Online Appendix to "Testing for Racial Bias in Police Traffic Searches"

Joshua Shea*

May 11, 2024

A Constraints in the bilinear programming problem

This section provides some examples of how to impose linear constraints in the bilinear program. This section also provides a numerical example motivating the monotonicity restriction (13) on the distributions of risk.

A.1 Imposing linear constraints

Consider the vector of variables $\mathbf{x} = (x_1, \ldots, x_K)'$. The monotonicity constraint

$$x_1 \leq x_2 \leq \cdots \leq x_K \tag{A.1}$$

may be written as

$$\begin{bmatrix} 1 & -1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{bmatrix} \mathbf{x} \leq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

To reverse the direction of monotonicity, simply reverse the inequalities. Linear constraints of the form

$$\sum_{k=1}^{K} a_k x_k \leq b$$

*Email: jkcshea@illinois.edu. Department of Economics, University of Illinois Urbana Champaign.

may be written as

$$\mathbf{a}'\mathbf{x} \leq b, \tag{A.2}$$

where $\mathbf{a} = (a_0, \ldots, a_K)'$.

To ensure that the search probabilities $\boldsymbol{\sigma}_r = (\sigma(g_0; r), \ldots, \sigma(g_K; r))$ that are being optimized over are consistent with being a cumulative distribution function of $T_i \mid R_i = r$ for $r \in \{w, m\}$, $\boldsymbol{\sigma}_r$ must be non-decreasing in index k, and each element must be in the unit interval. The non-decreasing property of $\boldsymbol{\sigma}_r$ takes the form of (A.1), and the bounds on each element of $\boldsymbol{\sigma}_r$ take the form of (A.2) (i.e., choose $\mathbf{a}$ to be a standard basis vector).

To ensure that each distribution of risk $\mathbf{p}_{r,z}$ is consistent with being a PMF, the elements of $\mathbf{p}_{r,z}$ must be in the unit interval and sum to 1. Both of these constraints take the form of (A.2). The researcher may also choose to impose monotonicity constraints on $\mathbf{p}_{r,z}$. These will take the form of (A.1).

If the researcher has a prior on how the average risk ranks across R_i and Z_i, the researcher can impose the ranking using linear constraints. To see how, write the average risk conditional on race and setting as

$$\begin{aligned}
\mathbb{E}[\mathit{Guilty}_i \mid R_i = r, Z_i = z] &= \sum_{k=1}^{K} g_k \mathbf{p}_{r,z,k} \\
&= \mathbf{g}'\mathbf{p}_{r,z},
\end{aligned}$$

where $\mathbf{g} = (g_0, \ldots, g_K)'$ is the vector of discretized risks. Then the ranking

$$\mathbb{E}[\mathit{Guilty}_i \mid R_i = r_1, Z_i = z_1] \leq \mathbb{E}[\mathit{Guilty}_i \mid R_i = r_2, Z_i = z_2]$$

takes the form

$$\begin{aligned}
& \sum_{k=1}^{K} g_k \mathbf{p}_{r_1,z_1,k} \leq \sum_{k=1}^{K} g_k \mathbf{p}_{r_2,z_2,k} \\
\iff & \sum_{k=1}^{K} g_k \mathbf{p}_{r_1,z_1,k} - \sum_{k=1}^{K} g_k \mathbf{p}_{r_2,z_2,k} \leq 0 \\
\iff & \mathbf{g}'(\mathbf{p}_{r_1,z_1} - \mathbf{p}_{r_2,z_2}) \leq 0.
\end{aligned}$$

This restriction has the same form as (A.2), with $\mathbf{a} = \mathbf{g}$ and $\mathbf{x} = \mathbf{p}_{r_1,z_1} - \mathbf{p}_{r_2,z_2}$.

A.2 Imposing integrality constraints

The BP framework nests earlier models in the literature where $Search_i = \mathbb{1}\{G_i \geq t(R_i)\}$ for some deterministic function t. These models effectively impose an integrality constraint on $\boldsymbol{\sigma}_r$ so that

$$\sigma(g_k; r) \in \{0, 1\} \text{ for } k = 1, \ldots K. \tag{A.3}$$

Under such a restriction, the BP program becomes a mixed integer program, which can also be solved to provable global optimality.

A.3 Motivating restrictions on the distribution of risk

In this section, I provide a simple numerical example for how the probability density function of risk for drivers stopped may be decreasing as risk increases, even though the officer may be more likely to stop drivers with higher risk.

Let $Stop_i \in \{0, 1\}$ denote the stop decision of an officer for driver i, and

$$\pi_{r,z}(g) \equiv \mathbb{P}\{Stop_i = 1 \mid R_i = r, Z_i = z, G_i = g\}$$

denote the probability that a driver is stopped conditional on her race R_i, the setting Z_i, and her risk G_i. Suppose for some $(r, z) \in \{w, m\} \times \mathcal{Z}$, that $\pi_{r,z}(g)$ is as shown in the top panel of Figure A.1. The officer has a 1% probability of stopping a driver with zero risk, and this probability monotonicically increase to 50% as risk increases to unity.

Let $f^{Pop}_{G|R,Z}(\cdot \mid r, z)$ denote the density of risk among the population of drivers, i.e., the distribution of risk unconditional on being stopped. Suppose that the population risk follows a beta distribution with shape parameters 1 and 9, as depicted in the middle panel of Figure A.1. This implies 10% of all drivers carry contraband.

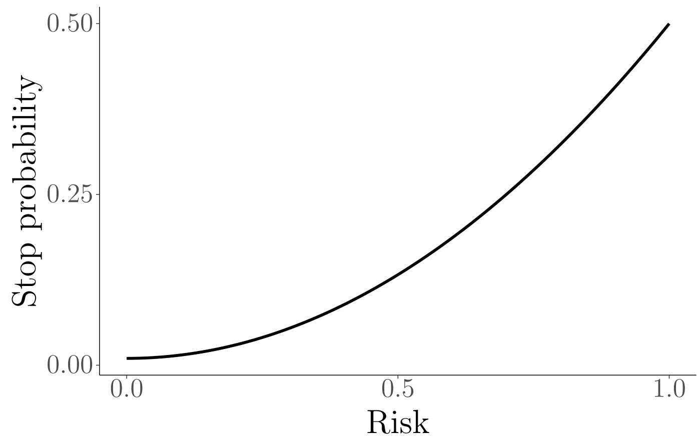
Then by Bayes' rule, the distribution of risk conditional on being stopped is

$$f_{G|R,Z}(g \mid r, z) = \frac{\pi_{r,z}(g)\ f^{Pop}_{G|R,Z}(g \mid r, z)}{\int_0^1 \pi_{r,z}(g')\ f^{Pop}_{G|R,Z}(g' \mid r, z)\ dg'},$$

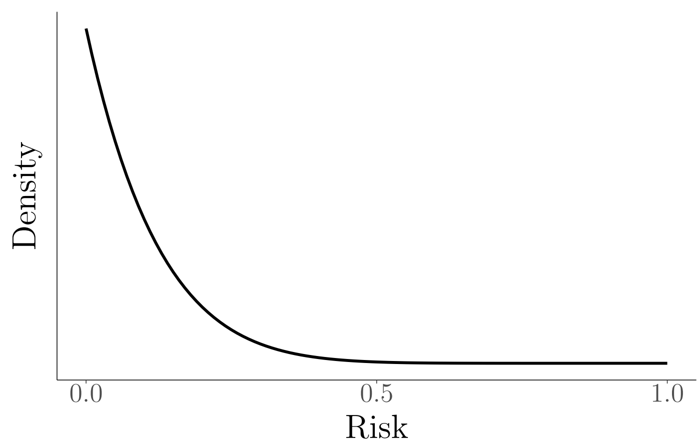
and is shown in the bottom panel of Figure A.1. Despite the officer being much more likely to stop high-risk drivers, the proportion of low-risk drivers in population is sufficiently large such that the density of risk conditional on being stopped is strictly decreasing.

Figure A.1: Monotone-decreasing density for risk of drivers stopped

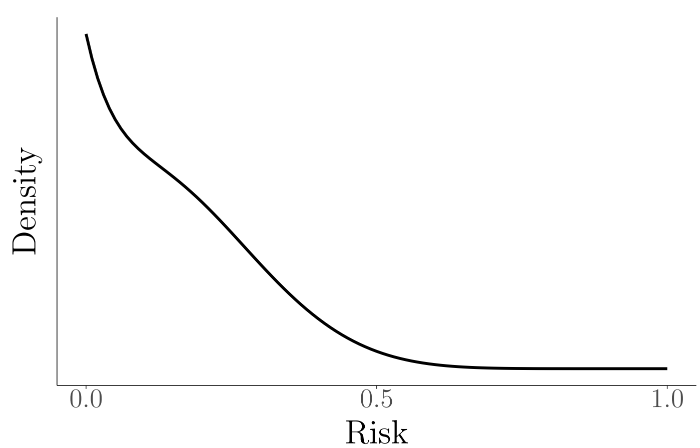
(a) Probability of stopping a driver



(b) Population distribution of risk



(c) Sample distribution of risk



B Modeling continuous risk using B-splines

In this section, I show how to adapt the methodology to allow for continuous distributions of risk.

Recall that the search and (unconditional) hit rates may be written as

$$\mathbb{P}\{Search_i = 1 \mid R_i = r, Z_i = z\} = \int_0^1 \sigma(g; r)\ dF_{G|R,Z}(g \mid r, z), \tag{B.4}$$

$$\mathbb{P}\{Hit_i = 1 \mid R_i = r, Z_i = z\} = \int_0^1 g\ \sigma(g; r)\ dF_{G|R,Z}(g \mid r, z). \tag{B.5}$$

Suppose there exists a density function $f_{G|R,Z}(\cdot \mid r, z)$ for $(r, z) \in \{w, m\} \times \mathcal{Z}$. Suppose also that $\sigma(g; r)$ and $f_{G|R,Z}$ can be modeled using B-splines. Then (B.4)–(B.5) can be written as bilinear terms

$$\mathbb{P}\{Search_i = 1 \mid R_i = r, Z_i = z\} = \boldsymbol{\sigma}_r' \mathbf{Q}_S \mathbf{p}_{r,z}, \tag{B.6}$$

$$\mathbb{P}\{Hit_i = 1 \mid R_i = r, Z_i = z\} = \boldsymbol{\sigma}_r' \mathbf{Q}_H \mathbf{p}_{r,z}, \tag{B.7}$$

where $\{\boldsymbol{\sigma}_r\}$, $\{\mathbf{p}_{r,z}\}$ are sets of parameters characterizing the officer's threshold distributions and the distributions of risk, respectively; and $\mathbf{Q}_S$ and $\mathbf{Q}_H$ are known matrices. Equations (B.6)–(B.7) follow from the fact that products of B-splines are also B-splines. Mørken (1991) provides the formula for calculating the coefficients for the products of two B-splines. To state this formula, it is necessary to first define several terms.

Following the notation in Mørken (1991), let k be a positive integer denoting the order of a spline, and $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots)$ be a non-decreasing sequence of real numbers denoting the knots of the spline. Then the B-spline $B_{i,k,\boldsymbol{\tau}}$ is defined using the recurrence relation

$$B_{i,k,\boldsymbol{\tau}}(x) \equiv \omega_{i,k,\boldsymbol{\tau}}(x) B_{i,k-1,\boldsymbol{\tau}}(x) + (1 - \omega_{i+1,k,\boldsymbol{\tau}}(x)) B_{i+1,k-1}(x),$$

where

$$\omega_{i,k,\boldsymbol{\tau}}(x) \equiv \begin{cases} \frac{(x-\tau_i)}{\tau_{i+k-1}-\tau_i} & \text{if } \tau_i < \tau_{i+k-1}, \\ 0 & \text{otherwise,} \end{cases} \tag{B.8}$$

and

$$B_{i,1,\boldsymbol{\tau}}(x) \equiv \begin{cases} 1 & \text{if } \tau_i \le x < \tau_{i+1} \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathcal{S}_{k,\boldsymbol{\tau}}$ denote the linear space spanned by the splines $\{B_{i,k,\boldsymbol{\tau}}\}$.

Let $\boldsymbol{\tau}'$ be a subsequence of $\boldsymbol{\tau}$. Then $\mathcal{S}_{k,\boldsymbol{\tau}'} \subseteq \mathcal{S}_{k,\boldsymbol{\tau}}$, and the B-splines $\{B_{j,k,\boldsymbol{\tau}'}\}$ are linear combinations of the B-splines $\{B_{i,k,\boldsymbol{\tau}}\}$ and can be written as

$$B_{j,k,\boldsymbol{\tau}'} = \sum_i \alpha_{j,k,\boldsymbol{\tau}',\boldsymbol{\tau}}(i) B_{i,k,\boldsymbol{\tau}}.$$

The coefficients $\{\alpha_{j,k,\boldsymbol{\tau}',\boldsymbol{\tau}}\}$ are discrete B-spline of order k on $\boldsymbol{\tau}$ with knots $\boldsymbol{\tau}'$ and satisfy the recurrence relation

$$\alpha_{j,k,\boldsymbol{\tau}',\boldsymbol{\tau}}(i) = \omega_{j,k,\boldsymbol{\tau}'}(\tau_{i+k-1})\alpha_{j,k-1}(i) + (1 - \omega_{j+1,k,\boldsymbol{\tau}'}(\tau_{i+k-1}))\alpha_{j+1,k-1}(i),$$

where $\omega_{j,k,\boldsymbol{\tau}'}$ is defined as in (B.8), and $\alpha_{j,1,\boldsymbol{\tau}',\boldsymbol{\tau}}(i) = B_{j,1,\boldsymbol{\tau}'}(\tau_i)$.

Suppose we have two splines, $f_1 \in \mathcal{S}_{k_1,\boldsymbol{\tau}_1}$ and $f_2 \in \mathcal{S}_{k_2,\boldsymbol{\tau}_2}$. Let $\mathcal{S}_{k,\boldsymbol{\tau}}$ denote the spline space containing the product of f_1 and f_2. The order of this spline space is $k = k_1 + k_2 - 1$. The knots of this spline space $\boldsymbol{\tau}$ contains all the distinct knots in $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$, with the multiplicity of the knots being determined as follows. For each knot τ in $\boldsymbol{\tau}$, let m_1 denote the multiplicity of τ in $\boldsymbol{\tau}_1$ and m_2 denote the multiplicity of τ in $\boldsymbol{\tau}_2$. Then the multiplicity of τ in $\boldsymbol{\tau}$ is

$$\widehat{m} = \begin{cases} \max(k_1 - 1 + m_2, k_2 - 1 + m_1) & \text{if } m_1 > 0 \text{ and } m_2 > 0, \\ k_1 - 1 + m_2 & \text{if } m_1 = 0 \text{ and } m_2 > 0, \\ k_2 - 1 + m_1 & \text{if } m_1 > 0 \text{ and } m_2 = 0, \\ 0 & \text{if } m_1 = 0 \text{ and } m_2 = 0. \end{cases} \tag{B.9}$$

Finally, let $P = \{p_1, \dots p_{k_1-1}\}$ be a set of $k_1 - 1$ integers from $I_{k-1} = \{1, \dots, k-1\}$. Let $Q = I_{k-1} \setminus P = (q_1, \dots, q_{k_2-1})$ be the set of the remaining $k_2 - 1$ integers. For a given integer i, define the knot vectors $\boldsymbol{\tau}^P$ and $\boldsymbol{\tau}^Q$ by

$$\boldsymbol{\tau}^P = (\dots, \tau_{i-1}, \tau_i, \tau_{i+p_1}, \tau_{i+p_2}, \dots, \tau_{i+p_{k_1-1}}, \tau_{i+k}, \tau_{i+k+1}, \dots), \tag{B.10}$$
$$\boldsymbol{\tau}^Q = (\dots, \tau_{i-1}, \tau_i, \tau_{i+q_1}, \tau_{i+q_2}, \dots, \tau_{i+q_{k_2-1}}, \tau_{i+k}, \tau_{i+k+1}, \dots). \tag{B.11}$$

Let Π denote the set of all subsets of I_{k-1} consisting of $k_1 - 1$ elements.

Theorem 1. *(Theorem 3.1 of Mørken (1991)) Let $f_1 = \sum_{j_1} c_{1,j_1} B_{j_1,k_1,\boldsymbol{\tau}_1}$ and $f_2 = \sum_{j_2} c_{2,j_2} B_{j_2,k_2,\boldsymbol{\tau}_2}$ be two given spline functions. Set $k = k_1 + k_2 - 1$ and construct the knot vector $\boldsymbol{\tau}$ according to (B.9). Then $f_1 f_2 \in \mathcal{S}_{k,\boldsymbol{\tau}}$ so that there exists coefficients $d_1, d_2, \dots$ such that $f_1(x) f_2(x) = \sum_i d_i B_{i,k,\tau}(x)$. Specifically, for a given i, the knot vectors $\boldsymbol{\tau}^P$ and $\boldsymbol{\tau}^Q$*

defined by (B.10)–(B.11) *satisfy* $\boldsymbol{\tau}_1 \subseteq \boldsymbol{\tau}^P$ *and* $\boldsymbol{\tau}_2 \subseteq \boldsymbol{\tau}^Q$, *and* d_i *is given by*

$$d_i = \sum_{P \in \Pi} \sum_{j_1} \sum_{j_2} c_{1,j_1} \alpha_{j_1,k_1,\boldsymbol{\tau}_1,\boldsymbol{\tau}^P}(i) c_{2,j_2} \alpha_{j_2,k_2,\boldsymbol{\tau}_2,\boldsymbol{\tau}^Q}(i) \bigg/ \binom{k-1}{k_1-1}.$$

It follows from Theorem 1 that the integral of $f_1 f_2$ can be written as a bilinear term,

$$\int f_1(x) f_2(x)\ dx = \sum_{j_1} \sum_{j_2} c_{1,j_1} c_{2,j_2} v_{j_1,j_2},$$

where

$$v_{j_1,j_2} = \sum_{i} \sum_{P \in \Pi} \alpha_{j_1,k_1,\boldsymbol{\tau}_1,\boldsymbol{\tau}^P}(i) \alpha_{j_2,k_2,\boldsymbol{\tau}_2,\boldsymbol{\tau}^Q}(i) \int B_{i,k,\boldsymbol{\tau}}(x)\ dx \bigg/ \binom{k-1}{k_1-1}$$

can be calculated. Equation (B.6) follows from letting f_1 be the threshold distribution $\sigma(\cdot; r)$, and letting f_2 be the density of risk $f_{G|R,Z}(\cdot \mid r, z)$. Equation (B.7) follows from the same reasoning, except f_1 is the threshold distribution scaled by risk, $g\sigma(g; r)$.[1]

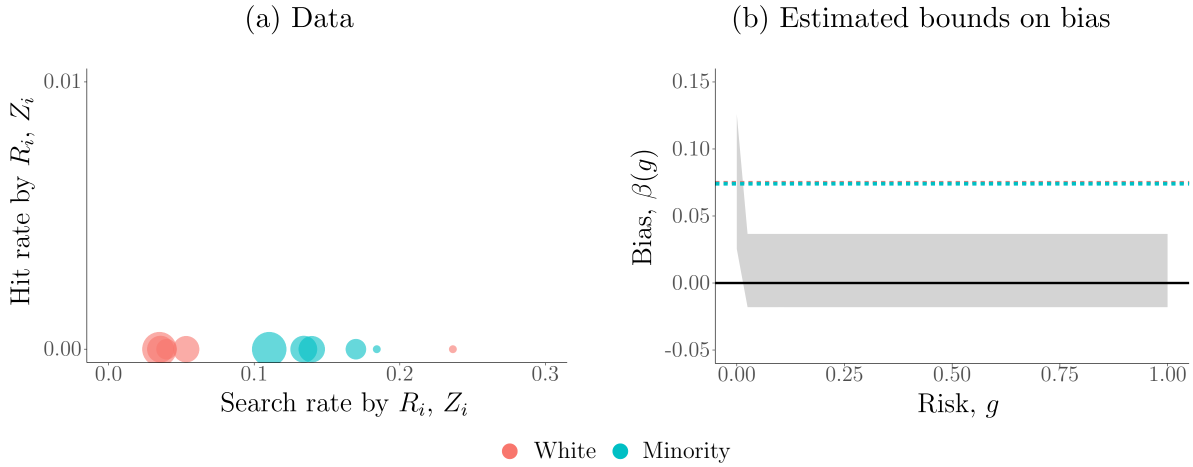
Shape restrictions on σ and $f_{G|R,Z}(\cdot \mid r, z)$ may be imposed through an auditing procedure (Shea and Torgovitsky, 2023). This procedure consists of first imposing the shape constraints on a coarse *constraint* grid over the domains of $\sigma(\cdot; r)$ and $f_{G|R,Z}(\cdot \mid r, z)$. The BP problems is then solved. Whether the solutions for $\sigma(\cdot; r)$ and $f_{G|R,Z}(\cdot \mid r, z)$ satisy the shape constaints is then checked on a much finer *audit* grid. Points in the audit grid where the shape constraints are violated are added to the constraint grid, and the BP problem is solved again. This procees is repeated until the shape constraints are satisfied on all points of the audit grid. This procedure avoids the computational and mathematical difficulties of determining whether the B-splines satisfy properties such as monotonicity, boundedness, and convexity (De Boor, 2001).

[1]If $\sigma(g; r)$ is a B-spline, then $g\sigma(g; r)$ will also be a B-spline.

C Estimates for biased officers

C.1 Estimates when averaging search and hit rates over $X_i \mid R_i = w$

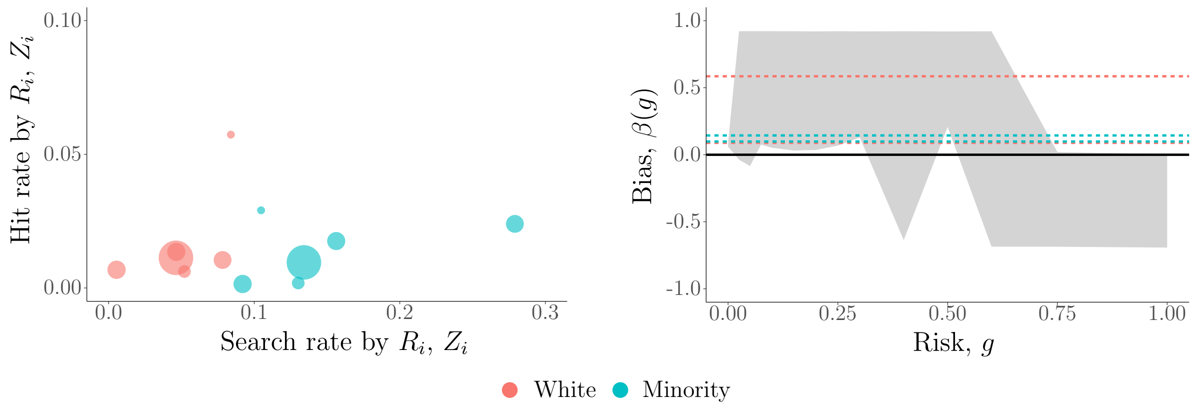
Figure C.2: Officer 8



Note: The size of the dots in the left panel represents the number of stops at each setting. The dashed lines in the right panel indicate the bounds on the average bias. If white drivers were treated as minority drivers, holding their risk constant, they would be searched between 7.50 and 7.53 percentage points more on average. If minority drivers were treated as white drivers, holding their risk constant, they would be searched between 7.36 and 7.46 percentage points less on average.

Figure C.3: Officer 23

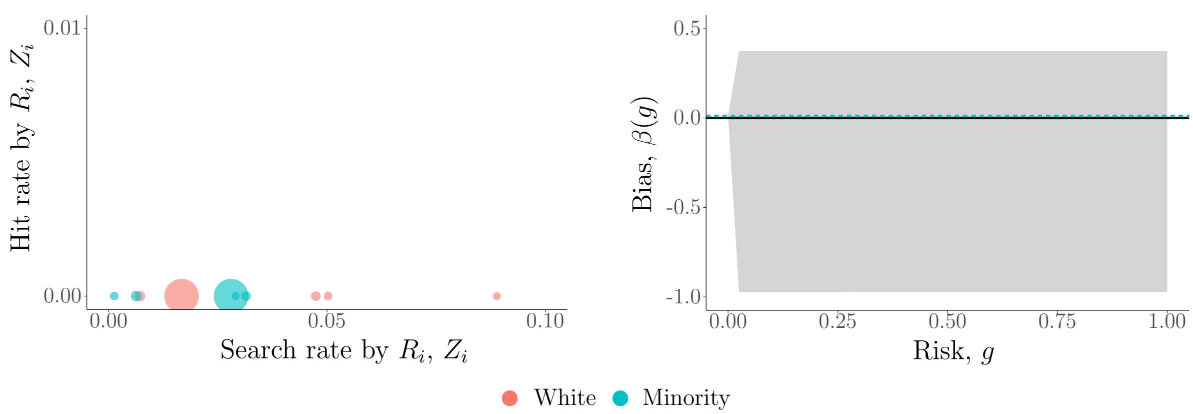
(a) Data (b) Estimated bounds on bias



Note: The size of the dots in the left panel represents the number of stops at each setting. The dashed lines in the right panel indicate the bounds on the average bias. If white drivers were treated as minority drivers, holding their risk constant, they would be searched between 8.85 and 58.54 percentage points more on average. If minority drivers were treated as white drivers, holding their risk constant, they would be searched between 9.87 and 14.39 percentage points less on average.

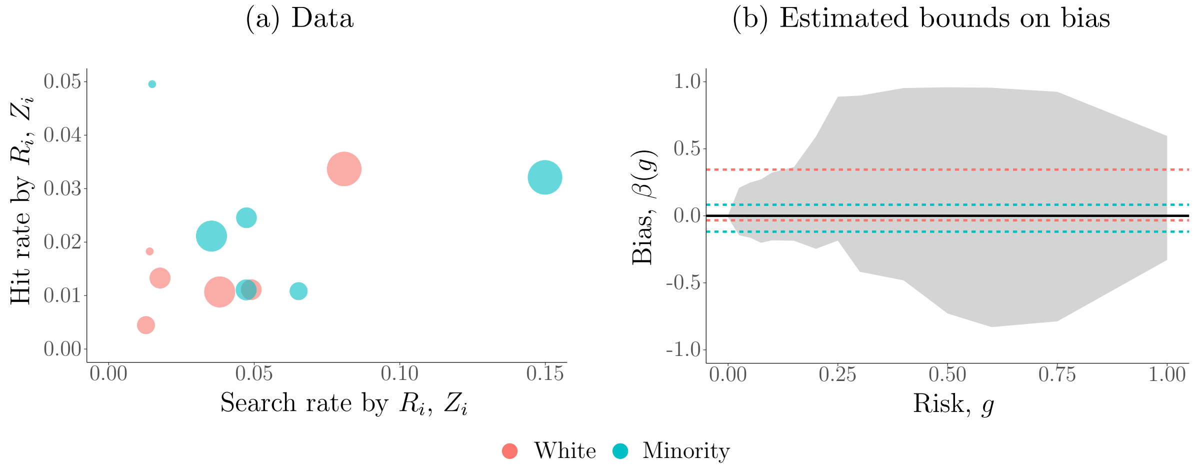
Figure C.4: Officer 35

(a) Data (b) Estimated bounds on bias



Note: The size of the dots in the left panel represents the number of stops at each setting. The dashed lines in the right panel indicate the bounds on the average bias. If white drivers were treated as minority drivers, holding their risk constant, they would be searched between 1.11 and 1.14 percentage points more on average. If minority drivers were treated as white drivers, holding their risk constant, they would be searched between 1.07 and 1.13 percentage points less on average.

Figure C.5: Officer 41

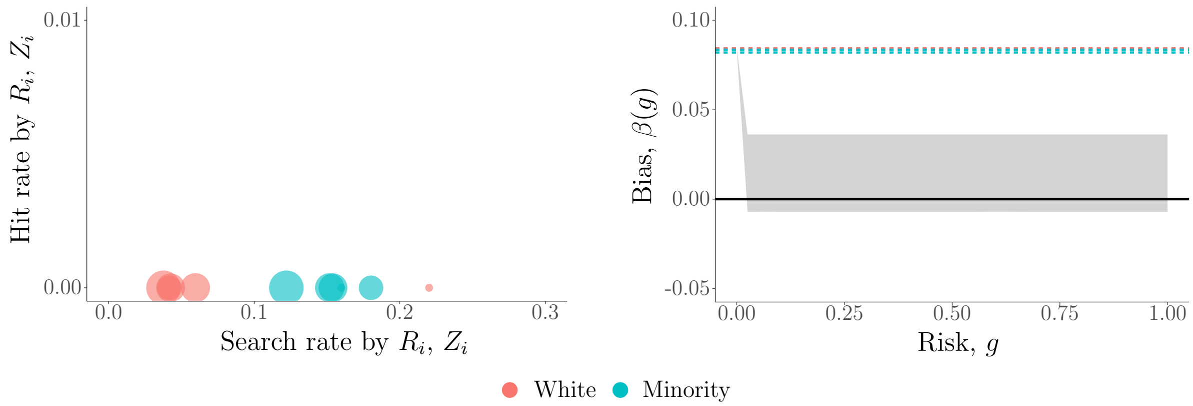


Note: The size of the dots in the left panel represents the number of stops at each setting. The dashed lines in the right panel indicate the bounds on the average bias. If white drivers were treated as minority drivers, holding their risk constant, they would be searched between 3.34 percentage points less and 34.49 percentage points more on average. If minority drivers were treated as white drivers, holding their risk constant, they would be searched between 11.82 percentage points less and 8.28 percentage points more on average.

C.2 Estimates when averaging search and hit rates over $X_i \mid R_i = m$

Figure C.6: Officer 8

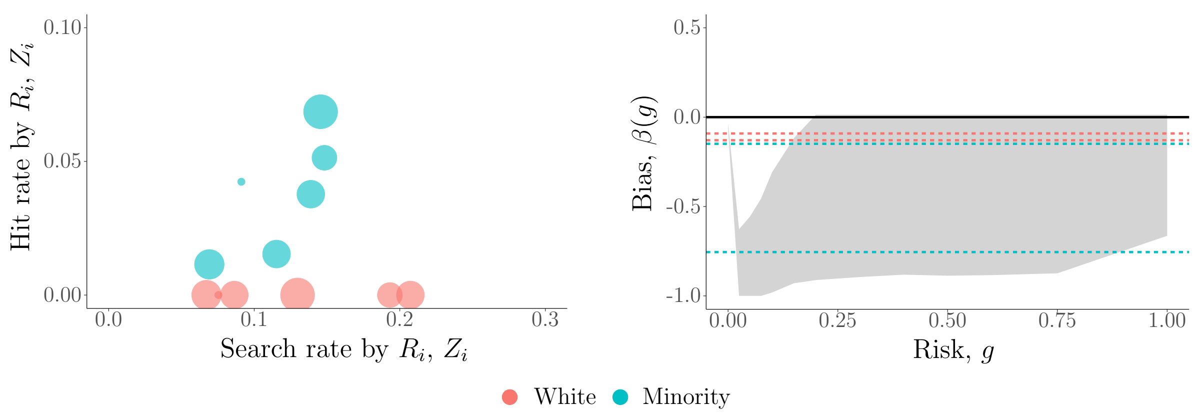
(a) Data (b) Estimated bounds on bias



Note: The size of the dots in the left panel represents the number of stops at each setting. The dashed lines in the right panel indicate the bounds on the average bias. If white drivers were treated as minority drivers, holding their risk constant, they would be searched between 8.37 amd 8.42 percentage points more on average. If minority drivers were treated as white drivers, holding their risk constant, they would be searched between 8.21 and 8.35 percentage points less on average.

Figure C.7: Officer 20

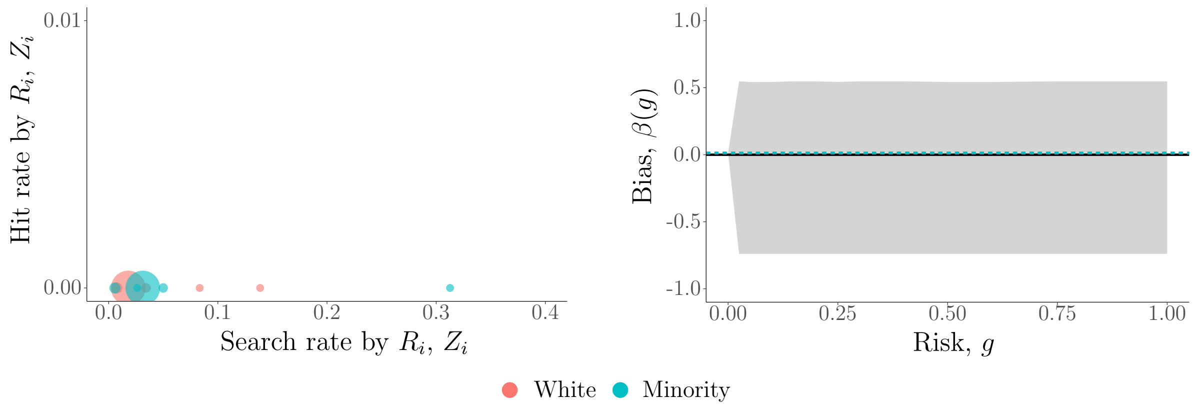
(a) Data (b) Estimated bounds on bias



Note: The size of the dots in the left panel represents the number of stops at each setting. The dashed lines in the right panel indicate the bounds on the average bias. If white drivers were treated as minority drivers, holding their risk constant, they would be searched between 9.22 and 12.91 percentage points less on average. If minority drivers were treated as white drivers, holding their risk constant, they would be searched between 17.26 and 74.75 percentage points more on average.

Figure C.8: Officer 35

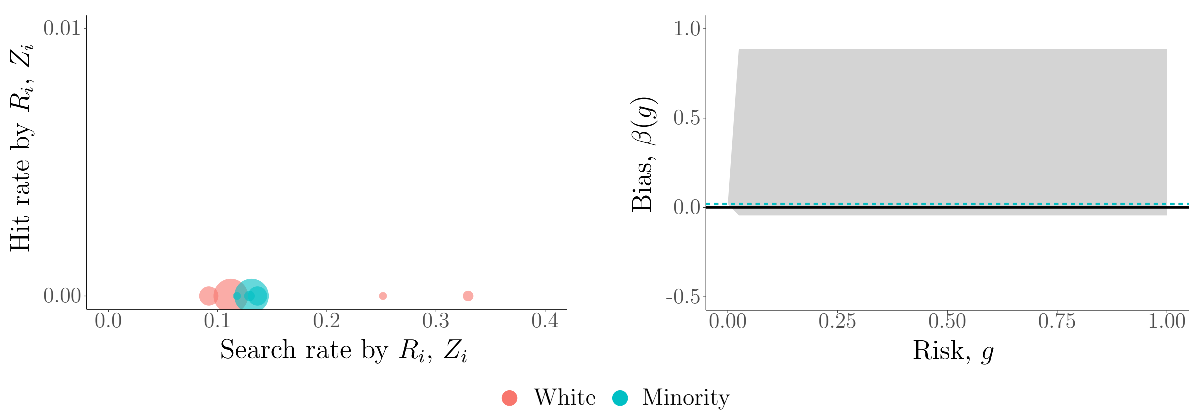
(a) Data (b) Estimated bounds on bias



Note: The size of the dots in the left panel represents the number of stops at each setting. The dashed lines in the right panel indicate the bounds on the average bias. If white drivers were treated as minority drivers, holding their risk constant, they would be searched between 1.33 and 1.34 percentage points more on average. If minority drivers were treated as white drivers, holding their risk constant, they would be searched between 1.31 and 1.34 percentage points less on average.

Figure C.9: Officer 43

(a) Data (b) Estimated bounds on bias



Note: The size of the dots in the left panel represents the number of stops at each setting. The dashed lines in the right panel indicate the bounds on the average bias. If white drivers were treated as minority drivers, holding their risk constant, they would be searched between 1.90 and 1.91 percentage points more on average If minority drivers were treated as white drivers, holding their risk constant, they would be searched between 1.90 and 1.99 percentage points less on average.

References

De Boor, C. (2001). *A Practical Guide to Splines, Revised Edition*, Volume 27. Springer-Verlag New York.

Mørken, K. (1991). Some Identities for Products and Degree Raising of Splines. *Constructive Approximation 7*, 195–208.

Shea, J. and A. Torgovitsky (2023). ivmte: An R Package for Extrapolating Instrumental Variable Estimates Away From Compliers. *Observational Studies 9*(2), 1–42.