Contagious Animosity in the Field:
Evidence from the Federal Criminal Justice System[*]

Brendon McConnell   Imran Rasul

January 2018

Abstract

The notion of ingroup-outgroup bias can be extended to multi-identity societies, where social preferences are defined over an ingroup and multiple outgroups. We use a quasi-experimental research design to recover the structure of social preferences across outgroups in a high stakes setting. In particular we investigate whether increased animosity towards Muslims post 9-11 had spillover effects on Black and Hispanic individuals in the Federal criminal justice system. Using linked administrative data tracking defendants from arrest through to sentencing, our core finding is that as 9-11 increased animosity towards Muslims, sentence and pre-sentence outcomes for Hispanic defendants significantly worsened. Outcomes for Black defendants were unchanged. The findings are in line with judges and prosecutors displaying social preferences characterized by *contagious animosity* from Muslims to Hispanics. To understand why increased animosity towards Muslims post 9-11 could spillover onto Hispanics, we draw on work in sociology to detail how Islamophobia and immigration have become gradually intertwined in American consciousness since the mid 1990s, but were most forcefully framed together in the aftermath of 9-11. We underpin a causal interpretation of our findings by providing evidence in favor of the identifying assumptions underlying the research design, and we narrow down the interpretation of the results as being driven by the structure of social preferences using evidence from decomposition analysis, and correlating sentencing differentials to judge characteristics, including their race/ethnicity. Our findings provide among the first field evidence of contagious animosity, that social preferences across outgroups are interlinked and malleable to circumstances. *JEL Classification: D91, J15.*

---

1   Introduction

Minority men are far more likely to come into contact with the Federal criminal justice system (CJS) than White men. Furthermore, decades of research have shown sentencing *outcomes* also vary by race and ethnicity. The central econometric challenge lies in establishing whether sentencing differentials are driven by unobserved heterogeneity correlated to defendant race/ethnicity, or whether they reflect discrimination. The question is of fundamental importance given equality before the law is a cornerstone of any judicial system, and because it is difficult to know whether and how to reduce sentencing disparities if their underlying cause remains unknown.

We advance this literature using three novel pillars of analysis to identify and measure the decisions of judges and prosecutors that determine racial/ethnic differential outcomes. The three building blocks underlying our analysis are, a modification of the notion of ingroup and outgroup bias in societies comprising multiple groups/identities, the construction of a novel research design built around this notion, and the exploitation of linked administrative records tracking defendants through all stages of the Federal CJS, to pin down the decisions of judges and prosecutors that drive sentencing differentials.

A vast literature examines the biological and evolutionary roots of ingroup bias, that is typically used to explain animus towards others. Individuals are assumed to have some social preference over the payoffs to their ingroup, σ_I, and their outgroup, σ_O, where $\sigma_I > \sigma_O$. The standard view is that such preferences are independent and immutable: $\text{corr}(\sigma_I, \sigma_O) = 0$.[1] However, there has been increasing attention on alternative formulations for social preference structures. A nascent body of laboratory evidence shows agents can display contagious altruism, so $\text{corr}(\sigma_I, \sigma_O) > 0$. In this view, positive altruism towards an outgroup fosters altruism towards the ingroup. A second scenario is one of parochial altruism, so that $\text{corr}(\sigma_I, \sigma_O) < 0$. Under this view, greater rivalry between groups fosters more cooperation within the ingroup.[2]

We apply this notion in the context of US society, a multi-identity setting. There is thus one ingroup and multiple outgroups so social preferences are described by $(\sigma_I, \sigma_{O1}, ...\sigma_{OJ})$. We ask does increased animosity towards one outside group $(\Delta\sigma_{Oj} < 0)$ drive social preferences towards another

---

[1]Ingroup bias is often regarded as a central aspect of human behavior whereby individuals aid members of a group they socially identify with, more than members of other groups they do not identify with as strongly [Tajfel *et al.* 1971]. Social psychologists have documented dimensions such as ethnicity, religiosity and political affiliation, as all being salient across contexts, in driving ingroup biases. In economics, ingroup biases have been studied in laboratory settings and show to emerge even in artificially created groups [Shayo 2009]. Bertrand and Duflo [2016] review field evidence on discrimination and ingroup biases in a variety of economic settings.

[2]Contagious altruism has been documented in public goods games in laboratory settings [Fowler and Christakis 2010, Suri and Watts 2011, Jordan *et al.* 2013]. The notion of parochial altruism goes back to Charles Darwin and has gained traction in economics, anthropology, political science and psychology [Alexander 1987, Boyd *et al.* 2003]. Much of this has however relies on self-reports or lab-in-field studies in the context of post-conflict societies [Bauer *et al.* 2016].

outside group ($\Delta\sigma_{Ok}$)?. The answer is no if social preferences across outgroups are independent. On the other hand, there can be *contagious animosity* across outgroups if $\Delta\sigma_{Oj}/\Delta\sigma_{Ok} > 0$, so hostility towards one outgroup drives hostility towards others. Alternatively, there might be *parochial animosity* if $\Delta\sigma_{Oj}/\Delta\sigma_{Ok} < 0$, so hostility towards one outgroup increases altruism towards other outgroups. While the study of ingroup-outgroup biases goes back decades, to the best of our knowledge, there has been little examination of spillover effects across outgroups [Tajfel *et al.* 1971, Bertrand and Duflo 2016]. The notion is important because it implies a malleability of outgroup biases, and that anti-discrimination policies against one outgroup can have positive or negative externalities on other outgroups.

We use the ideas of contagious/parochial animosity to construct a research design to examine racial and ethnic sentencing differentials in the Federal CJS: a high stakes and professional economic environment. This is a setting in which defendants are of multiple identities (by race, ethnicity, and perhaps other dimensions) and as documented later, the vast majority of Federal judges and prosecutors during our study period are White, so we view them as the ingroup. We consider 9-11 as an exogenously timed event that, in general, heightened the salience of insider-outsider differences in US society, and specifically, increased animosity towards Muslims: $\Delta\sigma_M < 0$ [Human Rights Watch 2002, Davis 2007, Woods 2011]. We use this exogenously timed shock towards one outgroup to measure spillover impacts on outcomes in the criminal justice system for other outgroups, namely for Black and Hispanic defendants. We thus identify the following elements in the overall structure of social preferences: $\Delta\sigma_B/\Delta\sigma_M$, $\Delta\sigma_H/\Delta\sigma_M$.

A priori, not all outgroups would necessarily be equally impacted through spillovers induced by the structure of social preferences. In particular, there are reasons why Hispanic defendants are closer to Muslims in social construct than other outgroups. To understand why, we draw on work in sociology to provide a detailed account of how Islamophobia and immigration have become gradually intertwined in American consciousness since the mid 1990s, but were most forcefully framed together in the aftermath of 9-11 [Romero and Zarrugh 2016]. Three channels are identified linking Islamophobia and Hispanics: (i) political rhetoric; (ii) policy; (iii) institutions. We further present quantitative evidence that post 9-11, anti-Hispanic sentiment rose.

We examine the impact of 9-11 on sentencing gaps across races/ethnicities using the Federal Justice Statistics Resource Center (*FJSRC*) dataset. This has two key strengths for the kind of analysis we conduct. First, it covers the universe of all male defendants up for sentencing from 1998 to 2003, so either side of 9-11 and totalling $230,000$ Federal criminal cases. It is nationally representative, covering cases from all 90 mainland US Districts, defendants of all ages, and all types of criminal offense. Such large and representative samples allow for both

Black-White and Hispanic-White differentials to be studied. Second, the *FJSRC* comprises four linked administrative data sources covering the time from a defendant's initial arrest and offense charge, and all subsequent stages of their processing through the Federal CJS, as shown in Figure 1. This linked administrative dataset thus allows pre-sentencing differential treatment arising from the behavior of prosecutors or legal counsel to be studied alongside the behavior of judges at sentencing. Furthermore, it enables us to pin down whether judges and prosecutors display similar kinds of social preference structures across outgroups ($\Delta\sigma_B/\Delta\sigma_M$, $\Delta\sigma_H/\Delta\sigma_M$), and to address long-standing challenges for empirical work on the CJS that is typically based on sentencing data only [Klepper *et al.* 1983].

The *FJSRC* data does not allow direct impacts of 9-11 on Muslim defendants to be studied because the records contain no identifier for religion. Even if they did, there are expected to be very few defendants of Muslim origin in the Federal criminal justice system in our study period.

To isolate the impact 9-11 had on sentencing outcomes, we compare between: (i) defendants who committed their last offense before 9-11 and were sentenced *before* 9-11 (the control group); (ii) defendants who *also* committed their last offense before 9-11, but were sentenced *after* 9-11 (the treated group). We construct a second difference in outcomes across race/ethnicity to estimate a difference-in-difference impact of 9-11 on sentencing outcomes. We base our sample on a ± 180 day sentencing window around 9-11 2001, where *all* defendants have committed their offense prior to 9-11, and hence entered Stage 1 of the Federal CJS timeline in Figure 1, but some were sufficiently far advanced along the timeline so as to come up for sentencing pre 9-11, while others had only just entered the timeline prior to 9-11, and so ended up being sentenced post 9-11.

The period we study is when sentencing guidelines are in place in the Federal CJS. These guidelines provide for *determinate* sentencing, mapping combinations of the severity of the offense and the defendant's criminal history into a specific sentencing range. Table A1 shows the full set of guideline cells. The guidelines do however allow judge's discretion to *downwards depart* from the recommended guideline cell, and so move in a Northerly direction in Table A1. This is the primary outcome of interest when studying judicial decision making, and is an important margin to consider. For example, Mustard [2001] documents that 55% of the Black-White sentencing differential is attributable to differences in downward departure.

Our core results are as follows. We first confirm that relative to Whites, Blacks and Hispanics sentenced pre 9-11 receive significantly longer prison sentences. For Hispanics sentenced post 9-11, sentencing differentials become further exacerbated through a specific channel: they become 13.5% less likely to receive a downward departure than Whites. The implied increase in sentence length for Hispanics is .736 months, corresponding to 18% of the conditional pre 9-11 differential

in sentence length. Placing a monetary value on this increased incarceration suggests the spillover effects from heightened animosity towards Muslims post 9-11 towards Hispanics, led to an increase of \$1547 in incarceration costs per Hispanic defendant. This maps to a large increase in total costs for the Federal CJS given the modal defendant in the study period is Hispanic.

Black-White sentencing differentials around 9-11 are unaffected along all sentencing margins. Overall, the results are consistent with judges displaying contagious animosity from Muslims to Hispanics ($\Delta\sigma_H/\Delta\sigma_M > 0$), while their social preferences are independent between Muslim and Black defendants ($\Delta\sigma_B/\Delta\sigma_M = 0$).

To underpin a causal interpretation, we provide evidence for the identifying assumptions underlying our research design. We first show the time a defendant spends in the CJS between when their last offense is committed and when they come up for sentencing is not impacted by 9-11. Hence there is no evidence of re-sequencing of cases by race/ethnicity post 9-11. Second, using data from other years to construct placebo 9-11 impacts, we show there are no natural race/ethnicity-time effects in sentencing differentials that occur around 9-11 each year. Third we show the estimates are robust to selection on unobservables, ruling out plausible changes in Hispanic-specific unobservable factors post 9-11 that could drive the main finding.

Our data and research design allow us to probe beyond judges' sentencing decisions. As has long been recognized [Klepper *et al.* 1983] a range of legal actors beyond judges are involved in the timeline of Federal criminal cases, and their behaviors: (i) can lead to differential treatment by race/ethnicity pre-sentencing; (ii) such differential treatment might not be detected in sentencing differentials. These concerns are heightened when sentencing guidelines are in place as these restrict the discretion of judges and might increase the power of prosecutors, especially in a system characterized by plea bargaining [Starr and Rehavi 2013]. We use the linked administrative data and our research design to move our 9-11 window to earlier stages of the case timeline on Figure 1, where key decisions by prosecutors are being made.

As with judges, the results on prosecutors' decisions are consistent with them displaying contagious animosity from Muslims to Hispanics ($\Delta\sigma_H/\Delta\sigma_M > 0$) and their social preferences being independent between Muslim and Black defendants ($\Delta\sigma_B/\Delta\sigma_M = 0$). More precisely, Hispanic defendants initially charged post 9-11 are 7.5pp more likely to receive an initial offense that carries a statutory minimum, and their statutory minimum sentence is 10.7 months longer. These impacts correspond to: (i) 60% of the pre 9-11 Hispanic-White gap in the the likelihood of an initial offense charge with a mandatory minimum; (ii) 77% of the pre 9-11 Hispanic-White gap in the statutory minimum sentence length. Indeed, these causal responses to 9-11 lead the Hispanic-

White differential on each margin to become as large as the pre 9-11 Black-White differential.[3]

Having established a causal spillover of 9-11 on Hispanic outcomes in the Federal CJS, our final set of results probe the data to narrow the interpretation of these widening Hispanic-White differentials. The aim is to explore whether the results can be explained through statistical discrimination, or whether they more likely reflect contagious animosity towards Hispanics.

We first present a Juhn-Murphy-Pierce decomposition of sentencing differentials for two cohorts of defendant: (i) those that come up for sentencing post 9-11, where Hispanics are significantly less likely to receive a downwards departures from judges; (ii) those whose initial offense charges are set by prosecutors post 9-11, where Hispanics are charged with offenses with significantly longer statutory minimum sentences. For both cohorts, the JMP decomposition shows Hispanic-White differences are largely driven by changes in unobserved drivers of outcomes; only negligible amounts of each cohort's unconditional DD in outcome can be attributed to either differences in their observables relative to Whites, or the sentencing prices of such observables. This helps to rule out explanations for the increased Hispanic-White differential based on the harshness with which certain offense types are dealt with post 9-11, offender characteristics including those that might perhaps closely predict recidivism such as the guideline cell they are assigned to, or explanations related to effort or allocation of legal counsel to defendants post 9-11. Overall, these decompositions suggest explanations for why Hispanic-White sentencing differentials worsen post 9-11 based on statistical discrimination do not easily fit the evidence.

Second, we analyze how judge characteristics correlate to our estimated Hispanic-White sentencing differential. We code characteristics of Federal judge's by district court, sourced from the *Biographical Directory of Federal Judges*. We document that in districts where there is a higher proportion of Hispanic Federal judges, the post 9-11 Hispanic-White sentencing differential for downward departures is significantly reduced. The fact that judge ethnicity correlates to the Hispanic-White sentencing differential is again *prima facie* evidence against the results being explained by statistical discrimination: if so, then *all* judges, irrespective of their own ethnicity should use defendant ethnicity as a marker for unobservable traits/latent types in determining sentencing outcomes. This is in the spirit of rank order tests used to distinguish statistical discrimination from animus in the literature using data on police arrests or on individual judges [Anwar and Fang 2006, Park 2017].

Both strategies suggest 9-11 had spillover effects on Hispanics through the behavior of judges

---

[3]On prosecutorial biases, Rehavi and Starr [2014] use related linked administrative data from the Federal CJS to show that prosecutor's initial offense charges account for half the Black-White sentencing gap. They do so for the period 2006-8, after sentencing guidelines have been abolished. We provider related results in the pre 9-11 period, when sentencing guidelines are in place.

and prosecutors, with them having social preferences displaying contagious animosity from Muslim to Hispanic outgroups, but independence between Muslim and Black outgroups.

Our analysis provides important contributions to two long-standing literatures: on ingroup and outgroup biases as drivers of human behavior, and on sentencing differentials in the CJS.

We provide among the first field evidence based on a quasi-experimental research design of the existence of contagious animosity. We do so in the high stakes and professional environment of the Federal CJS. Earlier work on sentencing differentials in other parts of the CJS has explicitly or implicitly framed the issue in terms of ingroup and outgroup biases [Bushway and Piehl 2001, Shayo and Zussman 2011, Abrams *et al.* 2012, Anwar *et al.* 2012, Rehavi and Starr 2014]. By allowing for multiple outgroups and developing the notion of contagious/parochial altruism, our work has the important implication that in multi-group societies, effective anti-discrimination policies targeting one group can have positive externalities onto other minority groups. Our analysis also helps address an appeal made in recent overviews of the economics of discrimination literature on the need to better bridge to the psychology literature on the origins of discriminatory behavior [Charles and Guryan 2011, Bertrand and Duflo 2016].[4]

The literature has studied three sources of racial/ethnic sentencing differential [Fischman and Schanzenbach 2012]: (i) judicial bias; (ii) prosecutorial bias; (iii) sentencing policies. The linked administrative data we use provides insights on the first two dimensions. We advance the literature by pinpointing the separate roles that judges and prosecutors have in driving the differential treatment of Hispanics in the Federal CJS post 9-11, and explaining the behavior of both through the structure of their social preferences across multiple outgroups.

The paper is organized as follows. Section 2 describes the Federal CJS, sentencing guidelines, and administrative data. Section 3 presents motivating evidence on long standing pre 9-11 sentiments against Hispanics, and then builds an evidence base to argue how 9-11, Islamophobia and immigration issues all became interlinked in the aftermath of 9-11. Sections 4 and 5 present our core findings on sentencing differentials, as driven by judges and prosecutors decision making respectively. Section 6 narrows the interpretation of increased sentencing differentials post 9-11 using decomposition analysis and judge characteristics. Section 7 concludes. The Appendix contains further data details and robustness checks.

---

[4]Shayo and Zussman [2011] provide evidence linking judicial outcomes and terrorism in Israel-Palestine. Using data on 1748 judicial decisions in Israeli small claims courts where the assignment to an Arab or Jewish judge is effectively random (and 31% of cases are heard by Arab judges), they find evidence for judicial ingroup bias, and that the extent of bias is associated with terrorism intensity in the vicinity of the court in the year before the ruling. Depew *et al.* [2017] use the random assignment of judges in juvenile court decisions in Louisiana between 1996 and 2016 to document negative ingroup judicial bias – so that Black defendants are more likely to be placed into custody rather than on probation, and receive longer sentences, if facing a Black judge. They suggest various mechanisms through which this might operate. Both papers emphasize ingroup and single outgroup differences.

2 The Federal Criminal Justice System

Criminal cases are filed in Federal court if prosecuted by a Federal agency or related to Federal law. This is a high stakes setting: cases heard in Federal courts tend to be more serious than those in State courts. In 2000 the three most frequent criminal offenses were for drug trafficking (40%), immigration (22%), and fraud (9%). Sentencing severity is harsher in Federal court: 88% (75%) of those convicted in Federal (State) court receive a custodial sentence, with the mean sentence being 67 (48) months in Federal (State) court.[5]

The primary legal actors determining outcomes in Federal criminal cases are judges, prosecutors and legal counsel. Judges in Federal courts are Presidential nominees, confirmed by Congress, and life appointees. The prosecution of Federal criminal cases in each of the 94 US District Courts is the responsibility of the US Attorney for that District, who is also a Presidential appointee reporting directly to the Attorney General. There are around 7 Federal judges per district, so close to 700 in total. They are among the most senior judges, and *a priori*, might be considered among those least susceptible to displaying contagious/parochial animosity across outgroups.

In 47% of Federal criminal cases, legal counsel is court appointed. Federal public defenders operate in 32% of cases, and 21% of defendants retain private counsel. This differs from State court cases where 68% of defendants have a public defender. Finally, jury trials in Federal courts occur only if a defendant pleads not guilty. In the Federal CJS this is rare: 96% of defendants plead guilty before they reach trial. By pleading guilty, the individual is convicted and only their sentence remains to be determined. Guilty pleas can be taken into account at sentencing, and such pleas can be Pareto improving for risk averse defendants and prosecutors. By pleading guilty, defendants give up the right to appeal except in capital cases (less than .1% of cases).

2.1 Timeline

Figure 1 shows the timeline for Federal criminal cases, as covered in the *FJSRC* data. Table A2 further details each stage. The first stage a defendant faces after having been arrested and formally charged with a Federal offense (Stage 0) is their initial court appearance where their defense counsel

---

[5]If both Federal and State courts have jurisdiction over a criminal act, prosecutors make case-by-case decisions on which court the defendant will be tried in, although the presumption is that Federal prosectors hold greater sway in such decisions given the greater resources at their disposal [Jeffries and Gleeson 1995]. The sorting of cases into systems is therefore an executive branch decision: judges and defense counsel have no formal role. The DD research design we use to estimate race/ethnic sentencing differentials eliminates cross sectional differences between defendants, by race/ethnicity, being sent to trial in the Federal system. Glaeser *et al.* [2000] provide a theoretical and empirical analysis of the sorting of cases into State and Federal systems. The difference in severity across courts is not driven by the composition of offenses: within offense type there is considerably harsher sentencing in Federal courts, reflecting the greater seriousness of such crimes.

is assigned (Stage 1). Bail is then determined (Stage 2), initial charges are filed by prosecutor's during arraignment (Stage 3), leading to the defendant's initial district court appearance (Stage 4), where they find out which judge they have been assigned to. Pre-trial motions take place at Stage 5, to determine what evidence can be used in trial. The defendant can then offer a plea (Stage 6), where 96% plead guilty, and defendant cooperation can be rewarded by prosecutors. The trial represents Stage 7, and sentencing occurs at Stage 8. In rare cases where a defendant pleads not guilty or for capital cases, they retain the right to appeal (Stage 9).

We first focus on sentencing (Stage 8). As 96% of defendants are already convicted, only their punishment is to be determined. This is where judges exercise discretion. Multiple legal actors are involved at earlier stages, and: (i) their behaviors can lead to differential treatment of defendants pre-sentencing; (ii) the presence of biases earlier in the timeline might not be detected in judicial sentencing differentials. In Section 5 we exploit the linked administrative data to consider earlier stages to pin point how other legal actors drive sentencing differentials, including the initial offense charges of prosecutors that have been shown to play an important role in Black-White sentencing gaps [Rehavi and Starr 2014]. In doing so, we measure whether the sentencing behavior of judges reinforces, offsets or ignores the behavior of prosecutors earlier in the timeline.

2.2   Linked Administrative Data

The *FJSRC* dataset comprises four linked administrative data sources covering the arrest/offense stage before an individual enters the Federal CJS (Stage 0), and all subsequent stages shown in Figure 1. We focus on male defendants so the sample covers $230,000$ Federal criminal cases that come up for sentencing from October 1998 to September 2003 across nearly all US districts. The Appendix provides further data details. To estimate Black-White and Hispanic-White sentencing differentials, we use two variables available at sentencing Stage 8. In the first, defendants are coded as Hispanic (41%) or non-Hispanic (59%). A separate race code then identifies defendants as white-race (71%), black-race (29%), other-race ($< .1\%$). We code Whites as white-race non-Hispanic; Blacks as black-race non-Hispanic; Hispanics as white- or black-race Hispanics. This implies 31% of defendants are White, 26% are Black and 43% are Hispanic.

The data details defendant demographics include age, highest education level, marital status and number of dependents. Legal controls include the type of defense counsel and other pre-sentence variables (such as whether the defendant is in custody), the Federal court district, and we use offense details to classify 31 offense types.[6] Most importantly, the data records the guideline

---

[6]These include kidnaping/hostage taking, sexual abuse, assault, bank robbery (including arson), drugs: trafficking, drugs: communication, drugs: simple possession, firearms: use (including burglary/breaking and auto theft), larceny, fraud, embezzlement, forgery/counterfeiting, bribery, tax offenses, money laundering, racketeer-

cell recommended to the judge in the pre-sentence report. This effectively proxies all case-specific factors the prosecution and legal counsel deem judges should factor into sentencing.

A concern when studying sentencing outcomes is that there can be selection of defendants into this stage of the CJS so the set of cases that reach sentencing might not be representative of the original population of arrested and charged defendants [Klepper *et al.* 1983]. As the *FJSRC* data comprises linked administrative sets covering arrest/offense Stage 0 through to sentencing Stage 8, we can estimate linkage rates for criminal cases across stages. In the Appendix we show that linkage rates are similar by race and ethnicity, and by offense type. The DD research design we use to estimate sentencing differentials eliminates cross sectional differences between defendants of different race/ethnicity (such as in linkage rates).

2.3   Federal Sentencing Guidelines

Federal sentencing guidelines were introduced in the Sentencing Reform Act of 1984 by the US Sentencing Commission (USSC). The goal was to alleviate sentencing disparities. This was to be achieved by the guidelines providing for *determinate* sentencing, whereby the discretion judges had over sentencing became more limited, and parole boards were abolished so that determined sentences matched the actual period of incarceration.

The sentencing guidelines are based on: (i) the severity of the offense; (ii) the defendant's criminal history. To run through a stylized example, an individual who commits a robbery is allocated a base level of 20 points. If a gun is involved an additional 5 points are awarded (if the individual had been a minimal participant in the robbery, 4 points would have been deducted). If the individual was found to be in obstruction of justice, an additional 2 points are awarded. Hence in this case the final score of the defendant on offense severity would be 23 points. There are six criminal history categories, each associated with a range of criminal history points. Criminal history points are based on each prior sentence of imprisonment (and vary with the length of that earlier imprisonment), whether the offenses was committed while under parole/release etc. Suppose the individual in the example above was assessed to have 7 criminal history points. The sentencing guidelines then stipulate they should be sentenced in the range of 70-87 months.

Table A1 shows the full set of guideline cells, mapping each combination of offense severity (1 to 43) and criminal history (scores 1 to 13, grouped into 6 bins) into a sentencing range. There are 43 x 6 = 258 guideline cells. These include those in Zone A, where the guidelines include zero sentence length, and cells in Zone D where the guidelines impose a life sentence.

---

ing (including gambling/lottery), civil rights offenses, immigration, pornography/prostitution, offenses in prisons, environmental, national defense offenses, antitrust violations, food and drug offenses, traffic violations and other smaller categories.

Between trial/conviction and sentencing (Stages 7 and 8), the pre-sentence report is drafted by prosecutors and legal counsel. This recommends a guideline cell. However, the guidelines still provide judges discretion to *downwards depart* from the recommended guideline cell, and move in a Northerly direction in the guideline cell Table A1. A judge can do so if they find mitigating circumstances of a kind not adequately taken into consideration by the USSC in formulating the sentencing guidelines. These circumstances include diminished capacity or rehabilitation after the offense but prior to sentencing, family responsibilities or prior good works. Downward departures may also be warranted "*[i]f reliable information indicates that the defendant's criminal history category substantially over-represents the seriousness of the defendant's criminal history or the likelihood that the defendant will commit other crimes.*" Judges are required to provide written explanations for the specific reason(s) for downward departing.

In our sample, judges grant downwards departure in 17% of cases. This results in a sentence below the original guideline range but they still lead to a custodial sentence in almost 90% of cases. Upwards departures are permitted but occur in less than 1% of cases. Judge-initiated downwards departures are the key sentencing outcome to consider because: (i) such decisions are cleanly attributable to judges; (ii) they are associated with reductions in sentence length.

The null hypothesis for our analysis is based on the USSC sentencing guidelines themselves that state that *"race, sex, national creed, religion and socioeconomic status"*, are factors that *"are not relevant in the determination of a sentence"* [*§5H1.10 of the sentencing guidelines].*[7]

3 Descriptives, 9-11, Research Design

3.1 Pre 9-11 Sentencing Differentials

We first examine pre 9-11 sentencing differentials along three margins of judicial decision making: (i) if a downward departure is granted; (ii) the number of guideline cells moved (including zero, using midpoints of guidelines cells to establish the guideline cell moved to in case of a downwards departure, and using the convention that a Northwards move of one cell corresponds to +1 cells moved); (iii) the sentence length (in months).

In Table 1, Columns 1, 3 and 5 show unconditional racial/ethnic differentials for each outcome. Black-White and Hispanic-White differentials are of statistical and economic significance. We next examine whether these differentials are robust to conditioning on a rich set of covariates including, the demographic characteristics of the defendant described earlier (X_i), the type of legal counsel

---

[7]The guideline cells were in operation until 2005. The Supreme Court's 2005 decision in US v. Booker found them to violate the Sixth Amendment right to trial by jury. The guidelines are now only considered advisory.

(L_i), offense type (OFF_{if}), the guideline cell they are assigned to in the pre-sentence report (G_{ig}), dummies for the Federal court district in which the case is considered (D_{id}), and dummies for fiscal year t, π_t. A key advantage of using the *FJSRC* data is that we can non-parametrically condition on the full set of guideline cells. This effectively proxies all case-specific factors that prosecutors and legal counsel deem judges should factor into their sentencing decision (such as whether a gun was used in the crime, the quality of drugs involved in drug offenses etc.). Such factors would otherwise typically be unobserved.

Columns 2, 4 and 6 show that conditioning on covariates, there are large changes in the Black- and Hispanic-dummy coefficient estimates $(\widehat{\delta}_B, \widehat{\delta}_H)$ on each margin. This is expected given defendants in each group differ on observables pre 9-11, as Table 2 shows. However, even conditional on covariates including the recommended guideline cell, we see that on two out of three margins, statistically significant Black-White and Hispanic-White sentencing differentials remain. Black and Hispanic defendants are significantly more likely to move fewer guideline cells, and have longer sentence lengths. A natural benchmark we use for the later analysis on any spillover impacts of 9-11 on outgroups, is the pre 9-11 sentencing gap, that is around 4 months for both outgroups relative to Whites, or around 10% of the White sentence length.

3.2   Sentiments Towards Outgroups

Our key aim is to understand whether judges and prosecutors in the Federal CJS display social preferences that are characterized by contagious or parochial animosity across outgroups. We do so by exploiting 9-11 as an exogenously timed increase in animosity towards one outgroup: Muslims. To see how this can spillover onto others, we first quantify pre 9-11 sentiments across various outgroups. We use data from the *National Election Survey* (NES). This collects 'thermometer readings' from survey respondents that are asked to report their attachment to various outgroups. Panel A of Figure 2 shows thermometer readings in 2000 expressed by White respondents. We normalize the readings by the sentiment Whites express towards other Whites (i.e. affinity with the ingroup σ_I). We see that nearly all outgroups have a negative thermometer reading ($\sigma_I > \sigma_{Ok}$ for most outgroups k). Most notably, sentiments towards Hispanics are *lower* than towards Blacks ($\sigma_I > \sigma_B > \sigma_H$). Indeed, this ranking across racial/ethnic groups is a well documented long run trend dating in thermometer readings in NES survey data going back to the 1970s [Davis 2007].

9-11 has been much documented to have increased animosity towards Muslims [Human Rights Watch 2002, Davis 2007, Woods 2011]. *A priori*, not all outgroups would potentially be impacted by any resulting contagious/parochial animosity. There are however reasons why Hispanics are closer to Muslims in social construct than other outgroups. To understand the link between 9-

11 and Hispanics, we draw on work in sociology by Romero and Zarrugh [2016]. They provide a detailed account of how Islamophobia and immigration have become gradually intertwined in American consciousness since the mid 1990s, but were most forcefully framed together in the aftermath of 9-11. They build an evidence base for this thesis by analyzing government reports, media accounts, non-governmental evaluations, statements by politicians, and other secondary sources. They argue that Islamophobia – or the extreme and irrational fear of Muslims and Islam – was deployed against Hispanics to garner political support, and justify increased surveillance and immigration enforcement. Romero and Zarrugh [2016] identify three channels linking Islamophobia and Hispanics: (i) political rhetoric; (ii) policy; (iii) institutions.

On political rhetoric, around 9-11 numerous politicians explicitly linked the events to immigration. Issues of security and threats to the nation were tied to immigration and specifically to the US-Mexico border.[8] On policy, immigration and terrorism issues have slowly become intertwined since the 1995 Oklahoma bombings. Two prominent legislative Acts linked immigration and terrorism pre 9-11: the Illegal Immigration Reform and Responsibility Act, and the Anti-Terrorism and Effective Death Penalty Act. Both became law in 1996, linking terrorism and immigration and broadening the set of Federal criminal cases subject to deportation. Post 9-11 the Patriot Act came into effect 45 days later, further increasing the link between terrorism and immigration through its near exclusive focus on immigration offenses. On institutions, the formation of the Department of Homeland Security (DHS) represented the first time terrorism and immigration agencies had been merged. The DHS merged 22 Federal agencies, and as such the culture of the joint bureaucracy changed.

All three channels led to claims that, *"the war on terror quickly turned into the war on immigrants"* [A.D.Romero, Executive Director, American Civil Liberties Union, Liptak 2003].

To provide quantitative evidence on impacts on Hispanics in the immediate post 9-11 period, Panel B of Figure 2 shows time series evidence from a Gallup Poll on immigration: this highlights a marked shift against immigration post 9-11. Panel C shows vandalism victimization rates, by race/ethnicity. The data show a spike in Hispanics reporting being victims of vandalism post 9-11, with the growth rates in victimization rates only slowly returning back to trend. Other studies show 9-11 worsened labor market outcomes for Hispanics [Orrenius and Zavodny 2009].

Taken together, these rhetorical, policy and institutional links between 9-11, immigration and Hispanics, leave open the possibility that outcomes for Hispanic defendants might plausibly be impacted in the aftermath of 9-11 if judges and prosecutors have social preferences across outgroups characterized by contagious/parochial animosity.

---

[8]This linkage occurred despite the thin connection between 9-11 and specific acts of illegal immigration (all the 9-11 hijackers entered the US legally).

3.3 Research Design

To isolate the impact 9-11 had on sentencing outcomes, we compare outcomes between: (i) defendants who committed their last offense before 9-11 and were sentenced *before* 9-11 (the control group); (ii) to defendants who *also* committed their last offense before 9-11, but were sentenced *after* 9-11 (the treated group). We then construct a second difference in outcomes across race/ethnicity to estimate a DD impact of 9-11 on criminal sentencing. Our working sample is based on a ±180 day sentencing window around 9-11 2001, where *all* defendants have committed their offense prior to 9-11, and hence entered the Federal CJS timeline in Figure 1, but some were sufficiently far advanced along so as to come up for sentencing pre 9-11, while others had only just entered the timeline prior to 9-11 and so ended up being sentenced post 9-11. To maintain comparability of both groups we restrict the sample further so that for those defendants sentenced before 9-11, their last offense was committed at least 180 days before 9-11.[9]

The working sample covers $40,228$ cases: 32% of defendants are White, 27% are Black, and 41% are Hispanic. Table 2 shows the exogenous characteristics of each group, for cases up for sentencing pre and post 9-11. The samples are well balanced on these exogenous defendant and legal characteristics. Where there is imbalance, the magnitude of differences is small. Hence given 9-11 was unanticipated, our evidence on contagious/parochial animosity is based on a sample of defendants and offenses that are representative of caseloads in the Federal CJS more broadly.

Figure 3 provides a graphical sense of the research design by plotting histograms of the dates of sentencing and last offense for treatment and control defendants, by group. Focusing first on the ingroup of White defendants in the top panel, the left hand histogram shows sentencing dates to be spread evenly around 9-11 as expected (with the control (treated) group entirely to the left (right) of 9-11). The right hand histogram shows the distribution of last offense dates, by treatment and control. By design, both treatment and control defendants committed their last offense before 9-11, the distribution of last offense dates in treatment and control follow a similar shape, but the distribution for the treated group is right-shifted relative to the control group. The remaining panels in Figure 3 show very similar patterns for sentencing and last offense dates for treated and control defendants in the two outgroups: Blacks and Hispanics.

---

[9]We keep cases in which: (i) guilty pleas are filed (that is so for 96% of defendants); (ii) three or fewer offenses were committed because for offenses that come up for sentencing from 01/10/2001 through to 30/09/2002, as we only observe the date of the first three offenses.

The difference-in-difference empirical specification is given by:

$$
\begin{aligned}
s_{iet} \;=\; & \alpha + \sum_e \delta_e Outgroup_e + \rho Post_t + \sum_e \phi_e (Outgroup_e \times Post_t) \qquad (1) \\
& + \beta X_i + \gamma L_i + \sum_f \omega_f OFF_{if} + \sum_g \gamma_g G_{ig} + \sum_d \lambda_d D_{id} + \epsilon_{iet},
\end{aligned}
$$

where s_{iet} is the sentencing outcome for individual i of outgroup e sentenced on day t based on a ± 180 sentencing day window around 9-11, $Post_t$ is a dummy equal to one if the defendant comes up for sentencing post 9-11, and all covariates (X_i, L_i, OFF_{if}, G_{ig}, D_{id}) are as defined earlier. ε_{iet} is clustered by group-Federal district.

We focus on downward departures as an outcome because: (i) this margin is cleanly attributable to the discretion of judges; (ii) in our research design we have one source of quasi-experimental variation: the exogenous timing of 9-11. Hence we cannot study impacts on judicial decision making *conditional* on being/not being downward departed. Below, we do discuss impacts on sentencing within guideline cells, and in the Appendix we provide descriptive evidence on how additional channels across guideline cells likely further reinforce sentencing differentials beyond those we can provide easily interpretable evidence on via the downwards departure channel.

3.4 Identifying Assumptions and Interpreting ϕ_e

Three assumptions underpin ϕ_e identifying a causal effect of 9-11 on sentencing outcomes for outgroup e. First, the time a defendant spends in the CJS between when their last offense is committed and when they come up for sentencing should not be differentially impacted by 9-11 across groups. This concern is partially ameliorated by the fact that there are proscribed periods of time between each stage of the Federal CJS, and restrictions on how long some stages can take (as shown in Figure 1). The evidence in Figure 3 further points to there being no queue jumping. We formally address the concern using survival analysis to predict the time a defendant spends in the CJS between the dates of last offense and sentencing by group. Second, we require there to be no race/ethnicity-time effects in sentencing differentials that occur naturally around 9-11 each year. We assess this concern using placebo checks using data from earlier years. Finally, we require there to be no missing covariates that determine sentencing outcomes, vary across groups *and* change post 9-11 2001 (but not in placebo years). We address this issue by estimating bounds on the key difference in differences terms accounting for selection on unobservables.

Under these assumptions, ϕ_e still need not be interpretable as reflecting contagious/parochial animosity, $\Delta\sigma_H/\Delta\sigma_M$, $\Delta\sigma_B/\Delta\sigma_M$. ϕ_e might reflect that judges anticipate changes in behavior of

defendants post 9-11, with these expectations differing across outgroups. For example, 9-11 might have altered labor market outcomes for minorities and this can affect recidivism rates differentially across groups; alternatively, judges might anticipate post 9-11 the police will reallocate resources in a way that differentially changes future detection probabilities by race/ethnicity. Taken together, such channels represent different forms of statistical discrimination, where stereotyping of defendants can lead to differential outcomes by race/ethnicity post 9-11, even though all defendants in the sample were already being processed in the Federal CJS by 9-11 2001. Of course, statistical discrimination is not legally permissible because sentencing differentials cannot be justified on the basis of statistical generalizations about group traits, irrespective of whether there is an empirical foundation for this (*JEB vs. Alabama ex rel TB, 511 US 127 1994*).

Given these alternative interpretations of ϕ_e have different welfare implications, we later use two strategies to probe the data to narrow the interpretation of the documented differentials: (i) decomposition analysis, to determine how much of the sentencing differential is attributable to unobservable factors and how much to changing sentencing prices on observables such as offense types and offender characteristics; (ii) correlating race/ethnicity sentencing differentials to judge characteristics, including their race/ethnicity, in the spirit of rank order tests used to distinguish statistical discrimination from animus in the literature using police arrest data [Anwar and Fang 2006, Park 2017].

4   Judges and Sentencing Outcomes

4.1   Downward Departures

Table 3 presents estimates of (1) for downward departures, the key margin of judicial discretion at sentencing. Column 1 shows that Hispanic-White sentencing gaps begin to open up on this margin post 9-11: relative to Whites, the likelihood Hispanics receive a downward departure falls significantly by 3.8pp (13.5%). In contrast, we see no such impact on Black defendants, on whom the post 9-11 impact for downward departures is a precisely estimated zero (and as shown at the foot of the Column, this is significantly different to the post 9-11 impact on Hispanics, $p = .042$). Recall that as shown in Table 1, no race/ethnic differentials existed pre 9-11 in rates of downward departure. The Hispanic-White sentencing differential only opens up post 9-11. If 9-11 sparked a rise in animosity towards Muslims, this pattern of results across outgroups is consistent with judges displaying contagious animosity from Muslims to Hispanics ($\Delta\sigma_H/\Delta\sigma_M > 0$), while their social preferences are independent between Muslim and Black defendants ($\Delta\sigma_B/\Delta\sigma_M = 0$).

Judges have to provide an explanation for downward departures: Columns 2 to 5 code these

into broad categories. The differential impact on Hispanics is driven by judges being less likely to downwards depart due to: (i) a belief that the criminal history of the defendant is overrepresented; (ii) other reasons. For both types of downward departure, the post 9-11 impact on Hispanics is significantly different from that on Blacks ($p = .042, .068$). There is no statistically significant shift in downward departures related either to plea bargains, or due to general mitigating circumstances.

We can straightforwardly convert these impacts on the propensity to downward depart into an implied change in expected sentence length. Denote the probability of being assigned to guideline cell g as p_g, the probability of being downward departed as p_d, and the expected sentence conditional on being sentenced within the range of guideline cell g as $E[s|g]$. The implied change in expected sentence length is,

$$\sum_g p_g \Delta p_d \left\{ E[s|g-4] - E[s|g] \right\}, \tag{2}$$

where we: (i) use the pre 9-11 empirical distribution of defendants (in a given outgroup) across guideline cells to measure p_g, (ii) assume that an individual moves four guideline cells (to $g - 4$) if downward departed (which is true for the median defendant pre 9-11); (iii) take the cell g midpoint to estimate $E[s|g]$. The foot of Column 1 in Table 3 shows the implied impact on Hispanic sentence lengths to be .736 months, corresponding to 18% of the conditional pre 9-11 Hispanic-White differential in sentence length.[10]

To monetize these sentencing impacts we note that: (i) the marginal annual cost per year of imprisoning a male prisoner is $29,000$ [Congressional Research Service 2013]; (ii) in the Federal system, the elasticity of incarceration with respect to sentence $\simeq .87$ [Rehavi and Starr 2014]. Combining these with our implied sentence impact suggests that 9-11 lead to an increase of 1547 in incarceration costs per Hispanic defendant, mapping to a large increase in total costs of the Federal CJS given that 40% of all defendants are Hispanic.[11]

---

[10]The formula for the implied sentence length impact is justified given the downward departure impact on Hispanics occurs across Regions of the guideline cell table in Figure A1. The impact for Hispanic defendants assigned to Region A (so with relatively low offense severity and criminal history scores) is $-.036$, while for Hispanic defendants in Regions B to D the impact is $-.037$, with both estimates being statistically significant from zero, and significantly different from the post 9-11 impacts on Blacks ($p = .033, .057$ respectively).

[11]Mueller-Smith [2016] uses over 2.6mn criminal cases in Texan State court data linked to individual administrative records on time in jail, unemployment insurance, public assistance benefits as well as on future criminal behavior, to estimate the total social cost generated by one year of incarceration to be between $56,000$ and $66,000$. This is double the figure we have used. An alternative benchmark is how sentencing differentials in the Federal CJS have been impacted by institutional reforms. For example, sentencing guidelines were abolished in 2005 following the Supreme Court's decision in US v. Booker. There is mixed evidence on what impact this abolition had on sentencing differentials. Fischman and Schanzenbach [2012] report no effects, while Yang [2015] uses individual matched judge and defendant data, finds Black sentences rise by two months as a result. Hence the magnitude of our main effect arising from contagious animosity corresponds to just over one third of this. Much of the sentencing boom in the State CJS has been attributed to moves towards determinate sentencing, which has been argued to more negatively impact outcomes for Blacks [Neal and Rick 2015].

The analysis so far conditions on two covariates determined during the case timeline: the offense type the defendant is charged with, and the guideline cell they are recommended to be placed in. This replicates earlier work in economics on sentencing outcomes, so conditional on all information available to judges at the point they make their key decision. An alternative approach, following Rehavi and Starr [2014], and more in line with legal studies, is to only condition on observables determined at the point a defendant enters the Federal CJS. To address this issue we exploit information from the arrest stage of the criminal time line (Stage 0): for the 67% of cases that can be linked back to the arrest stage we condition on over 400 codes corresponding to the precise offense the defendant was originally arrested for (rather than conditioning on the 31 offense type codes or 258 guideline cells based on prosecutor decisions during the timeline). Column 6 shows that conditional on original arrest codes, the Hispanic-White differential post 9-11 on downward departures remains significant, and is larger in absolute value at $-.046$pp. This impact remains statistically different than any post 9-11 impact on Black defendants ($p = .063$) and the implied sentence length impact is .889 months, corresponding to 30% of the conditional pre 9-11 Hispanic-White sentence differential.

Finally, if we consider sentence length directly as the outcome s_{iet}, then estimating (1) measures how 9-11 impacts *within cell* movements in sentence length by race/ethnicity. Doing so we find $\widehat{\phi}_H$ and $\widehat{\phi}_B$ to both be not statistically different from zero (this is the case whether the final or original arrest codes are used). Hence the channel through which sentence differentials are impacted is through the lower propensity to downward depart Hispanic defendants, not differential movements within the array of guidelines cells shown in Table A1.

4.2 Citizenship and Offense Type

There are two obvious reasons why Hispanic-White sentencing differentials might become exacerbated after 9-11, while Black-White differentials remain unchanged, and that have nothing to do with contagious animosity across outgroups. The first is that Hispanics constitute the majority of non-US citizen defendants. Punishments for non-citizens, such as deportation, differ from those available for citizens/resident legal aliens, and these might become harsher for non-citizens post 9-11. If so the Hispanic-White differential would just pick up this differential selection into citizenship status.

Column 1 of Table 4 addresses this concern by allowing the impact of ethnicity to vary between Hispanics citizens (US citizen, resident legal alien) and Hispanic non-citizens (illegal aliens, non-US citizen, status unknown).[12] For both groups of Hispanic, those that are sentenced post 9-11

---

[12]71% of defendants overall are classified as citizens, where 91% of non-citizens are Hispanic, so there is little

are significantly less likely to be downward departed. For Hispanic citizens the impact is a 2.8pp reduction in the likelihood of a downwards departure, corresponding to an implied higher sentence length of 17% of the pre 9-11 Hispanic citizen-White differential. For Hispanic non-citizens the impact is a 4.4pp reduction in downwards departure, an implied sentence length increase mapping to 16% of the pre 9-11 Hispanic non-citizen-White sentencing differential. There is no statistical difference between the two impacts ($p = .269$).

A second reason why Hispanic-White sentencing differentials might increase post 9-11 is that they are more likely to be charged with immigration offenses. If such offenses are more severely punished post 9-11, ϕ_H might just pick up that Hispanics are charged with immigration offenses at a greater rate than others. To address the issue, the remaining Columns of Table 4 split the sample by offense type (drug, immigration, other), while still allowing the impact of ethnicity to vary between Hispanic citizens and Hispanic non-citizens. For immigration offenses the vast majority of defendants in the Federal system are Hispanic (either citizens or non-citizens). Hence when examining those offenses we restrict the sample further to Hispanics only.[13]

We see that for Hispanics post 9-11: (i) Hispanic non-citizens are significantly less likely to receive downward departures for drug offenses (Column 2); (iii) on immigration offenses, there is little robust evidence that Hispanics, either citizen or non-citizens, experience a change in the likelihood of receiving a judicial downward departure, and this remains the case even if we focus exclusively on cases in border states (Columns 3 and 4); (iii) the lower likelihood of downward departures post 9-11 is largely driven by the impact on Hispanic citizens for other offenses (Column 5): these constitute around 40% of all offenses and often relate to firearms.[14]

4.3 Robustness and Support for Identifying Assumptions

Appendix Tables A3 to A5 conduct a battery of robustness checks on our core finding from Table 3. These show the result to be robust to: (i) alternative levels of clustering standard errors; (ii) excluding cases where perhaps because of prosecutor's decision making over the initial offense charges filed (Stage 3 in Figure 1), statutory minima or maxima bind partially over the range set by the guideline cell [Rehavi and Starr 2014]; (iii) estimating (1) separately for each group. We also use the fact the *FJSRC* data contains information on Hispanic origins and race (as described earlier, we combine both variables to construct our measure of ethnicity), to examine whether our

---

sense in splitting Black defendants by citizenship status.

[13] Specific immigration offenses due vary by citizenship though: over 90% of immigration offenses for citizens relate to smuggling, while for non-citizens, the most common immigration offense charge is illegal entry (76%).

[14] In line with our results, Mustard [2001] uses data on $77,000$ Federal criminal cases and documents that the Hispanic-White sentence gap is generated by those convicted of drug trafficking and firearm possession/trafficking.

findings pick up purely racial, rather than ethnic, sentencing differentials.

The Appendix also provides evidence in support of the three identifying assumptions required to interpret ϕ_e as measuring the causal impact of 9-11 on outcomes for outgroup e.

First, the time a defendant spends in the CJS between when their last offense is committed and when they come up for sentencing should not be differentially impacted by 9-11 across groups. To address this, Table A6 first shows the main results to be robust to controlling for time of offense, and Table A7 uses survival analysis to show the time a defendant spends in the CJS between their last offense and when they come up for sentencing is not impacted by 9-11 across groups.

Second, we require there to be no race/ethnicity-time effects in sentencing differentials that occur naturally around 9-11 each year. We address thus concern in using three pieces of evidence. First, we use data from earlier years to construct placebos 9-11 effects to check that there are no race/ethnicity-time effects in sentencing differentials naturally occurring around 9-11 each year. As Table A8 shows, the impact for Hispanics on downward departures only occurs post 9-11 in 2001, and not in earlier years. Second, we address concerns that the impacts we find might are driven by the Patriot Act, that was enacted 45 days after 9-11. Notwithstanding the earlier result that immigration offenses did not drive the main result, to shed further light on the matter we estimate a dynamic specific analogous to (1) that estimates impacts in 15-day windows post 9-11. As summarized in Figure A3, we document how impacts on judicial departures for Hispanics appear post 9-11 and pre Patriot Act, and for offenses unrelated to the Patriot Act. Third, we collate data on the date of confirmation of Bush-appointed US Attorneys (shown in Figure A3), to establish in Table A9 that none of the post 9-11 impacts we measure are driven by the share of time a Federal district spends under a Bush-appointed US Attorneys, that might otherwise signal a change in how the CJS views the trade-off between justice and social protection.

Finally, we require there to be no missing covariates that determine sentencing outcomes, vary across groups *and* change post 9-11 2001 (but not in placebo years). We address this issue by following Altonji *et al.* [2005] and Oster [2017] to estimate *bounds* on the treatment effect of $Outgroup_e$ accounting for selection on unobservables. The results in Table A10 show these bounds on $\widehat{\phi}_e$ to be tight. For them to include zero requires there to be unobserved factors that change for Hispanics post 9-11 that are orders of magnitude more predictive of sentencing outcomes than the covariates in (1), including the full set of guideline cell dummies.

5   Prosecutors and Pre-sentencing Outcomes

Federal prosecutors represent the second crucial actor whose decisions determine defendant outcomes. We extend our research design to examine this sequence of pre-sentence prosecutorial decision making. This enables us to provide insight on whether prosecutors, who around 9-11 were overwhelmingly White, display behaviors towards outgroups that are consistent with the pattern of results found for judges: contagious animosity from Muslims to Hispanics ($\Delta\sigma_H/\Delta\sigma_M > 0$), while social preferences are independent between Muslims and Blacks ($\Delta\sigma_B/\Delta\sigma_M = 0$).[15]

5.1   Initial Offense Charges

Prosecutors first decide the initial offense charge filed against defendants (Stage 3 in Figure 1). In the Federal criminal code, definitions of crimes often overlap, providing prosecutors discretion over initial charges. These charges are crucial because they determine: (i) if statutory minima/maxima sentences bind and take precedence over the guideline cell sentence range; (ii) outside options in plea bargaining (so defendants might plead to a lesser charge to avoid being charged with an offense with a mandatory minimum) [Yang 2016].[16]

In Table 5, we use the pre 9-11 sample to first document, by outgroup: (i) the frequency with which defendants receive an initial charge with a non-zero statutory minimum sentence; (ii) the length of statutory minimum sentence associated with their initial offense (setting initial offense charges without a statutory minimum to zero months).[17] We see that pre 9-11: (i) Blacks

---

[15]A recent study of State prosecutors by the *Women Donors Network* (using individual data assembled by the Center for Technology and Civic Life for 2014) found that: (i) 95% of elected prosector positions are held by Whites; (ii) the majority of states have no elected Black prosecutors. It is thus plausible the vast majority of Federal prosecutors in the early 2000s would have been White. A summary of the findings are available at http://wholeads.us/justice/wp-content/themes/phase2/pdf/key-findings.pdf (accessed May 13th 2016).

[16]Many forms of statutory minima exist and can have precedence over the minimum from the guideline cell. In 15.8% (3.6%) of cases the statutory minimum is above (below) the guideline minimum (maximum). Rehavi and Starr [2014] provide an example of how prosecutor's need to assess the strength of evidence, and characterization of ambiguous facts determine initial offense charges. This relates to the use of firearms in a burglary. If a gun is found in the car that transported a defendant to a burglary, the prosecutor must decide whether to allege the burglary legally qualified as a "crime of violence", that the gun qualified as a firearm, and that the defendant "carried" it "during and in relation to" the burglary. All these factors are necessary to trigger a five year mandatory sentence, and would run consecutively to the burglary sentence. Rehavi and Starr [2014] point out a lenient prosecutor might choose to "swallow the gun" and just charge the burglary. In drug cases, such statutory minima have also led to wide disparities in otherwise similar offenses, e.g. those relating to crack versus powder cocaine.

[17]Our coding of statutory minimum differs from the primary coding in Rehavi and Starr [2014]. They derive minima based on initial offense charges, while we use the realized mandatory minima as recorded from the *MFCS* part of the *FJSRC* data. To gauge the relationship between the two codings, we use the *AOUSC* stage of the *FJSRC* data to create a marker for whether there is a change in offense between the initial charge, and the conviction state using three, increasingly detailed, descriptions of offense: (i) most serious offense category (of which there are 51 distinct values); (ii) most serious offense (204 distinct values); (iii) primary offense charge (1543 distinct values). Of the defendant sample we can match from sentencing back to the arrest data, the coding of offenses was unchanged for 93.4% of cases under definition (i), 88.6% under (ii) and 81.6% under (iii). The qualitative pattern of results

are unconditionally 23.3pp more likely to be charged with an offense with a statutory minimum sentence length (Column 1); (ii) conditional on offender and legal counsel characteristics and Federal district, Blacks and Hispanics are significantly more likely to be charged with offenses with a statutory minimum (Column 2). The magnitudes of these differentials correspond to 76% (57%) increases over the baseline probability for White defendants. To control for the nature of offenses, we next exploit information from the arrest stage of the criminal time line (Stage 0). For the 52% of cases that can be linked back to arrest stage, we condition on a rich set of codes corresponding to the original offense the defendant was arrested for. The result in Column 3 shows that doing so, there remain significant Black-White and Hispanic-White differences in the likelihood of non-zero statutory minimum offense charge being given by Federal prosecutors.

Columns 4 to 6 document these differences in initial offense charges translate into a similar pattern of racial/ethnic differentials pre 9-11 for minimum sentence lengths. Blacks receive charges carrying minimum sentences that are conditionally 22 months longer than Whites, falling to 7.4 months in the subsample of cases linked with arrest offense codes. For Hispanics, prosecutors set initial charges with associated statutory minimums that are 14 months longer (or 63% higher) than for Whites, falling to 7.4 months in cases that can be linked with the arrest offense codes. This confirms that when sentencing guidelines are place, this margin is a key one along which prosecutor's actions determine pre-sentencing differentials across groups.[18]

We use our research design to examine whether 9-11, that increased animosity towards Muslims, had spillover effects on other outgroups in the Federal CJS through prosecutors' decisions and social preferences. To pinpoint the impact of 9-11 on prosecutors' behavior, we consider a narrow window covering a cohort of 3600 defendants *all* of whom entered the Federal system pre 9-11 but had their initial offense charges filed either side of 9-11. Taking the date of last offense to proxy for time of entry into the Federal CJS (Stage 1), we exploit the fact that the system requires defendants in (out of) custody to have their initial offense charges brought within 14 (21) days. This allows us to define two groups of defendant: (i) those whose last offense was committed 29 to 42 (43 to 63) days before 9-11 (depending on whether they are in custody or not) and so whose initial offense charge was determined prior to 9-11 (the control group); (ii) those whose last offense was committed 14 (21) days before 9-11 until the day before 9-11 and so their initial offense change would have been determined just after 9-11 (the treated group). We then estimate a specification analogous to (1) but where the outcomes are: (i) whether the defendant receives an initial charge with a non-zero statutory minimum sentence; (ii) the length of statutory minimum

---

presented in Table 5 and later using initial offense charges are robust to these alternative codings.

[18]Rehavi and Starr [2014] establish using similar linked data that prosecutor's initial offense charges account for half the Black-White sentencing gap in 2006-8, after sentencing guidelines had been abolished.

sentence associated with their initial offense. As before we do not condition on final offense type or the later determined guideline cell.[19]

The results are in Table 6: (i) Hispanic defendants initially charged post 9-11 are 7.5pp more likely to receive an initial offense that carries a statutory minimum corresponding to a 22% increase over the pre 9-11 period (an impact statistically different from that on Blacks, $p = .046$); (ii) their statutory minimum sentence is 10.7 months longer; (iii) there is no evidence that 9-11 impacts prosecutors' initial offense charges filed against Black defendants along either margin ($\widehat{\phi}_B = 0$ in Columns 1 and 2). The magnitude of these responses to 9-11 correspond to: (i) 60% of the pre 9-11 Hispanic-White gap in the the likelihood of an initial offense charge with a mandatory minimum; (ii) 77% of the pre 9-11 Hispanic-White gap in the statutory minimum sentence length. Indeed, these causal impacts of 9-11 leaves the overall post 9-11 Hispanic-White differential on each margin to be at least as large as the Black-White differential.

This pattern of results closely mirrors those found earlier for judges: they are consistent with the structure of social preferences across outgroups for prosecutors being such that there is contagious animosity from Muslims to Hispanics ($\Delta\sigma_H/\Delta\sigma_M > 0$), while their social preferences are independent between Muslims and Blacks ($\Delta\sigma_B/\Delta\sigma_M = 0$).

The next two Columns trace through the judicial sentencing impacts on this cohort of defendants (at Stage 8), thus providing novel evidence on the interlinkage between prosecutor and judge decisions. We compare defendants who all come up for sentencing post 9-11, but vary in whether their initial offense charge was filed pre or post 9-11. For Hispanics initially charged just after 9-11, the higher statutory minimum associated with their charge translates into significantly longer sentences of 9.3 months (and this impact is statistically different from that on Blacks, $p = .030$). The differential pre-sentencing treatment of this cohort of defendants represents additional large additional incarceration costs per defendant that we have not so far measured. The earlier costs were associated with the cohort that come up for sentencing around 9-11 (Table 3) whereas these results imply continuing *longer run* costs relating to the cohort of Hispanic defendants initially charged around 9-11, and come up for sentencing well after 9-11. Column 4 controls for the offense type and guideline cell assigned to. Doing so we find no difference in judicial sentencing outcomes for this cohort in sentence length. This implies conditional on all the information available to judges at sentencing, they do not offset the differential behavior of prosecutor's towards Hispanics around 9-11 with regards to initial offense charges.

---

[19]We remove those whose last offense was committed 15 to 28 (22 to 42) days before 9-11 to avoid mis-classifying individuals. If we try and condition on arrest offense codes, then the combination of a smaller sample and a rich set of arrest codes to control for mean that we lose precision, although the signs of all Post x Hispanic interactions remain as those shown.

5.2 Substantial Assistance

The next important pre-sentence decision of prosecutors is their granting defendants a substantial assistance departure: this can occur at the plea stage of the timeline (Stage 6) and allows Federal courts to refrain from imposing a sentence within the guideline cell range on the basis of substantial assistance provided by the defendant toward the prosecution of others, or in recognition of other forms of significant defendant cooperation. The discretion to file a motion for a substantial assistance departure rests solely with Federal prosecutors: they do not have to give reasons when they exercise discretion (unlike judges), with such decisions not being subject to significant appellate review [Fischman and Schanzenbach 2012]. Once such a motion is made, the sentencing judge determines if such a departure is warranted, and the degree of departure.[20]

Table 5 shows that pre 9-11, conditional on observables (where we condition on offense type but not on guideline cells), both outgroups are significantly less likely than White defendants to receive substantial assistance. In Table 6 we consider the impact of 9-11 on prosecutorial decisions on substantial assistance departures for two cohorts. In Column 5 we track the same cohort as in Column 1, for whom their initial charges were set either side of 9-11. We see that in this sample there are no subsequent impacts on the likelihood prosecutors granting substantial assistance departures. This rules out that the increase in statutory minimum sentence lengths associated with initial offense charges is being undone at a later stage through defendant cooperation in plea bargaining with prosecutors, leading prosecutors to request substantial assistance departures. This might have been so if post 9-11, prosecutors set tougher initial charges to Hispanic defendants with the aim of inducing more information during plea bargaining from them. This result further links back to the earlier evidence on judge's justifications for downward departures: we saw the reduction in downward departures for Hispanics was not being driven by reasons related to plea bargaining post 9-11 (Table 3, Column 3).

---

[20]The sentencing reduction for assistance to authorities is considered independently of any reduction for acceptance of responsibility. If the prosecutor wishes to sponsor a departure from the guideline range based on the defendant's cooperation, they must make a motion under §5K1.1. Such departures are identified in the *FJSRC* data. A departure from a statutory mandatory minimum penalty for cooperation requires a separate motion under 18 USC. §3553(e). These departures are not identified in the data. There has been some disagreement on whether mandatory minimum sentences set limits on the extent of departures. USSC guidelines state that upon motion of the government stating that the defendant has provided substantial assistance in the investigation or prosecution of another person, the court may depart from the guidelines. The appropriate reduction shall be determined by the court for reasons stated that may include, but are not limited to: (i) the court's evaluation of the significance and usefulness of the defendant's assistance; (ii) the truthfulness, completeness, and reliability of any information or testimony provided; (iii) the nature and extent of the defendant's assistance; (iv) any injury suffered, or any danger or risk of injury to the defendant or his family resulting from his assistance; (v) the timeliness of the defendant's assistance.

5.3 Pre-sentence Reports

The third key stage at which prosecutors influence pre-sentence outcomes is between trial and sentencing (Stage 7). In the Federal CJS defendants must come up for sentencing precisely 75 (90) days after trial if they are held in (out of) custody. The *FJSRC* data records whether a defendant is in custody after trial (66% of defendants are in custody), so we can recover the precise trial date for each defendant. We can therefore estimate the impact of 9-11 on outcomes between trial and sentencing: this is a critical period as it is when the pre-sentence report (PSR) is drafted, and it is a stage where the legal counsel of the defendant also has a key role.

To draft the PSR, the defendant's legal counsel first provides information on the defendant's life history to the (neutral) Probation Office. The defendant is then interviewed by a Probation Officer (PO), with defense counsel present. The PO collates information from this interview, forms submitted by the defense, and material provided by prosecutors, to prepare a draft PSR. This is shared with the defense counsel and prosecutors 35 days before sentencing. Either party can make factual/legal objections to the draft within 10 days of receipt. A fortnight before sentencing, the final PSR is presented to the judge. This describes the defendant's background and offense (including the impact on the victim). Most importantly, it reports a determined criminal history score and the offense severity and thus calculates the recommended guideline cell.

We can use our research design to assess whether 9-11 impacted suggested guideline cells in PSRs differently across outgroups. To do so we estimate a specification similar to before but with two changes. First, we split defendants into three groups: (i) those convicted and sentenced before 9-11 (the control group C); (ii) those convicted before 9-11, but sentenced after 9-11 ($T1$); (iii) those convicted and sentenced after 9-11 ($T2$). This three way split provides a clean comparison between the C and $T2$ group, where the latter have their PSR written *entirely after* 9-11. Second, as outcomes we consider the recommendations from the PSR: the criminal history score, the offense severity, and the minimum sentence recommended in the implied guideline cell.

Table 7 shows the results focusing on the comparison between C and $T2$. We find no evidence of differential impacts post 9-11 on criminal history scores for either outgroup. This is reassuring as this dimension of the guideline cell determination is least open to interpretation. On offense severity, we see a slight reduction for Hispanics, but overall for both groups we see no change in the minimum sentence in the guideline cell defendants are placed in. Given null impacts on five out of six margins, we conclude that prosecutor-legal counsel interactions at the PSR stage between trial and sentencing are *not* a major source of differential treatment of defendants by outgroup post 9-11. This suggests increased Hispanic-White sentencing gaps post 9-11 are not due to diminished effort on the part of legal counsel of Hispanic defendants at this stage of the case timeline.

6 Interpretation

The totality of evidence so far has shown a causal spillover impact of 9-11 on outcomes for non-Muslim minority groups in the high stakes and professional environment of the Federal CJS. One interpretation is that the changes in behavior of ingroup judges and prosecutors are driven by their social preference structures over outgroups. In particular, their behavior can be rationalized by them having contagious animosity from Muslims to Hispanics ($\Delta\sigma_H/\Delta\sigma_M > 0$), while social preferences are independent between Muslims and Blacks ($\Delta\sigma_B/\Delta\sigma_M = 0$). We now probe the data further using two very different approaches to rule out alternative interpretations of $\widehat{\phi}_e$.

6.1 Decomposition Analysis

We first present Juhn-Murphy-Pierce decompositions of sentencing differentials to understand whether they are being driven by changes in observables, sentencing prices for those observables, or unobserved factors. We focus on two cohorts of Hispanic defendant for whom 9-11 led to widening sentencing disparities relative to Whites: (i) in the cohort that come up for judicial sentencing just after 9-11, Hispanics are significantly less likely to be downward departed (Table 3); (ii) in the cohorts for whom prosecutors set initial offense charges just after 9-11, Hispanics receive charges with significantly longer statutory minimum sentence lengths (Table 6). The Juhn *et al.* [1993] (JMP) decomposition is implemented by first considering the following sentencing equation for White defendant i sentenced in period T: $s_{iT} = X'_{iT}\beta_T^W + u_T^W\theta_{iT}$, where β_T^W are sentence prices for Whites, θ_{iT} is a standardized residual capturing unobserved determinants of White sentences, and u_T^W is the standard deviation of this residual for Whites in period T. The Hispanic-White sentencing differential in period T is then, $\Delta s_T = s_T^H - s_T^W = \Delta X_T\beta_T^W + u_T^W\Delta\theta_T$. Given our DD research design we take a *second* difference over pre- to post 9-11 time periods ($T = 0$ to $T = 1$):

$$\Delta s_1 - \Delta s_0 = (\Delta X_1 - \Delta X_0)\beta_0^W + \Delta X_1(\beta_1^W - \beta_0^W) + (\Delta\theta_1 - \Delta\theta_0)u_0^W + \Delta\theta_1(u_1^W - u_0^W). \quad (3)$$

The $(\Delta X_1 - \Delta X_0)\beta_0^W$ component, or X-effect, measures the contribution to the DD in sentencing gaps of observables. Our research design ensures this component is small: as confirmed below, and analogous to what was shown in Table 2 that defendant observables are balanced pre- and post 9-11 by group. The $\Delta X_1(\beta_1^W - \beta_0^W)$ component, or β-effect, measures changes in sentencing prices pre- and post 9-11 for observables. For example, some offense types, such as those related to immigration, might be punished more harshly post 9-11 due to changes in expectation over defendant's future recidivism or detection probability. These impacts also capture changes in

the sentencing price of being in each recommended guideline cell. These recommendations embody case-specific information that prosecutors and legal counsel deem relevant for judge's sentencing decisions, such as whether a firearm is used, or for drug offenses, the quality of drugs etc.

While it is well understood that such decompositions do not represent formal tests for statistical discrimination [Charles and Guryan 2011], in our setting the usual concerns related to decomposition analysis for studying discrimination are partly ameliorated because: (i) the DD set-up provides common support in the cross-section of covariates across groups; (ii) the inclusion of guideline cell dummies allows us to capture many case-specific factors driving outcomes than would normally be measurable. With these issues in mind, the combined X- and β-effects can potentially encapsulate multiple channels through which statistical discrimination can operate, or channels through which post 9-11 sentencing might respond for reasons other than the structure of social preferences across outgroups of judges and prosecutors.

The $(\Delta\theta_1 - \Delta\theta_0)u_0^W$ component, or θ-effect, measures the change in Hispanic's position within the White residual sentencing distribution (measured at $T = 0$). Contagious animosity from Muslims towards Hispanics post 9-11 would increase Hispanic's average position in the White residual distribution. Finally, the $\Delta\theta_1(u_1^W - u_0^W)$ component, or u-effect, measures changes in the spread of the White sentencing residual from pre- to post 9-11, holding fixed the post 9-11 ethnic residual gap $\Delta\theta_1$. The θ- and u-effects reflect both social preferences towards outgroups and unobserved characteristics. *A priori* we might expect the θ-effect to predominantly reflect shifts in social preferences because it represents changes in the *position* of Hispanics in the White sentencing residual distribution, while the u-effect captures changes in the *spread* of this residual, that might be more reflective of omitted covariates whose importance changes over time by group.[21]

Table 8 shows the decomposition. In Column 1 we see that for the downward departure differential: (i) only 7% is attributable to observables (Row 4: X-effect + β-effect); (ii) 93% of the differential is due to unobservables (Row 5: θ-effect + u-effect); (iii) among the unobservable components, the θ-effect is by far the more important driver of the unconditional DD in downward departures, namely change in Hispanics' position within the White residual sentencing distribution (measured at $T = 0$) (Row 8); (iv) there is not much evidence of a change in the spread of the White residual: the u-effect is only .006 (Row 9).

Column 2 focuses on the cohort of defendants impacted by prosecutor's initial offense charges. In line with the earlier analysis, we do not control for offense type or guideline cells in (3). Column

---

[21]To check the validity of basing the JMP decomposition for downward departures off a linear probability model, we have also conducted cross-sectional decompositions in the pre- and post 9-11 periods separately, using both a Blinder-Oaxaca decomposition and the Fairlie [2005] extension of such decompositions to non-linear models. Constructing the implied difference-in-difference decomposition from either approach generates very similar conclusions as the JMP decomposition based on the LPM.

2 shows that: (i) based on observables, the Hispanic-White gap would be predicted to *fall* post 9-11 not rise (Row 4: X-effect + β-effect); (ii) unobservable factors entirely drive the Hispanic-White differential and among the unobservable components, the θ-effect is by far the more important driver of the DD in statutory minimum sentence lengths.

Figure 4 summarizes both decompositions, detailing further the X- and β-effects for covariates. For the cohort impacted by judicial decisions at sentencing: (i) as expected given the research design, each X-effect is small; (ii) sentencing prices on socio-demographic characteristics rise, and sentencing prices on Federal districts fall; (iii) sentencing prices on offense type or guideline cells are largely unchanged. For the cohort impacted by prosecutor decisions over initial charges: (i) again as expected given the research design, each X-effect is small; (ii) sentencing prices on the type of defense counsel and Federal districts fall.

The decomposition results suggest that the results for increased Hispanic-White differentials post 9-11 cannot be easily explained by changes in observables or the sentencing prices of those observables. This helps to further rule out explanations for the results based on the harshness with which certain offense types are dealt with post 9-11, offender characteristics including those that might perhaps closely predict recidivism such as the guideline cell they are assigned to, or explanations related to effort or allocation of legal counsel to defendants post 9-11. All this suggests explanations for why Hispanic-White sentencing differentials worsen post 9-11 based on statistical discrimination are not reconcilable with the actual evidence. Rather, the fact that both sentencing differentials are largely driven by the θ-effect, a fixed shift in an unobservable driver of outcomes, the evidence is most easily explained by the social preference structure across outgroups of judges and prosectors, being such that there is contagious animosity from Muslims to Hispanics.

6.2 Judge Characteristics

We next analyze how judge characteristics correlate to our measured sentencing differentials. In Federal court data, even though judges are randomly assigned, judge identifiers are typically unavailable (or only a subset of cases can be linked) because criminal cases considered are more serious and often of national importance. The *FJSRC* administrative data contains no information on judges, and there is no simple way to link judge and defendant identifiers for Federal criminal cases.[22] To make progress we have coded the characteristics of Federal judge's by district, sourced

---

[22]An important relevant exception is Yang [2015], who links individual judge data to Federal cases to examine how racial sentencing differentials are impacted once sentencing guidelines were struck down in *United States vs Booker* in 2005: she finds that increasing judicial discretion in sentence lengths increased average sentence lengths for Black defendants by 2 months. Recall that our core finding implied that the impact of 9-11 on Hispanic sentence lengths was .74 months. Hence the impact arising through social preference structures and contagious animosity around 9-11 is around one third of that arising from a institutional change in sentencing policy.

from the *Biographical Directory of Federal Judges.* This details the race/ethnicity, gender, and seniority of judges in 90 districts d, as well as whether they were appointed under a Democrat or Republican President. As described in the Appendix, we use this to construct judge characteristics at the district level ($\mathbf{J}_d$).

Similar to Guryan and Charles [2011], we then proceed in two steps. First, we estimate (1) allowing for a full set of interactions between each Federal district d and ($Hispanic_e \times Post_t$) to estimate the coefficient of interest: $\phi_{H,d}$. We do so for the likelihood of a downward departure. Figure A5 shows the spatial pattern of sentencing differentials we seek to explain, plotting $\widehat{\phi}_{H,d}$ for each district d. Second, we regress $\widehat{\phi}_{H,d}$ against $\mathbf{J}_d$ and other district characteristics, where observations are weighted by the share of defendants in district d in the working sample that are Hispanic, and robust standard errors are reported. Observations are weighted because the underlying regression from which each $\widehat{\phi}_{H,d}$ is estimated is based on individual observations, and this number varies by district. The weighted mean share of Hispanic (Black) judges in a district is 14% (7%); 17% of judges are women, 28% are of senior status, and 48% are appointed by Democrat Presidents. As there are only on average 7.5 judges per district, small changes in the composition of judges can significantly alter a defendant's probability to be sentenced by a minority judge.[23]

Table 9 shows the second stage results. In Column 1 we only control for judge race/ethnicity. We find that in districts where there are a higher proportion of Hispanic judges, the Hispanic-White sentencing differential, as measured by $\widehat{\phi}_{H,d}$, is significantly smaller. Column 2 shows this is robust to controlling for the seniority, gender, age and appointment characteristics of Federal district judges, as well as the share of the post 9-11 window the district spends under a Bush-appointed US Attorney. This suggests the Hispanic ethnicity of judges is not merely picking up them being Democrat appointees, and consistent with the evidence in Schanzenbach [2005], the presence of Democratic appointed judges has an independent correlation with Hispanic-White sentencing differentials.

Column 3 controls for the population shares of different ethnic groups in the district, as well the change (1990 to 2000) in population proportions for each group in the district. Doing so *increases* the coefficient on the district proportion of Hispanic judges from .200 to .548 (where both are significant at conventional levels) and this partial correlation becomes more precisely estimated. Hence the district proportion of Hispanic judges does not appear to be proxying for population

---

[23]Senior judges are partially retired and have greater discretion over their caseload. An individual becomes eligible for senior status at age 65 if one has served for at least 15 years. Judges are not required to take senior status at eligibility. When a judge elects to claim senior status, their seat opens up and the President can appoint a new judge to the lifetime appointment. Schanzenbach [2005] provides evidence that the absolute number of Hispanic Federal judges has been relatively constant over the period from 1990 to 2002; the rises in the number of Black and female judges are more pronounced.

characteristics of where the case is heard.

To more easily compare across covariates, Column 4 reports effect size estimates of each partial correlation. We see that a one standard deviation in the proportion of judges in the district of Hispanic origin increases $\widehat{\phi}_{H,d}$ by 3.2pp. This effect size is larger than the implied impact on the Hispanic-White sentencing differential of a one standard increase in the share of Democratically appointed judges. The effect size is comparable in absolute magnitude to the average effect across all districts, documented in Table 3 that post 9-11, Hispanic defendants are 3.8pp less likely to receive a downward departure.

The fact that judge ethnicity correlates to the Hispanic-White sentencing differential is *prima facie* evidence against the results being explained by statistical discrimination: if so, then *all* judges, irrespective of their own race/ethnicity should use defendant race/ethnicity as a marker for unobservable traits in determining sentencing outcomes. This is in the spirit of rank order tests used to distinguish statistical discrimination from animus in the literature using data on police arrests or on individual judges [Anwar and Fang 2006, Park 2017].[24] This interpretation is further reinforced by noting that there is little evidence to suggest that more experienced judges are correlated with smaller sentencing differentials (measured either through the senior status of judges or their age). This is counter to the Altonji and Pierret [2001] test of statistical discrimination exploiting the fact that with experience, decision makers learn about the true characteristics of agents and so become less reliant on proxies such as race or ethnicity.[25]

This evidence leaves standing one explanation for our results: the social preference structure of ingroup judges being such that there is contagious animosity from Muslims to Hispanic outgroups.

7 Conclusions

Ingroup bias is regarded as a central aspect of human behavior whereby individuals aid members of a group they socially identify with, more than members of other groups they do not identify with as strongly [Tajfel *et al.* 1971]. Much of this work has emphasized ingroup-outgroup differences, where individuals have some social preference over the payoffs to their ingroup, σ_I, and a single outgroup, σ_O, where $\sigma_I > \sigma_O$. We extend this notion to multi-identity contexts in which social preferences are defined over multiple outgroups $(\sigma_I, \sigma_{O1}, ...\sigma_{OJ})$. We then use a quasi-experimental

---

[24]Such hit-rate tests for racial bias in the context of arrest data have been devised to deal with the non-random selection of individuals into police stops. In our setting, such concerns over the infra-marginality problem of detecting bias are weaker because there is random matching of defendants to judges in the Federal CJS. In consequence, our results are reflective of some form of bias, whether it originates from animus, stereotypes or something else.

[25]To reiterate, a similar analysis cannot be conducted for Federal prosecutors. As with Federal judges, individual data on the ethnicity of Federal prosecutors (or legal counsel) is unavailable. Moreover, as described earlier, there would likely be little variation in prosecutor race/ethnicity to exploit around 2001.

research design around 9-11 to shed new light on the structure of social preferences across out-groups: $\Delta\sigma_{Oj}/\Delta\sigma_{Ok}$. Our research design allows us to investigate whether increased animosity towards Muslims in the aftermath of 9-11 had any spillover effects on Black and Hispanic individuals in the context of the high stakes and professional environment of the Federal criminal justice system. Using linked administrative data tracking defendants from arrest through to sentencing, we investigate the behavioral response to 9-11 of judges at sentencing, and of prosectors at multiple pre-sentencing stages of case timelines in the Federal CJS.

Our core finding is that as 9-11 sparked a rise in animosity towards Muslims, changes in sentence and pre-sentence outcomes for defendants are all in line with judges and prosecutors displaying social preferences characterized by *contagious animosity* from Muslims to Hispanics ($\Delta\sigma_H/\Delta\sigma_M > 0$). Hence increased animosity towards one outgroup has spillover effects of increased animosity towards another outgroup. In contrast, the social preferences of judges and prosecutors are independent between Muslim and Black defendants ($\Delta\sigma_B/\Delta\sigma_M = 0$). We underpin a causal interpretation of these findings by providing evidence in favor of the identifying assumptions that underlie our research design, and we further narrow down the interpretation of the results by ruling out that they are driven by statistical discrimination against Hispanic defendants, using multiple independent strategies.

Most generally, our findings provide among the first field evidence of contagious animosity, that social preferences across outgroups are inter-related and malleable to circumstances. This adds to a nascent body of work examining the endogenous structure of social preferences, that has so far typically been based on self-reported or observational data collected in post-conflict environments [Bauer *et al.* 2016]. An important implication of our findings is that anti-discrimination policies towards one specific outgroup can have important externalities on other outgroups.

More specifically, our findings provide new insights for long-standing research from across disciplines documenting large and persistent outcomes between Whites and minority groups in the criminal justice system. The key challenge has always been to establish whether such differentials are driven by unobserved heterogeneity correlated to defendant race/ethnicity, or whether they reflect discrimination. The question is of fundamental importance given equality before the law is a cornerstone of any judicial system, and because it is difficult to know whether and how to reduce disparities if their underlying cause remains unknown. Our analysis shows in the circumstances around 9-11, sentence and pre-sentence differentials across groups are malleable, and driven by judges and prosecutors having endogenous and interlinked social preferences across minority groups. As such our analysis helps address an appeal made in recent overviews of the economics of discrimination literature on the need to better bridge to the psychology literature on the origins

of discriminatory behavior [Charles and Guryan 2011, Bertrand and Duflo 2016].

On policy implications, our results suggest appointing more Hispanic judges to Federal district courts or as Federal prosecutors, might go some way towards reducing Hispanic-White sentencing differentials. Moreover, the fact we find no evidence of sentencing differentials arising from the drafting of pre-sentencing reports, a stage with the close involvement of the defendant's legal counsel, suggests that increasing accountability or legal counsel involvement at other stages might help mitigate biases. Our results suggest such polices would have impact on Hispanic defendants in periods when there are sharp rises in concerns over terrorism, or animosity towards Muslims for other reasons.

Two directions for future research are clear. First, in keeping with the earlier literature on ingroup bias, we have assumed homogeneity of preferences within groups, so all members of the ingroup have the same social preferences over outgroups. As judges are randomly assigned, our estimates reflect *average* sentencing differentials driven by the behavior of judges and prosecutors. This is in contrast to what is observed in labor market studies of discrimination: one of Gary Becker's key insights was that observed racial wage gap do not reflect the average level of employer discrimination, because minority employees can sort towards the least discriminating employer. If there is a sufficiently large share of minority workers relative to non-discriminating employers, the equilibrium wage gap reflects the tastes of the marginal employer. In our context, the lack of defendant-judge sorting is what leads us to measure average levels of animus.

Yet there is clearly much work to be done to understand within group heterogeneity and correlates of idiosyncratic variation in social preference structures within groups. A promising avenue in this specific context is to build on Yang [2015] and link individual judge data to Federal cases for our sample period. Utilizing such information would help shed light on individual characteristics correlated with the structure of social preferences, and so might have implications for how sentencing disparities could be mitigated through the initial selection or training of Federal judges.

Second, as Figure 2A makes clear, there are many potential outgroups one could define, over which there is a very rich set of social preferences structures to identify. Our work does so for $\Delta\sigma_H/\Delta\sigma_M$ and $\Delta\sigma_B/\Delta\sigma_M$, in one high stakes environment. There is no reason to expect contagious animosity/altruism to characterize all pairs. More broadly, there can be circumstances in which individuals have multiple identities, and other circumstances in which individuals can endogenously choose an identity in anticipation of the kinds of interlinked social preference structures we have documented. This opens up a wide array of research questions at the intersection of the formation of social preferences and the economics of identity.

A   Appendix

A.1   Data Sources

As described in Rehavi and Starr [2014], the four linked data sets are: (i) US Marshals Service (USMS) data, that covers the arrest/offense stage (Stage 0) and includes all persons arrested by Federal law enforcement agencies, persons arrested by local officials and then transferred to Federal custody, and persons who avoid arrest by self-surrendering; (ii) Executive Office for US Attorneys (EOUSA) data, covering initial appearance through to arraignment (Stages 1-3): these data come from the internal case database used by Federal prosecutors, and covers every case in which any prosecutor at a US Attorney's office opens a file; (iii) Administrative Office of the US Courts (AOUSC) data, covering initial district court appearances through to trial (Stages 4-7): these originate from Federal Courts and contain information on all criminal cases heard by Federal district judges, and any non-petty charge handled by a Federal magistrate judge; (iv) the Monitoring of Federal Criminal Sentences (MFCS) data set that is resourced from US Sentencing Commission (USSC), and covers sentencing Stage 8 [USSC MFCS 1999-2003]. This data set collects information on any case that results in conviction and sentencing for a non-petty offense. The Federal Justice Statistics Resource Center (FJSRC) data are then collected by the Bureau of Justice Statistics . We drop 4 out of 94 districts: Guam, Puerto Rico, N.Mariana Island and the Virgin Islands. We focus on male defendants up for sentencing from October 1998 to September 2003. We focus on this period because: (i) before October 1998 the data is less detailed; (ii) from October 2003 sentencing guidelines began to be reformed.[26]

The types of downward departure listed in the USSC sentencing guidelines and coded in the data are: (i) encouraged departure factors (those that take into factors such as coercion or duress, diminished capacity, or aberrant behavior of nonviolent offenders); (ii) discouraged departure factors (such as age, physical condition, family responsibilities, or prior good works); (iii) unmentioned factors that were not adequately considered by the guidelines (such as extraordinary rehabilitation after the offense but prior to sentencing). The last group are the most frequently cited type of downward departures (82% of the total), and this is so for all groups.

The data for judicial characteristics are sourced from the *Biographical Directory of Federal Judges*. To select the relevant judges to construct district-level judge characteristics, we used the data on commission and termination dates for each judge in the database, we restrict the sample to judges commissioned before the end of the working sample and those who terminated the bench

---

[26]More information on the *MFCS* data series can be found at,
http://www.icpsr.umich.edu/icpsrweb/NACJD/series/00083/studies?archive=NACJD&sortBy=7 (accessed 14th April 2016). The *FJSRC* data are available at https://www.bjs.gov/fjsrc/.

after the beginning of the sample.

The data on US Attorneys was sourced from https://www.congress.gov/ for nominations heard by the Senate Committee: Judiciary for the years 2001-2002. The sample consists of all US Attorney confirmations during this time period.

A.2   Linkage Rates

Figure A1 shows linkage rates across the four administrative data sets. We first consider cases observed at sentencing Stage 8, and estimate linkage rates to the *earlier* administrative records, as shown in Panel A of Figure A1 (right-to-left linkage rates). To prevent linkage rates being spuriously lowered due to case truncation, we consider cases up for sentencing in the final year of the *MFCS* data. We see that: (i) 90.2% of cases are also observed in the preceding administrative data (covering Stages 4-7); (ii) 84.7% of cases observed at sentencing can be further linked back to the two earlier administrative data sets (covering Stages 1-7); (iii) 75.1% of cases observed at sentencing can be linked back to arrest/offense stage. Linkage rates are quite similar across groups: 72% of records for White defendants up for sentencing can be linked all the way back to the arrest/offense stage; the corresponding rates for Black (Hispanic) defendants are 70% (81%). For drug offenses linkage rates back to the arrest/offense stage are 74-78% across groups, and for immigration offenses they are 71-85%. The fact that linkage rates are less than 100% implies either: (i) truncation of cases because some cases started before 1998 (our first year of data); (ii) linkage errors arising from the fact the *FJSRC* data originates from multiple agencies.

We next construct linkage rates from the arrest/offense stage through to sentencing, as shown in Panel B of Figure A1 (left-to-right linkage rates). The drawback is that only race is coded in the arrest/offense Stage 0 so when deriving these linkage rates we can only do so for white-race and black-race defendants (92% of those coded as Hispanic at sentencing are white-race). To again minimize linkage rates being spuriously lowered due to truncation, we consider cases where arrest/offense dates occur in the first year of the *FJSRC* data. The underlying administrative set from which the arrest/offense data are collected is from the US Marshals Service data, and this includes all persons arrested by Federal law enforcement agencies, persons arrested by local officials and then transferred to Federal custody, and persons who avoid arrest by self-surrendering. As Figure A1 shows, around 38% of such individuals actually enter the Federal CJS at Stage 1, and this rate is similar for white- and black-race individuals (38-39%). These rates reflect that in the majority of cases, either prosecutors do not pursue any case at all or that individuals are assigned to be tried in State courts. We see higher linkage rates for drug offenses, that do not vary much by race (54-55%), but for immigration offenses, black-race individuals are more likely to enter the

Federal CJS (45% versus 34%). Most importantly though, once an individual enters the Federal CJS at Stage 1, there remains a high linkage rate to the *subsequent* administrative data sets: (i) 84% of defendants in Stage 1 can be traced though to Stage 8; (ii) linkage rates are similar across races (84-86%), and across races for drug offenses (86-88%) and immigration offenses (76-82%).

A.3 Judicial Decision Making: Other Components

An alternative margin on judicial decision making we provide descriptive evidence on is the sentence length for those defendants sentenced within their original guideline cell G: $E[s|DD = 0]$. For defendant i assigned to cell G we define the within-guideline sentence as:

$$ws_{iG} = \frac{s_i - \min(s_{iG})}{\max(s_{iG}) - \min(s_{iG})} \in [0, 1], \tag{4}$$

where the sentence bounds for cell G are $\max(s_{iG})$, $\min(s_{iG})$. We then group ws_{iG} into those at the lower bound of the cell ($ws_{iG} = 0$), those strictly between the lower bound and midpoint ($0 < ws_{iG} < .5$), those at the midpoint ($ws_{iG} = .5$), those strictly between the midpoint and upper bound ($.5 < ws_{iG} < 1$) and those at the upper bound ($ws_{iG} = 1$). Panel A of Figure A2 graphs the density of the unconditional DD around 9-11 by ethnicity of these grouped within-guideline cell sentences, conditional on no downwards departure. We see that relative to White defendants, Hispanics are less likely to be at the lower bound of their guideline cell post 9-11 and more likely to be at the mid-point or upper bound.

Another sentencing outcome we provide descriptive evidence on is $E[s|DD = 1]$. This depends on the number of cells moved conditional on downwards departure as follows:

$$E[s|DD = 1] = \sum_d p(\textit{cells moved} = d)E[s|DD = 1, \textit{cells moved} = d), \tag{5}$$

where $p(\textit{cells moved} = d)$ is the probability that a defendant moves d cells conditional on receiving a downward departure. Panel B in Figure A2 graphs the density of the unconditional DD in cell movements conditional on downward departure, by group, using the convention that a Northwards move of one cell corresponds to +1 cells moved. This reveals that post 9-11 Hispanics are less likely that Whites to move five or more cells, and this mass gets shifted down to moving only two or three cells.

This evidence suggests both channels might be further impacting Hispanics around 9-11.

A.4 Robustness Checks

The main specifications cluster standard errors by group-district and so focus on geographically based unobservables that might be correlated by race/ethnicity. The alternative level of clustering we consider is at the level of week of sentencing x group, so placing more emphasis on *time-related* unobservables being correlated by race/ethnicity for sentencing outcomes. As Column 1 of Table A3 shows, the resulting standard errors are near identical to those in the baseline specification. The second check excludes cases where statutory minima or maxima bind partially over the range set by the guideline cell [Rehavi and Starr 2014]. This occurs in 19% of cases, but the estimated effects in Column 2 follow a similar pattern to those estimated in the main sample.

Table A4 shows the core results to be robust to estimating (1) separately for each group: the signs, significance and magnitude of estimates closely match the pooled specification, with there remaining an implied DD penalty of a 3.4pp reduction in the likelihood Hispanic defendants are granted downward departures if sentenced post 9-11 (Column 3). Table A5 estimates sentencing differentials by race, using a specification analogous to (1) but allow the post 9-11 impacts to vary by race, using the full set of race classifications in the *FJSRC* data. To establish the link between this split and our main specification, defendants previously coded as Hispanics are, in this specification, spread over those coded as white- or black-race, but with 92% of them being white-race. Strikingly, we find no evidence of racial sentencing differentials opening up post 9-11, relative to white-race defendants. Our main results thus point to spillover effects of contagious animosity onto Hispanics, rather than other racial groups or identities. The main document Hispanic-White ethnic sentencing differential is simply masked in this specification within the white-race impacts.

A.5 Evidence in Support of the Identifying Assumptions

A.5.1 Time in the Federal CJS

To underpin a casual interpretation of the results, we first examine the identifying assumption that the time a defendant spends in the Federal CJS between when they commit their last offense and when they come up for sentencing is not impacted by 9-11. Table A6 first addresses this concern by extending specification (1) to additionally control for the defendant's time in the CJS using two approaches: (i) include a series of dummies grouping the time between the last offense and sentence date; (ii) including a series of dummies grouping the last offense date. The main result is robust to using either approach (which is unsurprising given the descriptive evidence in Figure 3). A direct test of this identifying assumption is provided in Table A7 where we use OLS and survival models to estimate the time between last offense and sentencing date, and then test whether this

changes significantly by group, post 9-11. The survival models used are the nonparametric Cox and the log logistic model because it allows for a frailty parameter. Across specifications we find no robust evidence of a change in time defendants spend in the Federal CJS post 9-11, by group (Columns 1a-1c). Nor do we find any evidence of longer processing times for all defendants (the coefficient on $Post_t$ is not different from zero). These findings also hold for specific offense types (Columns 2a-4c).

A.5.2 Time Confounders

The second identifying assumption is that there are no natural time effects in race/ethnic sentencing differentials occurring around 9-11 each year. We use the $MCFS$ data on cases from earlier years (1999 onwards) to estimate placebo 9-11 impacts by race/ethnicity.[27] The results are shown in Table A8. These confirm that there are no natural race/ethnicity-time effects around 9-11. The impact for Hispanics on judicial downward departures only occurs post 9-11 in 2001, and not in earlier years. As shown at the foot of Column 1, taking account of any natural time trends in rates of downward departure for Hispanics occurring in all years, slightly increases the impact of 9-11 on Hispanics relative to our baseline estimate in Table 3: the implied DD impact in 2001 is to reduce judicial departures for them by 5.5pp.

A second time related concern is that a candidate time confounder for our main results is the introduction of the Patriot Act on the 26th of October 2001. This made important changes to how certain Federal offenses were treated (especially those related to immigration and money laundering), and might also have reflected different trade-offs and permanently altered objectives of the Federal CJS post 9-11. Of course the earlier results already documented impacts for non-Patriot Act offenses (such as drug offenses and other non-immigration offenses). However, to further examine how the Patriot Act relates to our earlier results, we estimate a modified specification based on (1) but that further splits the post 9-11 period into 15-day bins. This then gives three estimates on the differential impacts on Hispanic defendants post 9-11 and pre Patriot Act.

The results are shown in Figure A3, the graphs the estimated impact on Hispanics for non-Patriot Act offenses for the first three 15-days bins in the post 9-11 period so before the Patriot Act is introduced (the impacts for immigration offenses were shown earlier in Table 3). Although the estimates are noisy given the smaller sample sizes used to estimate each parameter, we see that each point estimate is negative and close to the baseline estimate (the dashed line).

The third time confounder is that over our sample period, President Bush was appointing

---

[27]The sample of criminal cases used are those $114,642$ cases for which sentencing occurs within a 6-month window of 9-11 in years 1998 to 2001 and: (i) if sentenced after 9-11, the last offense was committed prior to 9-11 each year; (ii) if sentenced before 9-11, the last offense was committed up to 6-months prior to 9-11 that year.

Federal US Attorneys. If such individuals have different preferences or views on the trade-off between justice and social concerns to those predominantly in place pre 9-11, this might in turn drive some of our main effects. Figure A4 shows the date of confirmation for Bush Appointed District Attorneys. As none are appointed pre 9-11, Federal districts spend varying shares of the post period under a Bush-appointed Attorney. In Table A9 we re-estimate our baseline results allowing for the post 9-11 impacts on each group to vary by the share of time the Federal district in which the case is heard spends under a Bush-appointed DA (as measured in deviation from mean). We find no evidence that our main finding on judicial downward departures is heterogeneous along this dimension.

A.5.3 Defendant Behavior and Other Missing Covariates

The third identifying assumption is there are no missing covariates that determine sentencing outcomes, vary across groups *and* change post 9-11 2001 (but not in placebo years). While the evidence presented earlier shows covariate balance pre- and post 9-11 by group, there could always be unobserved factors that changed post 9-11 2001 only for Hispanics. For example, the behavior of Hispanic defendants towards prosecutors and judges might have altered post 9-11, and this could be driving the results rather than contagious animosity of prosecutors and judges from Muslims towards Hispanics post 9-11.

To address whether our core DD results can reasonably be attributed to unobserved covariates correlated to race/ethnicity, we follow methods proposed in Altonji *et al.* [2005] and Oster [2017] to estimate *bounds* on the treatment effect of ethnicity on sentencing allowing for selection on unobservables (SoU). There are multiple potential origins of unobservables driving sentencing outcomes and varying by group (not just those arising from defendant behavior). The bounded treatment effect approach addresses the issue head on by assuming there are potentially many unobserved factors omitted from (1). This set of unobservables is denoted W_2, capturing a linear combination of j unobserved variables w_j^u, multiplied by their coefficients, $W_2 = \sum_{j=1}^{J_u} w_j^u \gamma_j^u$.

Key to this method is an assumption on how the unobserved and observed covariates driving sentencing outcomes relate to each other. Altonji *et al.* [2005] and Oster [2017] assume they relate through a proportional selection relationship where the coefficient of proportionality is denoted τ. It can then be shown that the true causal impact for ethnic group e, δ_e^*, depends on δ (and other factors): $\delta_e^* = \delta_e(\tau, .)$. Bounds on δ_e are then established by considering a range of plausible τ's. At one extreme, if $\tau = 0$ the unobserved covariates do not bias estimation in (1) and $\delta_e^* = \delta_e$. At the other extreme, Altonji *et al.* [2005] and Oster [2017] suggest equal selection ($\tau = 1$) as an appropriate upper bound on τ: intuitively, the set of unobservables cannot be *more* important

than the available covariates in explaining the treatment effect of ethnicity on sentencing outcomes. This is plausible in our context given we observe a rich set of defendant and legal characteristics including the recommended guideline cell. The bounds reported in Table A10 are $\delta_e(0) = \delta_e$ and $\delta_e(1)$, and we also report the coefficient of proportionality τ required for $\delta_e(\tau) = 0$.

The bounds in Column 1 of Table A10 show that allowing for SoU, there remains robust evidence of a post 9-11 Hispanic-White sentencing differential opening up on downward departures ($\delta_e \in [-.038, -.036]$): these treatment effect bounds are very tight. For there to be no Hispanic-White differential, $\tau = 27.7$ is required, so unobservables would need to be many times *more* important in explaining the Hispanic-White differential than the covariates in (1), including the recommended guideline cell. The remaining Columns in Table A10 also confirm tight treatment effect bounds on the main DD estimates for the types of downward departure driving the result: these relate to the judge's view of the classification of the criminal history of the defendant and other reasons, and not related to plea bargaining or general mitigating circumstances. Finally, Column 6 reiterates the core result is robust to SoU when we condition on initial arrest codes rather than those related to earlier decisions of prosecutors.

References

[1] ABRAMS.D, M.BERTRAND AND S.MULLAINATHAN (2012) "Do Judges Vary in their Treatment of Race?," *Journal of Legal Studies* 41: 347-83.

[2] ALEXANDER.R.D (1987) *The Biology of Moral Systems*, New York: Aldine De Gruyter.

[3] ALLPORT.G.W (1954) *The Nature of Prejudice*, Cambridge, MA: Perseus Books.

[4] ALTONJI.J.G, T.E.ELDER AND C.R.TABER (2005) "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy* 113: 151-84.

[5] ALTONJI.J AND C.R.PIERRET (2001) "Employer Learning and Statistical Discrimination" *Quarterly Journal of Economics* 116: 313-50.

[6] ANWAR.S AND H.FANG (2006) "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *American Economic Review* 96: 127-51.

[7] ANWAR.S, P.BAYER AND R.HJALMARSSON (2012) "The Impact of Jury Race in Criminal Trials," *Quarterly Journal of Economics* 127: 1017-55.

[8] BAUER.M, C.BLATTMAN, J.CHYTILOVA, J.HENRICH, E.MIGUEL AND T.MITTS (2016) "Can War Foster Cooperation?," *Journal of Economic Perspectives* 30: 249-74.

[9] BAYER.P, R.HJALMARSSON AND D.POZEN (2009) "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections," *Quarterly Journal of Economics* 124: 105-47.

[10] BERTRAND.M AND E.DUFLO (2016) "Field Experiments on Discrimination,", forthcoming in A.Banerjee and E.Duflo (eds.) *Handbook of Field Experiments*.

[11] BOYD.R, H.GINTIS, S.BOWLES AND P.J.RICHERSON (2003) "The Evolution of Altruistic Punishment," *Proceedings of the National Academy of Sciences* 100: 3531-5.

[12] BUSHWAY.S.D AND A.M.PIEHL (2001) "Judging Judicial Discretion: Legal Factors and Racial Discrimination in Sentencing," *Law and Society Review* 35: 733-64.

[13] CHARLES.K.K AND J.GURYAN (2011) "Studying Discrimination: Fundamental Challenges and Recent Progress," *Annual Review of Economics* 3: 479-511.

[14] CONGRESSIONAL RESEARCH SERVICE (2013) *The Federal Prison Population Buildup: Overview, Policy Changes, Issues and Options*, Report 7-5700, Washington DC: CRS.

[15] DAVIS.D (2007) *Negative Liberty: Public Opinion and the Terrorist Attacks on America*, Russell Sage Foundation.

[16] DEPEW.B, O.EREN AND N.MOCAN (2017) "Judges, Juveniles and In-group Bias," *Journal of Law and Economics* 60: 209-39.

[17] FAIRLIE.R.W (2005) "An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models," *Journal of Economic and Social Measurement* 30: 305-16.

[18] FISCHMAN.J.B AND M.M.SCHANZENBACH (2012) "Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums," *Journal of Empirical Legal Studies* 9: 729-64.

[19] FOWLER.J.H AND N.A.CHRISTAKIS (2010) "Cooperative Behavior Cascades in Human Social Networks," *Proceedings of the National Academy of Sciences* 107: 5334-8.

[20] GLAESER.E.L, D.P.KESSLER AND A.M.PIEHL (2000) "What Do Prosecutors Maximize? An Analysis of the Federalization of Drug Crimes," *American Law and Economics Review* 2: 259-90.

[21] HUMAN RIGHTS WATCH (2002) We Are Not the Enemy: Hate Crimes Against Arabs, Muslims, and Those Perceived to be Arab or Muslim after September 11, Human Rights Watch 6.

[22] JEFFRIES.J.C JR AND J.GLEESON (1995) "The Federalization of Organized Crime: Advantages of Federal Prosecution," *Hastings Journal* 46: 1095-134.

[23] JORDAN.J.J, D.G.RAND, S.ARBESMAN, J.H.FOWLER AND N.A.CHRISTAKIS (2013) "Contagion of Cooperation in Static and Fluid Social Networks," *PloS One* 8, e66199.

[24] JUHN.C, K.M.MURPHY AND B.PIERCE (1993) "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy* 101: 410-42.

[25] KLEPPER.S, D.NAGIN AND L-J.TIERNEY (1983) "Discrimination in the Criminal Justice System: A Critical Appraisal of the Literature," in Research on Sentencing: The Search for Reform, A.Blumstein, J.Cohen, S.E.Martin and M.H.Tonry (eds.) Vol. 2. Washington DC: National Academy Press.

[26] LIPTAK.A (2003) "For Jailed Immigrants, a Presumption of Guilt," *New York Times*, June 3.

[27] MUELLER-SMITH.M (2016) The Criminal and Labor Market Impacts of Incarceration, mimeo, University of Michigan.

[28] MUSTARD.D.B (2001) "Racial, Ethnic and Gender Disparities in Sentencing: Evidence from the US Federal Courts," *Journal of Law and Economics* 44: 285-314.

[29] NEAL.D AND A.RICK (2015) "The Prison Boom and Sentencing Policy," forthcoming, *Journal of Legal Studies*.

[30] ORRENIUS.P.M AND M.ZAVODNY (2009) "The Effects of Tougher Enforcement on the Job Prospects of Recent Latin American Immigrants," *Journal of Policy Analysis and Management* 28: 239-57.

[31] OSTER.E (2017) "Unobservable Selection and Coefficient Stability: Theory and Validation," forthcoming *Journal of Business Economics and Statistics*.

[32] PARK.K.H (2017) "Do Judges Have Tastes for Discrimination? Evidence from Criminal Courts," forthcoming, *Review of Economics and Statistics*.

[33] REHAVI.M.M AND S.B.STARR (2014) "Racial Disparity in Federal Criminal Sentences," *Journal of Political Economy* 122: 1320-54.

[34] ROMERO.L.A AND A.ZARRUGH (2016) "Islamophobia and the Making of Latinos as Terrorist Threats," forthcoming *Ethnic and Racial Studies*.

[35] SCHANZENBACH.M (2005) "Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics," *Journal of Legal Studies* 34: 57-92.

[36] SHAYO.M (2009) "A Model of Social Identity with an Application to Political Economy: Nation, Class and Redistribution," *American Political Science Review* 103: 147-74.

[37] SHAYO.M AND A.ZUSSMAN (2011) "Judicial Ingroup Bias in the Shadow of Terrorism," *Quarterly Journal of Economics* 126: 1447-84.

[38] STARR.S.B AND M.M.REHAVI (2013) "Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of *Booker*," *Yale Law Journal* 123: 2-80.

[39] SURI.S AND D.J.WATTS (2011) "Cooperation and Contagion in Web-based, Networked Public Goods Experiments," *PloS One*, 6, e16836.

[40] TAJFEL.H, M.G.BILLIG, R.P.BUNDY AND C.FLAMENT (1971) "Social Categorization and Intergroup Behavior," *European Journal of Social Psychology* 1: 149-78.

[41] USSC (1999-2003) Monitoring of Federal Criminal Sentences, 1999-2003 [Computer file], ICPSR version. Wash. DC: USSC [producer], 1999-2006. Ann Arbor, MI: ICPSR [distrib.].

[42] WOODS.J (2011) "The 9/11 Effect: Toward a Social Science of the Terrorist Threat," *Social Science Journal* 48: 213-33.

[43] YANG.C.S (2015) "Free At Last? Judicial Discretion and Racial Disparities in Federal Sentencing," *Journal of Legal Studies* 44: 75-111.

[44] YANG.C.S (2016) "Resource Constraints and the Criminal Justice System: Evidence from Judicial Vacancies," *American Economic Journal: Economic Policy* 8: 289-332.

Table 1: Pre 9-11 Sentencing Differentials in Judge's Decisions

Sample: Cases up for sentencing between 10/1/1998 and 09/10/2001
Standard errors in parentheses clustered by group-district

	Downward Departure		Cells Moved		Sentence Length	
	(1) Unconditional	(2) Conditional	(3) Unconditional	(4) Conditional	(5) Unconditional	(6) Conditional
Black	-.047**	-.008	-.509***	-.649***	42.2***	3.88***
	(.020)	(.006)	(.187)	(.070)	(3.41)	(.556)
Hispanic	.133**	.010	.156	-.768***	1.72	4.08***
	(.062)	(.015)	(.424)	(.100)	(4.17)	(.611)
Sentencing Outcome for Whites	.125		2.35		40.5	
Offender, Legal and District Controls	No	Yes	No	Yes	No	Yes
Offense Type Codes	No	Final	No	Final	No	Final
Guideline Cells	No	Yes	No	Yes	No	Yes
p-value: [Black = Hispanic]	.003	.118	.120	.166	.000	.750
Adjusted R-squared	.044	.242	.004	.217	.064	.743
Observations	130,895	130,895	130,895	130,895	130,895	130,895

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown in all Columns except 3 and 4 where a negative binomial specification is estimated. Standard errors are reported in parentheses, where these are clustered by group-district. The pre-9/11 sample of 130,895 Federal cases is used (those that come up for sentencing from 10/1/1998 to 09/10/2001). The dependent variable in Columns 1 and 2 is a dummy for whether the case receives a downwards departure. The dependent variable in Columns 3 and 4 is the number of cells moved (including zero), using midpoints of guidelines cells to establish the guideline cell moved to in case of a downwards departure. The dependent variable in Columns 5 and 6 is the sentence length (in months) including zero. In Columns 1, 3 and 5 we only condition on defendant group (White, Black, Hispanic). In Columns 2, 4 and 6 the following additional controls are included: fiscal year dummies, on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, the guideline cell, and Federal district dummies. The p-value at the foot of each Column is on the null that the coefficients on the Black and Hispanic dummy are equal against a two sided alternative.

Table 2: Balance Within Race/Ethnicity

Means, standard deviations in parentheses, p-values in brackets

	White			Black			Hispanic		
	Control: Pre 9-11	Treatment: Post 9-11	p-value	Control: Pre 9-11	Treatment: Post 9-11	p-value	Control: Pre 9-11	Treatment: Post 9-11	p-value
Sample Size	6137	6857		5162	5714		7749	8609	
Number Dependents	1.09	1.11	[.552]	1.67	1.71	[.396]	1.81	1.87	[.094]
	(1.42)	(1.41)		(1.84)	(1.82)		(1.77)	(1.79)	
Age	38.0	38.4	[.164]	31.5	32.0	[.039]	31.9	32.4	[.001]
	(12.2)	(12.0)		(9.21)	(9.26)		(9.27)	(9.19)	
Marital Status:									
Single	.335	.338	[.741]	.526	.541	[.201]	.329	.327	[.786]
Married or Cohabiting	.431	.436	[.710]	.340	.328	[.253]	.512	.509	[.668]
Divorced, Widowed or Separated	.213	.210	[.688]	.111	.112	[.807]	.103	.099	[.567]
Education:									
High School Graduate or Below	.659	.641	[.105]	.773	.778	[.642]	.820	.814	[.597]
Some College/College Graduate	.331	.351	[.087]	.221	.216	[.603]	.096	.092	[.424]
Defense Counsel:									
Privately Retained	.167	.165	[.833]	.078	.082	[.491]	.081	.077	[.565]
Court Appointed	.167	.174	[.389]	.165	.157	[.383]	.265	.282	[.092]
Federal Public Defender	.137	.132	[.402]	.156	.152	[.678]	.276	.259	[.223]
Other	.004	.005	[.670]	.005	.004	[.597]	.002	.001	[.154]

Notes: The sample refers to all cases for which sentencing occurs within a 6-month window of 9/11/2001. For those defendants sentenced after 9/11/2001 (treatment), the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001 (control), the last offense was committed at least 180 days prior to 9/11/2001. Means and standard deviations (in parentheses for continuous covariates) are shown. The p-values are tests of equality of the statistic within ethnic group across the two samples, based on an OLS regression that allows standard errors to be clustered by district.

Table 3: Judges' Decisions Around 9-11

Dependent Variable: Downward Departure Granted by Federal Judge

Standard errors in parentheses clustered by group-district

	(1) Baseline	(2) Reason: Criminal History Category Over Represented	(3) Reason: Pursuant to Plea Bargain	(4) Reason: General Mitigating Circumstances	(5) Reason: Other	(6) Initial Arrest Codes
Sentenced post 9-11*Hispanic	-.038***	-.013***	-.011	-.001	-.013*	-.046**
	(.013)	(.003)	(.007)	(.008)	(.007)	(.019)
Sentenced post 9-11*Black	-.013	-.005	.002	-.003	-.007	-.013
	(.008)	(.004)	(.003)	(.004)	(.005)	(.011)
Sentenced post 9-11	.006	.003	-.000	.001	.002	.003
	(.007)	(.002)	(.002)	(.004)	(.004)	(.009)
Offender, Legal and District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Offense Type Codes	Final	Final	Final	Final	Final	Arrest
Guideline Cells	Yes	Yes	Yes	Yes	Yes	No
p-value: [Post*B = Post*H]	.042	.037	.096	.795	.306	.063
Implied Sentence Length Impact (H)	.736					.889
% of Pre 9-11 Sentence Differential	18%					29.8%
Adjusted R-squared	.256	.042	.289	.068	.135	.257
Observations	40,228	40,228	40,228	40,228	40,228	26,852

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown in all Columns. Standard errors are reported in parentheses, where these are clustered by group-district. In Columns 1 to 5, the sample of 40,228 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. Columns 2 to 5 code downward departures into various broad categories of how judge's justify their decision to depart. In Column 6 the sample is restricted to those cases that can be linked back to arrest (Stage 0). The dependent variable throughout is a dummy for whether the case receives a downwards departure (where in Columns 2 to 5 this is modified based on the reasons given for departure). In all Columns we condition on defendant group (White, Black, Hispanic), whether the case comes up post 9-11, and interactions between the two, and the following additional controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the guideline cell, and Federal district dummies. In Columns 1 to 7 we control for the primary offense type. In Column 6 we instead control for arrest offense codes, but not guideline cells. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Black and post 9-11 x Hispanic dummy interactions are equal against a two sided alternative.

Table 4: Citizenship and Offense Type

Dependent Variable: Downward Departure Granted by Federal Judge

Standard errors in parentheses clustered by group-district

	(1) All Offenses	(2) Drug Offenses	(3) Immigration Offenses: Hispanics Only	(4) Immigration Offenses: Hispanics Only, Border States	(5) All Other Offenses
Sentenced post 9-11*Hispanic Citizen	-.028**	-.014	-.054	-.038	-.029**
	(.013)	(.019)	(.037)	(.049)	(.013)
Sentenced post 9-11*Hispanic Non-Citizen	-.044***	-.055*	.033	.017	-.007
	(.015)	(.031)	(.037)	(.048)	(.030)
Sentenced post 9-11*Black	-.013	-.001			-.018*
	(.008)	(.014)			(.010)
Sentenced post 9-11	.005	-.004			.009
	(.007)	(.013)			(.007)
Offender, Legal and District Controls	Yes	Yes	Yes	Yes	Yes
Offense Type Codes	Final	Final	Final	Final	Final
Guideline Cells	Yes	Yes	Yes	Yes	Yes
Implied Sentence Length Impact (H, Citizen)	.575 [17.2%]	.487 [8.2%]	.741	0.478	.371 [19.4%]
Implied Sentence Length Impact (H, Non-citizen)	.821 [15.9%]	1.42 [18.1%]	.424 [22.8%]	.422 [29.9%]	-.055 [-1.8%]
p-value: [Post*H Citizen= Post*H Non Citizen]	.269	.170	.237	.782	.509
Adjusted R-squared	.258	.298	.357	.342	.091
Observations	39,937	17,583	6,147	4,534	15,617

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown throughout. Standard errors are reported in parentheses, where these are clustered by group-district. The sample of 39,937 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001) and for which defendant citizenship is not missing. For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. Column 1 covers all offenses. Columns 2-5 are restricted to drug, immigration and other offenses respectively, where for immigration offenses, only Hispanic defendants are included and Column 4 further restricts the sample to US-Mexico Border States. The dependent variable is a dummy for whether the case receives a downwards departure. In all Columns we condition on interactions between Hispanic ethnicity, defendant citizenship (where citizens are defined as being US citizens or resident/legal aliens, and non-citizens are illegal aliens, non-US citizens and those for whom alien status is unknown), and whether the case comes up post 9-11, as well as each of these control variables alone. In all specifications the following additional controls are included: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, the guideline cell, and Federal district dummies. At the foot of each Column, the percentage reported in square brackets is the percentage of the pre 9-11 differential the implied sentence length impact corresponds to. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Hispanic Citizen and post 9-11 x Hispanic Non Citizen dummy interactions are equal against a two sided alternative.

Table 5: Pre 9-11 Sentencing Differentials in Prosecutors' Decisions

Sample: Cases up for Sentencing between 10/1/1998 and 09/10/2001
Standard errors in parentheses clustered by group-district

	Non-zero Statutory Minimum			Statutory Minimum			Substantial Assistance Departure	
	(1) Uncond.	**(2) Cond.**	**(3) Cond.**	**(4) Uncond.**	**(5) Cond.**	**(6) Cond.**	**(7) Uncond.**	**(8) Cond.**
Black	.233***	.168***	.051***	29.0***	21.6***	7.81***	-.002	-.025***
	(.022)	(.015)	(.006)	(2.66)	(1.69)	(.869)	(.019)	(.006)
Hispanic	.054	.126***	.056***	4.30	13.9***	7.37***	-.115***	-.090***
	(.036)	(.020)	(.009)	(4.03)	(2.14)	(.982)	(.018)	(.008)
Sentencing Outcome for Whites		.222			22.1			.218
Offender, Legal and District Controls	No	Yes	Yes	No	Yes	Yes	No	Yes
Offense Type Codes	No	No	Arrest	No	No	Arrest	No	Final
p-value: [Black = Hispanic]	0.000	.023	.508	.000	.000	.696	.000	.000
Adjusted R-squared	.040	.147	.495	.038	.136	.365	.023	.142
Observations	130,216	130,216	68,216	130,216	130,216	68,216	130,895	130,895

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown in all Columns except 5 and 6 where a negative binomial specification is estimated. Standard errors are reported in parentheses, where these are clustered by group-district. The pre-9/11 sample of 130,895 Federal cases is used (those that come up for sentencing from 10/1/1998 to 09/10/2001). The dependent variable in Columns 1 to 3 is a dummy for whether the initial charge filed by prosecutors has an associated mandatory minimum sentence length. The dependent variable in Columns 4 to 6 is the mandatory minimum sentence length (including zeroes for those without a minimum). The dependent variable in Columns 7 and 8 is whether the prosecutor grants a substantial assistance downwards departure. In Columns 1, 4 and 7 we only condition on defendant group (White, Black, Hispanic). In Columns 2, 3, 5, 6 and 8 the following additional controls are included: fiscal year dummies, on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); and Federal district dummies. In Columns 3 and 6 we additionally control for the primary offense type as measured at the arrest stage, while in Column 8 we additionally control for the primary offense type. The p-value at the foot of each Column is on the null that the coefficients on the Black and Hispanic dummy are equal against a two sided alternative.

Table 6: Prosecutors' Decision Making Around 9-11

Standard errors in parentheses clustered by group-district

	Prosecutor's Initial Charges		Judge's Sentencing		Prosecutor's Substantial Assistance Departure
	(1) Non-zero Statutory Minimum	(2) Statutory Minimum Length	(3) Sentence Length	(4) Sentence Length	(5) Same Cohort as in Col. (1)
Initial charges post 9-11*Hispanic	.075*	10.7**	9.33**	1.81	-.037
	(.042)	(5.34)	(4.65)	(2.65)	(.042)
Initial charges post 9-11*Black	-.010	.684	-5.39	.846	-.053
	(.048)	(7.50)	(7.36)	(3.66)	(.048)
Initial charges post 9-11	-.033	-5.96	-8.29**	-.873	.035
	(.033)	(3.90)	(3.94)	(2.34)	(.037)
Offender, Legal and District Controls	Yes	Yes	Yes	Yes	Yes
Offense Type Codes	No	No	No	Final	Final
Guideline Cell Dummies	No	No	No	Yes	No
p-value: [Post*B = Post*H]	.046	.172	.030	.755	.654
Adjusted R-squared	.171	.147	.190	.797	.180
Observations	3,612	3,600	3,612	3,612	3,612

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown in all Columns. Standard errors are reported in parentheses, where these are clustered by group-district. In Columns 1 to 5, the sample of Federal cases used is: (i) for those with initial charges after 9/11, defendants in (out of) custody committed their last offense between 14 (21) days before 9/11 and the day before 9/11; (ii) for those with initial charges before 9/11, defendants in (out of) custody committed their last offense between 42 (63) days before 9/11 and 38 (42) days before 9/11. In Column 6 the Natural Experiment sample of all Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). The dependent variable in Column 1 is a dummy for whether the defendant receives an initial charge with a non-zero statutory minimum sentence. The dependent variable in Column 2 is the length of statutory minimum sentence. The dependent variable in Columns 3 and 4 is the actual sentence length in months (as determined at the sentencing stage) and the dependent variable in Columns 5 and 6 is a dummy for whether the case receives a substantial assistance downwards departure at sentencing. In all Columns the following controls are included: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements) and Federal district dummies. In Column 4 the additional controls are offence type dummies and guideline cell dummies. In Column 5 the additional controls are offence type dummies. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Black and post 9-11 x Hispanic dummy interactions are equal against a two sided alternative.

Table 7: Pre-sentence Reports

OLS regression estimates
Standard errors in parentheses clustered by group-district

	(1) Criminal History Score	(2) Offense Severity Score	(3) Minimum Guideline Sentence
Convicted and Sentenced after 9-11 [T2]*Hispanic	.016	-.625***	-2.31
	(.047)	(.221)	(1.65)
Convicted and Sentenced after 9-11 [T2]*Black	.036	-.040	2.02
	(.055)	(.207)	(2.13)
Convicted and Sentenced after 9-11 [T2]	.048	.391***	2.57**
	(.036)	(.133)	(1.28)
Offender, Legal and District Controls	Yes	Yes	Yes
Offense Type Codes	Final	Final	Final
Adjusted R-squared	.253	.489	.326
Observations	40,228	40,228	40,228

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown in Columns 1 to 3. The natural experiment sample of 40,228 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. The dependent variable in Column 1 (2) is the criminal history score (offense severity score) reported in the pre-sentence report, and in Column 3 it is the lowest sentence in the recommended guideline cell. In all Columns we condition on defendant group (White, Black, Hispanic), whether the defendant is convicted before 9-11 but sentenced after 9-11 [treatment group T1], whether the defendant is convicted and Sentenced after 9-11 [treatment group T2], and interactions between the two treatment dummies and offender group, and the following additional controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, and Federal district dummies. The p-value at the foot of each Column is on the null that the coefficients on the Convicted before 9-11 but Sentenced after 9-11 [T1]*Hispanic dummy and Convicted and Sentenced after 9-11 [T2]*Hispanic dummy interactions are equal against a two sided alternative.

Table 8: Juhn-Murphy-Pierce Decompositions of Hispanic-White Differentials

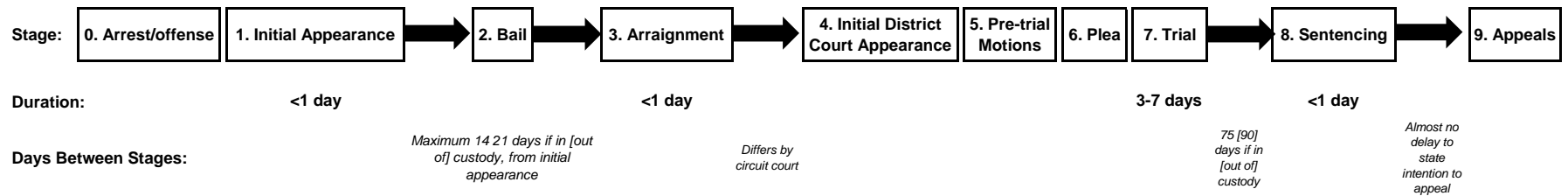|	Cohort 1: Judge Decisions (1) Downwards Departure	Cohort 2: Prosecutor Decisions (2) Statutory Minimum Length
1. Pre-9/11 (raw) differential	.158	-1.21
2. Post-9/11 (raw) differential	.117	6.18
3. Change in differential	-.041	7.40
4. Due to observables: X-effect + β-effect	-.003	-9.29
5. Due to unobservables: θ-effect + u-effect	-.038	16.7
6. Observable quantity: X-effect	.005	-2.72
7. Observable penalties: β-effect	-.008	-6.57
8. Unobservable quantities: θ-effect	-.044	22.2
9. Unobservable penalties: u-effect	.006	-5.54
X-Controls	Offender characteristics, defense counsel type, offense type dummies, guideline cell dummies, and Federal district dummies.	Offender characteristics, defense counsel type and Federal district dummies.

Notes: A Juhn-Murphy-Pierce [1993] decomposition, using a non-parametric procedure, is implemented. This decomposes the unconditional difference-in-difference for each sentencing outcome between Hispanics and Whites. In Column 1 this is based on Federal criminal cases in the Natural Experiment sample. Hence the decomposition is based on 29,352 cases for Hispanic or White defendants that come up for sentencing in a six month window either side of 9/11/2001. For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. The outcome in Column 1 is for whether any downward departure is received. In Column 2 the sample of Federal cases used is: (i) for those with initial charges after 9/11, defendants in (out of) custody committed their last offense between 14 (21) days before 9/11 and the day before 9/11; (ii) for those with initial charges before 9/11, defendants in (out of) custody committed their last offense between 42 (63) days before 9/11 and 38 (42) days before 9/11. The outcome in Column 2 is the length of statutory minimum sentence following from the initial offense charge. For both Juhn-Murphy-Pierce decompositions, Whites are chosen as the reference group.

Table 9: Judge Characteristics

Dependent Variable: Coefficient on post 9-11 x Hispanic x District dummy
Robust standard errors in parentheses
Observations weighted by district share of Hispanics in 2001

	(1) Race/Ethnicity	(2) Other Judge Characteristics	(3) District Population	(4) Effect Size
District Proportion Hispanic Judges	.225***	.204**	.554***	.032***
	(.073)	(.101)	(.207)	(.012)
District Proportion Black Judges	.272	.338	.097	.008
	(.217)	(.222)	(.207)	(.018)
District Proportion Senior Status Judges		-.066	.027	.004
		(.076)	(.090)	(.014)
District Proportion Male Judges		-.022	-.143	-.017
		(.095)	(.093)	(.011)
District Mean Judge Age		.006*	.004	.015
		(.003)	(.003)	(.014)
District Proportion Democratic President Appointees		.180**	.137**	.025**
		(.076)	(.066)	(.012)
District Proportion of Post-Period Window with Bush-Appointed US Attorney		.026	-.046	-.017
		(.027)	(.033)	(.013)
District Proportion Black (2000)			.275**	.032**
			(.127)	(.015)
District Proportion Hispanic (2000)			-.337*	-.034*
			(.184)	(.019)
Change in District Proportion Black (1990 - 2000)			-2.59**	-.027**
			(1.06)	(.011)
Change in District Proportion Hispanic (1990 - 2000)			-.100	-.002
			(.519)	(.011)
Mean of Dependent Variable		-.016		
Adjusted R-squared	.105	.172	.287	.287
Observations	88	88	88	88

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The results are based on the natural experiment sample (those that come up for sentencing in a six month window either side of 9/11/2001, where for those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. Each observation represents a single Federal court district and observations are weighted by the share of Hispanics in the district in the relevant sample of Federal criminal cases (the natural experiment or full sample). Robust standard errors are reported. The dependent variable is the coefficient on post 9-11*Hispanic*District from a difference-in-difference-in-difference regression for the Natural experiment sample period where in this first stage the full set of controls is included, and the dependent variable is whether a downwards departure is granted. The data for judicial characteristics are sourced from the *Biographical Directory of Federal Judges*. In order to select the relevant judges to construct characteristics for, we used the data on commission and termination dates for each judge in the database, and we restrict the sample to judges commissioned before the end of the natural experiment sample and those who terminated the bench after the beginning of the sample. Data for district level characteristics are from the 1990 and 2000 5% US census data. District proportions were constructed using the individual weights (perwt) provided by IPUMS. In Column 4, effect sizes on all covariates are reported.

Figure 1: Federal CJS Timeline

Stage:	0. Arrest/offense	1. Initial Appearance		2. Bail		3. Arraignment		4. Initial District Court Appearance	5. Pre-trial Motions	6. Plea	7. Trial		8. Sentencing		9. Appeals

Duration: <1 day <1 day 3-7 days <1 day

Days Between Stages:

Maximum 14 21 days if in [out of] custody, from initial appearance

Differs by circuit court

75 [90] days if in [out of] custody

Almost no delay to state intention to appeal

Notes: We use the Federal Justice Statistics Resource Center (FJSRC) data set for our analysis. This comprises information gathered from four linked administrative data sources. As described in Rehavi and Starr [2014], the four linked data sets are: (i) US Marshals Service (USMS) data, that covers the arrest/offense stage (Stage 0) and includes all persons arrested by Federal law enforcement agencies, persons arrested by local officials and then transferred to Federal custody, and persons who avoid arrest by self-surrendering; (ii) Executive Office for US Attorneys (EOUSA) data, covering initial appearance through to arraignment (Stages 1-3): these data come from the internal case database used by Federal prosecutors, and covers every case in which any prosecutor at a US Attorney's office opens a file; (iii) Administrative Office of the US Courts (AOUSC) data, covering initial district court appearances through to trial (Stages 4-7): these originate from Federal Courts and contain data on all criminal cases heard by Federal district judges, and any non-petty charge handled by a Federal magistrate judge; (iv) the Monitoring of Federal Criminal Sentences (MFCS) data set that is resourced from US Sentencing Commission (USSC), covering the sentencing Stage 8: this data set collects information on any case that results in conviction and sentencing for a non-petty offense. These data are collected by the Bureau of Justice Statistics.
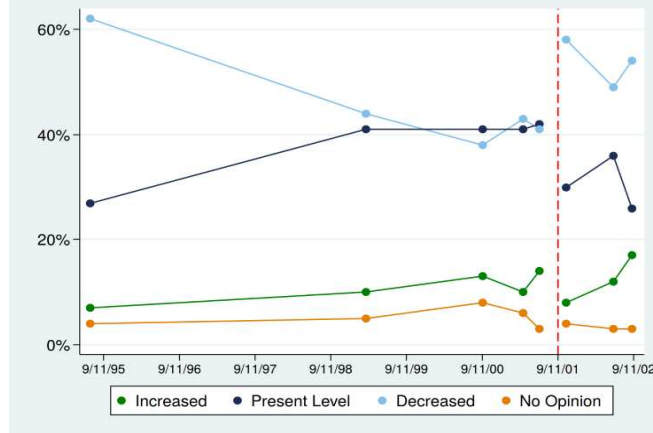
Figure 2: Pre and Post 9-11 Sentiments

A: Pre 9-11 Sentiments Towards Outgroups
NES 2000 Thermometer Reading (Normalized by White-White Rating)
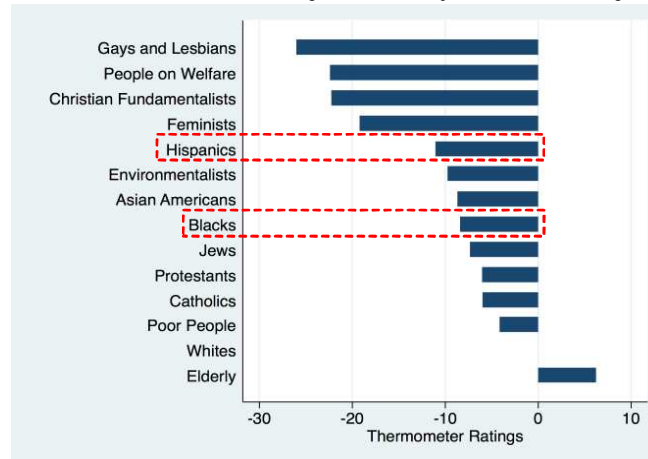
B: Sentiments Towards Immigrants Around 9-11
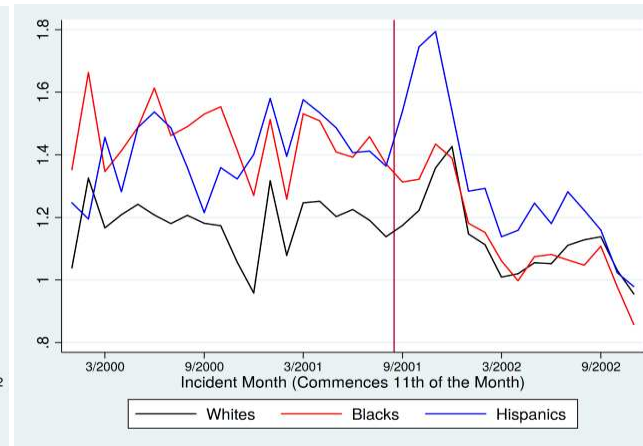### Gallup Poll on Immigration
Q: Should Immigration be Kept at Its Present Level, Increased or Decreased?

C: Crime Rates Around 9-11
Vandalism Victimization
*Growth Rate from Same Month in Previous Year*

Notes: Panel A and B is constructed from the National Election Survey, and are based on the thermometer ratings of White respondents only. Respondents were asked about their feelings towards many groups in American society, and to represent these opinions on a "feeling thermometer". Respondents were instructed: "If you don't know too much about a group or don't feel particularly warm or cold toward them, then you should place them in the middle, at the 50 degree mark. If you have a warm feeling toward a group or feel favorably toward it, you would give it a score somewhere between 50 degrees and 100 degrees, depending on how warm your feeling is toward the group. On the other hand, if you don't feel very favorably toward some of these groups--if there are some you don't care for too much--then you would place them somewhere between 0 degrees and 50 degrees." Thermometer readings in the raw data range from 0-97 (97-100 was top-coded at 97). Panel A shows the sample-weighted mean of various thermometer ratings are presented for White respondents in the year 2000, where the mean ratings have been normalized by White respondents' thermometer ratings for Whites. Panel B is based on a Gallup Poll that asks respondents, "Thinking more about immigration - that is, people who come from other countries to live here in the United States, in your view, should immigration be kept at its present level, increased or decreased?". The data was accessed via http://www.gallup.com/poll/1660/immigration.aspx. Panel B is based on data from the National Incident-Based Reporting System Extract Files. The outcome variable is vandalism victimization. The data was collapsed to the month level, where month was constructed to start on the 11th in order to align with 9/11/2001. In order to account for seasonal differences in victimization, the outcome variable is divided by its counterpart from the same month in the previous year, so can be interpreted as a growth rate.

Figure 3: Sentencing and Last Offense Dates

A. Sentencing Date   B. Date of Last Offense

Whites
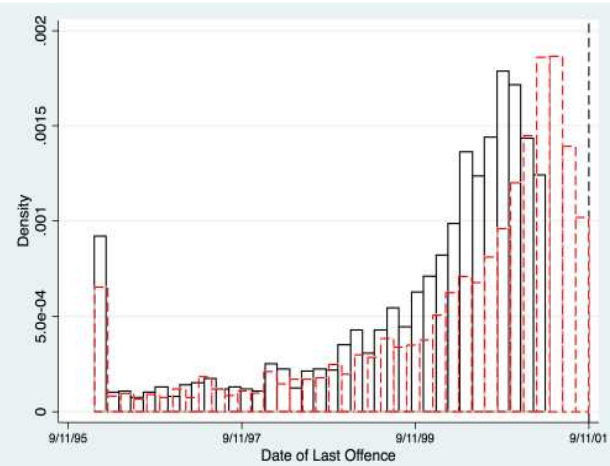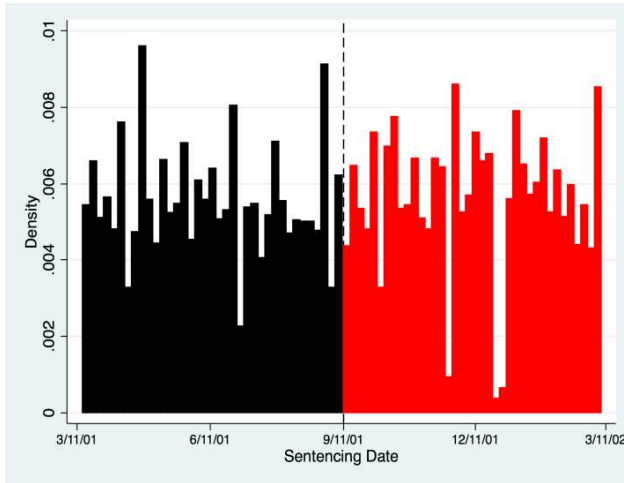
Blacks
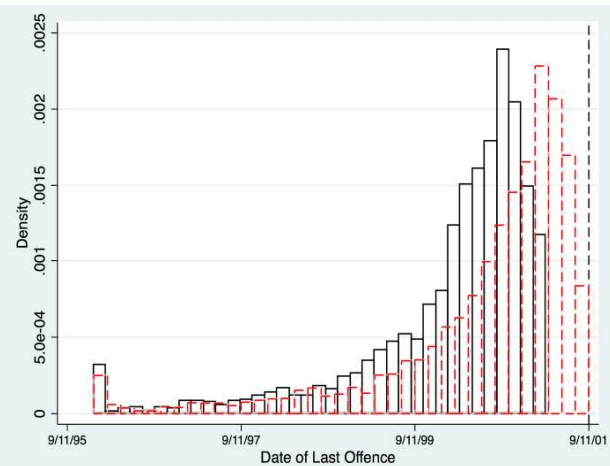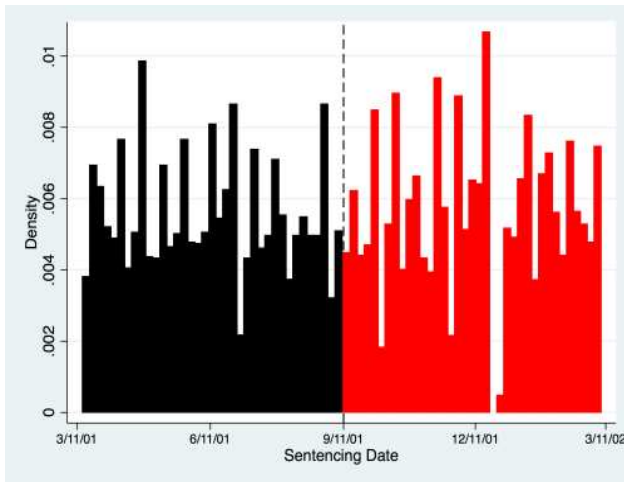
Hispanics
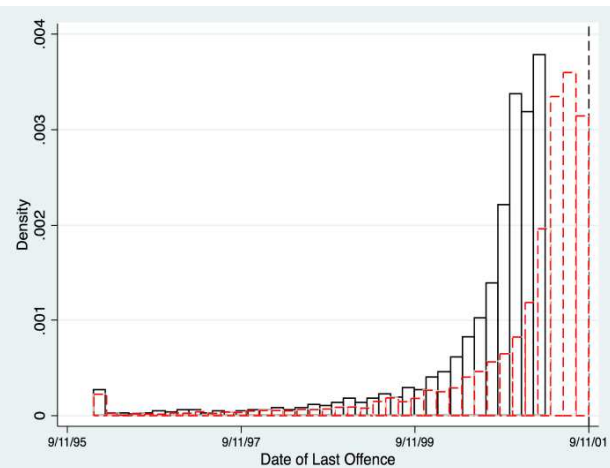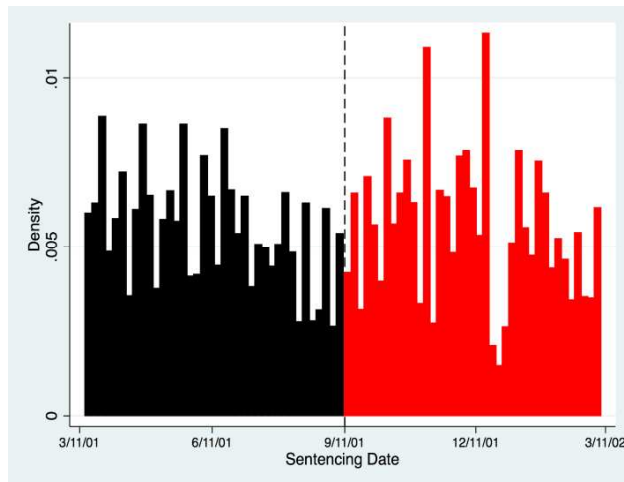
Notes: The left hand side figures show the distribution of dates of sentencing date, for each group: 9/11 is indicated by the vertical dashed line. The right hand side figures show the distribution of the dates of last offenses, by group. The first bar corresponds to a last offense date on or before 1st January 1996. The overlaid histograms are for those sentenced pre- and post-9/11. For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001.

Figure 4: Juhn-Murphy-Pierce Decompositions of Hispanic-White Differentials

Cohort 1: Judge Decisions
Downwards Departure
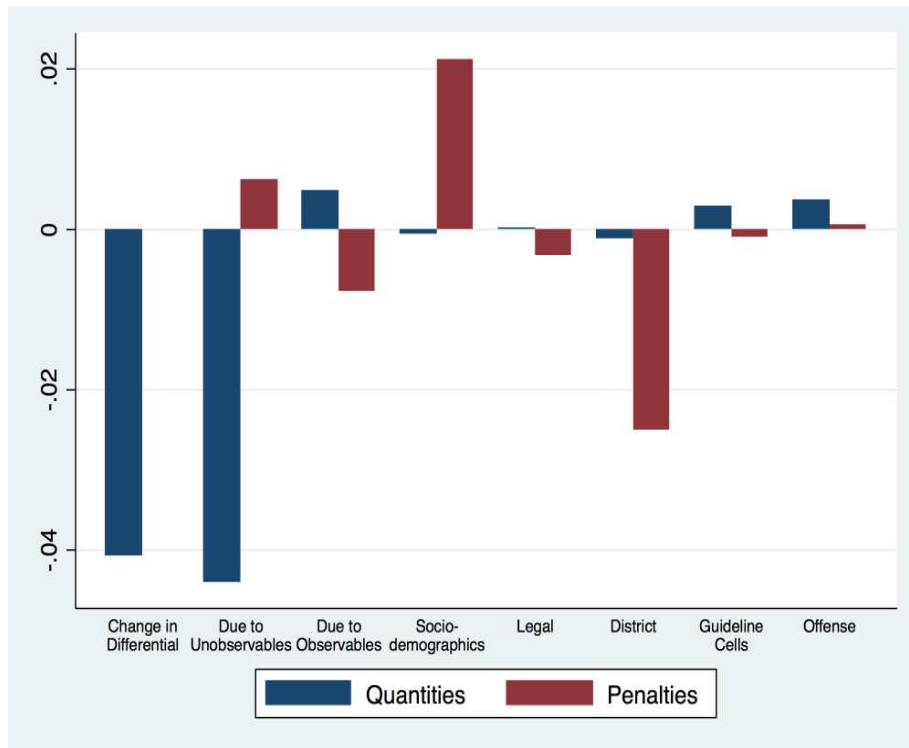
Cohort 2: Prosecutor Decisions
Statutory Minimum Sentence Length
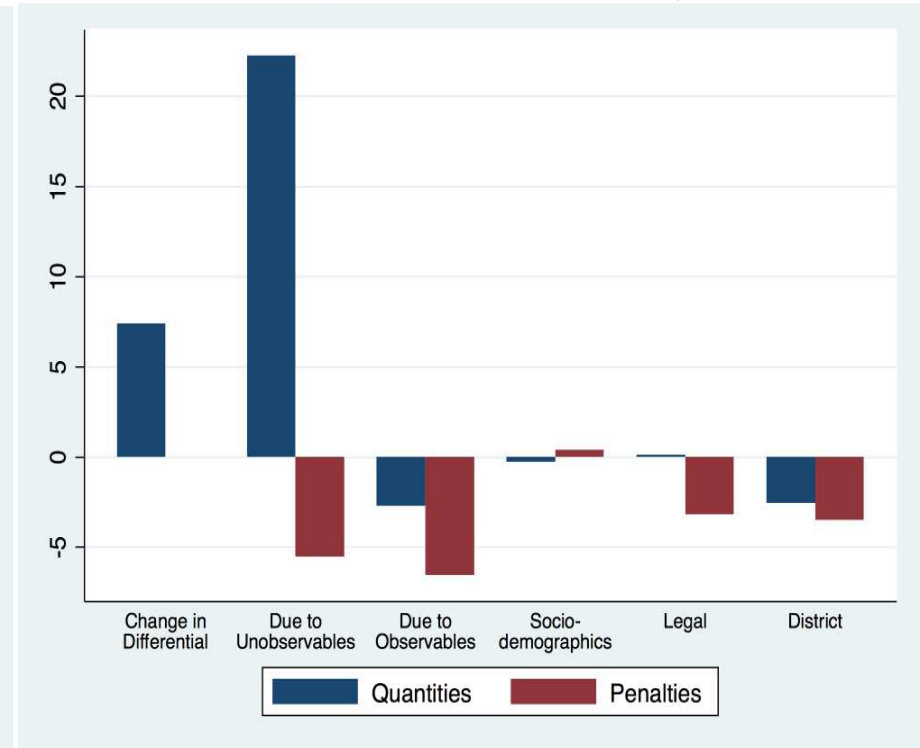
Notes: The graphs show key results from a Juhn-Murphy-Pierce [1993] decomposition, using a non-parametric procedure This decomposes the unconditional difference-in-difference for each sentencing outcome between Hispanics and Whites. In the left-hand graph this is based on Federal criminal cases in the Natural Experiment sample. Hence the decomposition is based on 29,352 cases for Hispanic or White defendants that come up for sentencing in a six month window either side of 9/11/2001. For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. The outcome in the left-hand graph 1 is for whether any downward departure is received. The controls in this decomposition are Offender characteristics, defense counsel type, offense type dummies, guideline cell dummies, and Federal district dummies. In the right-hand graph the sample of Federal cases used is: (i) for those with initial charges after 9/11, defendants in (out of) custody committed their last offense between 14 (21) days before 9/11 and the day before 9/11; (ii) for those with initial charges before 9/11, defendants in (out of) custody committed their last offense between 42 (63) days before 9/11 and 38 (42) days before 9/11. The outcome in the right-hand graph is the length of statutory minimum sentence following from the initial offense charge. The controls in this decomposition are offender characteristics, defense counsel type and Federal district dummies. For both Juhn-Murphy-Pierce decompositions, Whites are chosen as the reference group.

Table A1: Sentencing Guideline Cells (in months imprisonment)

	Offense Level	I (0 or 1)	II (2 or 3)	III (4, 5, 6)	IV (7, 8, 9)	V (10, 11, 12)	VI (13 or more)
Zone A	1	0-6	0-6	0-6	0-6	0-6	0-6
	2	0-6	0-6	0-6	0-6	0-6	1-7
	3	0-6	0-6	0-6	0-6	2-8	3-9
	4	0-6	0-6	0-6	2-8	4-10	6-12
	5	0-6	0-6	1-7	4-10	6-12	9-15
	6	0-6	1-7	2-8	6-12	9-15	12-18
	7	0-6	2-8	4-10	8-14	12-18	15-21
	8	0-6	4-10	6-12	10-16	15-21	18-24
Zone B	9	4-10	6-12	8-14	12-18	18-24	21-27
	10	6-12	8-14	10-16	15-21	21-27	24-30
Zone C	11	8-14	10-16	12-18	18-24	24-30	27-33
	12	10-16	12-18	15-21	21-27	27-33	30-37
Zone D	13	12-18	15-21	18-24	24-30	30-37	33-41
	14	15-21	18-24	21-27	27-33	33-41	37-46
	15	18-24	21-27	24-30	30-37	37-46	41-51
	16	21-27	24-30	27-33	33-41	41-51	46-57
	17	24-30	27-33	30-37	37-46	46-57	51-63
	18	27-33	30-37	33-41	41-51	51-63	57-71
	19	30-37	33-41	37-46	46-57	57-71	63-78
	20	33-41	37-46	41-51	51-63	63-78	70-87
	21	37-46	41-51	46-57	57-71	70-87	77-96
	22	41-51	46-57	51-63	63-78	77-96	84-105
	23	46-57	51-63	57-71	70-87	84-105	92-115
	24	51-63	57-71	63-78	77-96	92-115	100-125
	25	57-71	63-78	70-87	84-105	100-125	110-137
	26	63-78	70-87	78-97	92-115	110-137	120-150
	27	70-87	78-97	87-108	100-125	120-150	130-162
	28	78-97	87-108	97-121	110-137	130-162	140-175
	29	87-108	97-121	108-135	121-151	140-175	151-188
	30	97-121	108-135	121-151	135-168	151-188	168-210
	31	108-135	121-151	135-168	151-188	168-210	188-235
	32	121-151	135-168	151-188	168-210	188-235	210-262
	33	135-168	151-188	168-210	188-235	210-262	235-293
	34	151-188	168-210	188-235	210-262	235-293	262-327
	35	168-210	188-235	210-262	235-293	262-327	292-365
	36	188-235	210-262	235-293	262-327	292-365	324-405
	37	210-262	235-293	262-327	292-365	324-405	360-life
	38	235-293	262-327	292-365	324-405	360-life	360-life
	39	262-327	292-365	324-405	360-life	360-life	360-life
	40	292-365	324-405	360-life	360-life	360-life	360-life
	41	324-405	360-life	360-life	360-life	360-life	360-life
	42	360-life	360-life	360-life	360-life	360-life	360-life
	43	life	life	life	life	life	life

Table A2: Detailed Federal CJS Timeline

Stage		Who is involved	Description	Notes
1	Initial Appearance	Defendant, Federal Magistrate, Prosecutor (Assistant US Attorney), Assistant Federal Public Defender	If defendant cannot afford counsel, they fill out a financial affidavit, and are assigned to either a federal public defender or CJA panel counsel	A federal magistrate presides over proceedings until the defendant appears in district court (at Stage 4)
2	Bail	Defendant, Federal Magistrate, Prosecutor (Assistant US Attorney), defense Counsel, Pretrial Services	The bail hearing generally takes place within a week of the initial appearance, and depends on the case. Defendants seeking bail are then referred to Pretrial Services (neutral court employees, who interview the defendant and prepare a short life background and criminal history for the court). defense is present for this. Bail is then decided upon.	For "presumption" cases (drug dealing, bank robbery, child sex offenses), the govt. automatically gets 3 days to prepare for a bail hearing. If the govt. can prove the defendant is a flight risk, they get 3 days preparation time. The defense can ask for up to 5 days preparation time.
3	Arraignment	Defendant, Federal Magistrate, Prosecutor (Assistant US Attorney), defense Counsel, Federal Grand Jury	Happens within 14 (21) days from initial appearance for in-custody (out-of-custody) defendants. Defendant is arraigned on an indictment, which contains federal charges against him/her. Reviewed by grand jury. If sufficient evidence, jury "returns the indictment". After arraignment, magistrate adds the case to the district court calendar, and a district court judge is assigned. This judge will preside over the rest of the stages up to and including sentencing.	This is the stage where initial charges are filed, and so determines the statutory maximum and minimum for the offense.
4	Initial District Court Appearance	Defendant, District Court Judge, Prosecutor (Assistant US Attorney), defense Counsel	"Status" is decided: defense reviews the evidence ("discovery") in order to identify any motions. defense also discusses any pretrial dispositions (deals) with the prosecutor.	
5	Pretrial Motions	Defendant, Prosecutor (Assistant US Attorney), defense Counsel	Further prosecutor-defense interaction. The defendant's motion is sometimes called the moving papers or the opening brief. The prosecutor usually has one to three weeks to respond to the motion (the response is called an "Opposition"). The defense then typically has one or two weeks to respond to the Opposition (the defense response is called a "Reply"). One to two weeks after the Reply is filed, the court usually hears argument on the motion.	Modal pretrial motion is a suppression motion, where defense moves to suppress evidence or prevent the govt using it at trial.
6	Plea	Defendant, Prosecutor (Assistant US Attorney), defense Counsel	Guilt Plea is choice for large majority of case; either an open plea (no plea agreement) or with a plea agreement made with the prosecutor. Defense must inform defendant of every plea offer the prosecutor makes, and generally advises defendant on pros/cons of agreement. Defendant alone decides.	
7	Trial	Defendant, District Court Judge, Prosecutor (Assistant US Attorney), defense Counsel, Jury	The typical federal trial lasts 3-7 days. At the trial, the defendant has the right to testify – or to not testify, and if he or she does not testify, that cannot be held against the defendant by the jury. The defendant also has the right to "confront" (i.e., cross-examine) government witnesses, and can use the subpoena power of the court to secure evidence or witnesses for trial.	
8	Sentencing	Defendant, District Court Judge, Prosecutor (Assistant US Attorney), defense Counsel, Probation Office	If a defendant is convicted, sentencing takes place 75 (90) days later if the defendant is in (out of) custody. A defendant convicted of some offenses will likely be remanded into custody after trial. After a conviction, the defendant and his or her attorney complete forms relating to the defendant's life history and provide those to the (neutral) Probation Office. Several weeks after the conviction, the defendant will be interviewed by a Probation Officer, with defense counsel present. The Probation Officer will then take information from that interview, from the forms submitted by the defense, and from material provided by the government, and will prepare a draft presentence report. The draft presentence report (or PSR) is provided to defense counsel and the government 35 days before sentencing. The parties must make factual or legal objections to the report within 10 days of receipt. 14 days before sentencing, the final PSR is provided to the judge. This final PSR describes the defendant's background, describes the offense, and calculates the federal sentencing guidelines. It also includes a recommended sentence, and lists any unresolved objections. 7 days before sentencing, the parties submit sentencing memoranda to the court, arguing for their proposed sentences. 3 days later, the parties may submit replies to the sentencing memos. At the sentencing hearing, the district court judge must resolve any remaining objections to the PSR, make factual findings, and must consider the factors of the key sentencing statute, 18 USC § 3553(a). Before imposing the sentence, the court must permit the defendant to speak (or "allocute").	
9	Appeals	Defendant, District Court Judge, Supreme Court Judge	If the defendant did not waive the right to appeal in a plea agreement, the defense may appeal both the conviction and the sentence imposed. The public defender will continue to represent the defendant, for free, during the appeal. If the defendant does not win the appeal in their Circuit, he or she can file a petition for writ of certiorari with the Supreme Court of the United States. The public defender will continue to represent the defendant during the petition for certiorari and Supreme Court argument, if the writ is granted.	There is a very short period during which the defense must state its intention to appeal ("notice" an appeal), so the subject should be discussed immediately after sentencing.

Table A3: Robustness Checks on Sentencing Differentials Around 9-11

Dependent Variable: Downward Departure Granted by Federal Judge
Standard errors in parentheses, clustered by group-district in Column 2

	(1) Cluster on sentence week x group	(2) Excluding Cases Where Statutory Minima or Maxima Bind Partially
Sentenced post 9-11*Hispanic	-.038***	-.041***
	(.011)	(.011)
Sentenced post 9-11*Black	-.013	-.016*
	(.008)	(.008)
Sentenced post 9-11	.006	.009
	(.006)	(.007)
Offender, Legal and District Controls	Yes	Yes
Offense Type Codes	Final	Final
Guideline Cells	Yes	Yes
p-value: [Post*B = Post*H]	.022	.018
Adjusted R-squared	.256	.275
Observations	40,228	32,430

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown in all Columns. Standard errors are reported in parentheses, where these are clustered by sentence week x group in Column . In Column 1 the sample of 40,228 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. In Column 2 we exclude cases where statutory minima or maxima bind partially, namely if a statutory minimum is above the lower limit of the guideline cell or when the statutory maximum is below the upper limit. The dependent variable is a dummy for whether the case receives a downwards departure. In all Columns we condition on defendant group (White, Black, Hispanic), whether the case comes up post 9-11, and interactions between the two, and the following additional controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, the guideline cell, and Federal district dummies. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Black and post 9-11 x Hispanic dummy interactions are equal against a two sided alternative.

Table A4: Sentencing Differentials Around 9-11, by Group

Dependent Variable: Downward Departure Granted by Federal Judge
Standard errors in parentheses clustered by district

	(1) White	(2) Black	(3) Hispanic
Sentenced post 9-11	.004	-.008	-.030***
	(.006)	(.005)	(.011)
Difference with Whites		-.011	-.034***
		(.008)	(.013)
Difference with Blacks			-.023*
			(.012)
Offender, Legal and District Controls	Yes	Yes	Yes
Offense Type Codes	Final	Final	Final
Guideline Cells	Yes	Yes	Yes
Adjusted R-squared	.151	.074	.313
Observations	12,994	10,876	16,358

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown throughout. Standard errors are reported in parentheses, where these are clustered by district. The natural experiment sample of 40,228 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. In Column 1 only criminal cases involving White defendants are used. In Column 2 only criminal cases involving Black defendants are used. In Column 3 only criminal cases involving Hispanic defendants are used. The dependent variable is a dummy for whether the case receives a downwards departure. In all Columns we condition on whether the defendant is sentenced after 9-11 and the following controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, and Federal district dummies. In Column 2 we report the difference between the coefficient estimate between Blacks and Whites (and the corresponding standard error). In Column 3 we report the differences between the coefficient estimate between Hispanics and Whites, and Hispanics and Blacks (and the corresponding standard error).

Table A5: Sentencing Differentials Around 9-11, by Race

Dependent Variable: Downward Departure Granted by Federal Judge
Standard errors in parentheses clustered by district

	(1) Downward Departure
Sentenced post 9-11*Black	.009
	(.010)
Sentenced post 9-11*American Indian	-.037
	(.023)
Sentenced post 9-11*Asian/Pacific Islander	.034
	(.024)
Sentenced post 9-11*Multi-Racial	.004
	(.095)
Sentenced post 9-11*Other Race	-.118
	(.147)
Sentenced post 9-11	-.016*
	(.009)
Offender, Legal and District Controls	Yes
Offense Type Codes	Final
Guideline Cells	Yes
Adjusted R-squared	.254
Unadjusted R-squared	-
Observations	40,858

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown. Standard errors are reported in parentheses, where these are clustered by district. The natural experiment sample of 40,228 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. The dependent variable is a dummy for whether the case receives a downwards departure. We condition on defendant race, whether the case comes up post 9-11, and interactions between the two, and all the following additional controls are included: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, the guideline cell, and Federal district dummies.

Table A6: Time in the Federal CJS

Dependent Variable: Downward Departure Granted by Federal Judge

Standard errors in parentheses clustered by group-district

	(1) Include Dummies for 20 Groupings of Time Between Last Offense and Sentence Date	(2) Include Dummies for 20 Groupings of Last Offense Date
Sentenced post 9-11*Hispanic	-.035***	-.042***
	(.013)	(.012)
Sentenced post 9-11*Black	-.013	-.014*
	(.008)	(.008)
Sentenced post 9-11	.006	-.002
	(.007)	(.007)
Offender, Legal and District Controls	Yes	Yes
Offense Type Codes	Final	Final
Guideline Cells	Yes	Yes
p-value: [Post*B = Post*H]	.085	.015
Adjusted R-squared	.261	.257
Observations	40,228	40,228

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown. Standard errors are reported in parentheses, where these are clustered by group-district. The natural experiment sample of 40,228 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. The dependent variable is a dummy for whether the case receives a downwards departure. In all Columns we condition on defendant group (White, Black, Hispanic), whether the defendant is sentenced after 9-11 and interactions between this treatment dummies and offender ethnicity, and the following controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, and Federal district dummies. In Column 1 we additionally include dummies to group the days between last offense and sentencing date into 20 bins, and in Column 2 we instead additionally include dummies to group the date of last offense into 20 bins. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Black and post 9-11 x Hispanic dummy interactions are equal against a two sided alternative.

Table A7: Time Between Dates of Last offense and Sentencing

OLS and survival regression estimates; standard errors in parentheses, clustered by group-district

	All Offenses			Drug Offenses			Immigration Offenses			Other Offenses		
	(1a) OLS	(1b) Cox	(1c) Log logistic, Gamma Frailty	(2a) OLS	(2b) Cox	(2c) Log logistic, Gamma Frailty	(3a) OLS	(3b) Cox	(3c) Log logistic, Gamma Frailty	(4a) OLS	(4b) Cox	(4c) Log logistic, Gamma Frailty
Sentenced post 9-11*Hispanic	8.06	-.036	.033	12.7	-.074	.035	64.4*	-.078	.035	19.3	.014	.023
	(12.4)	(.030)	(.022)	(17.7)	(.056)	(.026)	(38.7)	(.097)	(.058)	(26.4)	(.062)	(.035)
Sentenced post 9-11*Black	13.9	-.021	.022	12.2	.003	.001	84.7	-.033	.047	16.4	-.034	.037
	(14.5)	(.029)	(.020)	(20.6)	(.053)	(.029)	(66.1)	(.202)	(.099)	(19.5)	(.039)	(.025)
Sentenced post 9-11	5.96	-.024	.007	3.44	-.037	.010	-61.4	.072	-.006	10.8	-.047*	.018
	(11.2)	(.020)	(.016)	(15.7)	(.045)	(.021)	(37.5)	(.090)	(.055)	(14.1)	(.025)	(.018)
Controls (incl. guideline cell)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value: [Post*B = Post*H]	.590	.617	.574	.975	.052	.175	.720	.806	.881	.910	.451	.700
Observations	40,228	40,228	40,228	17,722	17,722	17,722	6,790	6,790	6,790	15,716	15,716	15,716

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample of cases refers to those 40,228 cases for which sentencing occurs within a 6-month window of 9/11/2001. For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. In Columns 1a-1c, the full natural experiment sample is used. In Columns 2a-2c (3a-3c) (4a-4c) the sample is restricted to drug (immigration) (other) offenses. The dependent variable is the number of days between the date of the last offense and the sentencing date. In Columns 1a, 2a, 3a and 4a an OLS model is estimated. In Columns 1b, 2b, 3b and 4b a Cox proportional hazard model is estimated so that a negative coefficient means a lower hazard rate, and thus a longer duration. In Columns 1c, 2c, 3c and 4c a log-logistic model with a frailty parameter is estimated. In this model a positive coefficient implies a longer duration. In all Columns we condition on defendant group (White, Black, Hispanic), whether the defendant is sentenced after 9-11 and interactions between this treatment dummies and offender ethnicity, and the following controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); and Federal district dummies. offense type dummies are only controlled for in Columns 1a-1c. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Black and post 9-11 x Hispanic dummy interactions are equal against a two sided alternative.

Table A8: Placebo Checks on 9-11 Impacts

Dependent Variable: Downward Departure Granted by Federal Judge

Standard errors in parentheses clustered by group-district

	(1) Downward Departure
Sentenced post 9-11*Hispanic*2001	-.047***
	(.016)
Sentenced post 9-11*Hispanic	.008
	(.006)
Sentenced post 9-11*Black*2001	-.016
	(.010)
Sentenced post 9-11*Black	.002
	(.005)
Sentenced post 9-11*2001	.008
	(.008)
Sentenced post 9-11	-.003
	(.004)
DD Impact: POST*H*2001 - POST*H	-.055***
	(.021)
Confidence Interval	[-.096, -.013]
Offender, Legal and District Controls	Yes
Offense Type Codes	Final
Guideline Cells	Yes
Adjusted R-squared	.243
Unadjusted R-squared	-
Observations	114,642

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown. Standard errors are reported in parentheses, where these are clustered by group-district. The sample of cases used are those 114,642 cases for which sentencing occurs within a 6-month window of 9/11 in years 1998 to 2001. For those defendants sentenced after 9/11 each year, the last offense was committed prior to 9/11 that year, and if sentenced before 9/11 each year, the last offense was committed at least 180 days prior to 9/11 that year. The dependent variable is a dummy for whether the case receives a downwards departure. We condition on defendant group (White, Black, Hispanic) whether the case comes up post 9-11, and interactions between the two, and three way interactions between a post 9/11 dummy, a dummy for the 2001 NE period, and ethnicity. Throughout the following additional controls are included: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, the guideline cell, and Federal district dummies. At the foot of each Column we report the estimate of the common impact, the difference between the sentenced post-9/11 x 2001 interaction and the sentenced post-9/11 dummy, its standard error and confidence interval.

Table A9: Bush Appointed US Attorneys

Dependent Variable: Downward Departure Granted by Federal Judge
Standard errors in parentheses clustered by group-district

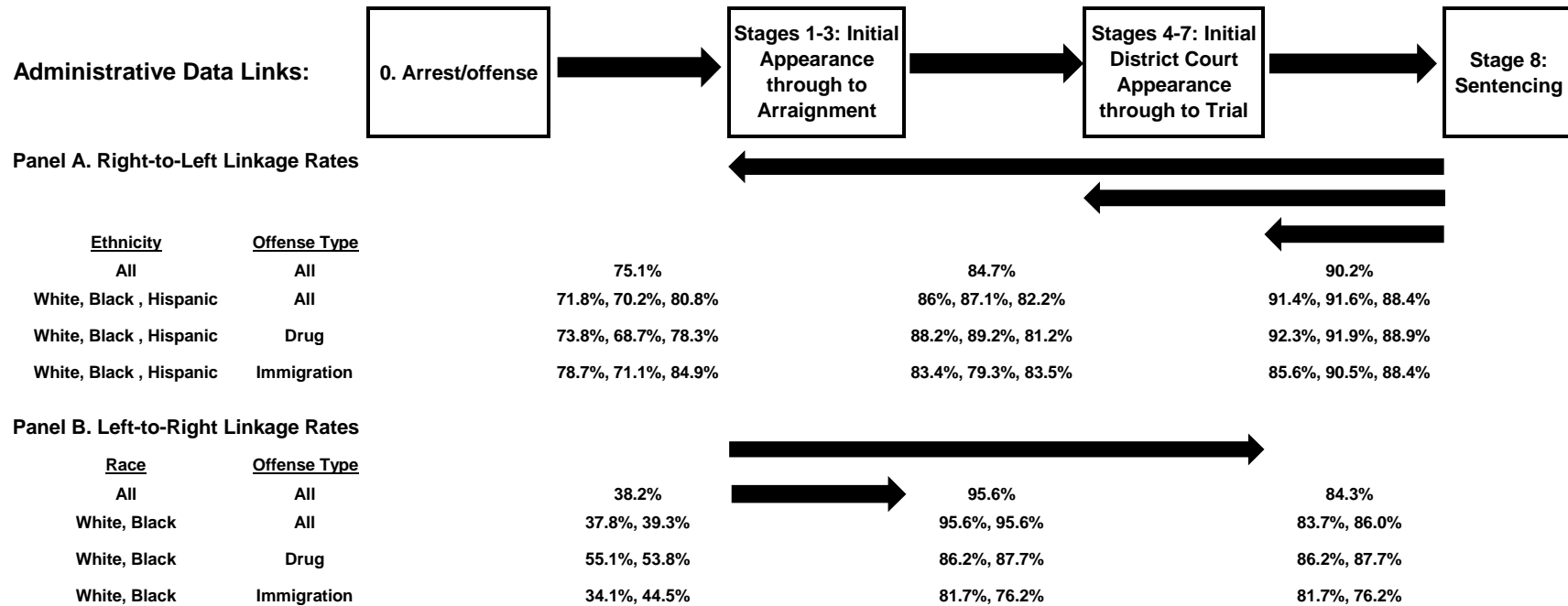|	(1) Downward Departure
Sentenced post 9-11*Hispanic	-.039***
	(.013)
Sentenced post 9-11*Hispanic*Post-period share under Bush US Attorney	.005
	(.028)
Sentenced post 9-11*Black	-.012
	(.009)
Sentenced post 9-11*Black*Post-period share under Bush US Attorney	.015
	(.018)
Sentenced post 9-11	.004
	(.007)
Sentenced post 9-11*Post-period share under Bush US Attorney	-.030
	(.018)
Offender, Legal and District Controls	Yes
Offense Type Codes	Final
Guideline Cells	Yes
Implied Sentence Length Impact (H)	.820
% of Pre 9-11 Ethnic Differential	20.1%
p-value: [Post*B = Post*H]	.022
Adjusted R-squared	.257
Observations	40,228

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown in all Columns. Standard errors are reported in parentheses, where these are clustered by ethnicity-district. The sample of 40,228 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. The dependent variable is a dummy for whether the case receives a downwards departure. We condition on defendant ethnicity (White, Black, Hispanic), whether the case comes up post 9-11, and interactions between the two, and the following additional controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, the guideline cell, and Federal district dummies. The share of time the district spends in the post period with a Bush appointed US Attorney is measured in deviation from mean. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Black and post 9-11 x Hispanic dummy interactions are equal against a two sided alternative.

Table A10: Treatment Effect Bounds on Judge's Decision Making Around 9-11

Dependent Variable: Downward Departure Granted by Federal Judge
Standard errors in parentheses clustered by ethnicity-district

	(1) Baseline	(2) Reason: Criminal History Category Over Represented	(3) Reason: Pursuant to Plea Bargain	(4) Reason: General Mitigating Circumstances	(5) Reason: Other	(6) Initial Arrest Codes
Sentenced post 9-11*Hispanic	-0.038***	-0.013***	-0.011	-0.001	-0.013*	-0.046**
	(0.013)	(0.003)	(0.007)	(0.008)	(0.007)	(0.019)
[Bounds: $\delta_{H(0)}$, $\delta_{H(1)}$]	[-.038, -.036]	[-.013, -.013]	[-.011, -.01]	[-.001, -.001]	[-.013, -.012]	[-.048, -.046]
τ required for coefficient of 0	27.680	-42.548	15.987	-4.931	16.336	-16.162
Sentenced post 9-11*Black	-0.013	-0.005	0.002	-0.003	-0.007	-0.013
	(0.008)	(0.004)	(0.003)	(0.004)	(0.005)	(0.011)
[Bounds: $\delta_{H(0)}$, $\delta_{H(1)}$]	[-.013, -.012]	[-.006, -.005]	[.002, .003]	[-.003, -.003]	[-.007, -.007]	[-.013, -.012]
τ required for coefficient of 0	23.713	-16.561	-1.552	-149.458	-22.410	23.907
Sentenced post 9-11	0.006	0.003	-0.000	0.001	0.002	0.003
	(0.007)	(0.002)	(0.002)	(0.004)	(0.004)	(0.009)
[Bounds: $\delta_{H(0)}$, $\delta_{H(1)}$]	[.006, .006]	[.003, .003]	[-.001, 0]	[.001, .002]	[.002, .002]	[.003, .004]
τ required for coefficient of 0	-15.843	-8.373	-0.805	-21.337	-6.170	-4.274
Offender, Legal and District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Offense Type Codes	Final	Final	Final	Final	Final	Arrest
Guideline Cells	Yes	Yes	Yes	Yes	Yes	No
p-value: [Post*B = Post*H]	.042	.037	.096	.795	.306	.063
Implied Sentence Length Impact (H)	.736					.889
% of Pre 9-11 Ethnic Differential	18%					29.8%
Unadjusted R-squared	0.264	0.052	0.296	0.077	0.143	0.267
R^{max}=min(1, 1.3 x unadjusted R-squared)	0.343	0.067	0.385	0.100	0.186	0.347
Adjusted R-squared	0.256	0.042	0.289	0.068	0.135	0.257
Observations	40,228	40,228	40,228	40,228	40,228	26,852

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown in all Columns. Standard errors are reported in parentheses, where these are clustered by ethnicity-district. In Columns 1 to 5, the sample of 40,228 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. Columns 2 to 5 code downdawrd departures into various broad categories of how judge's justify their decision to depart. In Column 6 the sample is restricted to those cases that can be linked back to arrest (Stage 0). The dependent variable throughout is a dummy for whether the case receives a downwards departure (where in Columns 2 to 5 this is modified based on the reasons given for departure). In all Columns we condition on defendant ethnicity (White, Black, Hispanic), whether the case comes up post 9-11, and interactions between the two, and the following additional controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the guideline cell, and Federal district dummies. In Columns 1 to 7 we control for the primary offense type. In Column 6 we instead control for arrest offense codes, but not guideline cells. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Black and post 9-11 x Hispanic dummy interactions are equal against a two sided alternative.

Figure A1: Linkage Rates Across Administrative Data Sets

Administrative Data Links:

| 0. Arrest/offense | → | Stages 1-3: Initial Appearance through to Arraignment | → | Stages 4-7: Initial District Court Appearance through to Trial | → | Stage 8: Sentencing |

Panel A. Right-to-Left Linkage Rates

Ethnicity	Offense Type			
All	**All**	75.1%	84.7%	90.2%
White, Black , Hispanic	**All**	71.8%, 70.2%, 80.8%	86%, 87.1%, 82.2%	91.4%, 91.6%, 88.4%
White, Black , Hispanic	**Drug**	73.8%, 68.7%, 78.3%	88.2%, 89.2%, 81.2%	92.3%, 91.9%, 88.9%
White, Black , Hispanic	**Immigration**	78.7%, 71.1%, 84.9%	83.4%, 79.3%, 83.5%	85.6%, 90.5%, 88.4%

Panel B. Left-to-Right Linkage Rates

| Race	Offense Type			
All	**All**	38.2%	95.6%	84.3%
White, Black	**All**	37.8%, 39.3%	95.6%, 95.6%	83.7%, 86.0%
White, Black	**Drug**	55.1%, 53.8%	86.2%, 87.7%	86.2%, 87.7%
White, Black	**Immigration**	34.1%, 44.5%	81.7%, 76.2%	81.7%, 76.2%

Notes: We use the Federal Justice Statistics Resource Center (FJSRC) data set for our analysis. This comprises information gathered from four linked administrative data sources. As described in Rehavi and Starr [2014], the four linked data sets are: (i) US Marshals Service (USMS) data, that covers the arrest/offense stage (Stage 0) and includes all persons arrested by Federal law enforcement agencies, persons arrested by local officials and then transferred to Federal custody, and persons who avoid arrest by self-surrendering; (ii) Executive Office for US Attorneys (EOUSA) data, covering initial appearance through to arraignment (Stages 1-3): these data come from the internal case database used by Federal prosecutors, and covers every case in which any prosecutor at a US Attorney's office opens a file; (iii) Administrative Office of the US Courts (AOUSC) data, covering initial district court appearances through to trial (Stages 4-7): these originate from Federal Courts and contain data on all criminal cases heard by Federal district judges, and any non-petty charge handled by a Federal magistrate judge; (iv) the Monitoring of Federal Criminal Sentences (MFCS) data set that is resourced from US Sentencing Commission (USSC), covering the sentencing Stage 8: this data set collects information on any case that results in conviction and sentencing for a non-petty offense. These data are collected by the Bureau of Justice Statistics.

Figure A2: Judicial Decision Making, Other Channels

A. Within Guideline Cell Sentence Length | Sentenced Within Guideline Cell
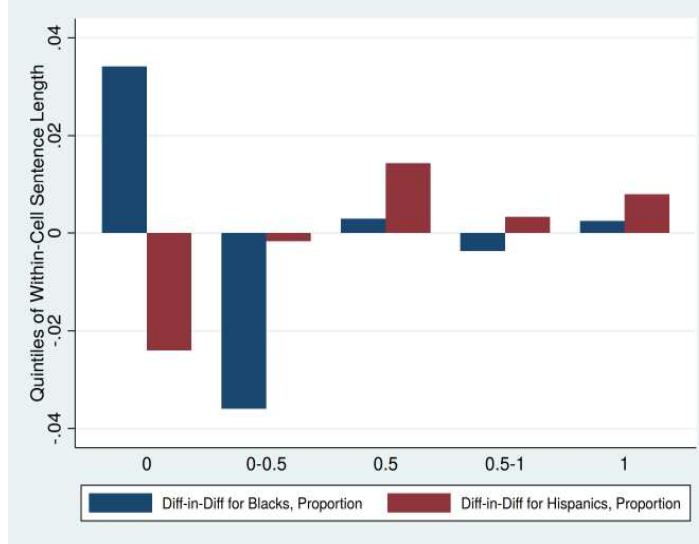

## B. Cell Movements | Downward Departure = 1
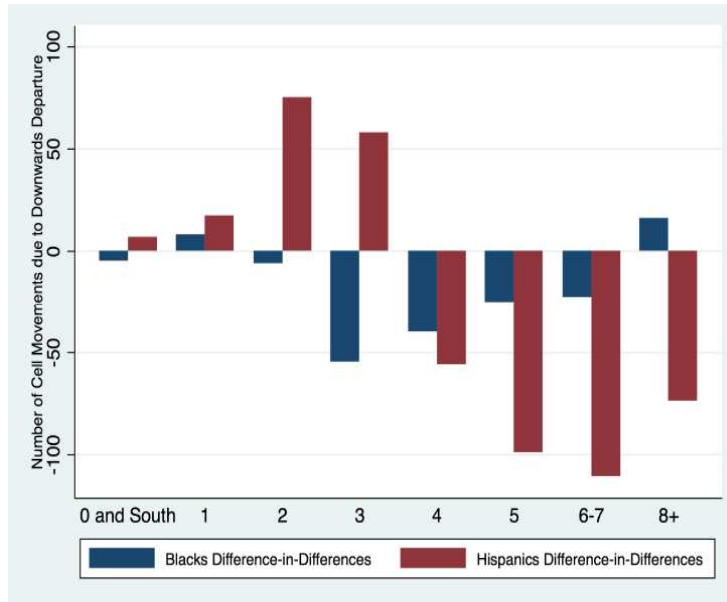
Notes: The data used to construct the figures is the Natural Experiment sample used throughout the paper. That is, the sample of 40,228 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001). For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. Panel A restricts the NE sample to those defendants who received a sentence within the range of their guideline cell. For these defendants, we calculate a variable that is within-cell position (based on the received sentence, and the lower and upper ranges of the respective guideline cells) and normalized these by the width of the cell in months. We then collapse this continuous variable into the quintiles displayed in the graph. The results presented are the difference-in-differences changes in the frequency of both blacks and Hispanics to lie within each of these quintiles, where White defendants are the reference group. Panel B restricts the data to those individuals who received a downwards departure. In absence of having information on the cell defendants were allocated to post-departure, we created a variable that compared sentence received with the recommend sentence length mid-points of less punitive guideline cells (cells lying to the north in Table A1) from their guideline cell based on their offense severity and criminal history. Defendants were allocated to a final cell that minimized the distance between received sentence and more northerly guideline cell sentencing mid-points. The dependent variable in Panel B counts the number of guideline cells moved to get from their initial cell to their allocated cell based on the algorithm described above. The results presented show the difference-in-differences changes in frequency of cell movements for both Blacks and Hispanics, where White defendants are the reference group.

Figure A3: Patriot Act
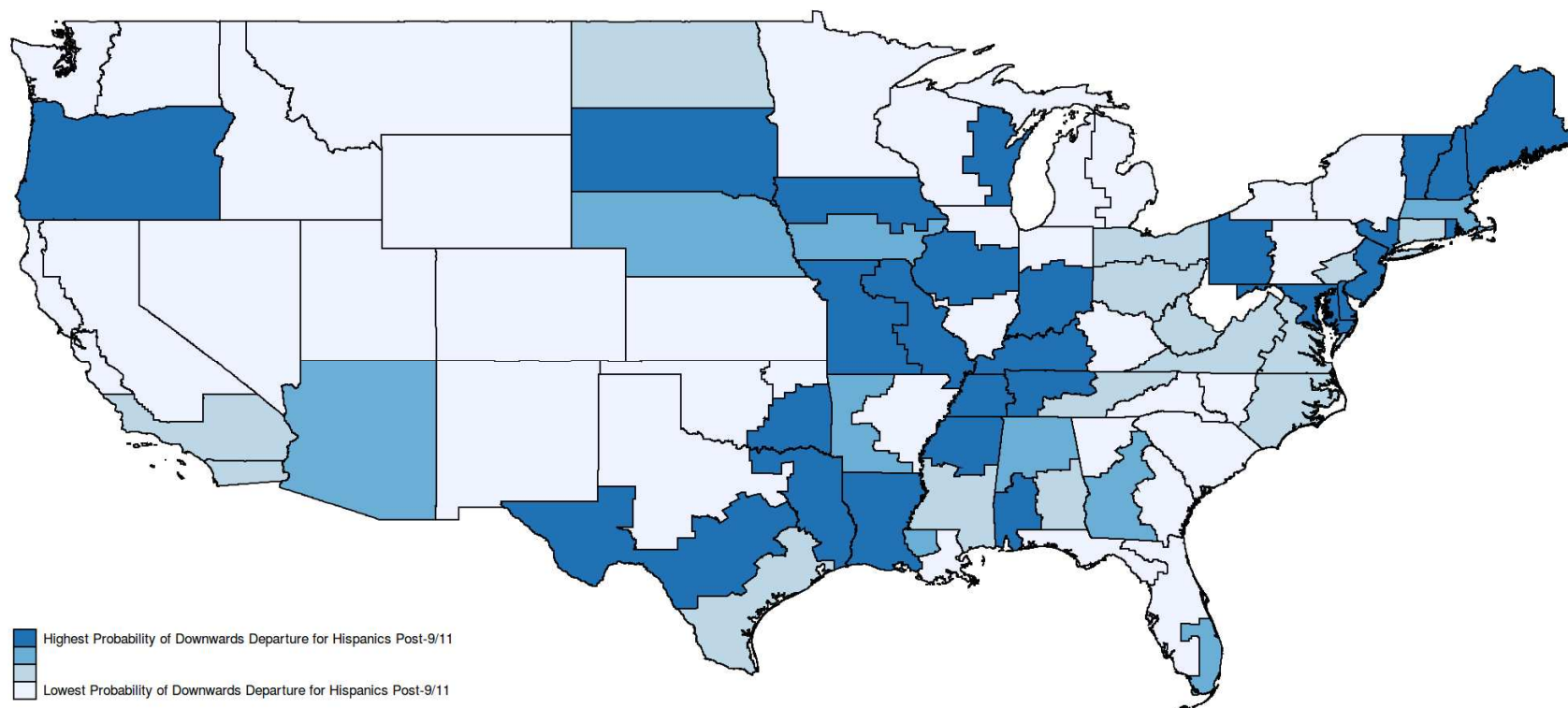### Hispanics: Non-PA Offences, Downwards Departure
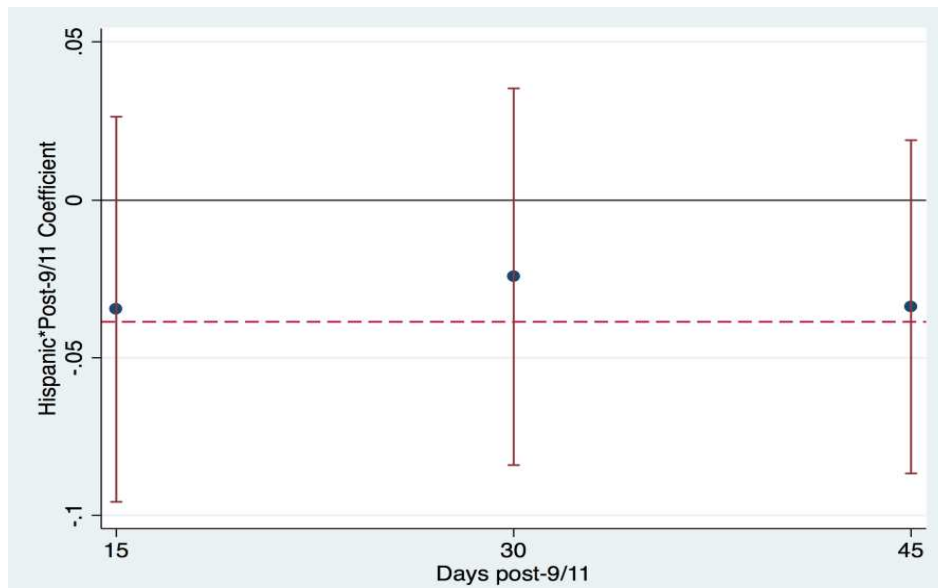
Notes: Panel A is based on the NE sample, where 40,228 Federal cases are used (those that come up for sentencing in a six month window either side of 9/11/2001). Panels B and C are based on the same sample, except that Patriot Act related offenses (Money Laundering and Immigration) are excluded, resulting in a sample of 32,930 cases. For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. The dependent variable in Panel A is a dummy for whether the case receives a downwards departure. The dependent variable in Panels B and C is a dummy for whether any prison sentence is given. In all three graphs the output is shown for results from a specific form of the main difference-in-differences regressions presented in the paper, where we divide the post-9/11 period into 15 day windows, and we show the coefficients for the first three such periods (and their associated standard error). In each Panel, the dashed line shows the corresponding estimate for the NE sample assuming a homogenous post impact. In the first panel, the regression coefficients for the Hispanic*post-9/11 terms are shown. In the remaining panels, the equivalent for post-9/11 is presented. In all regressions we condition on the following additional controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type, the guideline cell, and Federal district

Figure A4: Bush Appointed District Attorneys
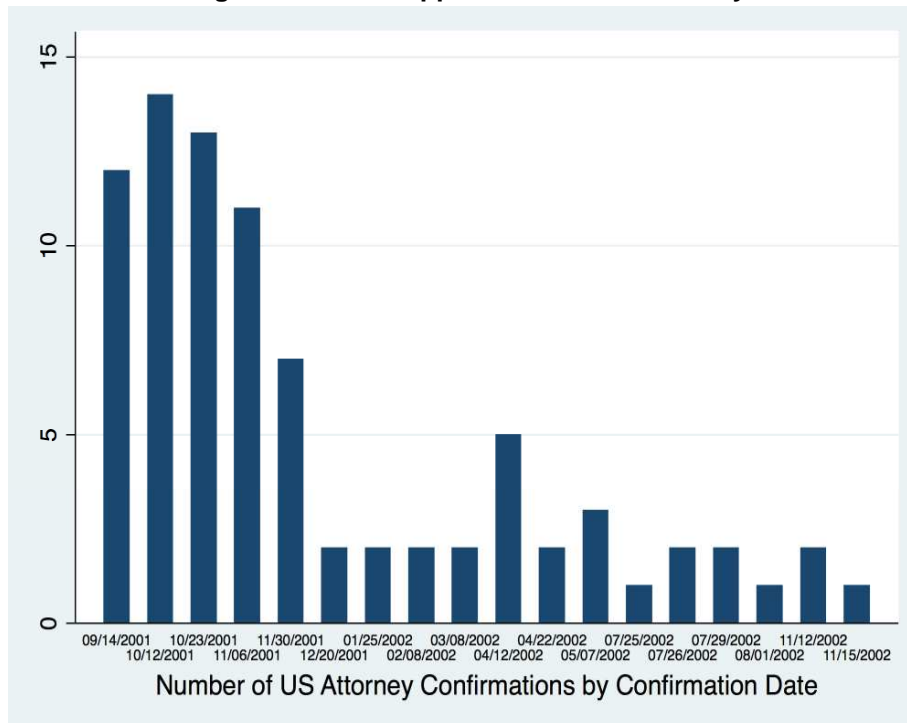
Notes: Data sourced from https://www.congress.gov/ for nominations heard by the Senate Committee: Judiciary for the years 2001-2002. The sample consists of all US attorney confirmations during this time period.

Figure A5: Spatial Pattern of Hispanic-White Sentencing Differentials

Highest Probability of Downwards Departure for Hispanics Post-9/11

Lowest Probability of Downwards Departure for Hispanics Post-9/11

Notes: For each Federal court district, we plot the coefficient on post 9-11*Hispanic*District from a difference-in-difference-in-difference regression for the Natural Experiment sample period where in this first stage the full set of controls is included, and the dependent variable is whether a downwards departure is granted. These coefficients are split into quartiles so that darker districts represent those where the probability of a downward departure is highest.