WEB APPENDIX – How Dark is Dark? Bright Lights, Big City, Racial Profiling

WC Horrace and SM Rohlin

A. Stop Data

Our police-citizen stop data come from the Syracuse Police Department. Prior to 2005 Syracuse police only recorded demographic data in the event of an arrest. If citizens were stopped and no arrest was made, no race information was recorded. In the end of 2004 the authors helped the Syracuse Police design a data collection instrument for stops that did not result in an arrest. The instrument is called "form 67," and it was put into practice in the first quarter of 2005. The form was recorded electronically for both types of stops: vehicular and pedestrian.\(^1\) The form was designed to isolate "discretionary stops," where police judgment was required as to whether or not a stop was warranted. Form 67 collects information on the date and time of the stop, address or location of the stop (recoded as longitude and latitude), demographic data on the citizen (name, age, race, etc.), the reason for the stop (e.g., obstructing traffic), the type of stop from a pull-down menu (crime investigation, drug investigation, gun investigation, local law, loitering, prostitution, suspicious person or vehicle, trespassing, unknown, or vehicle & traffic violation), whether or not the citizen wore gang colors or were known to be a gang member, whether or not a frisk (quick pat down for officer safety) or a search (looking for something illegal) was conducted, and whether or not the stop was at the officer's discretion.

The new forms were tested for a six-month period in the first half of 2005, and official data collection did not begin until July of 2005, so police had adequate time to learn how to properly use the form. In 2005 the police knew that the data were to be used in a racial profiling study over the sixth month period of July – December 2005, so to minimize any potential "Hawthorne effect" in the present study we purged any data collected in 2005.\(^2\) Therefore, the data start in January 2006 and end in December of 2009 (four years).

\(^1\) A small fraction of the stops were recorded on paper.
\(^2\) The Hawthorne effect occurs when subjects modify behavior in response to being studied. The 2005 study identified significant police bias towards blacks using the Knowles, Persico and Todd (2001) methodology. We did not test whether the Hawthorne effect was present in the purged data.
The Form 67 data were merged with Syracuse Police Discretionary Arrest Data for the same time period (2006-2009, inclusive). The Discretionary Arrest Data contain all the fields of the Citizen-Stop Data plus additional arrest information, including charges filed against the citizen. These data contain records for each charge filed in an arrest and the data were provided at the "charge level." Since there can be multiple charges in a single stop and arrest, the data were collapsed to the person-stop level. After merging and cleaning the data for miscoded or missing observations and filtering non-discretionary stops, we were left with 61,389 traffic stops over the period, of which 20,442 occurred during the evening intertwilight period. Roughly 1/3 of the traffic stops in Syracuse occur between 5:03 and 9:22 pm. The time distribution of the stops at 30 minute intervals is presented in the first row of Table 1. The relative frequency in each time interval is presented in the second row of Table 1. Of these 20,442 inter twilight stops, 11,419 were in daylight and 9,023 were in darkness. For purposes of the veil test of Grogger and Ridgeway (2006), daylight stops are tagged with $dark = 0$, and nighttime stops are tagged with $dark = 1$. During the inter twilight period the percentage of black drivers stopped changes from 48.90% in daylight to 51.11% in darkness.

The 20,442 inter twilight stops were tagged with census tract numbers, based on Syracuse census tracts defined by the US Census Bureau in 2000. The census data were used to calculate the population averages by race for each census tract in the City of Syracuse. The population percentages for the census tracts used in this study are 13.14% African-American and 82% Caucasian. A census tract is designated to be a "high black population area" if the black percentage is above 13.14%. There are a total of 125 census tracts in the stop data; of these tracts, 35 are black areas. During the inter twilight period there are 85 census tracts; of these tracts, 33 are black areas. Of the 20,442 inter twilight stops, 902 (4.41%) of them had stop locations that were unintelligible or were not specific enough to pinpoint an exact location. The remaining 19,540 inter twilight stops were "geo-coded" with the longitude and latitude coordinates.

---

3 A small number of stops occurred outside the city limits, and were not tagged with a census tract number. Only one of these stops appears in the 20,442 inter twilight stops.
4 Lack of an exact location precludes tagging with longitude and latitude coordinates, but it does not preclude tagging with a census tract number.
The Police Offense Crime Data contained counts of Part I offenses (murder, manslaughter, rape, robbery, aggravated assault, burglary, larceny, and motor vehicle theft) and drug offenses by census tract from 01/01/06 to 11/22/09. Part I crimes and drug crimes are considered some of the most violent crimes recorded, and there were 48,722 of them in Syracuse over the period. The data were merged with the US Census Bureau Data to determine crime rates in each census tract. The distribution of offenses per census tract over the period is as follows: the first quartile is 7 offenses, the median is 500 offenses, and the third quartile is 888 offenses. Census tracts with more than 500 offenses were deemed "high-crime" areas. Of the 85 census tracts in the intertwilight sample, 44 were high-crime areas.

B. Simulating Light Density

We use ArcGIS (a geographic information systems software package) to simulate the amount of light at each police-citizen encounter (and at any point in the city). ArcGIS does not simulate light per se; it simply allows users to simulate spatial distributions of any phenomena (e.g., crime, pollution, economic shocks) that propagate from a finite number of sources over a two-dimensional lattice, called a "raster map." First, we use an up-to-date aerial photograph of the city of Syracuse to identify streetlight locations (longitude and latitude pairs) throughout the city. Figure 1 displays the map of light locations. Based on conversations with an urban geographer, we simulate a measure of light density around each streetlamp using the "inverse distance weighted" (IDW) interpolation feature in ArcGIS. The feature uses a quadratic kernel weighting (Silverman, 1986, equation 4.5) and bandwidth selection based on a spatial variant of Silverman's Rule of Thumb (Silverman, 1986). In addition to the default values for IDW interpolation our urban geographer recommended a raster cell width of 30 feet and a "search radius" of 100 feet. The raster cell width is a measure of resolution of the resulting "light map" and the "search

---

6 The default unit of power in ArcGIS for each street lamp is 2, so we have assumed that each lamp produces the same amount of light. This is an assumption of convenience because measuring the light produced at each streetlight in the city was impractical. The actual values of the power or luminosity is unimportant in what follows.
radius" is the cutoff distance from the lamp where light goes to 0. Figure 2 provides a typical visual representation of the output and illustrates how the measure of light density intensifies as the number of streetlights in an area increases. The black dots represent streetlight locations, and blue represents the intensity of light with darker blue representing greater intensity of light. The lightest blue areas represent areas of no light from streetlights. At this resolution it is difficult to see the dissipation of light over the 100-foot search radius, but it is there. We also notice that at this resolution the light from the lamps is only approximately circular, and in some case produces an approximately 100ft by 100ft square of light.

Based on these lighting maps, each of the original 61,389 traffic stop locations (longitude and latitude pairs) is tagged with the newly created light intensity data. This is done whether or not the stop occurred at nighttime or daytime. In this sense we are treating our simulated light intensities as nighttime lighting, for in daylight streetlight intensities are irrelevant, either because the lights are turned off or because sunlight dominates street lighting during the day. Although we only have the location where the officer stopped the driver and not the location of where the officer first viewed the driver, we assume that the light densities of the two are highly correlated. We believe this to be a relatively weak assumption since the concentration of streetlights along a street is fairly uniform making the light density a block or two away similar. The mean value of the 61,389 light intensities is 0.00722 with a standard deviation of 0.0052. The median is 0.00721, and the interquartile range is [0.0021, 0.0106]. The absolute magnitudes of the light intensities are less important than the relative magnitude because the analysis is based on quantiles (relative light intensities) from the distribution. Using this citywide median light intensity is (0.00721) as a cutoff, we drop observations in the upper half of the distribution. Of the remaining (lower light intensity) traffic stops, 9,300 occur during the evening intertwilight period. Of these 5,283 occurred in daylight and 4,017 occurred at nighttime. This constitutes our refined sample with the daylight stops

---

7 We verified the 100ft cutoff with the urban geographer in the field using a light meter. We experimented with both the raster cell size of 30ft and the cutoff value of 100ft and found our results qualitatively robust to these selections.  
8 In some cases a precise location could not be determined, so some stops did not receive light density measures.  
9 In doing so we are dropping both nighttime and daylight stops. Dropping daytime stops ensures that stops receive equal weight in both daylight and darkness.
tagged as $darker = 0$ and the nighttime stops tagged as $darker = 1$. The time distribution of these stops at 30 minute intervals is presented in the third row of Table 1. The relative frequency in each time interval is presented in the fourth row of Table 1. The relative frequencies of our refined sample of 9,300 intertwilght stops is quite similar to the larger sample of 20,442 in the first and second rows of the table, and this may be evidence that our refinement is not inducing any bias in our results.

Table 2 presents the geographical distribution of stops over census tracts for the standard veil sample of 20,442 stops (first row) and for the refined sample of 9,300 stops (second row). For the standard sample veil sample the 20th percentile tract has 8 stops, the median has 120 stops, and the 80th percentile has 410 stops. For our refined sample the 20th percentile tract has 12 stops, the median has 74 stops, and the 80th percentile has 211 stops. Above the median number of stops, the number of stops at each percentile for the refined sample is about half that of the standard sample.

References


### Tables and Figures

**Table 1. Counts of Traffic Stops over the Intertwilight Period – Standard Veil & Refined Veil Samples**

<table>
<thead>
<tr>
<th>Sample/Time</th>
<th>5 - 5:30</th>
<th>5:30 - 6</th>
<th>6 - 6:30</th>
<th>6:30 - 7</th>
<th>7 - 7:30</th>
<th>7:30 - 8</th>
<th>8 - 8:30</th>
<th>8:30 - 9</th>
<th>9 - 9:30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>1,865</td>
<td>2,493</td>
<td>2,369</td>
<td>2,400</td>
<td>2,609</td>
<td>2,470</td>
<td>2,280</td>
<td>2,216</td>
<td>1,740</td>
</tr>
<tr>
<td>Rel. Freq.</td>
<td>9.12%</td>
<td>12.20%</td>
<td>11.59%</td>
<td>11.74%</td>
<td>12.76%</td>
<td>12.08%</td>
<td>11.15%</td>
<td>10.84%</td>
<td>8.51%</td>
</tr>
<tr>
<td>Refined</td>
<td>921</td>
<td>1,206</td>
<td>1,068</td>
<td>1,088</td>
<td>1,140</td>
<td>1,093</td>
<td>1,044</td>
<td>947</td>
<td>793</td>
</tr>
<tr>
<td>Rel. Freq.</td>
<td>9.90%</td>
<td>12.97%</td>
<td>11.48%</td>
<td>11.70%</td>
<td>12.26%</td>
<td>11.75%</td>
<td>11.23%</td>
<td>10.18%</td>
<td>8.53%</td>
</tr>
</tbody>
</table>

Relative frequencies (Rel. Freq.) of the standard veil and refined veil samples over time are similar.

**Table 2. Percentiles of Traffic Stops over Census Tracts – Standard Veil & Refined Veil Samples**

<table>
<thead>
<tr>
<th>Sample/percentile</th>
<th>Min</th>
<th>10th</th>
<th>20th</th>
<th>30th</th>
<th>40th</th>
<th>Median</th>
<th>60th</th>
<th>70th</th>
<th>80th</th>
<th>90th</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>22</td>
<td>63</td>
<td>120</td>
<td>212</td>
<td>322</td>
<td>410</td>
<td>662</td>
<td>1,126</td>
</tr>
<tr>
<td>Refined</td>
<td>1</td>
<td>3</td>
<td>12</td>
<td>25</td>
<td>47</td>
<td>74</td>
<td>114</td>
<td>159</td>
<td>211</td>
<td>335</td>
<td>592</td>
</tr>
</tbody>
</table>

Standard sample = 20,442 stops; refined sample = 9,300 stops; 85 census tracts.
Figure 1: Map of Streetlight Locations in the City of Syracuse, NY.
Figure 2: Visual example of the light density from streetlights.

Dots represent streetlights with darker blue areas containing more light.