Appendix A. Hispanic Adjustment Procedure to Produce "Clean" UCR White and Black Arrest Counts

Note: *Confounded* arrest counts refer to White, Black, or total arrest figures that include Hispanics; and *Clean* arrest counts refer to White, Black, or total arrest figures that do not include Hispanics.

Our adjustment method for removing the "Hispanic effect" from White and Black UCR arrest figures follows a straightforward procedure that is elaborated subsequently and includes the following four steps: 1) mimic UCR national estimates in the CA–NY data by adding Hispanic arrests into White and Black arrest categories to create *confounded White and Black CA–NY arrest figures*, 2) downward-adjust these *confounded White and Black CA–NY arrest figures* (to account for the relatively larger Hispanic population in CA–NY than in the nation as a whole), 3) use *clean* and *confounded White and Black CA–NY arrest figures* to create *correction factors* for UCR data, and 4) apply these *correction factors* to UCR arrest figures to estimate *clean* national counts of White and Black arrests that do not include Hispanics.

The first step in our adjustment procedure is to mimic UCR estimates by using the refined race and ethnicity information provided in CA–NY data to generate *confounded* CA–NY White and Black arrest counts that lump Hispanics into White and Black arrest categories as found in the UCR. Recall that, in contrast to national UCR arrest counts, the CA–NY data include a separate identifier for Hispanic arrestees and thus already provide *clean* White and Black arrest categories. Therefore, our first step is to reallocate the clean CA–NY Hispanic arrests into White and Black categories based on the share of the Hispanic population in California and New York that is considered White or Black, calculated as follows:

$$C_{ijk} = U_{ijk} + (H_{jk} \times R_{ik}) \qquad (A.1)$$

where C_{ijk} are the *confounded CA–NY arrest counts* for race group i (White or Black), offense j, and year k; U_{ijk} are the *clean* (original) *CA–NY arrest counts*; H_{jk} are the Hispanic arrest counts in the CA–NY data; and R_{ik} is the percentage of the CA–NY Hispanic population in race group i (White or Black) for year k.

Second, we recognize that the Hispanic effect on White and Black arrests counts is likely much stronger in CA–NY than in UCR data (because of the relatively greater presence of Hispanics in California and New York compared with the United States as a whole). Thus, we downward adjusted the size of the Hispanic effect on White and Black arrests in CA–NY data. Specifically, we multiplied the Hispanic arrest count in equation A.1 by

the ratio of the percent Hispanic in the U.S. White (or Black) population over the percent Hispanic in the CA–NY White (or Black) population. The following formula extends equation A.1 and illustrates this downward adjustment:

$$C_{ijk} = U_{ijk} \left[(H_{jk} \times R_{ik}) \times \left(\frac{P_{1ik}}{P_{2ik}} \right) \right] \qquad \text{(A.2)}$$

where P_{1ik} is the proportion of the race-group i population in the United States that is Hispanic for year k and P_{2ik} the proportion of the same race-group i population in California and New York that is Hispanic for the same year k.

Equation A.2 consists of two parts. The first part, described in equation A.1, involves the initial adjustment in which Hispanic arrests in CA–NY are reallocated into White and Black arrests to generate the *confounded CA–NY arrest counts* similar to those found in national UCR estimates. These counts then are multiplied by the second part of the equation (P_{1ik} / P_{2ik}), which downward adjusts the Hispanic reallocation into White and Black arrest counts that take into account the greater presence of Hispanics in California and New York relative to the United States as a whole. The end result proxies the confounded White and Black arrest counts found in national UCR estimates adjusting for the relatively greater presence of Hispanics in the CA–NY data.

Third, we compare the *clean* and *confounded* CA–NY arrest figures to create *correction factors* for removing the Hispanic effect on national UCR Black and White arrest figures, calculated as the ratio of *clean* over *confounded CA–NY arrest counts* (for each race, offense, and year), which is expressed as follows:

$$X_{ijk} = \frac{U_{ijk}}{C_{ijk}} \qquad \text{(A.3)}$$

where X_{ijk} is the *correction factor*, U_{ijk} is the *clean CA–NY arrest count*, and C_{ijk} is the *confounded CA–NY arrest count* for race group i, offense j, and year k.

Fourth, we apply the *correction factors* to the UCR arrest counts to estimate *clean UCR White and Black arrests* that exclude Hispanics, which is illustrated as follows:

$$Y_{ijk} = X_{ijk} \times A_{ijk} \qquad \text{(A.4)}$$

where Y_{ijk} are the *clean UCR arrest counts*; X_{ijk} are the *corrections factors* derived using the CA–NY data; and A_{ijk} are the *confounded (original) UCR arrest counts* for race group i, offense j, and year k. Combining

equations A.1 through A.4, which were previously described, yields the full equation for estimating *clean UCR White and Black arrest counts*, expressed as follows:

$$Y_{ijk} = \left\langle \frac{U_{ijk}}{\left\{ U_{ijk} + \left[(H_{jk} \times R_{ik}) \times \left(\frac{P_{1ik}}{P_{2ik}} \right) \right] \right\}} \right\rangle \times A_{ijk} \qquad (A.5)$$

A potential caveat of this technique is that it assumes the racial composition of CA–NY Hispanic arrestees matches that of the U.S. Hispanic population. Because of data limitations, we cannot assess the degree to which this assumption is true. However, it is worth noting that alternative adjustments using total U.S. race distributions of the Hispanic population are nearly identical to those obtained using CA–NY race/ethnic distributions. To test the robustness of our results, we used alternative adjustments in preliminary analyses to divide Hispanic arrests into "White" and "Black" categories, several of which placed higher shares of Hispanics in the "Black" category to account for the possibility that "Black-Hispanics" might be more likely to be arrested than "White-Hispanics." Although these alternative methods produced some variations in the Hispanic effects on White and Black arrest rates, substantive findings from the alternative adjustments matched those reported. Our results, therefore, seem robust across alternative adjustments for the "Hispanic effect" in White and Black arrest figures.

EXAMPLE OF CORRECTION PROCEDURE

The following illustrates use of our adjustment procedure to remove the "Hispanic effect" from national UCR estimates to produce "clean" White and Black homicide arrest counts for 1990.

Assume the following based on 1990 homicide arrest and population figures:

1. *CA–NY arrest figures:*

    a. 1,185 clean White homicide arrests (U_{ijk})
    b. 2,997 clean Black homicide arrests (U_{ijk})
    c. 2,244 clean Hispanic homicide arrests (H_{jk})

2. *Racial composition of CA–NY Hispanic population:*

    a. 91.44 percent of the CA–NY Hispanic population was White (R_{ik})
    b. 5.36 percent of the CA–NY Hispanic population was Black (R_{ik})

3. *Hispanic composition of White and Black populations in CA–NY:*

   a. Hispanics accounted for 23.25 percent of the White CA–NY population (P_{2ik})
   b. Hispanics accounted for 12.49 percent of the Black CA–NY population (P_{2ik})

4. *Hispanic composition of White and Black populations in the United States:*

   a. Hispanics accounted for 9.86 percent of the White U.S. population (P_{1ik})
   b. Hispanics accounted for 3.95 percent of the Black U.S. population (P_{1ik})

5. *UCR (confounded) arrest figures:*

   a. 7,942 White homicide arrests (A_{ijk})
   b. 9,952 Black homicide arrests (A_{ijk})

Substituting these values into equation A.1 yields *confounded* White and Black CA–NY arrest counts that include Hispanic arrests as follows:

$$
\begin{aligned}
\textit{White confounded arrest count} &= 1{,}185 + (2{,}244 \times .9144) \\
&= 1{,}185 + 2{,}052 \\
&= 3{,}237
\end{aligned}
$$

$$
\begin{aligned}
\textit{Black confounded arrest count} &= 2{,}997 + (2{,}244 \times .0536) \\
&= 2{,}997 + 120 \\
&= 3{,}117
\end{aligned}
$$

However, these White and Black confounded arrest counts are calculated without taking into account that the "Hispanic effect" in CA–NY data are inflated because of the relatively greater presence of Hispanics in California and New York than in the United States as a whole. Thus, we use equation A.2 to account for this effect and downward adjust our estimates as follows:

$$
\begin{aligned}
\textit{White confounded arrest count} &= 1{,}185 + ((2{,}244 \times .9144) \times (.0986/.2325)) \\
&= 1{,}185 + (2{,}052 \times .4240) \\
&= 1{,}185 + 870 \\
&= 2{,}055
\end{aligned}
$$

$$
\begin{aligned}
\textit{Black confounded arrest count} &= 2{,}997 + ((2{,}244 \times .0536) \times (.0395/.1249)) \\
&= 2{,}997 + (120 \times .3163) \\
&= 2{,}997 + 38 \\
&= 3{,}035
\end{aligned}
$$

Next, we derive White and Black correction factors by comparing *clean* and *confounded* CA–NY arrests using equation A.3 as follows:

$$White\ correction\ factor = 1{,}185/2{,}055$$
$$= .5766$$

$$Black\ correction\ factor = 2{,}997/3{,}035$$
$$= .9875$$

Finally, we apply our White and Black correction factors to national UCR estimates using equation A.4 as follows to produce estimates of *clean* White and Black national homicide arrests that have removed Hispanic counts:

$$White\ UCR\ clean\ arrest\ count = .5766 \times 7{,}942$$
$$= 4{,}579$$

$$Black\ UCR\ clean\ arrest\ count = .9875 \times 9{,}952$$
$$= 9{,}828$$

ALTERNATIVE ADJUSTMENT METHOD

To exhaust the data and address the validity of our adjustment procedure, we also employed an alternative Hispanic adjustment procedure to correct the national UCR White and Black arrests.[20] First, we calculated CA–NY Hispanic arrest rates. Second, we estimated U.S. Hispanic arrests (for each offense and year) by multiplying the CA–NY Hispanic arrest rate by the U.S. Hispanic population. Third, we divided our estimate of U.S. Hispanic arrests (for each year and offense) into "White-Hispanic" and "Black-Hispanic" arrests based on the proportion of the U.S. Hispanic population that is White or Black. Fourth, we created *clean* White and Black U.S. arrest figures (that exclude Hispanics) by subtracting our estimates of White-Hispanic and Black-Hispanic arrests from the original White and Black UCR arrest counts. The formula for this method is expressed as follows:

$$Y_{ijk} = A_{ijk} - \left\{ \left[\left(\frac{H_{jk}}{HPI_k} \right) \times HP2_k \right] \times P_{ik} \right\}$$

where Y is the clean UCR arrest counts for race group i, offense j, and year k; A is the confounded (original) UCR arrest counts; H is the Hispanic

---

20.　We thank Miles Harer for suggesting this adjustment method.

arrests in CA–NY; *HP1* is the Hispanic population in CA–NY and *HP2* is the Hispanic population in the United States; and *P* is the proportion of the U.S. Hispanic population that is White or Black.

Results obtained using the alternative adjustment method are nearly identical to those described, indicating that our findings are robust across various demographic techniques that might be used to adjust for the "Hispanic effect" on national UCR crime estimates. Additionally, although this alternative method is more methodologically straightforward, our adjustment procedure has several important advantages over the alternative method. First, our procedure adjusts for the "Hispanic effect" in California and New York before applying this correction to the U.S. crime figures, which enables it to account for effects of trending better in CA–NY Hispanic arrest estimates (e.g., the fact that Hispanic arrest trends in CA–NY have decreased dramatically in recent decades and might not be matched by similar declines in Hispanic crime throughout the rest of the United States). Thus, our method provides a more conservative estimate of the "Hispanic effect" on national arrest rates over time. Second, our adjustment is less susceptible to problems of misestimation and undercount in Hispanic population counts in census data. Specifically, our method corrects for Hispanic arrest counts before calculating White and Black rates, whereas the alternative method relies on Hispanic population counts from the U.S. Census (which have several well-documented problems, see Bean et al., 2001; Passel, Hook, and Bean, 2004) to create population-adjusted arrest rates to apply the adjustment.

Another alternative is to adjust the national arrest figures using a Hispanic rate that is set at the midpoint of White and Black rates. This midpoint represents a "ballpark" estimation that often is noted by commentators/analysts about race–ethnic differences in violent crime (i.e., Hispanic violence or crime levels relative to Whites and Blacks). Two main shortcomings are associated with the midpoint estimate; first, it is a ballpark estimate, and second, it is not offense specific but instead is referenced to violent crime in general and thus overlooks variation in the relative Hispanic effect by the type of violence or the type of crime. In contrast, our estimates are based on actual Hispanic violence figures representing two large states and by offense type, with the latter documenting considerable variation in the Hispanic effect by the type of violent crime.