Discrimination

Problem Set - Game Theory

GV482*

> "Michael Brown, unarmed, was shot 12 times by a police officer in Ferguson, Missouri, after Brown fit the description of a robbery suspect of a nearby store. Eric Garner, unarmed, was approached because officers believed he was selling single cigarettes from packs without tax stamps, and in the process of arresting him, an officer choked him and he died. Walter Scott, unarmed, was stopped because of a nonfunctioning third brake light and was shot eight times in the back while attempting to flee. Samuel Du Bose, unarmed, was stopped for failure to display a front license plate and while trying to drive away was fatally shot once in the head. Rekia Boyd, unarmed, was killed by a Chicago police officer who fired five times into a group of people from inside his police car. Zachary Hammond, unarmed, was driving away from a drug deal sting operation when he was shot to death by a Seneca, South Carolina, police officer. He was white. And so are 44 percent of police shooting subjects."
>
> Fryer (2019: 1211)

In the lecture, we have seen various forms of discrimination. Discrimination against minorities or against women. Discrimination by voters or by the bureaucracy. Discrimination due to beliefs, taste, or norms. There is one particular type of discrimination we have yet to talk about, one that has received a lot of attention over the 2020 Summer: discrimination by police force in the use of violence against minorities (especially, against African-Americans). We investigate this issue in this problem set and the next empirical one. As we will see, trying to uncover discrimination in police violence is a very difficult endeavour. Our goal in this week is to understand why with the help of

*Questions? Email d.foster@lse.ac.uk

some game theory.

Before starting on the problem, remember your first week of GV481 (for those who took that course). There, Mathilde briefly presented a paper by Johnson et al. (2019) who argued that there was no discrimination by majority officers since white and non-white officers were as likely to shoot a minority civilians. Basically, the authors were inferring discrimination by comparing the probability of being Hispanic or African-Americans conditional on being shot by a white officers with the probability of being Hispanic/African-American conditional on being shot by *a minority officer*. As Knox and Mummolo (2020a) pointed out, this is not enough to conclude whether discrimination is at play. Evidence of a lack of discrimination would require to show that there is no difference between the probability of being shot conditional on being a minority interacting with a white officer and the probability of being shot conditional on being a minority interacting with a minority officer. And to compute these conditional probabilities, Johnson et al. (2019) would have needed the likelihoods that a minority interacts with a white officer and with a non-white officer (also called, the base rate, the denominator in Bayes' rule formula).

The issue, well summarized in Knox and Mummolo (2020b), is that the data collection process by police departments make it very hard, if not impossible, to recover the actual number of interactions between officers and civilians. And this issue will be central to the problem set. In this problem set, we follow Knox, Lowe, and Mummolo (2020c) in defining the unit of analysis as an interaction between a police officer and a civilian.

Before proceeding, a few words of caution are in order. We will assume below that members of minority groups are more prone criminality and violent reaction than members of the majority group. This should not be misinterpreted. We are not saying that *everything else equals*, minority members are more likely to become criminals or engage in violence. This would be completely stupid, on top of being xenophobic or racist. Rather, we are saying that at the time of an interaction between a policy officer and an individual, everything else is **not** equal and, as result, minority members are more likely to be criminal or prone to violence. Factors such as poverty or past interactions with the policy are likely to explain the differences between the majority and minorities at the time of the interactions between an individual and a police officer (our unit of analysis, remember).[1] As

[1] The movie *La Haine*, for example, highlights how interactions with the police differ for majority group member in the rich center of Paris compared minorities in the suburb.

for the populism problem set, I include comments in gray (sometimes in parenthesis) to guide you through the problem set. Such comments will be absent from the next problem set dealing with democracy vs autocracy.

1 Preliminaries (not covered in the seminar)

This short section is meant to familiarize you with the problem. It raises two issues: (i) stopped white and black citizens may differ on unobservable characteristics and (ii) the distribution of these underlying characteristics may be distinct among whites and blacks. We look at these two issues separately and show that they push estimates of the effect of race in the use of force in separate dimension. We start with differences in underlying characteristics as it is the problem that Knox and Mummolo raise.

The table below shows the percentages of police stops resulting in the use of violence by the police for two groups of people (individuals with no criminal convictions and individuals with criminal convictions) according to their race: Whites or Blacks.

	Whites	**Blacks**
Group 1: No criminals	2%	4%
Group 2: Criminals	10%	12%

The researcher only observes the percentages of stopped individuals that experience police violence for each race. The researcher never observes whether the individuals being stopped has criminal or no criminal conviction. We assume that in the population α of individuals (independent of race) have no criminal conviction. We further assume that x_W of Whites that are stopped have no criminal conviction and x_B of Blacks that are stopped have no criminal conviction.

(a) Explain briefly why the real effect of being Black on police violence is 2%.

(b) Show that the researcher overestimates the effect of being Black on police violence if $x_B < x_W$ (i.e., Whites would be discriminated at the level of stop) and underestimates it if $x_B > x_W$ (i.e., Blacks would be discriminated at the level of stop as it is most likely).

We now look at the case when distributions of underlying characteristics differ among Whites and Blacks (again, as noted above, this can be due to past events). The table below again shows the percentages of police stops resulting in the use of violence by the police for two groups of people (individuals with no criminal convictions and individuals with criminal convictions) according to their race: Whites or Blacks.

	Whites	**Blacks**
Group 1: No criminals	4%	7%
Group 2: Criminals	10%	12%

The researcher only observes the percentages of stopped individuals that experience police violence for each race. The researcher never observes whether the individuals being stopped has criminal or no criminal conviction. We assume that in the population α_W of White individuals have no criminal conviction and $\alpha_B < \alpha_W$ of Black individuals have no criminal conviction. This possibility is not considered by Knox and Mummolo.

(c) Explain why we should think of the effects (rather than the effect) of race on police violence.

(d) Suppose that the proportion of individuals stopped from each group is equal to their proportion in the population (that is, there is no discrimination at the level of stops). Explain why the estimated effect is higher than any real effect of being Black on political violence whenever $\alpha_W \geq \frac{1}{6} + \alpha_B \frac{5}{6}$.

2 The model

We consider the interaction between a police officer P and a civilian I_g from group $g \in \{m, M\}$, with m indicating a a member of the minority and M a member of the majority. (Basically, we are going to compare how the police officer interacts with an individual from the majority group ($g = M$) and with an individual from a minority group ($g = m$).) The civilian is characterized by a two-dimensional type: a level of criminality $\theta^c_{I_g} \in \{-1, 1\}$, with $\theta^c = 1$ denoting a criminal, and $\theta^v_{I_g} \in \{-1, 1\}$, a "propensity for violent reaction" (for lack of a better term), with $\theta^v = 1$ indicating higher propensity for violence. (This means there are four types of individuals: (i) low criminality and low propensity for violence ($\theta^c_{I_g} = -1$ and $\theta^v_{I_g} = -1$), (ii) high criminality and low propensity for violence ($\theta^c_{I_g} = 1$ and $\theta^v_{I_g} = -1$), (iii) low criminality and high propensity for violence ($\theta^c_{I_g} = -1$ and $\theta^v_{I_g} = 1$), (iv) high criminality and high propensity for violence ($\theta^c_{I_g} = 1$ and $\theta^v_{I_g} = 1$; we do not assume that all criminals have a high propensity for violence, or vice versa.) At the time of the interaction, these types are treated as exogenous. However, as noted above, they could be the result of factors *outside* the interaction. As types are exogenous, the civilian is non-strategic in this setting.

The police officer makes two decisions. First, he decides whether to stop the civilian I_g: $a \in \{0, 1\}$, with $a = 1$ denoting the decision to stop her. Second, he decides how to react to in his interaction with the civilian: $v \in \mathbb{R}$, which we will refer to as the choice of violence. (We try to represent the police officer's overall action as a set of two decisions: first, should he stop the civilian? and, then, conditional on stopping the civilian to search her, the police officer must decide how to behave.)

When the police officer does not stop the civilian ($a = 0$), his payoff is then simply:

$$-\theta^c_{I_g}$$

This means that when the police officer does not stop a criminal, he suffers a loss of -1 (since a criminal has a type $\theta^c_{I_g} = 1$). In contrast, the police officer gets a payoff gain of 1 if he does not stop a non-criminal (since $\theta^c_{I_g} = -1$ for non-criminal). Basically, the police officer cares about type I (not stopping a criminal) and type II (stopping a non-criminal) errors.

When P stops the civilian ($a = 1$), his payoff is:

$$\theta^c_{I_g} + b^c_g - (v - (\alpha\theta^c_{I_g} + \theta^v_{I_g} + b^v_g))^2.$$

The first term ($\theta^c_{I_g}$) states that stopping a criminal increases the police officer's payoff, whereas stopping an innocent individual decreases the officer's payoff. This captures again the idea of avoiding type I and type II errors. The second term b^c_g represents the possible payoff bias in favour of stopping an individual from group g. This can be understood as a form of prejudice against (or for!) members of group g. We also assume, with the quadratic loss function, that there is an optimal reaction v to the behavior of the civilian, which is captured by $\alpha\theta^c_{I_g} + \theta^v_{I_g}$. (This is a strong assumption, but it captures the idea that there are appropriate reactions according to how the stop unravels.) The expression $\alpha\theta^c_{I_g} + \theta^v_{I_g}$ can be interpreted as higher value denoting more violent behavior. Note that the civilian's behavior is a function of her propensity for violent reaction as well as (possibly) her criminality status. The weight on the criminality status is $\alpha \in [-1, 1]$. If $\alpha = 0$, then criminality and civilian's reaction are independent. When $\alpha > 0$, criminals are more likely to be violent than non criminals, and inversely when $\alpha < 0$. For the police officer, the appropriate reaction may also be a function of an intrinsic bias (animus/prejudice) against (or for!) a member of group g: b^v_g. To avoid carrying out too many parameters, we assume $b^c_M = b^v_M = 0$ (note that the officer is biased against the minority of $b_m > 0$ and in favour if $b_m < 0$).

The police officer does not observe $\theta^c_{I_g}$ or $\theta^v_{I_g}$. His prior is that $Pr(\theta^c_{I_g} = 1) = \pi^c_g$ and $Pr(\theta^v_{I_g} = 1) = \pi^v_g$. We allow for the priors to differ between the two groups on both dimensions. Before deciding whether to stop the civilian, the police offer receives a signal $s^c_{I_g} \in \{-1, 0, 1\}$. The signal satisfies the following distribution conditional on the type: $Pr(s^c_{I_g} = 1|\theta^c_{I_g} = 1) = q^c$, $Pr(s^c_{I_g} = 0|\theta^c_{I_g} = 1) = 1 - q^c$, $Pr(s^c_{I_g} = -1|\theta^c_{I_g} = -1) = q^c$, and $Pr(s^c_{I_g} = 0|\theta^c_{I_g} = -1) = 1 - q^c$, with $q^c \in (0, 1)$. (Basically, we are assuming that the police officer does not make his decision at random. He first observes some hints about the criminality of the individual he encounters and then decides whether to stop her or not.) This signal could be understood as pure information or as subcategories of individuals. One issue for researcher, to which we will turn at the end of the problem set, is that the signal is observed by the police officer, but not recorded.

The distribution of the signal is summarized in Table 1 for convenience

Conditional on	$s^c_{I_g} = -1$	$s^c_{I_g} = 0$	$s^c_{I_g} = 1$
$\theta^c_{I_g} = 1$	0	$1 - q^c$	q^c
$\theta^c_{I_g} = -1$	q^c	$1 - q^c$	0

Table 1: Distribution of $s^c_{I_g}$ conditional on $\theta^c_{I_g}$

If the police officer stops the civilian, then he also receives a signal $s^v_{I_g} \in \{-1, 0, 1\}$ of $\theta^v_{I_g}$. The distribution of this signal conditional on the type is: $Pr(s^v_{I_g} = 1|\theta^v_{I_g} = 1) = q^v$, $Pr(s^v_{I_g} = 0|\theta^v_{I_g} = 1) = 1 - q^v$, $Pr(s^v_{I_g} = -1|\theta^c_{I_g} = -1) = q^v$, and $Pr(s^v_{I_g} = 0|\theta^v_{I_g} = -1) = 1 - q^v$, with $q^v \in (0, 1)$. (As above, we use this to capture the idea that when the officer decides how to act with a stopped citizen, he receives some partial information about this propensity for violence through the individual's behaviour.) The distribution is summarized in Table 2.

Conditional on	$s^v_{I_g} = -1$	$s^v_{I_g} = 0$	$s^v_{I_g} = 1$
$\theta^v_{I_g} = 1$	0	$1 - q^v$	q^v
$\theta^v_{I_g} = -1$	q^v	$1 - q^v$	0

Table 2: Distribution of $s^v_{I_g}$ conditional on $\theta^v_{I_g}$

Note that the two signals are independent so s^c reveals nothing about θ^v, and vice versa.

The game, in turn, proceeds as follows.

1. Nature draws $\theta^c_{I_g}, \theta^v_{I_g} \in \{-1, 1\}^2$.

2. P encounters civilian I_g and receives a signal $s^c_{I_g} \in \{-1, 0, 1\}$. He decides whether to stop her: $a \in \{0, 1\}$.

3. If P does not stop I_g, the game ends and payoffs are realized.

   If P stops I_g, he receives a signal $s^v_{I_g} \in \{-1, 0, 1\}$ and chooses $v \in \mathbb{R}$. The game ends and payoffs are realized.

We do not need to define an equilibrium concept since there is a single strategic actor: P who maximizes his expected payoff. Throughout, we impose the following assumptions. A minority of civilians from both groups are criminals and have a high propensity for violence: $\pi^c_g < 1/2$ and $\pi^v_g < 1/2$ for $g \in \{m, M\}$. Priors satisfy that minority members are weakly more likely to be criminal and prone to violence: $\pi^c_m \geq \pi^c_M$ and $\pi^v_m \geq \pi^v_M$. Though this is not very important, we assume that P stops I_g when indifferent.

Table 3 provides a point of reference by describing the model's main parameters and variables.

Variables	Definition
$\theta^c_{I_g} \in \{-1, 1\}$	Criminality of individual I_g
$\theta^v_{I_g} \in \{-1, 1\}$	Propensity for violence
π^j_g	Group-specific prior that $\theta^j_{I_g} = 1$, $j \in \{c, v\}$
α	Impact of $\theta^c_{I_g}$ on civilian's behavior in a stop
$s^c_{I_g} \in \{-1, 0, 1\}$	Signal of criminality received by P
$s^v_{I_g} \in \{-1, 0, 1\}$	Signal of violence propensity received by$P$
q^j	Probability that signal matches the type $\theta^j_{I_g}$, $j \in \{c, v\}$
$a \in \{0, 1\}$	decision whether to stop the civilian
$v \in \mathbb{R}$	Violence towards the civilian
b^c_m	Bias towards stop for/against minority
b^v_m	Bias towards violence for/against minority

Table 3: Main variables and parameters

For this sort of game, we are moving backward (it is not a form of backward induction since we have a single actor, but the logic is not too dissimilar). We, first, determine the behaviour (level of violence v) of the police officer in all possible cases (for all possible signals) if it were to stop the individual. We, then, determine in which cases (for which signals) the police officer stops an individual of each group. We will use these results to think about the difficulty to measure discrimination in police violence.

Part A - Police behaviour

In this first part, we think to determine the police officer's behavior as a function of his information: $v^*(s^c_{I_g}, s^v_{I_g})$. To do so, we first consider what the signals tell P about the civilian's level of criminality and propensity for violence.

Q1 In this question, we seek to understand what the police officer can infer from the signals. To answer these questions, you need to use Bayes' rule.

(a) Show that the officer's posterior after signal $s^c_{I_g} = 1$ satisfies: $Pr(\theta^c_{I_g} = 1 | s^c_{I_g} = 1) = 1$.

(b) Show that the officer's posterior after signal $s^c_{I_g} = -1$ satisfies: $Pr(\theta^c_{I_g} = 1 | s^c_{I_g} = -1) = 0$.

(c) Show that the officer's posterior after signal $s^c_{I_g} = 0$ satisfies: $Pr(\theta^c_{I_g} = 1 | s^c_{I_g} = 0) = \pi^c_g$.

(d) Show that the police officer's posterior regarding the propensity for violence ($\theta^v_{I_g}$) satisfies:

$$Pr(\theta^v_{I_g} = 1 | s^v_{I_g} = 1) = 1$$
$$Pr(\theta^v_{I_g} = 1 | s^v_{I_g} = 0) = \pi^v_g$$
$$Pr(\theta^v_{I_g} = 1 | s^v_{I_g} = -1) = 0$$

Q2: With this information, we are ready to compute the police officer's equilibrium behaviour conditional on a stop ($v^*(s^c_{I_g}, s^v_{I_g})$). Our goal is to fill the following table where each cell corresponds to a behaviour after a pair of signals.

	$s^c_{I_g} = 1$	$s^c_{I_g} = 0$	$s^c_{I_g} = -1$
$s^v_{I_g} = 1$	A		
$s^v_{I_g} = 0$			
$s^v_{I_g} = -1$			

Table 4: Equilibrium violence

For example cell A corresponds to the level of violence chosen by the police officer after seeing signal $s^c_{I_g} = 1$ and $s^v_{I_g} = 1$. On path (that is, in term of what we actually observes), some level of violence may never be realised because the individual is not stopped. However, the police officer's decision whether to stop an individual depends on the calculus of what he would have to do should he stop the citizen. Hence, we need to compute v^* for all possible scenarios.

(a) Explain briefly, but carefully why: $v^*(s^c_{I_g}, s^v_{I_g}) = E_{\theta^c_{I_g},\theta^v_{I_g}}(\alpha\theta^c_{I_g} + \theta^v_{I_g} | s^c_{I_g}, s^v_{I_g}) + b^v_g$.

Comment: The expression $E_{\theta^c_{I_g},\theta^v_{I_g}}$ means that we are taking the expectations over the random variables $\theta^c_{I_g}, \theta^v_{I_g}$, which are the unknown from P's perspective.

(b) Explain briefly, but carefully why we can rewrite: $v^*(s^c_{I_g}, s^v_{I_g}) = \alpha E_{\theta^c_{I_g}}(\theta^c_{I_g} | s^c_{I_g}) + E_{\theta^v_{I_g}}\theta^v_{I_g} | s^v_{I_g}) + b^v_g$.

Comment: The question asks which properties of expectations and the variables allow us to rewrite the choice of violence as such.

(c) Using your answer to **Q1**, explain briefly why

$$E_{\theta^v_{I_g}}(\theta^v_{I_g}|s^v_{I_g}=1)=1$$
$$E_{\theta^v_{I_g}}(\theta^v_{I_g}|s^v_{I_g}=0)=2\pi^c_g-1$$
$$E_{\theta^v_{I_g}}(\theta^v_{I_g}|s^v_{I_g}=-1)=-1$$
$$E_{\theta^c_{I_g}}(\theta^c_{I_g}|s^c_{I_g}=1)=1$$
$$E_{\theta^c_{I_g}}(\theta^c_{I_g}|s^c_{I_g}=0)=2\pi^v_g-1$$
$$E_{\theta^c_{I_g}}(\theta^c_{I_g}|s^c_{I_g}=-1)=-1$$

Based on our answers to **Q2**, we then have the following level of violence chosen by the police officer in all possible scenarios.

	$s^c_{I_g}=1$	$s^c_{I_g}=0$	$s^c_{I_g}=-1$
$s^v_{I_g}=1$	$\alpha+1+b^v_g$	$\alpha(2\pi^c_g-1)+1+b^v_g$	$-\alpha+1+b^v_g$
$s^v_{I_g}=0$	$\alpha+(2\pi^v_g-1)+b^v_g$	$\alpha(2\pi^c_g-1)+(2\pi^v_g-1)+b^v_g$	$-\alpha+(2\pi^v_g-1)+b^v_g$
$s^v_{I_g}=-1$	$\alpha-1+b^v_g$	$\alpha(2\pi^c_g-1)-1+b^v_g$	$-\alpha-1+b^v_g$

Table 5: Equilibrium violence

Part B - The decision whether to stop the civilian

Q3 We can now turn to the decision whether to stop the civilian. We denote $A_g(s^c_{I_g})$ P's expected utility from stopping I_g after receiving signal $s^c_{I_g}$ (remember that $s^v_{I_g}$ is received after stopping so cannot be condition his decision on it at this stage). We will compute it in this question. Note that this question is algebra intensive. It is not too hard, but it is quite tedious. If you hate algebra, you may want to skip it, though it is always good practice (for your exams, for your future life) to try to answer this sort of questions.

(a) Show that $-(v^*(s^c_{I_g}, s^v_{I_g}) - b^v_g - \theta^v_{I_g} - \alpha\theta^c_{I_g})^2 = -(E(\theta^v_{I_g}|s^v_{I_g}) - \theta^v_{I_g})^2 - \alpha^2(E(\theta^c_{I_g}|s^c_{I_g}) - \theta^c_{I_g})^2 - 2\alpha(E(\theta^v_{I_g}|s^v_{I_g}) - \theta^v_{I_g})(E(\theta^c_{I_g}|s^c_{I_g}) - \theta^c_{I_g})$.

(b) Show that $E_{\theta^c_{I_g},\theta^v_{I_g}}\big[-(E(\theta^v_{I_g}|s^v_{I_g})-\theta^v_{I_g})^2-\alpha^2(E(\theta^c_{I_g}|s^c_{I_g})-\theta^c_{I_g})^2-2\alpha(E(\theta^v_{I_g}|s^v_{I_g})-\theta^v_{I_g})(E(\theta^c_{I_g}|s^c_{I_g})-\theta^c_{I_g})|s^c_{I_g}, s^v_{I_g}\big] = -Var(\theta^v_{I_g}|s^v_{I_g}) - \alpha^2 Var(\theta^c_{I_g}|s^c_{I_g})$.
Comment: $Var(\theta|s)$ is the variance of θ conditional on the signal s. You should know the formula of the variance and see how it maps into what we are doing.

(c) Show that

$$\begin{aligned} Var(\theta^v_{I_g}|s^v_{I_g} = 1) =&0 \\ Var(\theta^v_{I_g}|s^v_{I_g} = 0) =&4\pi^v_g(1-\pi^v_g) \\ Var(\theta^v_{I_g}|s^v_{I_g} = -1) =&0 \end{aligned}$$

By a similar reasoning, we obviously have:

$$\begin{aligned} Var(\theta^c_{I_g}|s^c_{I_g} = 1) =&0 \\ Var(\theta^c_{I_g}|s^c_{I_g} = 0) =&4\pi^c_g(1-\pi^c_g) \\ Var(\theta^c_{I_g}|s^c_{I_g} = -1) =&0 \end{aligned}$$

Now, as noted above, P does not observer $s^v_{I_g}$ until he has stopped the civilian. Hence, we should not consider $Var(\theta^v_{I_g}|s^v_{I_g})$, but the expectation of the variance over different possible signals.
(d) Show that

$$E_{s^v_{I_g}}\left[Var(\theta^v_{I_g}|s^v_{I_g})\right] = (1-q^v)4\pi^v_g(1-\pi^v_g)$$

(e) We are now ready to solve **Q3**. Show that:

$$A_g(1) = 1 + b_g^c - 4\pi_g^v(1-\pi_g^v)(1-q^v) \tag{1}$$

$$A_g(0) = (2\pi_g^c - 1) + b_g^c - 4\alpha^2\pi_g^c(1-\pi_g^c) - 4\pi_g^v(1-\pi_g^v)(1-q^v) \tag{2}$$

$$A_g(-1) = -1 + b_g^c - 4\pi_g^v(1-\pi_g^v)(1-q^v) \tag{3}$$

Q4 We can now look at P's stopping decision. We first consider a civilian from the majority group M and then add some assumptions before turning to a civilian of the minority group m.

(a) Show that P stops a civilian of group M if and only if $s_{I_M}^c = 1$.

Absent additional assumptions, we cannot say much more. We will follow empirical researchers that document that minorities are more likely to be stopped than members of the majority (e.g., Gelman, Fagan, and Kiss, 2012).

(b) Suppose that the police officer chooses to stop I_m after signals $s_m^c = 1$ and $s_m^c = 0$ (but not after signal $s_m^c = -1$). Based on our set-up, what can we infer about b_m^c? About π_m^c?

Part C - The researchers' problem

In this part, we consider the problem of researchers who want to understand whether there is discrimination against minority in the use of violence. Building on the lecture, we say that there is discrimination against minorities in the use of violence if an individual from the minority group receives a different treatment v^* than an individual of the majority group with the same signals (s^c, s^v). Note that discrimination is equivalent to $b_m^v > 0$ (if taste-based) and/or $\pi_m^v > \pi_M^v$ (if statistical). Hence, we can say that there is no discrimination in the use of violence if and only if $b_m^v = 0$ and $\pi_m^v = \pi_M^v$.

From now on and until the end of the problem set, we will work under the assumption that minorities are stopped whenever $s_{I_m}^c = 1$ or $s_{I_m}^c = 0$. We will also assume that there are an infinite number of interactions between civilians and police officers under the model above (using our past problem sets, this means the proportion of stop is equal to the probability that an individual is stop, the average of a quantity is equal to the expectation of this quantity, etc.). Of these interactions, researchers only observe those when a stop occurs.

As a starting point, we will assume that the researchers observe v^* (given by Table 5) whenever a stop occurs. We will consider under which conditions, observable levels of violence permit to determine whether the police discriminates against the minority in its use of violence.

Q5 In this question, we compare the average violence against members of group M and members of group m, which we denote $\overline{V}_M$ and $\overline{V}_m$, respectively. We, first, need to compute these expected level of violence.

(a) We start with a civilian from the majority group. Recall that a citizen from group M is stopped only if $s^c_{I_M} = 1$. Show that

$$\overline{V}_M = \alpha + (2\pi^v_M - 1) \tag{4}$$

(b) We now turn to $\overline{V}_m$, which will prove slightly more complicated. Recall that a member of group m is stopped whenever she produces signal $s^c_{I_m} = 1$ or $s^c_{I_m} = 0$. Explain why the proportion of minority individuals with signal $s^c_{I_m} = 1$ *among* those arrested is: $\frac{\pi^c_m q^c}{\pi^c_m q^c + (1-q^c}$.

(c) Show that the average level of violence against the minority members who are arrested after signal $s^c_{I_m} = 1$ is: $\alpha + (2\pi^v_m - 1) + b^v_m$.

(d) Show that the average level of violence against the minority members who are arrested after signal $s^c_{I_m} = 0$ is: $\alpha(2\pi^c_m - 1) + (2\pi^v_m - 1) + b^v_m$.

(e) Using your answers to **(b)**-**(d)**, show that

$$\overline{V}_m = \alpha + (2\pi^v_m - 1) + b^v_m - 2\alpha(1 - \pi^c_m)\frac{1 - q^c}{\pi^c_m + (1 - q^c)} \tag{5}$$

(f) Suppose the researchers find $\overline{V}_m \geq \overline{V}_M$ and concludes that officers discriminate against minorities in the use of violence. Under which condition, is this inference correct? Justify your answer and provide some intuition for it.

Comment: To answer this question, suppose there were no discrimination according to the discussion above, could we still find $\overline{V}_m \geq \overline{V}_M$? If so, under which condition?

Unfortunately, the researchers' problem is even more severe. Researchers do not observe the level of violence in stops that did not occur and, further, only observe the level of violence if police officers engage in certain behaviours. In our set-up, this can be understood as researchers observing $v_g^*(\cdot)$ if and only if $v_g^*(\cdot) > 0$. That is, observations are left-censored (or truncated). Let's denote $\overline{V}_g^+$ the average observed violence against members of group g. To limit the number of cases, suppose that $\alpha < 1 - 2\pi_m^v$.

Q6 In this question, we try to understand what type of inference researchers can make given the observability problem they face. To do so, we suppose in this question only that police officers do not have prejudice ($b_m^v = 0$) and hold the same prior for the minority group as for the majority group ($\pi_m^v = \pi_M^v$) when it comes to the propensity for violence (we do not exclude discrimination at the level of stops). We want to understand the implication for the observed average level of violence.

(a) Show that when it comes to the interactions between a police officer and individuals from the majority group, the researchers observe the following quantity (appropriately modifying Table 5):

	$s_{I_M}^c = 1$	$s_{I_M}^c = 0$	$s_{I_M}^c = -1$
$s_{I_M}^v = 1$	$\alpha + 1$	Not stopped	Not stopped
$s_{I_M}^v = 0$	0	Not stopped	Not stopped
$s_{I_M}^v = -1$	0	Not stopped	Not stopped

Table 6: Observed violence against majority

(b) Explain why the average observed level of violence against members of group M is

$$\overline{V}_M^+ = \pi_M^v q^v (\alpha + 1) \tag{6}$$

Comment: Think about the proportion of stopped individuals for which you observe some violence.

(c) Show that when it comes to the interactions between a police officer and individuals from the minority group, the researchers observe the following quantity (appropriately modifying Table 5):

	$s^c_{I_m} = 1$	$s^c_{I_m} = 0$	$s^c_{I_m} = -1$
$s^v_{I_m} = 1$	$\alpha + 1$	$\alpha(2\pi^c_m - 1) + 1$	Not stopped
$s^v_{I_m} = 0$	0	0	Not stopped
$s^v_{I_m} = -1$	0	0	Not stopped

Table 7: Observed violence against minority

(d) Explain why the average observed level of violence against members of group m is

$$\overline{V}^+_m = \pi^v_m q^v \left(\alpha + 1 - 2\alpha(1 - \pi^c_m) \frac{1 - q^c}{\pi^c_m q^c + (1 - q^c)} \right) \tag{7}$$

(e) A researcher finds $\overline{V}^+_m \geq \overline{V}^+_M$. Is this enough to conclude that there is discrimination in the use of violence? If not, what additional condition is necessary? Compare to your answer in **Q5(f)**.

Q7 In this question, we assume that police officers have some prejudice against members of the minority: $b^v_m > 0$. Further, we assume that this bias satisfies the following properties (to limit the number of cases): $-\alpha < b^v_m - (1 - 2\pi^v_m) < \alpha(1 - 2\pi^c_m)$ and $b^v_m < 2\pi^v_m$. Further, we suppose that $\alpha > 0$ and still maintain $\alpha < 1 - 2\pi^v_m$.

(a) Show that the observed level of violence against minority members now satisfy

	$s^c_{I_g} = 1$	$s^c_{I_g} = 0$	$s^c_{I_g} = -1$
$s^v_{I_g} = 1$	$\alpha + 1 + b^v_m$	$\alpha(2\pi^c_m - 1) + 1 + b^v_m$	Not stopped
$s^v_{I_g} = 0$	$\alpha + (2\pi^v_m - 1) + b^v_m$	0	Not stopped
$s^v_{I_g} = -1$	0	0	Not stopped

Table 8: Observed violence against minority

(b) Show that the average observed violence against minority members satisfies

$$\begin{aligned}\overline{V}_m^+ =& \frac{\pi^c q^c}{\pi^c q^c + (1-q^c)}\big(\pi_m^v q^v(\alpha + 1 + b_m^v) + (1-q^v)(\alpha + (2\pi_m^v - 1) + b_m^v)\big) \\ &+ \frac{1-q^c}{\pi^c q^c + (1-q^c)}\pi_m^v q^v(\alpha(2\pi_m^c - 1) + 1 + b_m^v)\end{aligned} \tag{8}$$

Give an interpretations of the quantities in $\overline{V}_m^+$.

Comment: For the interpretation, it can be useful to think in term of compliers and always takers. A complier is an individual who suffers some (recorded) level of violence only because she is a member of the minority. An always taker is an individual who would always be on the receiving end of violence. Note that we have two sorts of compliance problems: at the level of stops and at the level of the use of violence.

(c) Can you rank $\overline{V}_m^+ - \overline{V}_M^+$ (with $\overline{V}_M^+$ given by Equation 6 since it is unaffected by discrimination) and $\overline{V}_m - \overline{V}_M$ (i.e., which difference is bigger)? Provide some intuition for your answer. Discuss the implications of your finding.

Q8 In this last question, we connect this problem set with the empirical problem set you will solve next week. We do not data on the level of violence (though, we have some data on different degrees of violence), but we know whether a police officer used violence or not. Hence, we will consider the probability that violence is used against an individual of group g. We again assume that researchers observe violence only if the level is high enough to be recorded: $v_g^* > 0$. We denote $\overline{\Pi}_g^+$ the probability that violence is recorded in a stop of an individual of group $g \in \{m, M\}$. To limit the number of cases, we continue to assume that $\alpha < 1 - 2\pi_m^v$.

(a) Explain why $\overline{\Pi}_M^+ = \pi_M^v q^v$.

(b) Show that if $b_m^v = 0$ and $\pi_m^v = \pi_M^v$, then $\overline{\Pi}_m^+ = \overline{\Pi}_M^+$. Explain why the probability of violence is more appropriate to uncover discrimination in the use of force.

In what follows, we assume that police officers have some prejudice against members of the minority: $b_m^v > 0$. Further, we assume that this bias satisfies the following properties (to limit the number of cases): $-\alpha < b^v_m - (1 - 2\pi^v_m) < \alpha(1 - 2\pi^c_m)$ and $b^v_m < 2\pi^v_m$, with $\alpha > 0$.

We are interested in recovering the Average Treatment effect on the Treated conditional on a stop. That is, we will focus on the probability of violence, conditional on being stopped, for a minority member in an interaction compared to what would have happened if the member were to be from the majority group. We denote this quantity $ATT_{a=1}$ (in this we follow, Knox, Lowe, and Mummolo's (2020) with a small difference as they use the variable M to denote that an individual is stopped).

(c) Explain carefully why

$$ATT_{a=1} = \frac{\pi^c_m q^c}{\pi^c_m q^c + (1 - q^c)}(1 - q^v) + \frac{1 - q^c}{\pi^c q^c + (1 - q^c)}\pi^v_m q^v \tag{9}$$

Comment: Remember what we did in the lecture to capture the ATT. You take the proportion from the treated group (here treated is a minority individual-officer interaction) and compare the probability of violence in the treatment group and in the control group (the control group is a majority individual-officer interaction).

(d) Show that under the assumptions,

$$\overline{\Pi}^+_m = \pi^v_m q^v + \frac{\pi^c_m q^c}{\pi^c_m q^c + (1 - q^c)}(1 - q^v)$$

Comment: You should use Table 8 to answer this question.

(e) Show that $\overline{\Pi}^+_m - \overline{\Pi}^+_M = ATT_{a=1} - \frac{1-q^c}{\pi^c q^c + (1-q^c)}\pi^v_m q^v + (\pi^v_m - \pi^v_M)q^v$. Give some interpretation for the two last terms. What are their consequences for an empirical analysis using differences in the probability of the (recorded) use of violence?