individuals in a given county during a particular year. Information on the foreign-born is contained in the 1970, 1980, 1990, and 2000 U.S. Censuses. The data used in this paper has been obtained from *USA Counties* for the 1980, 1990, and 2000 Censuses, and from the National Historical Geographic Information System (NHGIS) for the 1970 Census.<sup>3</sup>

**Total Population** is defined as the total number of residents in a given county during a particular year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

**Population per Square Mile** is defined as the total number of residents per square mile in a given county during a particular year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

---

<sup>3</sup> *USA Counties* is available online at <<http://censtats.census.gov/usa/usa.shtml>> and data can be obtained from NHGIS at <<http://www.nhgis.org/>>.

**Fraction Female** is defined as the number of female residents divided by the total number of residents in a given county during a particular year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

**Median Age** is defined as the as the age in years of the median person in the age distribution within a given county during a particular year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

**Region** is defined as a set of indicator variables for each of the four Census Bureau Regions: West, Midwest, South, and Northeast.<sup>4</sup>

**Median Household Income** is defined as the as the income, inflated to 2000 dollars (using the Consumer Price Index for All Urban Consumers), of the median household in the income distribution within a given county during a particular year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

**Fraction Families Below Poverty Level** is defined as the as number of families below the poverty level applicable to a particular year divided by the number of families for whom poverty status has been determined in a given county and that year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been

---

<sup>4</sup> See <[http://www.census.gov/geo/www/us\\_regdiv.pdf](http://www.census.gov/geo/www/us_regdiv.pdf)> for precise definitions of each Census Region.

obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

**Payroll per Capita** is defined as the as the total private non-farm annual payroll, inflated to 2000 dollars (using the Consumer Price Index for All Urban Consumers), divided by the total number of residents in a given county and a particular year. It is based on information collected by the US Census Bureau. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. The necessary information is not available for 1970 in the *City and County Data Books* contained in ICPSR Study No. 2896.

**Unemployment Rate** is defined as the percentage of the civilian labor force that is not employed in a given county during a particular year. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1977 *City and County Data Book* contained in ICPSR Study No. 2896. The information in these data sets is based on official publications by the U.S. Census Bureau.

**Police Expenditure per Capita** is defined as all direct general expenditures for police protection by local government, inflated to 2000 dollars (using the Consumer Price Index for All Urban Consumers), dived by the total number of residents in a given county and a particular year. It is based on information collected by the US Census Bureau. For the years 1977, 1982, 1987, 1992, 1997 and 2002 it has been obtained from *USA Counties*, and linearly interpolated. The necessary information is not consistently available the *City and County Data Books* contained in ICPSR Study No. 2896 for years preceding 1977.

**Fraction Institutionalized** is defined as the number of inmates in institutions, such as prisons, jails, and mental institutions, dived by the total number of residents in a given county and a particular year. It is based on information collected by the US Census

Bureau. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from ICPSR Study No. 2896.

**Median Value of Housing Units** is defined as the as the value, inflated to 2000 dollars (using the Consumer Price Index for All Urban Consumers), of the median housing unit in the value distribution within a given county during a particular year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

**Median Rent** is defined as the as the rent, inflated to 2000 dollars (using the Consumer Price Index for All Urban Consumers), of the median housing unit in the rent distribution within a given county during a particular year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

**Fraction of Housing Units Owner Occupied** is defined as the number of housing units that are occupied by their respective owners divided by the total number of housing units in a given county during a particular year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

**Fraction of Housing Units Vacant** is defined as the number of housing units that are vacant divided by the total number of housing units in a given county during a particular year. It is based on information collected by the US Census. For the years 1980, 1990,

and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

**New Building Permits per Existing Unit** is defined as the number of building permits issued in a particular year divided by the total number of housing units in a given county during that year. It is based on information collected by the US Census. For the years 1980, 1990, and 2000 it has been obtained from *USA Counties*. For 1970 it has been taken from the 1972 *City and County Data Book* contained in ICPSR Study No. 2896.

### *C. Construction of Instrument*

County level information on the number of immigrants from different source countries by decade (based on the respective US Census) is distributed by the National Historical Geographic Information System (NHGIS). Data for the years 1960, 1970, 1980, 1990, and 2000 have been obtained from this source, and have been merged with the previously described data.

In creating the instrument the set of countries in the raw data has been aggregated up into nine groups: Northwestern Europe, Eastern Europe, Southern Europe, Asia, Mexico, South and Central America, Africa, Canada, and all other countries. Aggregation is required as only a subset of source countries is consistently identified in the raw data, whereas especially in earlier years information on other countries has already been aggregated. Moreover, aggregation has the advantage of lessening measurement error, which is almost surely present in the number of immigrants from any individual country. The total number of immigrants in each group for the US as a whole as well as for each county separately has been determined for each year, i.e.  $\sum_c immigrants_{c,g,t}$  and

$immigrants_{c,g,t}$ , where  $t$  indexes years,  $c$  counties, and  $g$  denotes one of those nine source country groups. County  $c$ 's *predicted* total number of immigrants in year  $t$  is defined as

$$\hat{immigrants}_{c,t} = \sum_g \left[ \left( \sum_c immigrants_{c,g,t} \right) \left( \frac{immigrants_{c,g,t-20}}{\sum_c immigrants_{c,g,t-20}} \right) \right].$$

The predicted change in the natural logarithm of the number of immigrants from year  $t - 10$  to  $t$ , e.g. the actual instrument used in the paper, then equals

$$\Delta \log \left( \hat{immigrants}_{c,t} \right) = \log \left( \hat{immigrants}_{c,t} \right) - \log \left( \hat{immigrants}_{c,t-10} \right).^5$$

---

<sup>5</sup> Observations for which this number could not be determined due to missing information on the composition of immigrants have been assigned a value of .5. The results are quantitatively and qualitatively robust to this imputation.

## REFERENCES

U.S. Department of Justice (2004). *Uniform Crime Reporting Handbook*. Washington, D.C.