Using high-frequency location data to evaluate racial bias in policing

Pradhi Aggarwal* Alec Brandon Ariel Goldszmidt Justin Holz
John A. List Ian Muir Gregory Sun Thomas Yu

December 9, 2022

Abstract

Prior research finds that, conditional on an encounter, minority civilians are more likely to be punished by police than white civilians. An open question is whether the actual *encounter* is related to race. Using high-frequency location data of rideshare drivers operating on the Lyft platform in Florida, we estimate the effect of driver race on traffic stops and fines for speeding. Estimates obtained across traditional and machine learning approaches show that, relative to a white driver traveling the same speed, minorities are 24 to 33 percent more likely to be stopped for speeding and pay 23 to 34 percent more in fines. We find no evidence that these estimates can be explained by racial differences in accident and re-offense rates. Our approach provides key insights into the total effect of civilian race on outcomes of interest and highlights the methodological import of combining high-frequency data and machine learning to evaluate critical social issues.

*Researchers Alec Brandon and Justin Holz, with support from the Lyft economics team, which was run by John List and included Pradhi Aggarwal, Ariel Goldszmidt, Ian Muir, Gregory Sun, and Thomas Yu, directed the empirical analyses in this paper. All analyses conducted by the economics team were solely for purposes of this paper and the results were shared with the researchers to form the basis of the conclusions discussed herein. These conclusions are those of the authors in their individual capacities and any errors are their own.

1 Introduction

A large collection of empirical research finds that, conditional on an encounter, police officers are more likely to enforce a law, conduct a search, or use force when a civilian belongs to a racial minority group (West, 2018; Fryer, 2018; Goncalves and Mello, 2021; Hoekstra and Sloan, 2021). Less, however, is known about the unconditional effect of civilian race on encouters with police (Durlauf and Heckman, 2020; Knox et al., 2020). Using a sample randomly selected from over 40 billion observations of driving speed of drivers on the rideshare platform Lyft, we estimate the effect of a driver's race on whether they are stopped by the police and how much they are fined for speeding.

Prior research on speeding finds that, conditional on being stopped, there are substantial effects of driver race on the punishment for speeding. Makowsky and Stratmann (2011) find that minority drivers are 3 to 12 percent more likely to be cited for speeding than white drivers. Anbarci and Lee (2014) and Goncalves and Mello (2021) find that, relative to white drivers, minority drivers are 10 to 20 percent less likely to receive leniency from a ticketing officer. Anwar et al. (2021) finds that, as a result of police behavior, Black drivers are 55 percent more likely to be convicted of a misdemeanor for speeding than white drivers. Yet a fundamental limitation of these studies is that race could also influence the probability of being stopped by police. When officers stop minorities irrespective of speed, but only the worst offending white drivers, analysis conditional on being stopped compares fundamentally different groups, leading to inferential difficulties (Knox et al., 2020).

We overcome this limitation by reviewing driving data from Lyft records in the state of Florida. A key part of our exploration relates to the fact that to operate on the platform, drivers must use a smartphone that can communicate their location in realtime. Combining this information with administrative data on driver race and police stops for speeding, which we were able to secure by filing a Freedom of Information Act request, we can directly measure the effect that driver race has on the probability of being stopped for speeding.

To do so, we consider two strategies for specifying the relationship between speeding, driver race, and other features that might be related to a speeding stop. The first strategy uses fixed effects to describe the features other than race that could be related to a speeding stop. These fixed effects capture the speed a driver travels, the location they were traveling in, and other features of the car and driver. The second strategy uses recent advances in machine learning methods to select the appropriate specification of controls from our measurement of driver speed, location, and other features. Inference with the machine learning model is then conducted on the effect of driver race that accounts for the degrees of freedom used to determine the specification of controls.

Across the two strategies, we find that minority[1] drivers are 24 to 33 percent more likely to receive a speeding ticket for traveling the *exact same* speed as white drivers. These differences amount to minority drivers paying 23 to 34 percent more in fines for the same level of speeding as white drivers. Importantly, both of these differences are highly statistically significant. Further analyses consider whether these racial differences can be explained by police punishing minority drivers more harshly because of differences in re-offense or accident rates. These analyses find no evidence to support these alternative explanations: recidivism rates across minority and white drivers are isomorphic; likewise, accident rates are also statistically indistinguishable.

Our overall findings offer two main contributions to the literature on civilian race and policing.[2] First, our findings speak to prior research on the effects of civilian race on encounters with police. Due to limitations of their data, these studies can only estimate the effect of civilian race on the intensity of an encounter (e.g., fine amount) conditional on an encounter actually occurring (Makowsky and Stratmann, 2011; Anbarci and Lee, 2014; Goncalves and Mello, 2021; Anwar et al., 2021). Our findings provide the first estimates of the extensive and intensive margin effect of civilian race on police encounters for speeding.

[1] In our analysis, we use the term 'minority' to mean Asian and Pacific Islander, Hispanic, or Black drivers.

[2] See Ba and Rivera (2019), MacDonald and Braga (2019), Rim et al. (2019) Devi and Fryer Jr (2020), Holz et al. (2020), Owens (2020), Ba et al. (2021), Owens and Ba (2021), and Adger et al. (2022) for recent papers studying policing.

Second, our use of high-frequency location data provides a proof of concept for future research on civilian race and police encounters that stem from traffic stops. Traffic stops are one of the most common settings of civilian encounters with police (Fryer Jr, 2019). Traffic stops are also the most common first step in encounters that escalate to outcomes that have been widely studied in the literature, such as fines, searches, incarceration, and use of force (Langton et al., 2013; Fryer Jr, 2019). Recent technological advances allow researchers to observe huge samples of high-frequency location data (Chen et al., 2019; Chen and Pope, 2020; Cai et al., 2022). We demonstrate how this type of data combined with modern empirical techniques can be used to overcome a fundamental limitation of prior research on civilian race and police encounters.

Our findings also contribute to the literature on recidivism (Kuziemko, 2013; Mueller-Smith, 2015; Bhuller et al., 2020; Norris et al., 2021; Rose and Shem-Tov, 2021; Goncalves and Mello, 2017). While this literature commonly finds evidence that minorities recidivate at higher rates than non-minorities, we find no racial differences in speeding behavior following tickets. This finding, combined with our evidence on racial selection into police encounters, suggests that measuring recidivism using subsequent convictions conflates re-offense rates with the level of policing; a finding of import for academics and policymakers alike.

Our findings also offer concrete prescriptions for policymakers and business leaders interested in reducing racial inequities. For policymakers, our findings suggest that, relative to police officers, automated technologies such as speeding cameras could help reduce selective enforcement of traffic regulations. For business leaders our findings indicate that race-blind approaches to disincentivizing risky behavior may not be so blind. For example, car insurance rates typically increase when drivers are cited for speeding, but our findings indicate that such citations are not blind to driver race. Accounting for race in the relationship between citations and insurance rates could help diminish the impact of racial differences in the enforcement of speeding regulations.

The remainder of this paper is organized as follows. Section 2 presents background on our data and the Lyft setting. Section 3 presents the results of our analysis of the effect of race on police punishment, Section 4 evaluates the implications of alternative interpretations of our results and Section 5 concludes.

2 Data Construction and Summary Statistics

Our analysis relies on several sources of data. First, we obtain records of traffic violations and accidents in Florida via Freedom of Information Act requests of the Florida Court Clerks and Comptrollers. These records allow us to observe the approximate time of each speeding citation and accident, the associated fine of each citation, as well as the driver's license number of each driver involved.

Second, we obtain information on drivers and their driving speed from Lyft. To operate on the Lyft platform, drivers must submit an application and pass a background check. These applications allow us to observe the name, birthdate, driver's license number, a headshot, and information on each driver's motor vehicle. If an application is approved, drivers can begin offering rides by logging into an application on their smartphone. This application provides high-frequency communications of a driver's location, which allows us to infer their driving speed. Subsequently, we can infer when a driver was stopped, as described in the Appendix. For Florida drivers from August 2017 to August 2020, this dataset contains over 40 billion observations. To make the computation more tractable, we keep all location pings where a driver was cited and randomly sample 0.05% of uncited pings, resulting in nearly 20 million observations.

Third, we obtain information on the speed limits drivers face from the Florida Department of Transportation's (FDOT) Open Data Hub. This data allows us to measure each road segment's speed limit and other road characteristics detailed in the Appendix. In combination with the location pings, the speed limit data allow us to measure periods of speeding.

Fourth, to overcome concerns with misreported race by police Luh (2022), we obtain records that assist in the inference of driver race from the Florida State Election Board. Because Florida is a Voting Rights Act state, voter records include self-reported race. We can link these records to drivers with the information provided to Lyft on driver name, gender, and birthdate. These voter records allow us to directly observe the race of 45.7% percent of drivers in our sample. For the drivers who are not in the voter records, we build a race prediction model as described in the Appendix. The results of this analysis allow us to assign drivers to two groups: white and minority drivers, where minority includes Asian and Pacific Islander, Black, and Hispanic.[3]

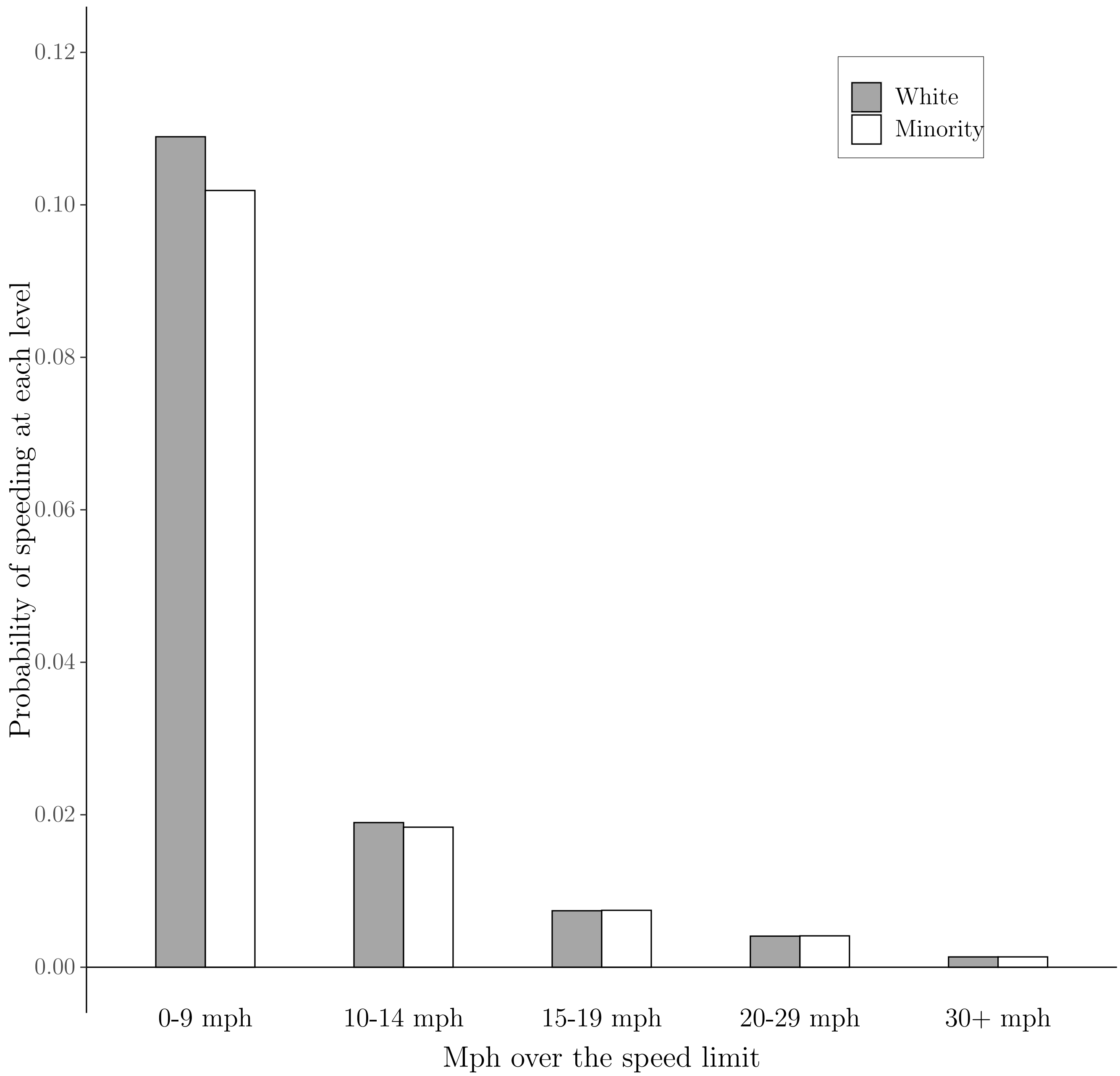
Figure 1 provides a summary of speeding by race. We plot the share of pings in different buckets of driving speed over the limit. Across each bucket, we see that drivers rarely travel more than 10 miles per hour over the legal limit. Furthermore, we see that, relative to the speed limit, white and minority drivers travel similar speeds. This is especially true for speeding more than 10 miles per hour over the limit, which dramatically increases the financial penalty.

3 Results

We use two empirical strategies to estimate the effect of driver race on the extent and intensity of encounters with police. The first uses a battery of fixed effects to hold all else equal when comparing white and minority drivers. We refer to this strategy as the "FE Model". The fixed effects are as follows: we control for time with year, month, day of week, and hour of day fixed effects. We account for geography with fixed effects for whether a driver was operating in a given geohash of approximately 1,200 by 600 meters. These controls allow us to account for effects driven by overpolicing in certain locations. We also account for road features with measures of lane count and log average annual daily traffic.

[3]We limit our analyses to these three groups due to lack of sufficient data on other groups, and we recognize that it may not be exhaustive of all those who identify as persons of color. We also acknowledge that each community has its own unique history and experience of racism in the United States.

Figure 1: Speeding Behavior of Lyft Drivers in Florida by Race Group



Note: This figure plots the proportion of pings transmitted to Lyft in different speeding buckets. Speeding buckets are constructed by comparing the driving speed received by Lyft to the speed limit reported by the FDOT.

To control for features of a car, we use manufacturer and state of vehicle registration fixed effects, as well as controls for vehicle age and vehicle age squared. To account for features of a driver other than race, we use data observed on driver age, age squared, and gender. We cannot use driver fixed effects because speeding citations are rare. We also control for the different phases of driving on Lyft's platform with fixed effects for whether a driver was waiting to be offered a trip, driving to pick up a rider, or providing a ride.

The second strategy we utilize selects the appropriate controls using double machine learning techniques (Chernozhukov et al., 2017). We refer to this strategy as the "DML Model". The identification of the DML model is conceptually similar to that of the FE model and assumes that we have access to a sufficiently rich set of controls to account for all sources of differences in police enforcement other than driver race. However, rather than the controls entering the model linearly, the DML approach uses a machine learning algorithm to automatically learn the specification through which the controls enter into our regression model of stops.

Therefore, relative to the FE model, the DML model can be thought of as implementing the same basic approach as the FE model, but with a more flexible and data-driven specification. All variables that we use for the FE model also enter into our DML model with two exceptions. First, the DML model uses raw coordinates of longitude and latitude rather than fixed effects for each geohash. Second, rather than using hour of day, day of week, and month fixed effects, we supply the ML model with the exact second of week and day of year at which each observation occurred.[4] This allows the DML model to automatically partition space and time based on the data, rather than relying on geohashes and hours, which are inherently arbitrary partitions.

Figure 2 plots our main findings. The top two panels respectively plot the average number of citations and fines for white and minority drivers after accounting for variation explained by the control variables in a given model. The bottom two panels respectively

[4]This can be thought of as implementing a more continuous control for weekly and yearly seasonality. Our temporal controls also continue to include year, as in our specification of the FE model.

plot the difference in citations and fines for white and minority drivers. Across each panel we see that, regardless of the model estimated, minority drivers are more likely to be cited for speeding and pay 23 to 34 percent more in fines. Next we consider additional analyses that can help sharpen the interpretation of these estimates.

4 Alternative Interpretations

Having shown that differences in speeding behavior cannot explain racial inequities in citations and fines, we now consider two potential non-race related explanations for our findings. First, we investigate whether the differences in an officer's treatment of minorities stem from a desire to reduce accidents. To do this, we evaluate whether motorists drive less safely, conditional on speed. Figure 3 examines the relationship between driver race, speed, and accidents. Across each speed bucket we see no statistically significant differences in accident rates between white and minority drivers on Lyft's platform. These results suggest that officers are not ticketing motorists because of differences in minority driving behavior that is not captured by the driver's speed.

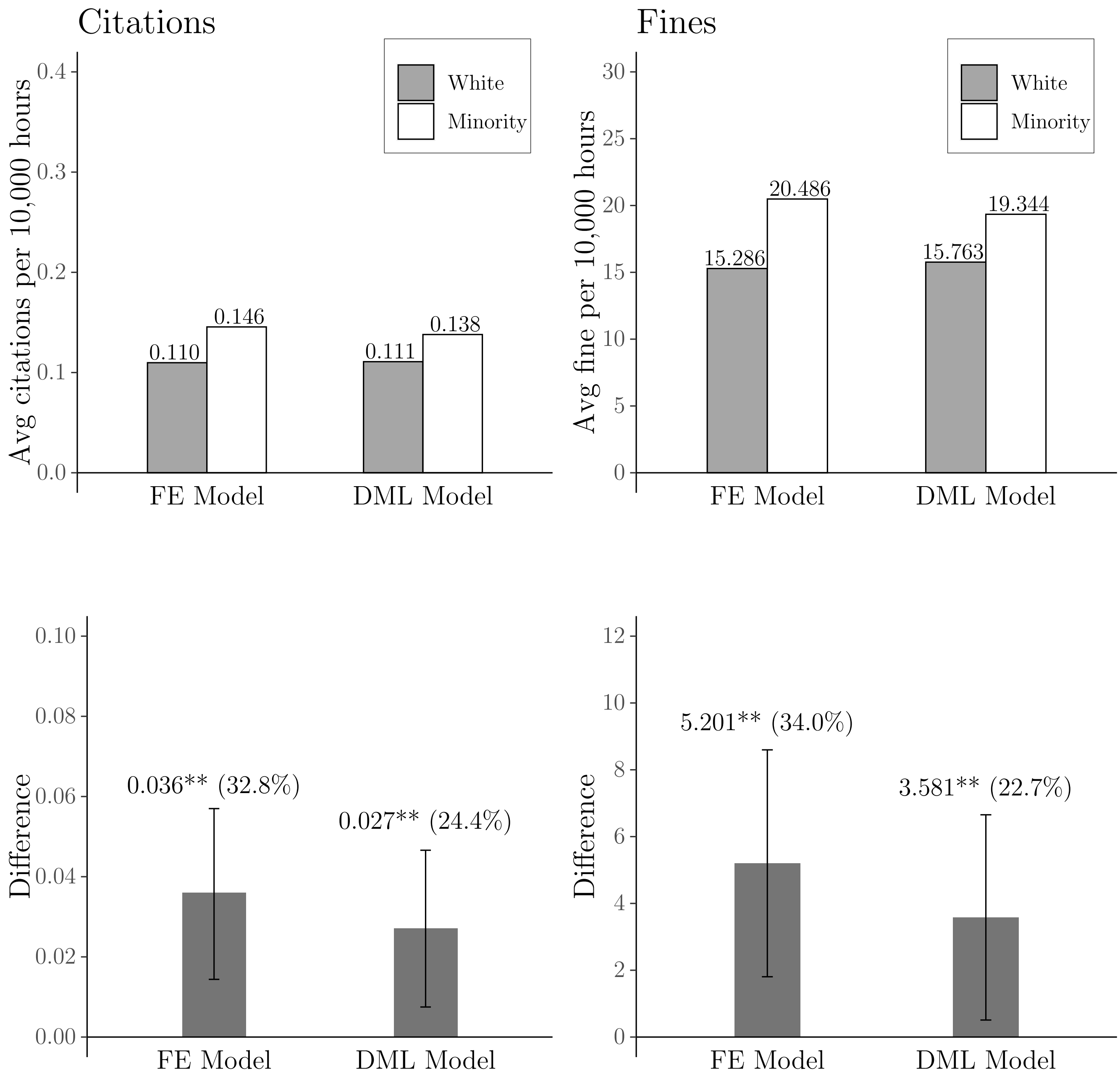
Second, we consider whether officers ticket minorities more than non-minorities because they make decisions based on their expectations of the motorist's future driving behavior. There are two main theories why an officer's beliefs about the future could matter. Punishment reduces the expected marginal social costs of future speeding behavior or seeks to punish those most likely to re-offend (Goncalves and Mello, 2017).[5] Officers may ticket minorities at a higher rate in order to reduce the rate of future crime if they believe that minorities are more responsive to speeding tickets. On the other hand, several papers on the deterrence effect of criminal punishments document that minorities recidivate at higher rates, suggesting that an efficiency motive might be driving the officer's decisions.[6]

We measure re-offense by examining whether a ticket reduces drivers' portion of time

[5] Prior work suggests that punishing speeding motorists can reduce future accidents (Makowsky and Stratmann, 2011; Luca, 2015; Goncalves and Mello, 2017).

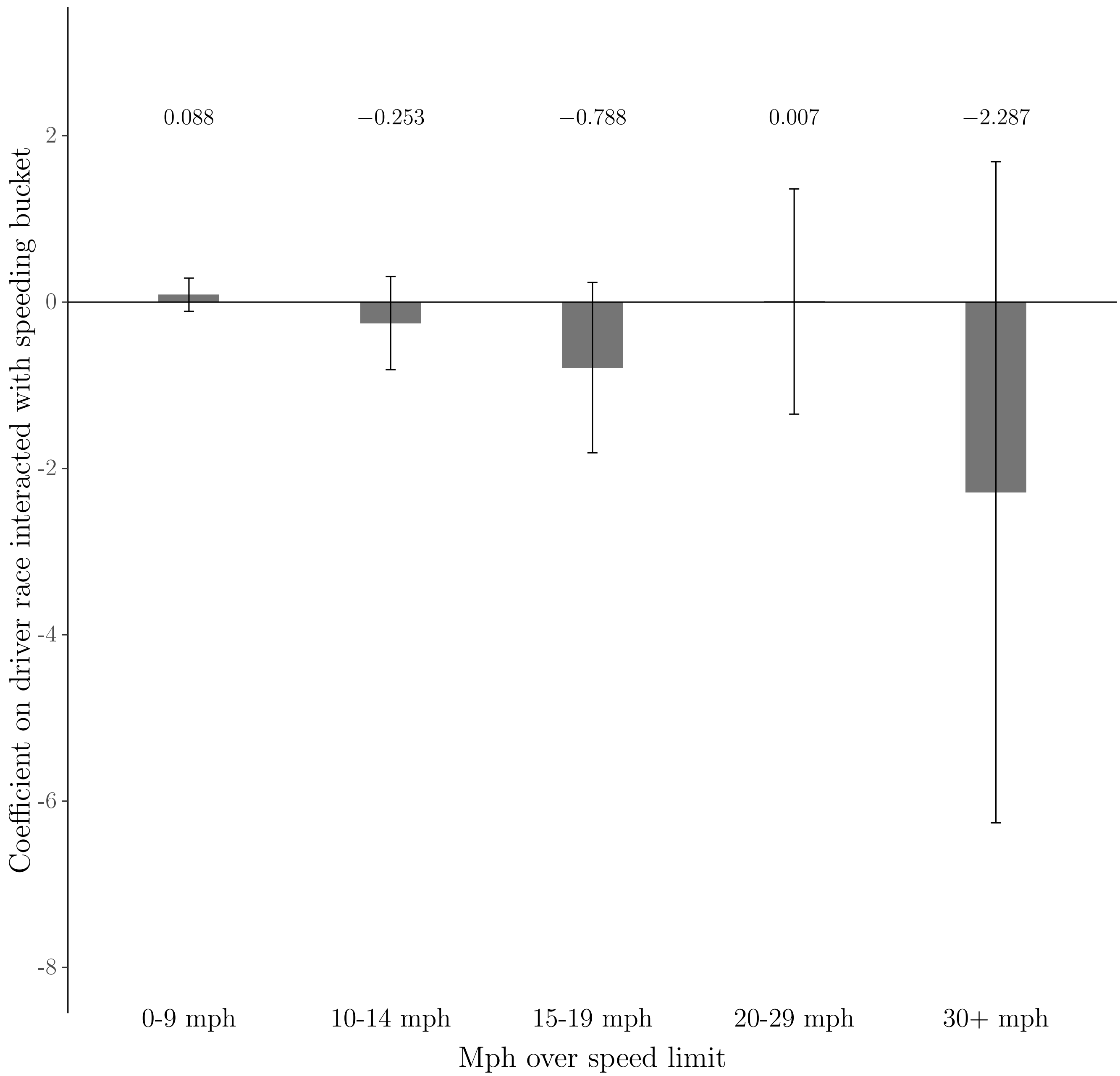
[6] See Kuziemko (2013); DOJ (2018); Goncalves and Mello (2017, 2021).

Figure 2: Effect of Driver Race on Police Enforcement of Speed Limits



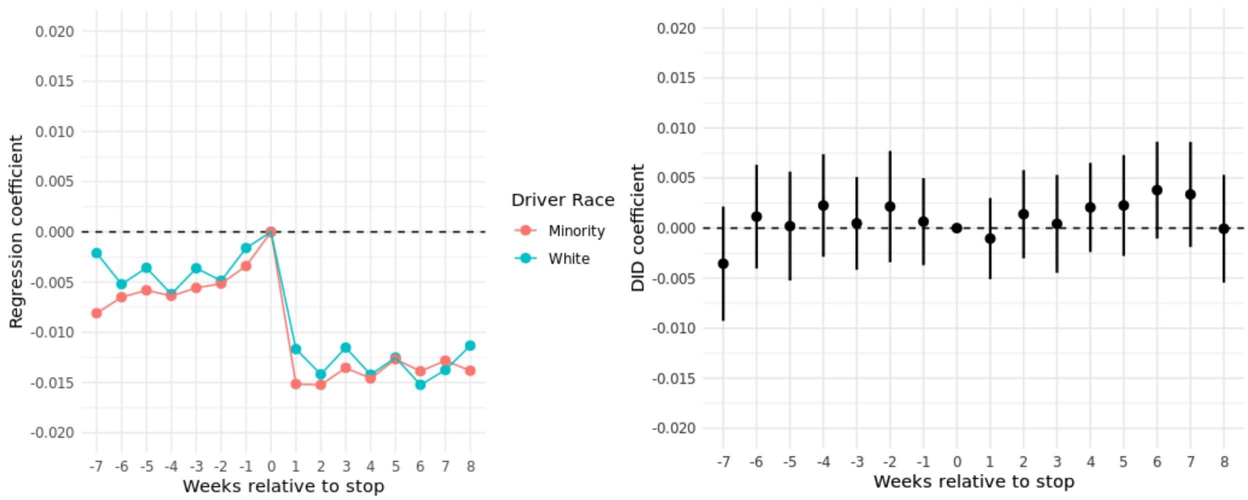
Note: This figure plots the citation and fines associated with each race group of drivers.

Figure 3: Effect of Driver Race on Relationship Between Driving Speed and Accidents



Note: This figure plots the differences-in-differences term for the effect of driver race on accident rates at each speeding bucket, along with 95-percent confidence intervals.

Figure 4: Effect of Citation on Reoffenses



(A) Re-offense by Race

(B) Racial Differences in Reoffenses

driving 10 miles per hour over the speed limit. We estimate these effects using a two-way fixed effects regression model. We now use stop region, which roughly corresponds to county because drivers change their geohash often within a week, but rarely change their region within or across weeks. We also control for driver and vehicle characteristics, the calendar month and year of stop, and the region of stop fixed effects. These results appear in Figure 4.

Panel A of Figure 4 shows that ticketing drivers reduces their speeding behavior in the subsequent eight weeks. However, there are no differences in the portion of time spent speeding by race. Given this evidence, it is unlikely that differences in citations by race arise from officers' responding out of either a deterrence or efficiency motive. Together, these analyses provide no support for alternative interpretations of our main findings. Police cite and penalize minority drivers on Lyft's platform far more than white drivers. These differences cannot be explained by differences in driver behavior.

Moreover, these results highlight a key advantage of high-frequency location data. Previous work (see fn.6) on the deterrence effect of criminal punishments has focused on recidivism, that is, the probability of being punished again for the same crime. Unlike these studies, our data allow us to examine whether punishments reduce *reoffenses* rather than subsequent arrests or convictions. In contrast with the recidivism literature, we find no differences in re-offences by race. The similarity in re-offense rates by race, together with our results from Section 3 suggest that researchers should take caution when using recidivism as a measure of re-offense.

Recidivism conflates civilian behavior with officer behavior. Minorities face higher levels of policing than white drivers, leading to a higher probability of punishment for the same behavior. Thus, higher recidivism rates for minority drivers are more likely to be driven by police behavior than civilian behavior. Moreover, reduced recidivism rates from punishments that change the level of policing an offender receives, such as probation, electronic monitoring, and criminal registries, may be driven by changes in the probability that the crime is detected rather than the probability that the crime is committed.

5 Discussion

Does civilian race influence the extent and intensity of encounters with police? There is tremendous interest in these questions amongst social scientists, policymakers, and the public at large. Yet, prior research only estimates the intensity after conditioning on the existence of an encounter. While such estimates can be informative, the parameters that animate discussions about race and policing are the unconditional effect of civilian race on the extent and intensity of encounters with police: for the same set of behaviors, does race affect the chances of a police encounter?

Using data on the population of active drives on the Lyft platform in Florida, we estimate the unconditional effect of civilian race on stops and fines for speeding. Across two empirical strategies that control for driving speed, we find that drivers who belong to a racial minority group are 24 to 33 percent more likely to be stopped for speeding and pay 23 to 34 percent more in fines. Our complementary analyses on re-offense and accident rates fail to find evidence that these racial differences in the extent and intensity of encounters with police can be explained by non-race related factors.

While our findings are not guaranteed to generalize beyond drivers on Lyft's platform or Florida, this concern can be empirically evaluated by making use of the research design we develop with broader samples of high-frequency location data. Recent research makes use of such samples to examine geographic mobility and racial differences in voting wait time (Chen et al., 2019; Chen and Pope, 2020; Cai et al., 2022). Combining these types of samples with our research design could greatly advance scientific understanding of race effects in policing and provide further justification for policy interventions to ameliorate these effects.

References

Adger, C., M. Ross, and C. Sloan (2022). The effect of field training officers on police use of force.

Anbarci, N. and J. Lee (2014). Detecting racial bias in speed discounting: Evidence from speeding tickets in boston. *International Review of Law and Economics 38*, 11–24.

Anwar, S., P. Bayer, R. Hjalmarsson, and M. L. Mizel (2021). Racial disparities in misdemeanor speeding convictions.

Ba, B. A., D. Knox, J. Mummolo, and R. Rivera (2021). The role of officer race and gender in police-civilian interactions in chicago. *Science 371*(6530), 696–702.

Ba, B. A. and R. Rivera (2019). The effect of police oversight on crime and allegations of misconduct: Evidence from chicago. *U of Penn, Inst for Law & Econ Research Paper* (19-42).

Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020). Incarceration, recidivism, and employment. *Journal of Political Economy 128*(4), 1269–1324.

Cai, W., J. Gaebler, J. Kaashoek, L. Pinals, S. Madden, and S. Goel (2022). Measuring racial and ethnic disparities in traffic enforcement with large-scale telematics data. *PNAS Nexus 1*(4), pgac144.

Chen, M. K., K. Haggag, D. G. Pope, and R. Rohla (2019). Racial disparities in voting wait times: evidence from smartphone data. *The Review of Economics and Statistics*, 1–27.

Chen, M. K. and D. G. Pope (2020). Geographic mobility in america: Evidence from cell phone data. Technical report, National Bureau of Economic Research.

Chernozhukov, V., D. Chetverikov, M. Demier, E. Duflo, and C. Hansen (2017). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Devi, T. and R. G. Fryer Jr (2020). Policing the police: The impact of" pattern-or-practice" investigations on crime. Technical report, National Bureau of Economic Research.

DOJ (2018).

Durlauf, S. N. and J. J. Heckman (2020). An empirical analysis of racial differences in police use of force: A comment. *Journal of Political Economy 128*(10), 3998–4002.

Fryer, R. G. (2018). Reconciling results on racial differences in police shootings. In *AEA papers and proceedings*, Volume 108, pp. 228–33.

Fryer Jr, R. G. (2019). An empirical analysis of racial differences in police use of force. *Journal of Political Economy 127*(3), 1210–1261.

Goncalves, F. and S. Mello (2017). Does the punishment fit the crime? speeding fines and recidivism. *Speeding Fines and Recidivism (October 27, 2017)*.

Goncalves, F. and S. Mello (2021). A few bad apples? racial bias in policing. *American Economic Review 111*(5), 1406–41.

Hoekstra, M. and C. Sloan (2021). Does race matter for police use of force? evidence from 911 calls.

Holz, J. E., R. G. Rivera, and B. A. Ba (2020). Peer effects in police use of force. Technical report, Working Paper.

Knox, D., W. Lowe, and J. Mummolo (2020). Administrative records mask racially biased policing. *American Political Science Review 114*(3), 619–637.

Kuziemko, I. (2013). How should inmates be released from prison? an assessment of parole versus fixed-sentence regimes. *The Quarterly Journal of Economics 128*(1), 371–424.

Langton, L., M. R. Durose, et al. (2013). *Police behavior during traffic and street stops, 2011*. US Department of Justice, Office of Justice Programs, Bureau of Justice ....

Luca, D. L. (2015). Do traffic tickets reduce motor vehicle accidents? evidence from a natural experiment. *Journal of Policy Analysis and Management 34*(1), 85–106.

Luh, E. (2022). Not so black and white: Uncovering racial bias from systematically misreported trooper reports. *Available at SSRN 3357063*.

MacDonald, J. and A. A. Braga (2019). Did post-floyd et al. reforms reduce racial disparities in nypd stop, question, and frisk practices? an exploratory analysis using external and internal benchmarks. *Justice Quarterly 36*(5), 954–983.

Makowsky, M. D. and T. Stratmann (2011). More tickets, fewer accidents: How cash-strapped towns make for safer roads. *The Journal of Law and Economics 54*(4), 863–888.

Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. *Unpublished working paper 18*.

Norris, S., M. Pecenco, and J. Weaver (2021). The effects of parental and sibling incarceration: Evidence from ohio. *American Economic Review 111*(9), 2926–63.

Owens, E. (2020). The economics of policing. *Handbook of Labor, Human Resources and Population Economics*, 1–30.

Owens, E. and B. Ba (2021). The economics of policing and public safety. *Journal of Economic Perspectives 35*(4), 3–28.

Rim, N., R. Rivera, and B. A. Ba (2019). In-group bias and the police: Evidence from award nominations. *U of Penn, Inst for Law & Econ Research Paper* (20-02).

Rose, E. K. and Y. Shem-Tov (2021). How does incarceration affect reoffending? estimating the dose-response function. *Journal of Political Economy 129*(12), 3302–3356.

West, J. (2018). Racial bias in police investigations. *Retrieved from University of California, Santa Cruz website: https://people. ucsc. edu/˜ jwest1/articles/West_RacialBiasPolice. pdf*.

Supplementary Appendix

6 Detailed Description of Data

6.1 Traffic Stop Data

Our data cover the universe of citations written by the Florida Highway Patrol from August 30th 2017 to August 25th, 2020. We make the following restrictions to the data set that reduce the number of observations:

1. The stop is due to a speeding citation.
2. The observation is not missing a driver's license number.
3. The stop occurs to a driver while they are driving and the Lyft application is online.

6.2 Merge to Lyft Data

A unit of observation in the Lyft data set is a driver-phase. There are three potential phases: the driver is searching for a passenger (P1), the driver is driving to a passenger (P2) or the driver is driving with a passenger to the passengers destination (P3). We match each stop with a driver-phase in the following way:

1. We attempt to match the stop to a period such that the recorded stop time occurs within the period start and end times.
2. If we are unable to find a match, we instead find a weaker match where the stop time occurred within 5 minutes of the period start or end time.
3. If neither match is found, we drop the stop.

This process results in 1,423 citations. For each citation, we consider all location pings that occur in the 3 minutes leading up to the citation. We match each citation to a single ping using the following criteria in order of preference:

1. The ping where the road speed limit matches the police reported speed limit
2. The ping with the fastest speed
3. The ping that occurs closest to the estimated stop time

6.3 Accident Data

The information on crash reports during our sample period comes from the Florida Department of Transportation's (FDOT) administrative records of known crashes. FDOT collects this information during a police response or investigation. The data also contains the date and time of the incident, information on injuries to individuals involved in the crash, and an estimated dollar value of property damage.

FDOT identifies the drivers involved in the accident by their driver's license number. During the sample period, we observe 8,243 accidents that we are able to match to Lyft drivers while they are online on the Lyft platform. Note that not all of these drivers receive a speeding citation in our dataset. We use the same process as described in the previous section to match accidents to Lyft location pings.

6.4 Race Prediction Model

We build a race prediction model with four components, drawing from similar models in the literature (Luh, 2021). First, we train a convolutional neural network (CNN) to predict race from profile pictures of Lyft users. These profile pictures are obtained from the 5,531 users in Florida who responded to Lyft's 2019 Economic Impact Report (EIR) survey and reported their race. Second, we train a long short-term memory (LSTM) network using Florida voter records for the general population, excluding drivers on the Lyft platform.

The LSTM network is then fine-tuned on data from the EIR respondents. Third, we train a Bayes classifier using public census data on common last names associated with geographic census blocks and race, as well as information about phone manufacturer, carrier and locale settings from EIR respondents. Finally, we use an XGBoost classifier that combines the predictions from first three models to produce a final race prediction. Overall our model achieves 87.6% accuracy.

6.5 Speed Limit Data

We obtain speed limit and other road characteristics information from FDOT's Open Data Hub. Our models include controls for speed limit, road lane count, and log average annual daily traffic.

7 Descriptive Analysis of Sample and Driving Patterns

Table 1: Sample characteristics

	2017 FL population aged 18+	Cited FL drivers	Lyft drivers	Cited Lyft drivers
Portion female	0.517	0.414	0.289	0.198
	(0.500)	(0.493)	(0.453)	(0.398)
Average age	47.7	37.5	40.4	40.4
		(15.1)	(12.6)	(12.3)
Portion White	0.549	0.665	0.276	0.253
	(0.498)	(0.472)	(0.447)	(0.435)
Portion Black	0.154	0.183	0.252	0.242
	(0.361)	(0.387)	(0.434)	(0.428)
Portion Hispanic	0.247	0.060	0.454	0.482
	(0.431)	(0.237)	(0.498)	(0.500)
Portion Other	0.050	0.093	0.018	0.023
	(0.218)	(0.290)	(0.133)	(0.150)
Portion minority	0.451	0.335	0.724	0.748
	(0.498)	(0.472)	(0.447)	(0.434)
Number of drivers	12,336,038	7,172,871	222,833	1,408

Note: Portion of each race for cited FL drivers is based on the police's definition of driver race. Standard deviation in parentheses.

Table 2: Control variables used

Category		Variable
Driver Behavior	1.	speed over speed limit
Time	2.	year fixed effects
	3.	month fixed effects
	4.	day of week fixed effects
	5.	hour of day fixed effects
Geography	6.	geohash6 fixed effects
Vehicle	7.	vehicle age and age squared
	8.	vehicle make fixed effects
	9.	vehicle state registration = FL
	10.	indicator for missing state registration
Driver	11.	driver age and age squared
	12.	driver sex fixed effects
	13.	indicator for missing driver age
Road	14.	lane count
	15.	log average annual daily traffic (AADT)
	16.	indicator for missing lane count
	17.	indicator for missing log AADT
Lyft phase	18.	phase fixed effects (p1, p2, p3)

Table 3: Summary statistics on speeding behavior

	White	Black	Hispanic	API	Minority	Total
Number of drivers	61,510	56,134	101,240	3,954	161,328	222,838
	(27.6%)	(25.2%)	(45.4%)	(1.77%)	(72.4%)	
Average speed relative to	−22.8	−22.5	−23.9	−24.1	−23.4	−23.2
speed limit	(14.4)	(28.6)	(12.6)	(9.67)	(19.7)	(18.4)
Average portion of time	0.836	0.842	0.868	0.856	0.859	0.853
not speeding	(0.163)	(0.154)	(0.133)	(0.149)	(0.141)	(0.148)
Average portion of time	0.130	0.119	0.100	0.111	0.107	0.113
0-9mph over speed limit	(0.145	(0.134)	(116)	(0.127)	(0.123)	(0.130)
Average portion of time	0.021	0.023	0.018	0.020	0.020	0.020
10-14mph over speed limit	(0.059)	(0.060)	(0.049)	(0.063)	(0.053)	(0.055)
Average portion of time	0.007	0.009	0.008	0.007	0.008	0.008
15-19mph over speed limit	(0.034)	(0.039)	(0.031)	(0.029)	(0.034)	(0.034)
Average portion of time	0.004	0.005	0.004	0.004	0.005	0.004
20-29mph over speed limit	(0.028)	(0.027)	(0.025)	(0.025)	(0.026)	(0.026)
Average portion of time	0.002	0.002	0.001	0.002	0.002	0.002
30+mph over speed limit	(0.024)	(0.020)	(0.016)	(0.024)	(0.018)	(0.020)
Average citations per 10,000 speeding hours	0.141	0.158	0.125	0.177	0.136	0.136
Average fines per 10,000 speeding hours	19.1	20.8	18.5	25.2	19.3	19.2

Note: Standard deviations in parentheses

Table 4: Summary Statistics on control variables

	White	Black	Hispanic	API	Minority	Total
Number of drivers	61,510	56,134	101,240	3,954	161,328	222,838
	(27.6%)	(25.2%)	(45.4%)	(1.77%)	(72.4%)	
Average portion of P1 time	0.519	0.459	0.525	0.511	0.502	0.507
	(0.270)	(0.261)	(0.241)	(0.272)	(0.251)	(0.256)
Average portion of P2 time	0.144	0.158	0.118	0.146	0.133	0.136
	(0.156)	(0.159)	(0.132)	(0.163)	(0.144)	(0.148)
Average portion of P3 time	0.336	0.382	0.355	0.342	0.364	0.356
	(0.228)	(0.227)	(0.210)	(0.230)	(0.217)	(0.221)
Portion of female drivers	0.283	0.387	0.244	0.122	0.291	0.289
Average driver age	44.0	37.5	39.8	40.8	39.0	40.4
	(14.2)	(11.5)	(11.5)	(12.2)	(11.6)	(12.6)
Portion of FL registration	0.906	0.929	0.967	0.921	0.953	0.940
Average road lane count	2.31	2.37	2.41	2.37	2.40	2.37
	(0.385)	(0.380)	(0.341)	(0.388)	(0.357)	(0.367)
Average log avg AADT	10.0	10.1	10.1	10.1	10.1	10.1
	(0.504)	(0.470)	(0.416)	(0.480)	(0.438)	(0.458)
Average vehicle age	5.29	5.46	4.65	5.20	4.94	5.04
	(3.56)	(3.46)	(3.29)	(3.52)	(3.38)	(3.43)

Note: Missing observations are omitted for the respective outcomes
Standard deviations in parentheses

Table 5: Effect of driver race on citations by speeding bucket

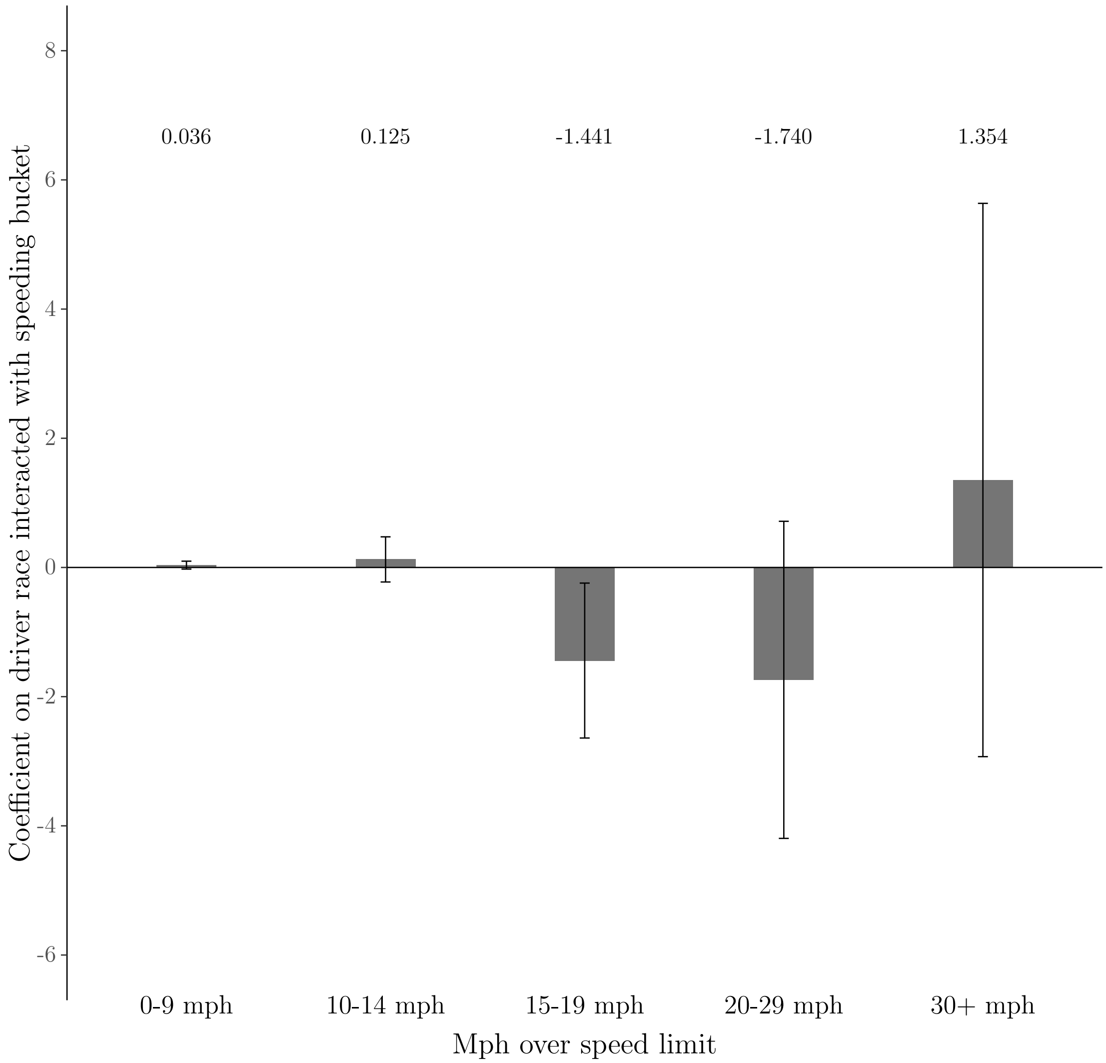
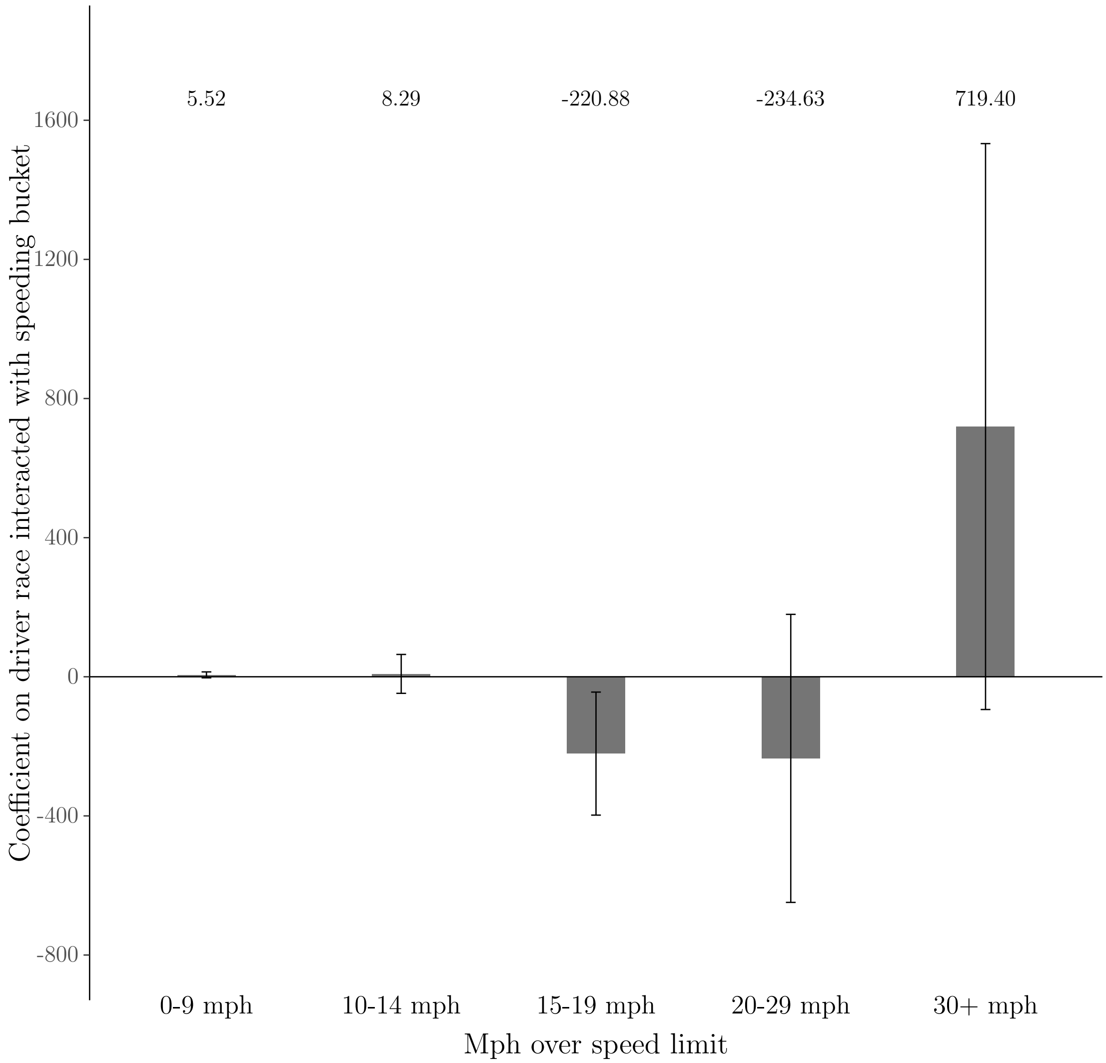


Table 6: Effect of driver race on fines by speeding bucket



8 Robustness Checks

When matching each citation to the GPS ping that is most likely the instance that the driver was caught speeding, we consider all location pings that occur in the 3 minutes leading up to the citation. We rerun our analysis changing this threshold to 1 minute and 5 minutes, respectively, leading up to the citation. Our results are robust to these changes.

Table 7: Robustness check with altered definition of cited ping (FE Model)

Threshold for cited ping		1 minute	3 minutes	5 minutes
Citations	Average citations per 10,000 hours for white drivers	0.108	0.110	0.110
	Average citations per 10,000 hours for minority drivers	0.144	0.146	0.146
	Difference between race groups	0.037**(33.9%)	0.036**(32.8%)	0.037**(33.3%)
Fines	Average fine per 10,000 hours for white drivers	14.9447	15.286	15.217
	Average fine per 10,000 hours for minority drivers	20.382	20.486	20.598
	Difference between race groups	5.437**(36.4%)	5.201**(34.0%)	5.381**(35.4%)

Table 8: Robustness check with altered definition of cited ping (DML Model)

Threshold for cited ping		1 minute	3 minutes	5 minutes
Citations	Average citations per 10,000 hours for white drivers	0.112	0.110	0.112
	Avg citations per 10,000 hours for minority drivers	0.138	0.138	0.138
	Difference between race groups	0.026**(23.5%)	0.027**(24.4%)	0.026**(23.2%)
Fines	Average fine per 10,000 hours for white drivers	15.744	15.763	15.662
	Average fine per 10,000 hours for minority drivers	19.223	19.344	19.346
	Difference between race groups	3.479**(22.1%)	3.581**(22.7%)	3.684**(23.5%)