HOW DARK IS DARK? BRIGHT LIGHTS, BIG CITY, RACIAL PROFILING

William C. Horrace and Shawn M. Rohlin*

Abstract—Grogger and Ridgeway (2006) use the daylight saving time shift to develop a police racial profiling test that is based on differences in driver race visibility and (hence) the race distribution of traffic stops across daylight and darkness. However, urban environments may be well lit at night, eroding the power of their test. We refine their test using streetlight location data in Syracuse, New York, and the results change in the direction of finding profiling of black drivers. Our preferred specification suggests that the odds of a black driver being stopped (relative to nonblack drivers) increase 15% in daylight compared to darkness.

I. Introduction

GIVEN the large racial disparities in the U.S. criminal justice system, police racial profiling is an important policy issue. Outside of actual field experiments or qualitative assessments, there are essentially two accepted ways to quantitatively test for the existence of police racial profiling from data on traffic stops: outcome-based approaches (Knowles, Persico, & Todd, 2001) and benchmarking analysis (Steward, 2004; Weiss & Grumet-Morris, 2005).[1] Outcome-based approaches assume that police adjust vehicle search frequencies to maximize arrests (or some other outcome measure). To apply the approach to police stop data, one simply calculates arrest rates by race conditional on a vehicle search and then compares the conditional arrest rates to determine differentials across race. The idea is that in the absence of racial bias, arrest rates will be the same across races. However, if police are biased against a particular race, then that race will be searched more frequently without arrest and have a lower arrest rate. The approach is based on a taste-based discrimination model (Becker, 1957) where police search vehicles to maximize arrests and drivers choose to carry or not carry contraband in response to police search intensities. In equilibrium, police oversearch a particular race if the marginal cost of searching that race is lower than that of other races (i.e., if police have a taste for discrimination against that race). This is in contrast to statistical discrimination (Phelps, 1972; Arrow, 1973), which is differential police behavior by race that is justified based on the statistical likelihood that there are differences in criminal behavior by race. Statistical discrimination may exist in the context of arresting maximizing behavior and in the presence or absence of taste-based discrimination.

Benchmarking analysis compares police stop rates by race to some population benchmark across races. If the stop frequency for any race exceeds the benchmark, that is evidence of profiling of that race. Ideally, the benchmark would be race percentages of the at-risk population (i.e., violating the law and observed by police), but this is almost never observable. For example, in the earliest applications of benchmarking analysis, stop rates for African American drivers were compared to the resident population percentage of blacks living in urban areas, and invariably stop rates for black drivers were much higher than those of the resident population. This was used as evidence of police racial profiling. Criticisms abound (Dominitz, 2003), but most obvious is that in large urban areas, not all African Americans drive, so the resident population percentage of blacks is a poor proxy for the at-risk population of black drivers.[2] All attempts at improving benchmark analysis focus on refining measures of the at-risk population of drivers (Zingraff et al., 2000; Alpert, Smith, & Dunham, 2003; McConnell & Scheidegger, 2001; Lamberth, 1994), and this paper contributes to the literature in the same way. While the benchmarking literature is fairly agnostic on the reasons for racial profiling, in the absence of racial prejudice, the practice could be attributable to optimal police behavior in a model of statistical discrimination.[3] Hence, we use the terms *racial profiling* and *statistical discrimination* synonymously.

An arguably excellent refinement to the benchmark approach is the study of Grogger and Ridgeway (2006), which develops a test for racial profiling "behind a veil of darkness." The method exploits the exogenous variability in the visibility of driver race between daylight and darkness. The idea is that driver visibility is limited during darkness, making it difficult for police to use race as a criterion in traffic stops. Therefore, differences in the race distribution of stops between darkness and daylight may be evidence of profiling.[4] The testing methodology differences out the at-risk population of drivers by assuming that the racial mix of the at-risk population does not change between daylight and darkness; they call this a constant relative risk assumption. Of course, this assumption may not hold in general, as work schedules and, hence, traffic patterns vary by race (Hamermesh, 1996). To ensure the assumption is not violated, they focus their empirical analysis on stops made during the evening intertwilight period

[1] This paper is not intended to survey the literature, so there are many good papers that we will not cite.

[2] An advantage of outcome-based approaches is that the at-risk population is irrelevant. However, outcome-based approaches are not without their shortcomings. Criticisms are in Anwar and Fang (2006), Antonovics and Knight (2009), Sanga (2009, forthcoming), and Ayres (2001).

[3] Even if the driving behaviors of blacks and whites are identical, police may still be more inclined to stop a particular race ceteris paribus if that race is more likely to be in violation of some other law (e.g., an expired license or registration) in addition to the violation that justified the initial stop (e.g., a broken taillight).

[4] The veil test has a local average treatment effect (LATE) interpretation. See Imbens and Angrist (1994).

(approximately 5:00 p.m. to 9:30 p.m.), when it may be either dark or light depending on the time of year due to the daylight saving time (DST) change and the tilt of the earth.[5] They also embed the testing exercise in a logit regression, allowing the test to condition on clock time and other covariates that help the assumption hold. They apply the test to data from Oakland, California, and conclude that there is no racial profiling of black drivers and that there may be (in fact) profiling of nonblack drivers in Oakland.

This paper uses the same approach to analyze police stops in Syracuse, New York, between 2006 and 2009 but with two data-driven improvements. First, our data include only discretionary stops. That is, dispatched and warrant stops are excluded, so we examine only behaviors when police use their own judgment to justify a traffic stop. Second, we improve the power of Grogger and Ridgeway's (2006) veil test by simulating ambient light from streetlights throughout the city and thereby refine their definition of darkness in the test.[6] The ideal test is based not on daylight and darkness, but on the visibility (or lack thereof) of the driver. Daylight and darkness serve only as proxies for visibility, and the power of the test is positively related to the correlation between daylight and visibility (or darkness and lack of visibility).[7] Using geographic information system (GIS) software, we determine streetlight and traffic stop locations and restrict the test to (condition the test on) stop locations with little or no ambient street lighting at night.[8] That is, the proposed test is restricted to stops where nighttime visibility is lowest regardless of whether the stop occurred in light or dark. Therefore, the samples and interpretations (conditioning arguments) of the two tests are different. We find the veil test to be sensitive to nighttime ambient light in Syracuse, and our refinement changes the results in the direction of finding racial profiling of black drivers.

Few papers have exploited ambient light for identification of economic relationships. Recently Doleac and Sanders (2013) use changes in DST to make inferences on crime. Like Grogger and Ridgeway (2006), they discuss the possible effects of street lighting on their results, but do not measure or exploit it in their analysis of crime. Smith (2014) explores the effect of DST on fatal accidents in daylight and darkness but ignores the effects of streetlights. Henderson, Storeygard, and Weil (2012) develop a statisti-

cal framework to use satellite data on nighttime lighting to analyze income growth measures in a unique way.[9] Our analyses use lighting features of these papers as a proxy for visibility to identify results on police behaviors. We have found no other analyses that simultaneously exploit both these lighting features (DST and streetlights) to identify economic relationships.

The next section reviews the Grogger and Ridgeway (2006) methodology and discusses our refinements. Section III discusses the data, which are in many ways ideally suited for testing for racial profiling behind a veil of darkness. Section IV provides the results, which show that streetlights matter in the veil test. Section V summarizes and concludes.

II.   Veil of Darkness Methodology

Grogger and Ridgeway's ideal test of racial profiling is based on the event that the race of an at risk driver is either visible (V) or not visible ($\bar{V}$). To be at risk, a driver must be violating the law and observed by police. Because race visibility per se is not observed by the econometrician, the ideal test is infeasible. Let S be a binary random variable indicating the event of a police traffic stop, and let B be a binary random variable indicating that a driver is black and at risk of being stopped, so that $\bar{B}$ indicates that a driver is not black and at risk. Hence, the random variables B and $\bar{B}$ represent the race distribution of the at-risk population of drivers. Let t be continuous clock time: then an ideal test is based on $K_{ideal}(t)$ in,

$$\frac{\mathrm{P}(S|t,V,B)}{\mathrm{P}(S|t,V,\bar{B})} = K_{ideal}(t) \times \frac{\mathrm{P}(S|t,\bar{V},B)}{\mathrm{P}(S|t,\bar{V},\bar{B})}. \quad (1)$$

The ratios on the left- and right-hand sides of equation (1) are the relative risks of a black driver being stopped when race is visible and when it is not visible, respectively. In the absence of profiling $K_{ideal}(t) = 1$ for each t, so the risk of a black driver being stopped is independent of visibility. Hence, $K_{ideal}(t) \neq 1$ captures the extent to which visibility affects the risk of a black driver being stopped and may suggest profiling.

While race visibility is a function of many factors, a feasible version of the veil test incorporates darkness as a proxy for unobserved race invisibility. Let $dark$ be a binary random variable for darkness. Substituting $dark = 0$ and $dark = 0$ for V and $\bar{V}$ (respectively) and applying Bayes' rule in equation (1):

$$K'(t) = \frac{\mathrm{P}(B|t,S,dark=0)\mathrm{P}(\bar{B}|t,S,dark=1)}{\mathrm{P}(\bar{B}|t,S,dark=0)\mathrm{P}(B|t,S,dark=1)}$$
$$\times \frac{\mathrm{P}(B|t,dark=1)\mathrm{P}(\bar{B}|t,dark=0)}{\mathrm{P}(\bar{B}|t,dark=1)\mathrm{P}(B|t,dark=0)}. \quad (2)$$

---

[5] Stops are rare during the morning intertwilight period, so Grogger and Ridgeway ignore them, as do we.

[6] By "simulating ambient light," we mean that we do not have actual measures of light at each location in the city, but we do have the physical location of each streetlight. Assuming an average light intensity for each light and a radial rate of decay, we simulate the relative light intensity at any location. Details are in the online appendix. We also considered the effects of rain on visibility, but there were too few rainy stops for credible inference.

[7] Grogger and Ridgeway (2006) admit that streetlights enhance race visibility and erode the power of their test.

[8] In doing so, we make the implicit assumption that the location of the observed violation and the stop location are approximately the same. We discuss this assumption in the online appendix.

[9] We considered using satellite imagery to develop a measure of ambient light. However, the image resolution was too coarse to produce a reliable light measurement for an individual streetlight. Satellite imagery is useful only for analyzing larger swaths of land.

The first ratio in equation (2) relates race to darkness for the distribution of stopped drivers. The second ratio relates race to darkness for the distribution of at-risk drivers. If the race distribution of at-risk drivers is independent of darkness, this ratio equals 1. This is Grogger and Ridgeway's assumption of constant relative risk, and if it holds, it implies their veil test:

$$K_{GR}(t) = \frac{P(B|t, S, dark = 0)P(\bar{B}|t, S, dark = 1)}{P(\bar{B}|t, S, dark = 0)P(B|t, S, dark = 1)}. \quad (3)$$

If darkness has a race blinding effect,

$$P(V|dark = 0) > P(V|dark = 1), \quad (4)$$

and if there is discrimination against black drivers (i.e., $K_{ideal}(t) > 1$), then Grogger and Ridgeway show that $K_{ideal}(t) \geq K_{GR}(t) > 1$. In fact, they prove that $K_{GR}(t)$ is strictly decreasing in $P(V|dark = 1)$, so if this probability equals 0, the veil test will be ideal (have greatest power).

Grogger and Ridgeway limit their test to the evening intertwilight range (about 5:00 p.m. to 9:30 p.m.), ensuring that the time period under study is both dark and light over the course of the year.[10] Daylight saving laws play an important role in ensuring that the intertwilight range is large enough to include sufficient data to reliably perform the test. To operationalize the test, take logarithms of equation (3):

$$\ln K_{GR}(t) = \ln \frac{P(B|S, t, dark = 0)}{1 - P(B|S, t, dark = 0)}$$
$$- \ln \frac{P(B|S, t, dark = 1)}{1 - P(B|S, t, dark = 1)}. \quad (5)$$

Then the veil test may be performed with a logit regression of B on the darkness binary variable and an additive spline function of time, which may be interacted with the darkness variable—that is,

$$\frac{P(B = 1|S, t, dark)}{1 - P(B = 1|S, t, dark)}$$
$$= \exp\{\beta_1 dark + \gamma_1' s(t) + \gamma_2' dark \times s(t)\}, \quad (6)$$

where β_1 is a scalar coefficient on the darkness binary variable, $s(t)$ is an additive spline vector over continuous time, and γ_1 and γ_2 are vectors of coefficients on $s(t)$ and its interaction with the darkness variable, respectively. Plugging equation (6) into equation (5) provides the veil test statistic:

$$\ln K_{GR}(t) = -\beta_1 - \gamma_2' s(t). \quad (7)$$

Different versions of the test can be performed for different zero restrictions on γ_1 and γ_2 in equations (6) and (7). Versions of the test that condition on other covariates are straight-

forward extensions of equation (6). In our empirical analyses, we follow Grogger and Ridgeway and include covariates related to different geographical areas where stops occur. This helps control for differential exposure to police across neighborhoods. It also helps ensure that the constant relative risk assumption holds. We sometimes include binary variables for stops in census tracts with relatively large populations of black residents and for census tracts with relatively high crime incidences; we may also include a set of binary variables for census tract number.[11]

Like any other test of racial profiling, the veil methodology has its limitations. First, there are always variables that are observable to police but unobservable to econometricians, and this is particularly troublesome when unobservables are correlated with race. However, this could be a shortcoming of any benchmarking study regardless of the methodology employed. Second, while one may take steps to ensure that the constant relative risk assumption holds, there are ways in which it may still be violated. For example, there may be seasonality in the race distribution of the at-risk population. Simply put, winter driving and summer driving may be different, so it may induce changes in the race distribution of drivers and could be interpreted as racial profiling.[12] Third, the power of the test is "reduced by anything that reduces the correlation between visibility and darkness."[13] For example, vehicle characteristics that are correlated with race will reduce the power of the veil test. We note that the veil test allows for nonreporting of stops to occur as long as race-specific reporting rates do not vary between dark and light.[14]

A. The Refined Veil Test

Our refinements to the veil test are designed to ensure that $P(V|dark = 1)$ is smaller, so that equation (3) has better power and is closer to the ideal test (where V is observed). That is, we refine the darkness binary variable by collecting streetlight location data and simulating nighttime light intensity at every stop location in the city. (We discuss the light simulation methodology in the online appendix.) Given the distribution of nighttime light intensities of each stop (regardless of when the stop occurred), we limit the sample to stop locations with lower levels of nighttime ambient light, so that $P(V|dark = 1)$ is smaller and the theoretical correlation between V and our darkness variable is stronger. In particular, let *darker* be our more refined bin-

---

[10] In Syracuse, the latest occurrence of the end of civil twilight (dark) falls in July at 9:22 p.m., and the earliest occurrence falls in December at 5:03 p.m.

[11] We also performed the test conditioning on other covariates like the race, gender, and experience of the officer; age and gender of the driver; and outcome of the traffic stop (e.g., ticket or arrest). Our conclusions are robust to the inclusion of these variables, so we report only results that control for differential police exposure across areas in the city: black area, high-crime area, or census tract number.

[12] Our conclusions are robust to inclusion of a binary variable for a stop that occurred during winter.

[13] Grogger and Ridgeway (2006, p. 886).

[14] Grogger and Ridgeway provide an excellent discussion on nonreporting, which occurs when a stop is made but is not recorded and does not appear in the data. That is, the police intentionally withhold the data.

TABLE 1.—COUNTS OF TRAFFIC STOPS OVER THE INTERTWILIGHT PERIOD: STANDARD VEIL AND REFINED VEIL SAMPLES

Sample/Time	5:00–5:30	5:30–6:00	6:00–6:30	6:30–7:00	7:00–7:30	7:30–8:00	8:00–8:30	8:30–9:00	9:00–9:30
Standard veil sample	1,865	2,493	2,369	2,400	2,609	2,470	2,280	2,216	1,740
Relative frequency	9.12%	12.20%	11.59%	11.74%	12.76%	12.08%	11.15%	10.84%	8.51%
Refined veil sample	921	1,206	1,068	1,088	1,140	1,093	1,044	947	793
Relative frequency	9.90%	12.97%	11.48%	11.70%	12.26%	11.75%	11.23%	10.18%	8.53%

Relative frequencies of the standard veil and refined veil samples over time are similar.

ary variable for darkness (i.e., conditional on lower ambient street lighting).[15] Therefore, the event $darker = 0 \Leftrightarrow dark = 0$, so that $\Pr(V|t, darker = 0) = \Pr(V|t, dark = 0)$. However, the event $darker = 1 \Rightarrow dark = 1$, so our refined test assumption is:

$$\Pr(V|darker = 0) = \Pr(V|dark = 0) > \Pr(V|dark = 1)$$
$$> \Pr(V|darker = 1), \qquad (8)$$

leading to our refined veil test based on the odds ratio,

$$K_*(t) = \frac{\mathrm{P}(B|S, t, darker = 0)\mathrm{P}(\bar{B}|S, t, darker = 1)}{\mathrm{P}(\bar{B}|S, t, darker = 0)\mathrm{P}(B|S, t, darker = 1)}. \qquad (9)$$

Since the veil test is strictly decreasing in $\mathrm{P}(V|dark = 1)$, it must be true that $K_*(t) > K_{GR}(t)$ in the population. However, this may not hold in practice because the sample of the refined test is necessarily different (smaller) than that of the standard veil test. Nonetheless, if the data admit a sufficiently large number of stops with lower levels of nighttime ambient light, then equation (9) should produce a more powerful test. That is, the probability of rejecting the null hypothesis given that it is false should be greater for the statistic in equation (9) than for the statistic in equation (3).

Similar to the Grogger and Ridgeway veil test, our refined veil test may be embedded in a logit regression with the *darker* variable, a time spline, and their interaction. The refined veil test will suffer from the same problems as the standard veil test, but it is unlikely to suffer from any additional problems. In particular, nonrandomness in the placement of streetlights will not affect its validity in relation to the standard test; it will affect only its interpretation. Even if nonrandom placement did affect the validity of the refined test, this would have to be balanced against its improved power. Even then, nonrandom placement may not be cause for concern. First, empirical evidence on the correlation between streetlight placement and criminal activity is mixed. For example, Welsh and Farrington (2007, table 2) summarize the results of eight U.S. studies conducted from 1974 to 1998 in urban environments. Half the studies conclude that street lighting had no significant effects on local criminal activity. Second, we limit our analysis to traffic stops, so the majority of the reasons for being "at risk" are related to driver behavior (e.g., speeding, improper signaling), which is

likely to be independent of neighborhood amenities or criminality. Finally, the distributional cutoff of nighttime light intensities for our refined measure of darkness is based on the global distribution of the light intensity at all times and in all stop locations in the city, while our refined analysis is based on the local light intensity of a subset of these stops (about half of them). This may induce some exogenous variability in our local measure of darkness.[16]

III. **Data**

For a complete description of the data, see the online appendix. Our data are a merging of 2006–2009 Syracuse police stop data with 2000 U.S. census tract data. The final data consist of 20,442 traffic stops that occur during the evening intertwilight period.[17] This is about one-third of the 61,389 traffic stops in Syracuse over the period. 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 intertwilight stops, 11,419 were in daylight and 9,023 were in darkness. For purposes of the veil test of Grogger and Ridgeway in equation (3), daylight stops are tagged as $dark = 0$ and nighttime stops as $dark = 1$. During the intertwilight period, the percentage of black drivers stopped changes from 48.90% in daylight to 51.11% in darkness.

For our refined test in equation (9) we simulate the nighttime light intensity distribution for all stops in the city (regardless of whether they occurred in daylight, darkness, or the intertwilight period), and drop observations above the median light intensity. (See the online appendix for details of the light intensity simulation.) Of the remaining (lower light intensity) traffic stops, 9,300 occur during the evening intertwilight period. Of these, 5,283 occur in daylight and 4,017 at nighttime. This constitutes our refined sample with daylight stops tagged as $darker = 0$ and nighttime stops tagged as $darker = 1$. The time distribution of these stops at 30 minute intervals is presented in the third

---

[15] We simulate the distribution of nighttime ambient light for all traffic stops and drop those with light intensities above the median. The remaining stops are assigned values of $darker = 1$ if they occur at nighttime or $darker = 0$ if they occur during daylight. This ensures equal weighting of daytime and nighttime stops in the test.

[16] The argument is similar to that used in the spatial sorting model of Bayer and Timmins (2007) in which activity outside the location of interest is a valid instrument for activity in the location of interest. While we have not followed their exact identification strategy, it is in this spirit that the cutoff may induce some exogenous variability in the refined darkness measure.

[17] Although we have pedestrian stops in our data, we use traffic stops in the spirit of the veil test. In fact, we redid the entire analysis for pedestrian stops, and none of our tests could reject the hypothesis of equal distributions of race in daylight and in dark for either measure of darkness. This may be interpretable as a placebo test, confirming the efficacy of the veil methodology for these data.

TABLE 2.—STANDARD VEIL ($\ln K_{GR}$) AND REFINED VEIL ($\ln K_*$) TESTS

Covariates in Addition to the Darkness Indicator	$\ln K_{GR}$ (*dark*)	$\ln K_*$ (*darker*)
No Additional Covariates	−0.0885***	0.0510
	(0.0282)	(0.0419)
	[20,442]	[9,300]
Time Spline	0.0019	0.1348**
	(0.0314)	(0.0470)
	[20,442]	[9,300]
Time Spline &High Black Population Indicator	0.0099	0.1187**
	(0.0322)	(0.0481)
	[20,317]	[9,240]
Time Spline & High Crime Indicator	0.0083	0.1417***
	(0.0316)	(0.0474)
	[20,303]	[9,245]
Time Spline & Census Tract Indicators	−0.0677**	0.0977*
	(0.0345)	(0.0521)
	[20,430]	[9,295]

Each cell contains the estimate of $\ln K_{GR}$ or $\ln K_*$ (standard error) and [sample size]. Time Spline is a third-order spline in clock time with eight internal knots. High Black Population is a census tract above the city average of 13.14% black. High Crime is a census tract above the city median of 500 crimes over the period. Significant at *90%, **95%, ***99%.

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 intertwilight stops are quite similar to those of 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 the results that follow.

IV. Results

The analysis focuses on black/nonblack stop differences, so Caucasian, Hispanic and other demographic groups are aggregated into "nonblack" stops. Sample sizes were too small for Hispanic and other demographic groups to justify separate analyses. The objective is to compare the results of the veil test in equation (3) and our refined veil test in equation (9). We find that the refinement mostly changes the test results in the direction of finding racial profiling of blacks. We begin with the results in table 2, which contains two columns of test statistics: one for the standard veil test with heading "$\ln K_{GR}$ (dark)" and a second for our refined veil test with heading "$\ln K_*$ (darker)." The statistics are based on various logit regressions of the black binary variable (B) oneither *dark* (resulting in $\ln K_{GR}$) or *darker* (resulting in $\ln K_*$) and additional covariates, as indicated in the table. The additional covariates may include an eight internal knot, third-order spline of continuous time, a binary variable for a stop that occurs in a census tract with a relatively high black population, a binary variable for a stop that occurs in a relatively high-crime census tract, or a set of binary variables for a stop that occurs in a particular census tract in the city.[18]

The first row in table 2 contains the results of the veil test with No Additional Covariates (i.e., logit on the *dark* or *darker* variable alone). The veil test statistic ($\ln K_{GR}$) based on *dark* is a significant −0.0885, indicating discrimination against nonblack drivers. However, our refined test statistic ($\ln K_*$) based on *darker* is an insignificant 0.0510, indicating unbiased policing in Syracuse. Therefore, our refined indicator of nighttime stops changes the conclusion of the test. In fact, the difference $\ln K_* - \ln K_{GR}$ is significant at the 99% level, so the refinement has a significant impact on the test.[19] The remaining entries in the first row of table 2 are the standard errors (in parentheses) and the sample sizes (in brackets) for each test.[20] The sample sizes for the refined test are about half that of the standard test (compare 9,300 to 20,442), because our refined measure of darkness is based on only stops below the median level of light intensity for nighttime stops.

The test results in table 2 are more compelling when we condition on continuous clock time. Our eight internal knot, third-order spline in the logit regression in equation (6) is

$$\gamma_2' s(t) = \sum_{k=1}^{8} \gamma_{2,k}(t - \tau_k)_+^3 + \gamma_{2,9}t + \gamma_{2,10}t^2 + \gamma_{2,11}t^3, \quad (10)$$

with knots $\tau_k = 30 \times k$ (corresponding to 30 minute internals over the $t = 1, \ldots, 259$ minute intertwilight period) and where $(x)_+ = x$ for $x > 0$ and $(x)_+ = 0$ otherwise. For the Time Spline regressions in the second row of table 2, the usual veil test statistic ($\ln K_{GR}$) is an insignificant 0.0019, and our refined veil statistic ($\ln K_*$) is a significant 0.1348. That is, conditional on clock time, the usual veil test of Grogger and Ridgeway suggests no profiling, while the refined test suggests profiling of blacks. This makes clear the importance of controlling for clock time in the veil test. Results are similar when (in addition to the time spline) we include binary variables for High Black Population ($\ln K_{GR}$ insignificant 0.0099, and $\ln K_*$ significant 0.1187) and High Crime ($\ln K_{GR}$ insignificant 0.0083, and $\ln K_*$ significant 0.1417). The logit with the "high-crime" binary variable is our preferred specification, as it includes the spline, produces the starkest contrast between the competing tests, and maintains 95% significance of our refined test statistic. Our refined test suggests that the odds of a black driver being stopped (relative to a nonblack driver) increase $e^{0.1417} = 1.152$ times in daylight as compared to darkness. When we add binary variables for Census Tract, the usual veil test is a significant −0.0677 and the refined test is a significant 0.0977. Regardless of the specification in table 2, our streetlight refinement changes the test results in the direction of finding racial profiling of black drivers. While we do not report them, our results are robust to the addition of other covariates like race, gender, and experience of the

---

[18] All analyses are done in STATA. The choice of eight internal knots is to place knots at half-hour increments between 5:03 p.m. and 9:22 p.m. Table 2's results do not condition on the interaction between the darkness variables and the spline, so the test statistics do not vary with time. Time-varying tests are performed in the sequel.

[19] In table 2 the difference between the two tests is always significant at the 90% level or higher.

[20] Standard errors are calculated using the LINCOM command in Stata.

Figure 1.—Standard Veil Test Statistic: Eight Internal Knot, Third-Order Spline
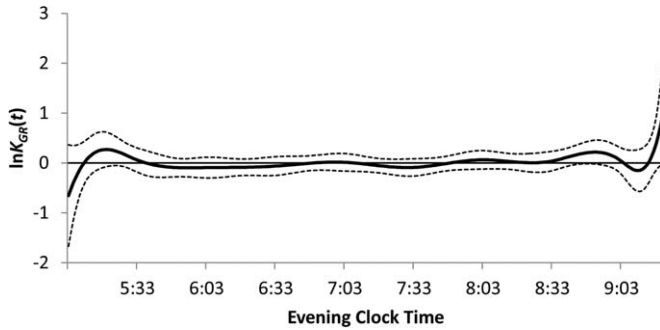
Heavy curve: $\ln K_{GR}$ with spline and no additional covariates; dashed curves: 95% confidence interval. $\ln K_{GR} > 0$ indicates profiling of black drivers. Sample size: 20,442.

Figure 2.—Refined Veil Test Statistic: Eight Internal Knot, Third-Order Spline
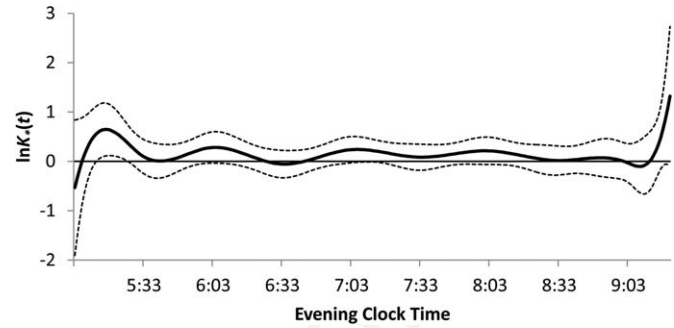
Heavy curve: $\ln K_*$ with spline and no additional covariates. dashed curves: 95% confidence interval. $\ln K_* > 0$ indicates profiling of black drivers. Sample size: 9,300.

officer; age and gender of the driver; and outcome of the stop (e.g., ticket or arrest).

Figure 1 illustrates the results for the veil test in equation (7) with no zero restrictions, so it is based on logit conditional on the *dark* variable, the spline, and their interaction. Figure 2 illustrates the same results for the refined veil test using the *darker* variable. In both figures, the horizontal axis is evening clock time (t) over the intertwilight period. The heavier curves in figures 1 and 2 are $\ln K_{GR}(t)$ and $\ln K_*(t)$, respectively; the dashed curves are 95% confidence intervals. In both figures, the test statistics are largely insignificant, but this is not surprising, as in both cases, the likelihood ratio test fails to reject the hypothesis $\gamma_2 = 0$, which provides support for privileging the more restrictive specifications in table 2.

V.   Conclusion

For a sample of Syracuse Police stop data, we demonstrate that the veil of darkness test of Grogger and Ridgeway is sensitive to the correlation (or lack thereof) of darkness and visibility. Based on their arguments, a more refined test may be required in urban environments where streetlights may erode this correlation and the power of the veil of darkness test. This is borne out in our analysis of Syracuse stops between 2006 and 2009, where ignoring streetlights often leads to conclusions of no racial profiling, while our refined test leads to conclusions of racial profiling of blacks. These different conclusions are not necessarily contradictory, as their interpretations (conditioning arguments) are different. In fact, the different conclusions suggest that controlling for heterogeneity of nighttime lighting may be important. How this difference arises is hard to say; it could be due to differential police behaviors in poorly lit areas, but it could be due to differential driving behaviors or some other unobservable features of poorly lit areas, despite our attempts to control for these differentials in the analysis. Nonetheless, this paper demonstrates that accounting for heterogeneity in nighttime ambient lighting may be important for the veil test. Like Grogger and Ridgeway, we

also show empirically that controlling for clock time is important. Our preferred specification suggests that the odds of a black driver being stopped (relative to nonblack drivers) increase 15% in daylight as compared to darkness.

For future work, it may be useful to perform an outcome-based test on these data and compare the results to those of the veil test. Also, our refined measure of darkness is continuous, but we treat it as binary. Developing a test based on a continuous definition of darkness may also prove useful.

REFERENCES

Alpert, Geoffrey P., Michael R. Smith, and Roger G. Dunham, "Toward a Better Benchmark: Assessing the Utility of Not-at-Fault Traffic Crash Data in Racial Profiling Research," Paper presented the Confronting Racial Profiling in the 21st Century: Implications for Racial Justice conference, Boston (2003).
Antonovics, Kate, and Brian G. Knight, "A New Look at Racial Profiling: Evidence from the Boston Police Department," THIS REVIEW 91 (2009), 163–177.
Anwar, Shamena, and Hanming Fang, "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *American Economic Review* 96 (2006), 127–151.
Arrow, Kenneth, "The Theory of Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets* (Princeton, NJ: Princeton University Press, 1973).
Ayres, Ian, *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination* (Chicago: University of Chicago Press 2001).
Bayer, Patrick, and Christopher Timmins, "Estimating Equilibrium Models of Sorting Across Locations," *Economic Journal* 117 (2007), 353–374.
Becker, Gary S., *The Economics of Discrimination* (Chicago: University of Chicago Press, 1957).
Doleac, Jennifer L., and Nicholas J. Sanders, "Under the Cover of Darkness: How Ambient Light Influences Criminal Activity," College of William and Mary working paper (2013).
Dominitz, Jeff, "How Do the Laws of Probability Constrain Legislative and Judicial Efforts to Stop Racial Profiling?" *American Law and Economics Review* 5 (2003), 412–432.
Grogger Jeffrey T., and Greg Ridgeway, "Testing for Racial Profiling in Traffic Stops from behind a Veil of Darkness," *Journal of the American Statistical Association* 101 (2006), 878–887.
Hamermesh, Daniel S., *Workdays, Workhours, and Work Schedules: Evidence for the United States and Germany* (Kalamazoo, MI: W. E. Upjohn Institute for Employment Research, 1996).
Henderson, Vernon, Adam Storeygard, and David Weil, "Measuring Economic Growth from Outer Space," *American Economic Review* 102 (2012), 994–1028.

Imbens, Guido, and Joshua Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62 (1994), 467–475.

Knowles, John, Nicolo Persico, and Petra Todd, "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy* 109 (2001), 203–229.

Lamberth, John, "Revised Statistical Analysis of the Incidence of Police Stops and Arrests of Black Drivers/Travelers on the New Jersey Turnpike between Exits or Interchanges 1 and 3 from the Years 1988 through 1991, "Temple University working paper (1994).

McConnell, Elizabeth H., and Amie R. Scheidegger, "Race and Speeding Citations: Comparing Speeding Citations Issued by Air Traffic Officers with Those Issued by Ground Traffic Officers," paper presented at the annual meeting of the Academy of Criminal Justice Sciences, Washington, DC (April 4–8 2001).

Phelps, Edmund, "The Statistical Theory of Racism and Sexism," *American Economic Review* 62 (1972), 659–661.

Sanga, Sarath, "Reconsidering Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy* 117 (2009), 1155–1159.

——— "Does Officer Race Matter?" *American Law and Economics Review* (forthcoming).

Smith, Austin C., "Spring Forward at Your Own Risk: Daylight Saving Time and Fatal Vehicle Crashes," University of Colorado Boulder working paper (2014).

Steward, Dwight, *Racial Profiling: Texas Traffic Stops and Searches* (Austin, TX: Steward Research Group, 2004).

Weiss, Alexander, and Aviva Grumet-Morris, "Illinois Traffic Stop Statistics Act Report for the Year 2004" (Evanston, IL: Northwestern University Center for Public Safety, 2005).

Welsh, Brandon, and David Farrington, *Improved Street Lighting and Crime Prevention: A Systematic Review* (Stockholm: Swedish National Council for Crime Prevention 2007).

Zingraff, Mathew T., Mary Mason, William Smith, Donald Tomaskovic-Devey, Patricia Warrent, Harvey L. McMurray, and C. Robert Fenlon, "Evaluating North Carolina State Highway Patrol Data: Citation, Warnings, and Searches in 1998," report submitted to North Carolina Department of Crime Control and Public Safety and North Carolina State Highway Patrol (2000).