Hispanic-White Sentencing Differentials in the Federal Criminal Justice System*

Brendon McConnell Imran Rasul

June 2017

Abstract

In the Federal criminal justice system (CJS), large Hispanic-White differences in sentencing outcomes exist. We examine the malleability of factors that might drive such differences, such as biases against Hispanics. To do so, we exploit 9-11 as an exogenously timed cue heightening the salience of insider-outsider divisions, that might have been especially impactful on Hispanics given interlinkages drawn between terrorism, border security and immigration. Exploiting linked administrative data covering criminal cases from arrest through to sentencing, we use a DiD research design based on defendants all of whom were arrested pre 9-11, but who came up for sentencing either side of 9-11. We find that among those sentenced post 9-11, Hispanic-White judicial sentencing differentials are exacerbated relative to these sentenced pre 9-11, while Black-White sentencing differentials are unaffected. Our data and research design allows us to further document the differential treatment of Hispanic defendants by prosecutors in pre-sentencing stages of the CJS, such as with regards to the initial offense charges set. To shed light on the origins of these judicial and prosecutorial sentencing differentials, we use decomposition analysis to show the vast majority of sentencing differentials are driven by unobservables rather than sentencing prices on observables such as offense type or criminal history. Furthermore, we document that in districts with a higher proportion of Hispanic judges, Hispanic-White differentials are significantly reduced, consistent with judges' ingroup biases influencing their decision making. Our results provide insights into the magnitude, channels and malleability of Hispanic-White sentencing differentials in the professional and high-stakes Federal CJS. *JEL Classification: J15, K14.*

*We gratefully acknowledge financial support from the Dr. Theo and Friedl Schoeller Research Center for Business and Society, and the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS (grant number RES-544-28-5001). We thank Oriana Bandiera, Patrick Bayer, Daniel Bennett, Marianne Bertrand, Pietro Biroli, Dan Black, David Card, Kerwin Charles, Steve Cicala, Gordon Dahl, Brad DeLong, Ben Faber, Rick Hornbeck, Randi Hjalmarsson, Emir Kamenica, Kevin Lang, Neale Mahoney, Alan Manning, Olivier Marie, Ioana Marinescu, Michael Mueller-Smith, Emily Owens, Daniele Paserman, Steve Pischke, Steven Raphael, Jesse Rothstein, Anna Sandberg, Johannes Schmeider, Robert Topel and numerous seminar and conference participants for valuable comments. All errors remain our own. Author affiliations and contacts: McConnell (Southampton, brendon.mcconnell@gmail.com); Rasul (University College London, i.rasul@ucl.ac.uk).

1 Introduction

Ethnic minority men are far more likely to come into contact with the Federal criminal justice system (CJS) than White men. Beyond the frequency of contact with the criminal justice system, decades of research have further shown sentencing *outcomes* also vary by ethnicity. While much of this has focused on Black-White differences, Hispanics are now the modal defendant in the Federal CJS and the group whose incarceration rate is growing fastest. As a result, Hispanic men are four times as likely to go to prison during their lives as Whites [Starr and Rehavi 2013]. It has become increasingly noted that Hispanics are a minority group that have been somewhat understudied relative to others in the economics of discrimination literature more generally [Charles and Guryan 2011], despite Hispanics being ever more prominent in the political, legal and cultural life of America. This is an evidence gap we start to fill in this paper.

The central econometric challenge lies in understanding whether sentencing differentials are driven by unobserved heterogeneity correlated to defendant ethnicity, or whether they reflect discrimination. The question is important given that equality before the law is a cornerstone of any judicial system, and because it is difficult to know what could be done to reduce sentencing disparities if their underlying causes remain unknown. Our contribution is to combine linked administrative data with a research design to examine the malleability of factors that might drive such differences, such as bias against Hispanics in the CJS. We use this to provide insights into the magnitude, channels and malleability of Hispanic-White sentencing differentials in the Federal criminal justice system.

We use the *Monitoring of Federal Criminal Sentences* (MFCS) data set: this comprises information from 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 shown in Figure 1. A prominent set of papers have used State or lower court data to estimate the causal impact of sentence length on criminal and economic outcomes. These have exploited the random assignment of judges to cases, so that variation in the harshness of judges leads to exogenous variation in the sentences received by defendants [Kling 2006, Abrams *et al.* 2012, Aizer and Doyle 2015, Mueller-Smith 2016]. In Federal court data, even though judges are randomly assigned, judge identifiers are typically unavailable because criminal cases considered are more serious and often of national importance.

The key advantages of Federal criminal court data however relate to being able to tackle long-standing challenges for empirical work on the CJS [Klepper *et al.* 1983]: (i) it is nationally representative, covering cases from all 90 mainland US Districts, defendants of all ages, and all types of criminal offense; (ii) the linked administrative data allows pre-sentencing differential

treatment arising from the behavior of prosecutors or legal counsel to be studied alongside the behavior of judges; (iii) large samples allow for both Black-White and Hispanic-White differentials to be studied: our data covers 230,000 Federal criminal cases occurring between 1998 and 2003.

Our research design allows us to make progress on the malleability of factors that driving Hispanic-White sentencing differentials in the US Federal CJS. A candidate explanation behind such differentials is ingroup bias against Hispanics. There is a vast literature examining the biological and evolutionary roots of ingroup bias, and they have been extensively documented in lab and field settings [Shayo 2009, Bertrand and Duflo 2016]. There is also evidence for ingroup bias in judicial contexts including US State courts [Bushway and Piehl 2001, Shayo and Zussman 2011, Abrams *et al.* 2012, Anwar *et al.* 2012, Rehavi and Starr 2014]. Our analysis sheds new light on whether such biases partly determine Hispanic-White sentencing differentials in the high stakes environment of the Federal criminal justice system, and whether such biases are malleable in a setting where decisions are made by professional and experienced judges, prosecutors and legal counsel, and the universe of criminal offenses and district courts can be studied.

We consider 9-11 as an exogenously timed event that heightened the salience of insider-outsider differences in US society [Human Rights Watch 2002, Davis 2007, Woods 2011]. To understand the link between 9-11 and Hispanics, 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 [Romeo and Zarrugh 2016]. We present qualitative and quantitative evidence that post 9-11, among average Americans, anti-immigration and anti-Hispanic sentiment rose and was somewhat persistent. Our analysis then sheds light on whether such sentiments might also bias the decision making of judges and prosecutors against Hispanics post 9-11 in the high stakes and professional environment of the Federal criminal justice system, thus further exacerbating existing Hispanic-White sentencing differentials in this period.

To isolate the impact the event had on sentencing outcomes in the Federal CJS, we compare sentencing outcomes 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 ethnicities to estimate a difference-in-difference (DiD) impact of 9-11 on sentencing. 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, where these guidelines provide for *determinate* sentencing. Table A1 shows the full set of guideline cells, mapping combinations of the severity of the offense and the defendant's criminal history into a specific sentencing range. 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 it 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.

We first confirm that relative to Whites, Hispanics sentenced pre 9-11 receive significantly longer prison sentences on average, and that these differences are unlikely to be explained solely by unobserved heterogeneity across defendants correlated to their ethnicity. Hispanics sentenced post 9-11 when the salience of insider-outside divisions is heightened, sentencing differentials become even further exacerbated through a specific channel: Hispanics become 13.5% less likely to receive a downward departure than Whites. The implied impact on Hispanic sentence lengths to be .736 months, corresponding to 43% of the unconditional pre 9-11 Hispanic-White sentence differential, or 18% of the conditional pre 9-11 differential. Placing a monetary value on this increased incarceration suggests the heightened salience of insider-outsider differences post 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.

Black-White sentencing differentials around 9-11 are unaffected along all sentencing margins.

To underpin a causal interpretation of these results, we provide evidence for two identifying assumptions. 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. Second, using data from other years to construct placebo 9-11 impacts, we show there are no ethnicity-time effects in ethnic sentencing differentials that occur naturally around 9-11 each year.

As has long been recognized [Klepper *et al.* 1983] a range of legal actors beyond judges are involved in the Federal CJS, and their behaviors: (i) can lead to differential treatment by 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 [Starr and Rehavi 2013]. We address the issue by combining the linked administrative data with our research design to consider decisions made at earlier stages of the case timeline on Figure 1, where we move our 9-11 window to when these other decisions are being made. We examine: (i) prosecutor decisions over which initial

offense charge to file; (ii) the initiation of substantial assistance departures by prosecutors, that are often given in recognition of defendant cooperation; (iii) prosecutor-legal counsel interactions in drafting pre-sentence reports that provide judges with a recommended guideline cell. Echoing the findings of Rehavi and Starr [2014], we find using our research design around 9-11, 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, 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 responses to 9-11 leaves the Hispanic-White differential on each margin to overall become at least as large as the Black-White differential.

Having established the key decisions of Federal judges and prosecutors that drive differential outcomes across ethnicities, and shown these margins to be malleable to outside events, we then probe the data to understand the *origins* of the documented differentials using two strategies: (i) decomposition analysis; (ii) correlating ethnic sentencing differential to Federal judge characteristics, including their ethnicity.

We use a Juhn-Murphy-Pierce decomposition of sentencing differentials for two cohorts of Hispanic defendant: (i) the cohort that come up for sentencing just post 9-11, who are significantly less likely than Whites to receive a downwards departures from judges; (ii) the cohort whose initial offense charges are set by Federal prosecutors post 9-11, who are charged with offenses with significantly longer statutory minimum sentences. For both cohorts, the JMP decomposition of sentencing outcomes shows these differences are largely driven by changes in unobserved drivers of sentencing outcomes; only negligible amounts of each cohort's unconditional DiD in outcome can be attributed to either the DiD in their observables relative to Whites, or the sentencing prices of such observables. This helps to rule out explanations for the 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.

On judge characteristics, we analyze how they correlate to our estimated Hispanic-White sentencing differential. We hand-coded 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 judges, the Hispanic-White sentencing differential for downward

departures is significantly reduced, conditional on other judge characteristics and demographic characteristics of the Federal district. 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]. As with the decomposition analysis, these results run counter to statistical discrimination explaining our findings, and rather suggests 9-11 primed judges to display ingroup biases [Schanzenbach 2005, Abrams *et al.* 2012].

The literature has studied three sources of ethnic sentencing differential [Fischman and Schanzenback 2012]: (i) judicial bias; (ii) prosecutorial bias; (iii) sentencing policies. Our central contribution is to provide new insights for Hispanic-White sentencing differentials on the first two dimensions in the context of the high stakes and professional environment of the Federal CJS by combining linked administrative data with a novel research design. We show that 9-11 cues ingroup biases determine sentencing outcomes and so highlight the fact that biases driving differential outcomes by ethnicity are malleable. We further advance the literature by pinpointing the separate roles that judges and prosecutors have in driving the differential treatment of Hispanic defendants in the Federal CJS [Shayo and Zussman 2011, Abrams *et al.* 2012, Rehavi and Starr 2014]. By showing 9-11 potentially cued the salience of inside-outsider divisions in American society and this impacted decision making in the Federal CJS, 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].[1]

The paper is organized as follows. Section 2 describes the Federal CJS, sentencing guidelines,

[1]Salience theory of judicial decisions [Bordala *et al.* 2015] provides a theoretical underpinning to judicial bias. This has a central premise that the evaluation of choices occurs in a comparative context: hence in evaluating a range of options, attention is drawn to unusual, extreme or salient attributes. In terms of evidence, work has documented judicial decision making being impacted by contextual factors [Rachlinski 1996, Kelman *et al.* 1996, Guthrie *et al.* 2001], behavioral biases (such as anchoring and gamblers fallacy) [Englich *et al.* 2006, Chen *et al.* 2014], extraneous factors (such as caseload sequencing and lunch breaks) [Danzinger *et al.* 2011], and media reports biasing juries [Philippe and Ouss 2016]. Fewer papers have linked such biases to ethnic sentencing differentials [Abrams *et al.* 2012, Anwar *et al.* 2012, Eren and Mocan [2016]. Rachlinski *et al.* [2009] present evidence on racial bias from implicit association tests on judges ($N = 133$). They find judges harbor implicit racial biases and these can influence judgements, but that these biases can also be offset given sufficient motivation. Their sample is too small to look beyond Black-White differences. On prosecutorial biases, Rehavi and Starr [2014] use similar linked administrative data from the FCJS 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 replicate and extend their findings to the pre 9-11 period, when sentencing guidelines are in place.

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 highly interlinked in the immediate aftermath of 9-11, and this might plausibly have cued bias towards Hispanics among decision makers in the Federal CJS. Sections 4 and 5 presents our core findings on ethnic sentencing differentials, as driven by judicial and prosecutorial decision making respectively. Section 6 presents evidence on the origins of these sentencing differentials using decomposition analysis and judge characteristics. Section 7 concludes. The Appendix contains further data details and robustness checks.

2 The Federal Criminal Justice System

Criminal cases are filed in Federal court if an individual is prosecuted by a Federal agency or breaks a Federal law. 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 DiD research design we use to estimate Hispanic-White sentencing differentials eliminates cross sectional differences between defendants, by ethnicity, being sent to trial in the Federal system.[2]

As criminal cases heard in Federal courts tend to be more serious than those in State courts, the types of offense considered differ from those in State courts. For example, in 2000 the three most frequent criminal offenses filed in Federal courts were for drug trafficking (40%), immigration (22%), and fraud (9%), while at the State level the most frequent criminal cases related to drug sales (19%), other drug offenses (18%), and assault (10%). 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.[3]

The legal actors determining sentencing outcomes in Federal criminal cases are judges, prosecutors, the defendant's legal counsel, and juries. Judges in Federal courts are nominated by the

[2] Glaeser *et al.* [2000] provide a theoretical and empirical analysis of the sorting of cases into State and Federal systems, exploiting the gradual increase in drug related offenses falling under the remit of Federal courts. Their model highlights that when taking on cases Federal prosecutors balance the social costs of crime with private career concerns. They find evidence suggesting Federal prosecutors are more likely to take to trial high-human capital criminals, consistent with both the social costs motives (as they have more resources than state prosecutors) but also career concerns (because of the prestige of pursuing such criminals, and the possibility for greater learning on the job as they are then up against good public defenders).

[3] 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.

President, confirmed by Congress, and appointed for life (in contrast, State court judges can be elected, appointed or a combination). There are just over 7 Federal judges per district, so that there are around 700 in total: they are among the most senior judges in the country, and *a priori*, might be considered among those least susceptible to biased judgments. 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.

Legal counsel in Federal courts differs from State courts: 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 (that represent less than .1% of cases) [Alesina and La Ferrara 2015].

2.1 Timeline

Figure 1 shows the timeline of Federal criminal cases. 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 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), given this is where judges exercise their discretion over defendant outcomes, and as 96% of defendants are already convicted, only their punishment remains to be determined. The ethnic sentencing differentials we measure in relation to judicial decision making, are conditional on defendant's reaching sentencing Stage 8. This includes conditioning on the guideline cell recommended to the judge in the pre-sentence report drawn up by the defense counsel and prosecutor between trial and sentencing.

Multiple legal actors are involved at earlier stages, and: (i) their behaviors can lead to differen-

tial treatment of defendants pre-sentencing; (ii) the presence of biases earlier in the timeline might not be detected in judicial sentencing differentials. This might especially be so when sentencing guidelines are in place as these restrict the discretion of judges and potentially increase the power of prosecutors [Starr and Rehavi 2013]. In Section 5 we exploit the linked administrative data to consider earlier stages to pin point how other legal actors drive ethnic 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]. A novel aspect of our analysis is that it allows us to measure whether the sentencing behavior of Federal judges reinforces or offsets the behavior of other legal actors with regards to Hispanic-White sentencing differentials.

2.2 Linked Administrative Data

We use the *Monitoring of Federal Criminal Sentences* (MFCS) data set for our analysis. This comprises information gathered from 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 [USSC MCFS 1999-2003]. The Appendix provides further data details.

To estimate Hispanic-White and Black-White sentencing differentials, we use two variables available at the sentencing Stage 8 in the *MFCS* data. In one variable, defendants are classed as either Hispanic (41%) or non-Hispanic (59%). A separate race code then separately identifies defendants as white-race (71%), black-race (29%), other-race ($< .1\%$). Whites are then coded as white-race and non-Hispanic; Blacks as black-race and non-Hispanic; Hispanics as white- or black-race and Hispanic. This implies 31% of defendants are ethnically White, 26% are Black and 43% are Hispanic.[4]

The *MFCS* data contains a rich set of information for each criminal case: defendant demographics include their 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), and offense details are recorded that we use to classify the offense into 31 various types.[5] Most importantly, the data records the guideline cell recommended to the

[4]The other-race classifications include American Indian/Alaskan Native, Asian/Pacific Islander, multi-racial and other. The *MFCS* data thus does not contain an identifier for Arabs nor Muslims, and so those groups are not the focus of our study (even if such identifiers existed, the numbers of such defendants would be miniscule, corresponding to less than .1% of criminal cases). Using our coding, 92% of Hispanics are white-race.

[5]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, racketeering (including gambling/lottery), civil rights offenses, immigration, pornography/prostitution, offenses in prisons,

judge in the pre-sentence report. This effectively proxies all case-specific factors the prosecution and legal counsel deem judges should factor into their sentencing decision. Finally the data record the Federal court district of sentencing. Table A3a shows the sample descriptives for the *MCFS* full sample of cases, as well as the working sample we use for our analysis based on the 94% of cases in which there is no missing data on the core covariates.

2.3 Linkage Rates

A concern when studying sentencing outcomes is that there can be selection of defendants into this stage of the CJS [Klepper *et al.* 1983]: as the result of actions of various legal actors through the case timeline, the set of cases that reach sentencing might not be representative of the original population of arrested and charged defendants. As the *MCFS* 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. We first consider cases observed at sentencing Stage 8, and estimate linkage rates to the *earlier* administrative records, as shown in Panel A in the lower part of Figure 1 (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 our *MCFS* data. We see that: (i) 90% of cases are also observed in the preceding administrative data (covering Stages 4-7); (ii) 85% of cases observed at sentencing can be further linked back to the two earlier administrative data sets (covering Stages 1-7); (iii) 75% of cases observed at sentencing can be linked back to arrest/offense stage. Linkage rates are quite similar across ethnicities: 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 ethnicities, 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 *MCFS* data originates from multiple agencies.

We next construct linkage rates from the arrest/offense stage through to sentencing, as shown in Panel B in the lower part of Figure 1 (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 our *MCFS* data. The underlying

environmental, national defense offenses, antitrust violations, food and drug offenses, traffic violations and other smaller categories.

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. 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%).

To reiterate, the difference-in-difference research design we utilize to estimate ethnic sentencing differentials eliminates cross sectional differences between defendants of different ethnicity (such as in linkage rates) among those assigned to be tried in the Federal system.

2.4 Federal Sentencing Guidelines

Federal sentencing guidelines were introduced in the Sentencing Reform Act of 1984 by the US Sentencing Commission (USSC). The explicit goal of the reform was to alleviate sentencing disparities that research had indicated were prevalent in the Federal CJS. This was to be achieved by the guidelines providing for *determinate* sentencing, whereby: (i) the discretion judges had over penalties imposed at the sentencing stage became more limited; (ii) parole boards were abolished so that determined sentences matched the actual period of incarceration far more closely.[6]

The USSC 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

[6]This is in contrast to the prior system of indeterminate sentencing, in which a sentence with a maximum (and perhaps a minimum) was pronounced by a judge, but the actual time served in prison was determined by a parole commission after the sentence began. As part of the same reforms, such parole on Federal cases was abolished. The notion that the majority of a Federal court sentence should be served is also something that has become strengthened by other Federal laws, such as truth-in-sentence (TIS) laws, that further eliminate or restrict parole and/or remissions. In 1994, a Federal TIS law stated that to qualify for TIS Federal funding, offenders must serve at least 85% of the sentence for qualifying crimes before becoming eligible for parole. As of 2008, 36 states qualified for this additional funding.

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 would then stipulate they should be sentenced in the range of 70-87 months.

Table A1 shows the full set of guideline cells, mapping each possible combination of offense severity (1 to 43) and criminal history (scores 1 to 13, grouped into 6 bins) into a sentencing range. Hence there are 43 x 6 = 258 guideline cells. These include those in Zone A on Table A1, where the guidelines include zero sentence length, and cells in Zone D where the guidelines impose a life sentence. Accounting for the empirical distribution of offense severity and criminal histories, the expected width of a guideline cell is 15 months, and the sentencing range within a guideline cell therefore corresponds to around 25% of the minimum sentence [Schanzenbach 2005].

Between trial/conviction and sentencing (Stages 7 and 8), the pre-sentence report is drafted by prosecutors and legal counsel, and this specifies a recommended guideline cell. However, the sentencing guidelines still provide *judges* discretion over which guideline cell to ultimately place a defendant in. They allow a judge to *downwards depart* from the recommended guideline cell, and so 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.[7]

In our sample of 230,000 Federal criminal cases from October 1998 to September 2003, judges grant downwards departure in 17% of cases. Downward departures result 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 typically associated with reductions in sentence length; (iii) they are likely correlated to the prison conditions under which incarceration is served, and this in turn might impact recidivism and other future behaviors through the accumulation of criminal capital [Bayer *et al.* 2009]. The null hypothesis for our analysis is based

[7] In Section 5 we separately examine substantial assistance departures: these originate from the prosecution and are given on the basis of the defendant providing substantial assistance toward the prosecution of others.

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].*[8]

3 Descriptives, 9-11, Research Design

3.1 Pre 9-11 Sentencing Differentials

We first present motivating evidence on pre 9-11 ethnic sentencing differentials in the Federal CJS, so cases up for sentencing between October 1st 1998 and September 10th 2001. Descriptively, we consider three margins of judicial decision making: (i) if a downward departure is granted; (ii) the number of guideline cells moved (including zero, using the convention that a Northwards move of one cell corresponds to +1 cells moved, and using midpoints of guidelines cells to establish the guideline cell moved to in case of a downwards departure); (iii) the sentence length (in months).

In Table 1, Columns 1, 3 and 5 show unconditional ethnic differentials in these outcomes. We see that Black-White and Hispanic-White differentials are typically 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 (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 fiscal year dummies $\pi_{t\in y(t)}$ where $y(t)$ is the set of days in fiscal year y. A key advantage of using the *MCFS* 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.). These factors would otherwise typically be unobservable to the econometrician. We thus estimate the following OLS specification for individual i of ethnic group e sentenced on day t:

$$s_{iet} = \alpha + \sum_{e} \delta_e Ethnic_e + \beta X_i + \gamma L_i + \sum_{f} \omega_f OFF_{if} + \sum_{g} \gamma_g g_{ig} + \sum_{d} \lambda_d D_{id} + \pi_{t\in y(t)} + \epsilon_{iet}. \quad (1)$$

s_{iet} is the sentencing outcome. Columns 2, 4 and 6 show that there are of course large changes in the Black and Hispanic dummy coefficient estimates ($\widehat{\delta}_B$,$\widehat{\delta}_H$) as we move from the unconditional

[8]The guideline cells were in operation from 1987 until 2005. The Supreme Court's 2005 decision in US v. Booker found the guidelines violated the Sixth Amendment right to trial by jury. The guidelines are now only considered advisory. Much of the sentencing boom in the State CJS has been attributed to moves towards determinate sentencing, which is argued to more negatively impact outcomes for Blacks [Neal and Rick 2015].

to conditional specifications along each margin. This is to be expected given defendants differ in observables by ethnicity (as Table A3a shows). However, even once we condition on rich set of 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. On all margins, the point estimates for Hispanics are *larger* in absolute value than for Blacks. A natural benchmark we use for the later analysis on the impacts of 9-11 is the pre 9-11 sentencing gap, that is around 4 months for both ethnic groups relative to Whites, or around 10% of the White sentence length. Pre 9-11, the data suggests no ethnic sentencing differential on one judicial decision margin: the likelihood of receiving a downward departure.[9]

Of course the central concern that has plagued the literature is whether these differentials are due to unobserved heterogeneity across defendants of different ethnicities or not. To assess whether these differentials can reasonably be attributed to unobserved heterogeneity, 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 such selection on unobservables (SoU). The origins of such unobserved heterogeneity that drive sentencing outcomes and vary by ethnicity can of course stem from many sources such as: the characteristics of those arrested by the police and assigned to be tried in the CJS; (ii) the behavior of judges, prosecutors, legal counsel and defendants, during the various stages of the CJS. The fact that unobserved heterogeneity might stem from so many sources is the core reason research on sentencing differentials has been deadlocked. 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 making 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 the conditional specification (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

[9] As a point of comparison, it is worth noting that around the same pre 9-11 period, in the State CJS Blacks received around 20% longer (conditional) sentence lengths [Bushway and Piehl 2001].

of defendant and legal characteristics including the recommended guideline cell. For the two sentencing outcomes estimated using OLS, the bounds reported in Table 1 are $\delta_e(0) = \delta_e$ and $\delta_e(1)$, and we also report the coefficient of proportionality τ required for $\delta_e(\tau) = 0$.[10]

The bounds in Column 6 of Table 1 show that allowing for SoU: (i) there is *no* robust evidence of a Black-White sentencing differential: for there to be no Black-White differential on the sentence length margin, $\tau = .315$ is required, and this is entirely plausible given the covariates conditioned on in (1). In sharp contrast, for Hispanic-White sentencing differentials there remains robust evidence of a gap in sentence length ($\delta_e \in [4.08, 4.84]$); $\delta_e(\tau) = 0$ requires $|\tau| > 1$, so unobservables would need to be *more* important in explaining the Hispanic-White differential than the observables conditioned on in (1), that includes the recommended guideline cell.

To be clear, this does not rule out there being discrimination against Blacks anywhere in the Federal CJS. Rather the estimated bounds highlight that conditional Black-White differences could go to zero if unobservable characteristics of Black defendants driving sentencing outcomes, are correlated to their observed covariates to a plausible degree ($\tau \leq 1$). This is not the case for Hispanic-White sentencing gaps, for which the evidence suggests can only by ruled out by the omission of covariates under highly implausible conditions (the sign of τ varies across sentencing margins and $|\tau| > 1$). The remainder of the paper examines a candidate explanation for these robust Hispanic-White sentencing differences: ingroup biases that lead to 'outsiders' to US society (Hispanics) being treated differently to 'insiders' to US society (Whites), and exploits 9-11 as an event to provide insights on whether such underlying factors are indeed malleable.

3.2 Pre 9-11 Sentiments Towards Hispanics

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

[10] The procedure assumes the true data generating process is, $s_{iet} = \sum_e \delta_e^* Ethnic_e + \Theta \mathbf{W}_1 + \mathbf{W}_2 + \epsilon_{iet}$, so that specification (1) omits $\mathbf{W}_2$, a linear combination of j unobserved variables w_j^u, multiplied by their coefficients, $W_2 = \sum_{j=1}^{J_u} w_j^u \gamma_j^u$. Altonji *et al.* [2005] and Oster [2017] assume unobserved and observed covariates relate through a proportional selection relationship, $\tau \frac{\sigma_{1E}}{\sigma_{11}} = \frac{\sigma_{2E}}{\sigma_{22}}$, where $\sigma_{iE} = Cov(\mathbf{W}_i, Ethnic_e)$, $\sigma_{ii} = Var(\mathbf{W}_i)$, and τ is the coefficient of proportionality. Denoting the coefficient from the unconditional regression of s_{iet} on the $Ethnic_e$ dummies as $\mathring{\delta}_e$, and the corresponding R^2 as $\mathring{R}$, and defining the R^2 from the conditional regression (1) as $\tilde{R}$, Oster [2017] shows the true treatment effect is:

$$\delta_e(\tau) = \delta_e^* - \tau \left[\mathring{\delta}_e - \delta_e^* \right] \frac{R_{max} - \tilde{R}}{\tilde{R} - \mathring{R}}, \tag{2}$$

where R_{max} is defined as the R^2 from the hypothetical regression of sentencing outcomes on $Ethnic_e$, $\mathbf{W}_1$ and $\mathbf{W}_2$. Oster [2017] selects $R_{max} = 1.3\tilde{R}$ so the set of interest for δ_e, from $\tau = 0$ to $\tau = 1$, is $[\delta_e, \delta_e(min(1.3\tilde{R}, 1), 1)]$, as reported in Table 1.

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]. Field evidence on discrimination and ingroup biases in a variety of economic settings also exists [Bertrand and Duflo 2016].

In the context of American society, it has long been argued that Hispanics have a perceived 'foreignness' to the US [Huntingdon 2004]. More broadly, Hispanics are often argued to confront forms of racial framing through the lens of migrant 'illegality' so that regardless of legal status, often thought of as 'illegal aliens' [De Genova 2002].

To quantify such sentiment against Hispanics, we use data from the *National Election Survey* (NES) that collects 'thermometer readings' from survey respondents asked to report their attachment to various groups (as well as towards political candidates). Panel A of Figure 2A shows these readings for 2000, the last *NES* survey year pre 9-11. We see that American's affection for Latinos was *lower* than for African Americans. Panel B confirms this is a long term trend: using earlier *NES* surveys back to 1992, we see that in each and every year Americans report lower affinity with Hispanics than with Blacks (indeed, Davis [2007, p203] notes this ranking had been true in the *NES* data from 1976). Finally, we see this ranking also remains unchanged (with animosity growing towards Hispanics) in all post 9-11 survey years.

3.3 9-11, Islamophobia and Immigration

9-11 has been documented to have increased xenophobia among American society in its immediate aftermath [Human Rights Watch 2002, Davis 2007, Woods 2011]. To understand the link between 9-11 and Hispanics, we draw on work in sociology by Romeo and Zarrugh [2016], henceforth RZ. 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 government agencies, as well as other secondary sources. Ultimately, they argue that Islamophobia – or the extreme and irrational fear of Muslims and Islam – was deployed against Hispanics to garner political support, create fear, and justify increased surveillance and immigration enforcement. RZ 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.[11] 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, explicitly linking terrorism and immigration and broadening the set of Federal criminal cases subject to deportation.[12] Of course, post 9-11 the Patriot Act, that came into effect some 45 days later, further increased 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 [US Congress 2002], 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 the impacts on Hispanics in the post 9-11 period, Panel A of Figure 2B shows time series evidence from a Gallup Poll on immigration: this highlights a marked and persistent shift against immigration among poll respondents after 9-11. Panel B shows vandalism victimization rates, by ethnicity. Again, we see a spike in vandalism against Hispanics after 9-11, and growth rates only slowly returning back to trend. Other papers have shown 9-11 also worsened labor market outcomes for Hispanics [Orrenius and Zavodny 2006].

Beyond these society-wide impacts on attitudes post 9-11, a body of work in psychology is informative on potential individual reactions to 9-11. This work documents how anxiety increases individual's sensitivity to risk, and that in societies with a high threat, individuals might become oversensitive to danger signals [Gadarian and Albertson 2014]. Moreover, studies in cognitive psychology suggest stress and anxiety are associated with biased information processing, where individuals tend to pay more attention to threatening information [Eysenck 1992, Yiend and Mathews 2001], and where anxiety heightens attention to threat and prioritizes the processing of threat cues [Mathews 1990]. A key issue for our study is whether such cognitive mechanisms also impact the behavior of judges and other actors in the Federal CJS.[13]

[11] A typical statement was that, "*everything that happened that infamous day in NYC was a direct result of how our immigration system has failed*" [Rep. Elton Gallegy, Taley 2001]. 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).

[12] When signing the AEDPA, President Clinton remarked, "*[AEDPA] makes a number of major, ill-advised changes in immigration laws having nothing to do with fighting terrorism*" [Johnson 2003]. He was partly referring to the Act limiting judicial review for immigration decisions.

[13] The evidence provided in Shayo and Zussman [2011] suggests that might be so: using data from 1700 cases filed in Israeli small claims courts and exploit the random assignment of cases to Jewish or Arab judges, to provide evidence of judicial ingroup bias. By further exploiting the random timing and location of terrorist attacks in Israel, they document a short-lived difference in case outcomes depending on the ethnicity of defendants, plaintiffs and judges. To the best of our knowledge this is the only current evidence linking the salience of insider-outsider

3.4 Research Design

We consider 9-11 as an exogenously timed event that heightened the salience of insider-outsider differences and so could potentially have cued ingroup biases against Hispanic defendants. The analysis is thus informative of whether factors that might drive White-Hispanic sentencing differentials, such as biases against Hispanics in the CJS, are malleable. To isolate the impact the event had on sentencing outcomes in the Federal CJS, we compare outcomes between: (i) defendants who committed their last offense before 9-11 and were sentenced *before* 9-11 (a 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 ethnicities to estimate a difference-in-difference (DiD) impact of 9-11 on criminal sentencing. Our natural experiment (NE) 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.[14]

Table A3b shows the descriptives for the NE sample of $40,228$ cases, where 32% of defendants are White, 27% are Black, and 41% are Hispanic (an ethnic composition near identical to the full sample). Moreover, there are few differences in descriptives relative to the full sample (shown in Table A3a). Given 9-11 was unanticipated, our evidence is based on a sample of defendants and offenses that are representative of caseloads in the Federal criminal justice system more broadly. More substantively, this implies the DiD estimate is identified from a set of criminal cases committed pre 9-11 that are representative of cases passing through the Federal CJS in other times. This is one building block for the external validity of the natural experiment.

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 groups, by ethnicity. Focusing first on 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 groups. By design, both groups committed their last offense before 9-11, the distribution of last offense dates in the two groups follow a similar shape, but the distribution for the treated

differences to ingroup biases in judicial outcomes.

[14] 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, in the *MCFS* data we only observe the date of the first three offenses.

group is right-shifted relative to the control group. The remaining panels of Figure 3 show very similar patterns for sentencing and last offense dates for treated and control groups among Black and Hispanic defendants.

The DiD specification we estimate is:

$$\begin{aligned} s_{iet} &= \alpha + \sum_e \delta_e Ethnic_e + \rho Post_t + \sum_e \phi_e (Ethnic_e \times Post_t) \\ &\quad + \beta X_i + \gamma L_i + \sum_f \omega_f OFF_{if} + \sum_g \gamma_g G_{ig} + \sum_d \lambda_d D_{id} + \epsilon_{iet}, \end{aligned} \tag{3}$$

where s_{iet} is the sentencing outcome for individual i of ethnic group 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 are as defined earlier. ε_{iet} is clustered by ethnicity-district. The partial correlation with ethnicity, δ_e, now captures any cross sectional differences between defendants of ethnicity tried in the Federal CJS (such as differential sorting of defendants into the Federal system, or differential linkage rates across stages within the Federal CJS), and the difference-in-difference coefficient of interest is ϕ_e. To be clear, this measures the DiD in sentencing outcomes conditional on the case reaching sentencing Stage 8.

We also note that the decomposition analysis presented later, confirms the observable characteristics of defendants (X_i, L_i, OFF_{if}), by ethnicity, are very similar for those in the pre- and post 9-11 sample periods. This covariate balance thus ensures any differential outcomes post 9-11 across ethnicities are not driven simply by pre- and post- differences in observables of defendants of different ethnicities.

The reasons we focus on downward departures as an outcome is twofold: (i) because this margin of sentencing is most cleanly attributable to the discretion of Federal judges; (ii) ultimately in our research design, we have one credible source of quasi-experimental variation: the exogenous timing of 9-11. Hence we cannot study impacts on judicial decision making *conditional* on being downward departed or not. However, in the Appendix we provide more descriptive evidence on how these additional channels likely further reinforce Hispanic-White sentencing differentials post 9-11 beyond those we can provide DiD evidence on via the downwards departure channel.

We later study the behavior of prosecutors and legal counsel at earlier stages of the case timeline to further measure if there is differential treatment of defendants by ethnicity pre-sentencing. That moves us closer to the alternative way to measure discrimination in the CJS long debated among legal scholars, conditioning on factors that make defendants otherwise equal at the point they enter the Federal CJS in Stage 1 [Starr and Rehavi 2013].[15]

[15] We note that 9-11 can impact sentencing outcomes for *all* defendants irrespective of their ethnicity, as measured

3.5 Identifying Assumptions and Interpreting ϕ_e

Three assumptions underpin ϕ_e identifying a causal treatment effect of ethnicity on sentencing outcomes. 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 impacted by 9-11. 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 address the concern more formally using survival analysis to predict the time a defendant spends in the CJS between the dates of last offense and sentencing. Second, we require there to be no ethnicity-time effects in ethnic sentencing differentials that occur naturally around 9-11 each year, say because types of criminal offense vary around the year and are correlated with defendant ethnicity. We formally 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 ethnic groups *and* change post 9-11 2001 (but not in placebo years). If all three assumptions hold, then on average there is no change in unobserved heterogeneity between treatment and control groups by defendant's ethnicity and ϕ_e measures the causal impact of ethnicity on sentencing differentials in the ± 180 day sentencing window around 9-11. As defendants do not anticipate 9-11, this estimate has external validity for the magnitude of ethnicity sentencing differentials in other times.

ϕ_e can reflect differences in outcomes post 9-11 driven through multiple channels. Judges might anticipate changes in behavior of defendants post 9-11, with these expectations differing across defendants by ethnicity. For example, 9-11 might have altered labor market outcomes for minorities and this can affect recidivism rates differentially across ethnic groups; alternatively, judges might anticipate post 9-11 the police will reallocate resources in a way that differentially changes future detection probabilities by ethnicity. Taken together, such channels represent different forms of statistical discrimination, where stereotyping of defendants by ethnicity can lead to differential outcomes by 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*). ϕ_e also partly captures true ethnic discrimination against ethnic group e post 9-11, and this might be especially impactful on Hispanics given the event heightened the salience of insider-outsider differences. Given these alternative interpreta-

by ρ. This can arise, for example, either because anticipated changes in recidivism/detection probabilities are the same for all defendants post 9-11, or because society faces different liberty-security trade-offs post 9-11 [Davis 2007].

tions of ϕ_e have different welfare implications, we later use two strategies to probe the data to understand the origins 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 ethnic sentencing differential to Federal judge characteristics, including their ethnicity, that is somewhat in the spirit of rank order tests used to distinguish statistical discrimination from animus in the literature using data on police arrests [Anwar and Fang 2006, Park 2017].

4 Judicial Sentencing Decisions

4.1 Downward Departures

Table 2 presents estimates of (3) where our key focus is on the granting of downward departures, as that is the primary form of discretion that judges exercise at sentencing. In Column 1 we see that Hispanic-White sentencing gaps become significantly larger when the salience of insider-outside divisions is heightened: 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 is significantly different to Hispanics, $p = .042$).

As described in Section 2, judges have to provide an explanation for downward departures: Columns 2 to 5 then code these explanations into the most common broad categories. We see that 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 accurately represents either the seriousness of that history or the likelihood the defendant will commit other crimes; (ii) other reasons. There is no statistically significant shift in downward departures related either to plea bargains, or due to general mitigating circumstances.

We can convert our baseline causal impact on the likelihood of a downward departure into an implied change in expected sentence length for Hispanics as follows. 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 in guideline cell g as $E[s|g]$. The implied *change* in expected sentence length is then given by,

$$\sum\nolimits_g p_g \left\{ -\Delta p_d E[s|g] + \Delta p_d . E[s|g-4] \right\}, \tag{4}$$

where we use the pre 9-11 empirical distribution of Hispanic defendants across guideline cells to

proxy p_g, assume that if an individual receives a downward departure, they move four guideline cells (which is the case for the median defendant downward departed pre 9-11) and so are sentenced in cell $g-4$, and take the midpoint sentence in each guideline cell as an estimate for $E[s|.]$. The foot of Column 1 in Table 2 shows the implied impact on Hispanic sentence lengths to be .736 months, corresponding to 18% of the conditional pre 9-11 Hispanic-White sentence differential shown in Table 1. If the behavioral response of judges to 9-11 is driven by the heightened salience of insider-outsider divisions, then by this benchmark, this mechanism leads to a non-trivial increase over the pre 9-11 Hispanic-White sentencing differential.[16]

Of course we could also examine the impact on sentence length directly. However, given that 80% of cases result in no downward departure, such effects are never precisely measured. Using quantile regression estimates for sentence length conditional on the full set of covariates in (3), we find that post 9-11 Hispanics have an increase in their median sentence by .632 months that is statistically significant at the 5% level (with no significant impact among Black defendants).

To place a monetary value on these sentencing impacts coming solely through changes in the propensity to downward depart, we start by noting that: (i) the marginal annual cost per year of imprisoning a male prisoner of $\$29,000$ [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 the heightened salience of insider-outsider differences post 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.[17]

One concern is that we have conditioned on two classes of outcome endogenously determined during the Federal timeline: the offense type the defendant is charged with, and the guideline cell they are recommended to be placed in. We have done so in order to mirror 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 in line with legal studies on discrimination, is to only condition on observables determined at

[16]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).

[17]An alternative benchmark can be based on Mueller-Smith [2016]: he 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 $\$66,000$. If we apply even the lower bound estimate to our sample of defendants in the Federal CJS, then as Mueller-Smith [2016] makes clear, sentencing differentials would need to have substantial deterrence effects for them to have welfare-neutral impacts.

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 subset of cases that can be linked from prosecutor stages back to the arrest stage we can 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). As Figure 1 showed, linkage rates to back to arrest data are imperfect: we can link back 67% of cases in the NE sample to exploit this arrest data. The result in Column 6 shows that accounting for original arrest codes, the Hispanic-White differential on downward departures remains, 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.

Linking our findings with the established literature on labor market discrimination, a key insight of Gary Becker's work is that the observed racial wage gap will not reflect the average level of employer discrimination. The reason is that 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, not the average level of discrimination in the labor market. This contrasts sharply with what we can infer in the case of criminal sentencing: as defendants cannot sort over sentencing judges, and judges cannot turn down cases they are assigned to, our estimates reflect the *average* ethnic sentencing differentials driven by judicial behavior in the Federal CJS.

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 the salience of insider-outsider differences. The first relates to the fact 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 3 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).[18]

We see that for both groups of Hispanic, those that are sentenced post 9-11 are significantly less

[18] 71% of defendants overall are classified as citizens, where 91% of non-citizens are Hispanic, so there is little sense in splitting Black defendants by citizenship status.

likely to receive a judicial downward departure, all else equal. For Hispanic citizens the impact is a 2.8pp reduction in the likelihood of a downwards departure, corresponding to an implied sentence length increase of .58 months that maps to 17% of the pre 9-11 Hispanic citizen-White sentencing differential. For Hispanic non-citizens the impact is a 4.4pp reduction in downwards departure, an implied sentence length increase of .82 months that maps to 16% of the pre 9-11 Hispanic non-citizen-White sentencing differential. There is no statistical difference between the two impacts (p-value=.269)

The second reason why Hispanic-White sentencing differentials might increase post 9-11 is that Hispanics are more likely to be charged with immigration offenses than other defendants. 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. We address the issue in the remaining Columns of Table 3 by splitting the NE 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.[19]

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: these non-drug and non-immigration offenses constitute around 40% of all offenses and often relate to firearm offenses (Column 5).[20]

4.3 Robustness Checks and Support for the Identifying Assumptions

As described in the Appendix, Tables A4 to A6 conduct a battery of checks on our core finding from Table 2, Column 1. These show the result to be robust to: (i) alternative levels of clustering of the 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 (3) separately for

[19]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%).

[20]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.

each ethnicity. Finally, we use the fact the *MCFS* 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 findings pick up racial, rather than ethnic, sentencing differentials.

In Appendix Tables A7 to A10 we provide evidence in support of the underlying identifying assumptions. On the assumptions related to the time spent in the Federal CJS around 9-11, we use 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.

To address concerns related to time confounders we present three sets of evidence. First, we use data from earlier years to construct placebos 9-11 effects to check that there are no ethnicity-time effects in ethnic sentencing differentials that occur naturally around 9-11 each year. We find that when doing so (as shown in Table A9), the impact for Hispanics on judicial downward departures only occurs post 9-11 in 2001, and not in earlier years. Indeed, 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 1.

Second, we address concerns that some of the impacts we find might be driven by the passage of the Patriot Act, that was enacted 45 days after 9-11. Notwithstanding the earlier result that immigration offenses did not appear to drive the main result, to shed further light on the matter we estimate a dynamic specific analogous to (3) that estimates impacts in 15-day windows post 9-11. We use this to document how impacts on judicial departures for Hispanics appear post 9-11 and pre Patriot Act. Third, we collate data on the date of confirmation of G.W.Bush-appointed US Attorneys, to establish 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.

5 Prosecutorial Decisions

Federal prosecutors represent a second crucial legal actor whose decisions determine defendant outcomes. As shown in Figure 1, their key decisions occur at early stages in case timelines. This analysis therefore more closely measures ethnic differentials conditional on factors that make defendants otherwise equal at the point of entry into the Federal CJS. We extend our analysis to examine this sequence of prosecutorial decision making to understand the extent to which their behavior drives pre-sentencing *differential treatment* of defendants by ethnicity, and whether such behaviors are also malleable by outside events.

5.1 Prosecutor's Initial Offense Charges

The first critical decision prosecutors have discretion over is the initial offense charges 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].[21]

To begin with, we use the pre 9-11 sample to consider, by ethnicity: (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). For each outcome we then estimate a specification analogous to (3) but do not condition on offense type, or guideline cell (the former because the offense charge might go across offense type boundaries, and the latter because it is determined later in the timeline). We use this to present conditional ethnic differentials, and to examine whether these differentials are robust to accounting for selection on unobservables, using the same approach as in Table 1 for judicial decisions.

Table 4 presents the results. On initial offense charges we see that pre 9-11: (i) Blacks 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 ethnic differentials correspond to 76% (57%) increases over the baseline probability for White defendants. Both impacts are robust to accounting for selection on observables: the implied bounds do not include zero, and the implied τ required for the bound to be at zero is larger than one in absolute value for both ethnicities. One concern is that the nature of the offense is not controlled for in Column 2. To address this issue we exploit information from the arrest stage of the criminal time line (Stage 0): for the subset of cases that can be linked from prosecutor stages back to the arrest stage we can condition on a rich

[21] 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.

set of codes corresponding to the precise offense the defendant was originally arrested for. We can link back 52% of cases to exploit this arrest data. The result in Column 3 shows that doing so: (i) there remain significant Black-White differences in the likelihood of non-zero statutory minimum offense charge being given by Federal prosecutors, although now the SoU bounds just include zero, and the implied τ required for the bound to be at zero is just under one; (ii) Hispanic-White differentials remain statistically significant and robust to SoU.

We document a similar pattern of ethnic differentials pre 9-11 for minimum sentence lengths (Columns 4 to 6). Pre 9-11 Federal prosecutors set initial offense charges such that the actual length of statutory minimum sentences is significantly higher for Black and Hispanic defendants, and this remains so even when we consider the subsample of cases that can be matched back to the rich set of arrest offense codes (Column 6). This confirms that when sentencing guidelines are place, this margin is a key one along which prosecutor's actions determine ethnic differentials: exactly the point established by Rehavi and Starr [2014]. The magnitude of the effect is such that conditional on observables related to the offender, legal counsel and district, Blacks receive charges carrying minimum sentences that are 22 months longer than Whites: this is near double the minimum sentence for Whites. The same is true for Hispanics: prosecutors set initial charges with associated statutory minimums that are 14 months longer (or 63% higher) than for White defendants, falling to 7.4 months in the subsample of cases that can be linked with the arrest offense codes. Both conditional impacts are robust to accounting for selection on observables.[22]

We next examine whether the events of 9-11, that heightened the salience of insider-outsider differences, lead to these ethnic differentials being widened for Hispanic (but not for Black) defendants who were already being processed in the Federal CJS on 9-11. 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 (a 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 (a treated

[22]Rehavi and Starr [2014] establish using similar linked administrative data that prosecutor's initial offense charges account for half the Black-White sentencing gap in the period 2006-8, after sentencing guidelines had been abolished and judges are not required to issue sentences within the guidelines. We thus establish that their findings replicate in the pre 9-11 sample period, when sentencing guidelines are always in place.

group). We then estimate a specification analogous to (3) but where the outcomes considered are: (i) whether the defendant receives an initial charge with a non-zero statutory minimum sentence; (ii) the length of statutory minimum sentence associated with their initial offense. As before we do not condition on final offense type (OFF_{if}) or the later determined guideline cell.[23]

The results are shown in Table 5: (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 (and this impact is 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 prosecutor's 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 responses to 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.

The next two Columns trace through the judicial sentencing impacts on this same cohort of defendants (at Stage 8), and so allow us to provide novel evidence on the interlinkage between prosecutorial and judicial decisions. We thus 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. We see that for Hispanics who were 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 of ingroup biases that we have not so far measured. The earlier costs were associated with the cohort that come up for sentencing around 9-11 (Table 2) whereas these results imply continuing *longer run* costs of ingroup biases that relate to the cohort of Hispanic defendants initially charged around 9-11, and come up for sentencing well after 9-11. In Column 4 we then control 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.

[23] 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

Apart from the judge-initiated downward departures studied earlier, another form of downward departure originates from Federal prosecutors and are referred to as 'substantial assistance departures'. These 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 Schanzenback 2012]. Once such a motion is made, the sentencing judge determines if such a departure is warranted and if so, they determine the degree of the departure.[24]

To examine this margin of Federal prosecutor's decision making, we repeat the analysis in Tables 4 and 5 for substantial assistance departures. In Table 4, Columns 5 and 6 show that pre 9-11: (i) unconditionally, Hispanic defendants are significantly less likely to receive substantial assistance departure than White defendants; (ii) this difference remains significant conditional on observables (where we condition on offense type as that has been determined by Stage 6, but we do not condition on guideline cells as those are determined in Stage 7, as described below). The magnitude of the ethnic differential is that Hispanics are 9pp less likely to receive substantial assistance, corresponding to a 41% reduction relative to the likelihood for White defendants pre 9-11. This gap is robust to accounting for selection on unobservables.

In Table 5 we then consider the impact of 9-11 on prosecutorial decisions on substantial assistance departures for two cohorts. In Column 5 we track the same cohort of defendants considered earlier 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 de-

[24]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 within the *MCFS* data. A departure from a statutory mandatory minimum penalty for cooperation requires a separate motion under 18 U.S.C. §3553(e) – these kinds of departure are not identified in the *MCFS* data. There has been some disagreement among the circuit courts as to how to determine the extent of a departure, and whether mandatory minimum sentences set limits on the extent of the departure. The 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 who has committed an offense, 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, consideration of the following: (i) the court's evaluation of the significance and usefulness of the defendant's assistance, taking into consideration the government's evaluation of the assistance rendered; (ii) the truthfulness, completeness, and reliability of any information or testimony provided by the defendant; (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.

partures. This helps rule out that the earlier increase in statutory minimum sentence lengths associated with initial offense charges was being undone at a later stage of the timeline through defendant cooperation in plea bargaining with prosecutors, thus leading prosecutors to request substantial assistance departures.[25] Finally, in Column 6 we consider the full NE sample of cases. Comparing defendants up for sentencing around 9-11, we see that post 9-11 there is no evidence of any change in the likelihood Hispanics or Blacks being granted a substantial assistance departure.

5.3 Pre-sentence Reports

The third key stage at which Federal prosecutors can 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 *MCFS* data records whether a defendant is in custody after trial (66% of defendants are remanded in custody in the NE sample), so we can recover the precise trial date for each defendant. This allows us to estimate the impact of 9-11 on prosecutor-legal counsel interactions between trial and sentencing: this is a critical period because it is when the pre-sentence report (PSR) is drafted.

More precisely, the defendant's legal counsel provides information on the defendant's life history to the (neutral) Probation Office. The defendant is 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 Federal prosecutors, to prepare a draft PSR. This is provided to 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 and hence sentence range.

We now assess whether 9-11 impacted the suggested sentencing guideline cells differently across defendants by their ethnicity, as a result of the prosecutor-legal counsel interactions when preparing the PSR. We estimate a specification similar to (3) 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 key recommendations from the PSR: the criminal history score, the offense severity, and the minimum

[25] Our data does not cover the details of plea bargains. We only note that over 95% of defendants plead guilty (pre and post 9-11, for all ethnicities).

sentence recommended in the implied guideline cell (hence unlike in (3), we obviously do not condition on the guideline cell).

Table 6 shows the results focusing on the clean comparison between the C and $T2$ group of defendants: we find no evidence of differential impacts post 9-11 on either Hispanic nor Black defendants for five out of six dimensions of the PSR. Reassuringly, we find null impacts on criminal history scores: this is as expected as this is the dimension of the guideline cell determination that is least open to interpretation. In short, prosecutor-legal interactions at the PSR stage between trial and sentencing are *not* a source of differential treatment of defendants by ethnicity post 9-11 when insider-outsider differences are most salient. These results suggest any increased Hispanic-White sentencing gaps post 9-11 are not due to diminished effort on the part of legal counsel to Hispanic defendants. Indeed, the point estimates in the first row of Table 6 suggest if anything marginally improved outcomes in pre-sentence reports for Hispanics post 9-11. This is notable because unlike the other stages discussed, it is a stage at which the defendant's defense counsel is involved and shapes sentencing outcomes.[26]

6 Origins of Sentencing Differentials

Having established the malleability of the decisions of Federal judges and prosecutors driving differential outcomes across ethnicities, we now probe the data to understand the origins of the documented differentials. We do so using two strategies: (i) decomposition analysis; (ii) correlating ethnic sentencing differential to Federal judge characteristics, including their ethnicity.

6.1 Decomposition Analysis

Our analysis has highlighted two cohorts of Hispanic defendant for whom 9-11 led to widening sentencing disparities relative to Whites: (i) for those cohorts that come up for judicial sentencing just after 9-11, Hispanics are significantly less likely to receive downward departures (Table 2); (ii) for those cohorts for whom prosecutors set initial offense charges just after 9-11, Hispanics receive charges associated with significantly longer statutory minimum sentence lengths (Table 5). To rule out some potential drivers of these differentials, we use a Juhn *et al.* [1993] decomposition to split the raw DiD in sentencing outcomes into those attributable to: (i) changes in the observable characteristics of defendants; (ii) changes in the returns to these observables (or changes in the

[26] Two further points are of note from Table 6. First, for those defendants in $T1$ we also find no impacts on these PSR outcomes: these are harder to interpret because these PSRs will be drafted both pre- and post 9-11.Second, there is evidence in Table 6 of a *common* impact of having the PSR written after 9-11: significantly higher offense severity scores are recommended, and the consequent minimum sentence in the guideline cell significantly rises by 2.6 months.

sentence 'price' of observables); (iii) changes in unobservables. The JMP decomposition is implemented by first considering the following sentencing equation for White defendant i sentenced in period T: $s_{iT} = X'_{iT}\beta_T^W + \sigma_T^W \theta_{iT}$, where β_T^W are sentence prices for Whites, θ_{iT} is a standardized residual capturing unobserved determinants of White sentences, and σ_T^W is the standard deviation of this residual for Whites in period T. The Hispanic-White sentencing differential in period T is then, $\Delta s_T = s_T^H - s_T^W = \Delta X_T \beta_T^W + \sigma_T^W \Delta\theta_T$. Given our DiD research design we take a *second* difference over time periods, considering how the ethnic sentencing gap changed pre- to post 9-11 ($T = 0$ to $T = 1$):

$$\Delta s_1 - \Delta s_0 = (\Delta X_1 - \Delta X_0)\beta_0^W + \Delta X_1(\beta_1^W - \beta_0^W) + (\Delta\theta_1 - \Delta\theta_0)\sigma_0^W + \Delta\theta_1(\sigma_1^W - \sigma_0^W). \quad (5)$$

The $(\Delta X_1 - \Delta X_0)\,\beta_0^W$ component, or X-effect, measures the contribution to the DiD in sentencing gaps of observables. Our research design is such that this component should be small: this is confirmed below and is line with defendant observables being balanced pre- and post 9-11 by ethnicity. 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 all these 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, g_{ig}. 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 DiD set-up provides common support in the cross-section of covariates across ethnicities; (ii) the inclusion of guideline cell dummies allows us to capture many more 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 justifiably respond.

The $(\Delta\theta_1 - \Delta\theta_0)\sigma_0^W$ component, or θ-effect, measures the change in Hispanic's position within the White residual sentencing distribution (measured at $T = 0$). Shifts in discrimination against Hispanics post 9-11 would lead to an increase in Hispanic's average position in the White residual distribution. Finally, the $\Delta\theta_1(\sigma_1^W - \sigma_0^W)$ component, or σ-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-effect and σ-effect reflect both discrimination and unobservable offense and defendant

characteristics. *A priori* we might expect the θ-effect to predominantly reflect shifts in ethnic discrimination because it represents changes in the *position* of Hispanics in the White sentencing residual distribution, while the σ-effect captures changes in the *spread* of this residual, that is less clear would be driven by ethnic discrimination.

Table 7 shows the JMP decomposition for Hispanic-White sentencing gaps for the two cohorts identified above. On judge's sentencing decisions, the decomposition for downward departures (the margin along which ethnic sentencing differentials change post 9-11) is based on a LPM.[27]

Column 1 shows that: (i) only 7% is attributable to observables (Row 4: X-effect + β-effect); (ii) 93% of the Hispanic-White differential is due to unobservables (Row 5: θ-effect + σ-effect); (iii) among the unobservable components, the θ-effect is by far the more important driver of the unconditional DiD 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 σ-effect is only .006 (Row 9).

Column 2 focuses on the cohort of defendants impacted by prosecutor's initial offense charges. For this continuous outcome the application of the JMP decomposition is straightforward, and in line with the earlier regression evidence we do not control for offense type or guideline cell in the set of X's. Column 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 DiD in statutory minimum sentence lengths.

Figure 4 graphically summarizes both decompositions, detailing further the X- and β-effects for covariates. This reiterates that for each cohort, the bulk of the raw differential is due to unobservable factors. For the cohort impacted by judicial decisions at sentencing, we see that: (i) as expected given the DiD design, each X-effect is small; (ii) sentencing prices on socio-demographic characteristics (highest education level, marital status, age and number of dependents) rise, and sentencing prices on Federal districts fall. For the cohort impacted by prosecutorial decisions over initial charges, we see: (i) again as expected given the DiD design, each X-effect is small; (ii) sentencing prices on the type of defense counsel and Federal districts fall. We examine further this variation in Hispanic-White sentencing differentials across Federal districts below.

Taken together the decomposition results suggest that for both sets cohorts of Hispanic defendant for whom 9-11 led to greater sentencing disparities relative to Whites, neither disparity

[27] To check the validity of basing the JMP decomposition off a linear probability, 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.

is easily explained by changes in observables or the sentencing prices of those observables. This especially helps to further rule out explanations for the 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. Taken together, this suggests explanations for why Hispanic-White sentencing differentials worsen post 9-11 based on statistical discrimination do not easily fit the evidence.

6.2 Judge Characteristics

We now analyze how judge characteristics correlate to our measure of sentencing differentials. The administrative data contains no information on judges, and there is no simple way to link judge and defendant identifiers for Federal criminal cases. To make progress we have hand-coded the characteristics of Federal judge's by district, sourced from the *Biographical Directory of Federal Judges*. This details the 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 thus construct judge characteristics at the district level ($\mathbf{J}_d$).

Similarly to Guryan and Charles [2011], we then proceed in two steps. First, we estimate (3) 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 a downward departure is given as this is the margin along which ethnic sentencing differentials further open up post 9-11. Figure 5 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 NE 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. In contrast to Federal prosecutors, there are a substantial share of judges from minority backgrounds. 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.[28]

[28]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

Table 8 shows the second stage results. In Column 1 we only control for judge ethnicities: we see 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. This is in line with judges displaying ingroup bias towards defendants along the lines of insider-outsider divisions [Schanzenbach 2005, Abrams *et al.* 2012].

Column 2 shows this finding to be 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, all else equal.

Column 3 controls for the population shares of different ethnic groups in the district, as well the change (1990 to 2000) in the proportion of the population from each ethnic 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 characteristics of where the Federal criminal case is heard. Moreover, the partial correlation of the proportion of the district population that is Hispanic in 2000, or the change in the Hispanic population share between 1990 and 2000 in the district, are both negative. This is contrary to the contact hypothesis, that states that interpersonal contact is an effective ways to reduce prejudice between majority and minority group members [Allport 1954].

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 be using 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]. This interpretation is further reinforced by noting the robust evidence across specifications of a partial correlation between judges appointed under Democrat Presidents and the Hispanic-white sentencing differential on downward departures across districts. There is little evidence to suggest that more experienced judges are correlated with smaller ethnic sentencing differentials (measured either through the senior status of judges or their age). As such, this is counter to the Altonji and Pierret [2001] test of statistical discrimination that exploits the fact that with experience, decision mak-

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 considerably more pronounced.

ers (judges/employers) learn about the true characteristics of agents (workers/defendants) and so become less reliant on proxies such as ethnicity.

To more easily make comparisons across covariates, Column 4 standardizes 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 2 that post 9-11, Hispanic defendants are 3.8pp less likely to receive a downward departure.

To examine the external validity of these correlations outside of the window around 9-11, the next Column repeats the exercise but first estimates $\widehat{\delta}_{H,d}$ from the sample of 130,000 Federal criminal cases pre 9-11 from October 1998 to September 2001. We continue to report all coefficients as effect sizes to aid comparability. Strikingly, in the full sample we also see evidence of ingroup bias: a one standard deviation in the proportion of district judges of Hispanic origin increases the Hispanic-White sentencing differential for downward departures, $\widehat{\delta}_{H,d}$, by .063, that is actually slightly larger than the effect size estimate based on the NE sample estimates.

Finally, we note that 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. However, even if it were available for our sample period, 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, and so there is no variation in prosecutor ethnicity to exploit.[29]

7 Conclusions

A large body of literature across disciplines documents that for similar offenses, Blacks and Hispanics face a higher probability of arrest, conviction and harsher penalties conditional on conviction. If historic trends continue, then among the 2001 birth-cohort, one in three Black men and one in six Hispanic men can expect to spend time in prison during their lives [CEA 2016]. The central challenge lies in understanding whether such differential outcomes in the criminal justice system by ethnicity are driven by unobserved heterogeneity across defendants, correlated to their ethnicity,

[29] A summary of the findings are available at http://wholeads.us/justice/wp-content/themes/phase2/pdf/key-findings.pdf (accessed May 13th 2016).

or whether they reflect true discrimination. The primary reason research on sentencing differentials has been deadlocked is because the origins of such unobserved heterogeneity can stem from so many sources such as: (i) the characteristics of those arrested by the police and assigned to be tried in the CJS; (ii) the behavior of judges, prosecutors, legal counsel and defendants, during the various stages of the CJS. We tackle these issues by developing a difference-in-difference research design that addresses the first empirical concern, and by exploiting linked administrative data to tackle the second issue. We do so in the high stakes environment of the Federal criminal justice system, where decisions are made by professional and experienced judges, prosecutors and legal counsel, and the universe of criminal offenses and district courts can be studied.

The key contribution of our analysis has been to provide insights into the magnitude, channels and malleability of Hispanic-White sentencing differentials in the Federal criminal justice system. This is important given Hispanics are: (i) the modal defendant in the Federal CJS; (ii) the group whose incarceration rate is growing fastest; (iii) a relatively understudied minority group in the economics of discrimination literature [Charles and Guryan 2011], despite Hispanics being ever more prominent in the political, legal and cultural life of America. We build an evidence base that is consistent with these differentials being driven by the ingroup biases of judges and prosecutors, whereby 'outsiders' (Hispanics) are treated differently to 'insiders' (Whites), and that these biases are malleable and responsive to outside events. Going beyond the existing literature we document evidence of such ingroup biases driving the behavior of Federal judges at sentencing, and the behavior of prosecutor's when setting initial offense charges.

Finally, we document that the implied sentencing impacts driven by behavioral responses of judges and prosecutors to 9-11 represent a significant widening of pre 9-11 Hispanic-White sentencing differentials. If 9-11 makes salient insider-outsider differences, then such implicit biases might drive ethnic sentencing differentials in other times. 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 offer the suggestion that appointing more Hispanic judges to Federal district courts or as Federal prosecutors, might go some way towards reducing Hispanic-White sentencing differentials. Although judges are already required to justify a downward departure, there might scope for increasing scrutiny when such departures are (surprisingly) *not* granted. Increased scrutiny of prosecutors when they set initial charges might also be considered. Moreover, the fact we find no evidence of ethnic sentencing differentials at the pre-sentencing stage, 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.

A natural next step would be to use such linked administrative data to understand the origins of Black-White sentencing differentials in the Federal CJS. There is of course a vast literature in social psychology suggesting stereotyping of Blacks might lie at the root of such differences; laboratory experiments provide foundational evidence for this based on visual processing [Eberhardt *et al.* 2004], and recent field experiments also highlight the role that limited attention might play in driving discrimination [Bartos *et al.* 2016]. The challenge lies in developing credible research designs in the context of the criminal justice system that cause the strength of such factors underpinning the origins of discrimination to vary across time or space in a manner orthogonal to other characteristics of criminal cases. Given the social and economic consequences of how the criminal justice system is differentially experienced by individuals of different ethnicities, we hope our findings here on Hispanic-White sentencing gaps encourage others to also take up this challenge.

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] AIZER.A AND J.DOYLE (2015) "Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-assigned Judges," forthcoming, *Quarterly Journal of Economics.*

[3] ALESINA.A AND E.LA FERRARA (2015) "A Test of Racial Bias in Capital Sentencing," *American Economic Review* 104: 3397-433.

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

[5] 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.

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

[7] 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.

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

[9] BARTOS.V, M.BAUER, J.CHYTILOVA AND F.MATEJKA (2016) "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review* 106: 1437-75.

[10] 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.

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

[12] BORDALO.P, N.GENNAIOLI AND A.SHLEIFER (2015) "Salience Theory of Judicial Decisions," *Journal of Legal Studies* 44: 7-33.

[13] 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.

[14] CEA (2016) *Economic Perspectives on Incarceration and the Criminal Justice System*, Washington DC.: Council of Economic Advisors Report.

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

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

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

[18] DANZINGER.S, J.LEVAV AND L.AVNAIM-PESSO (2011) "Extraneous Factors in Judicial Decisions" *PNAS* doi:10.1073/pnas.1018033108.

[19] DE GENOVA.N (2002) "Migrant "Illegality" and Deportability in Everyday Life," *Annual Review of Anthropology* 31: 419-47.

[20] EBERHARDT.J.L, P.A.GOFF, V.J.PRUDIE AND P.G.DAVIES (2004) "Seeing Black: Race, Crime and Visual Processing," *Journal of Personality and Social Psychology* 87: 876-93.

[21] EREN.O AND N.MOCAN (2016) Emotional Judges and Unlucky Juveniles, NBER WP 22611.

[22] EYSENCK.M.W (1992) *Anxiety: The Cognitive Perspective*, Mahwah, NJ: Lawrence Erlbaum.

[23] 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.

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

[25] GADARIAN.S.K AND B.ALBERTSON (2014) "Anxiety, Immigration, and the Search for Information," *Political Psychology* 35: 133-64.

[26] 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.

[27] GUTHRIE.C, J.RACHLINSKI AND A.WISTRICH (2001) "Inside the Judicial Mind," *Cornell Law Review* 86: 777-830.

[28] 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.

[29] HUNTINGTON.S (2004) "The Hispanic Challenge," *Foreign Policy* 141: 30-45.

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

[31] JOHNSON.K.R (2003) "September 11 and Mexican Immigrants: Collateral Damage Comes Home," *DePaul Law Review* 52: 849-70.

[32] 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.

[33] KELMAN.M, Y.ROTTENSTREICH AND A.TVERSKY (1996) "Context-Dependence in Legal Decision Making," *Journal of Legal Studies* 25: 287-318.

[34] 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.

[35] KLING.J (2006) "Incarceration Length, Employment, and Earnings," *American Economic Review* 96: 863-76.

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

[37] MATHEWS.A (1990) "Why Worry? The Cognitive Function of Anxiety," *Behaviour Research and Therapy* 31: 57-62.

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

[39] 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.

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

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

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

[43] PHILIPPE.A AND A.OUSS (2016) "No Hatred or Malice, Fear or Affection": Media and Sentencing, mimeo TSE.

[44] RACHLINSKI.J (1996) "Gains, Losses, and the Psychology of Litigation," *Southern California Law Review* 70: 113-86.

[45] RACHLINSKI.J.J, S.L.JOHNSON, A.J.WISTRICH AND C.GURTHRIE (2009) "Does Unconscious Racial Bias Affect Trial Judges?," Cornell Law Faculty Publications Paper 786.

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

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

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

[49] 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.

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

[51] 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.

[52] 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.

[53] 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.].

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

[55] YIEND.J AND A.MATHEWS (2001) "Anxiety and Attention to Threatening Pictures," *Quarterly Journal of Experimental Psychology* 54: 665-81.

A Appendix

A.1 Data Sources

The data used were obtained from the Inter-university Consortium for Political and Social Research and are part of the *MFCS* series, derived from cases received by the USSC. 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) US Sentencing Commission (USSC) data, 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. We drop 4 out of 94 districts: Guam, Puerto Rico, N.Mariana Island and the Virgin Islands.

We focus on male defendants that come 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.[30]

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 ethnicities.

The data for judicial characteristics are sourced from the *Biographical Directory of Federal Judges*. To select the relevant judges to construct the 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 natural experiment sample and those who

[30] More information on the data series can be found at,
http://www.icpsr.umich.edu/icpsrweb/NACJD/series/00083/studies?archive=NACJD&sortBy=7 (accessed 14th April 2016).

terminated the bench after the beginning of the sample. We perform an analogous sample cut of judges relevant for the pre 9-11 sample specifications.

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 Judicial Decision Making: Other Components

The first additional component we provide descriptive evidence on is $E[s|DD = 0]$, that we measuring using information on sentence length for those defendants sentenced within their original guideline cell g. For defendant i assigned to cell g we define the within-guideline sentence as:

$$ws_{ig} = \frac{s_i - \min(s_{ig})}{\max(s_{ig}) - \min(s_{ig})} \in [0, 1]\,, \tag{6}$$

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 ($ws_{ig} = 0$). Panel A of Figure A1 graphs the density of the unconditional DiD 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.

The other additional component we provide descriptive evidence on is $E[S|DD = 1]$. This depends on the number of cells moved conditional on downwards departure as follows:

$$E[s|DD = 1] = \sum\nolimits_d p(cells\ moved = d)E[s|DD = 1, cells\ moved = d), \tag{7}$$

where $p(cells\ moved = d)$ is the probability that a defendant moves d cells conditional on receiving a downward departure. Panel B in Figure A1 graphs the density of the unconditional DiD in cell movements conditional on downward departure, by ethnic 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.

In short, this descriptive evidence suggests both channels might be further impacting Hispanics around 9-11, but our research design does not allow to econometrically identify these impacts given we have one source of quasi-experimental variation in the timing of 9-11.

A.3 Robustness Checks

The main specifications cluster standard errors by ethnicity-district and so focus on geographically based unobservables that might be correlated by ethnicity for sentencing outcomes. The alternative level of clustering we therefore consider is at the level of week of sentencing x ethnicity, so placing more emphasis on *time-related* unobservables being correlated by ethnicity for sentencing outcomes. The resulting standard errors are near identical to those in Table 2 in most cases (Table A4, Column 1).

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 follow a similar pattern to those estimated on the NE sample (Table A4, Column 2). In Section 5 we explicitly examine whether post 9-11, prosecutor's change their decisions over the initial offense charges to file at Stage 3 post 9-11 differentially across ethnicities. Table A5 shows the core results to be robust to estimating (3) separately for each ethnicity: the signs, significance and magnitude of estimates matches closely the pooled specification, with there remaining an implied DiD penalty of a 3.4pp reduction in the likelihood Hispanic defendants are granted downward departures if sentenced post 9-11 (Column 3).

On racial sentencing differentials, Table A6 shows the results, where we estimate a specification analogous to (3) but allow the post 9-11 impacts to vary by race, using the full set of race classifications in the *MCFS* data. To establish the link between this split and what we have previously used, it is important to note that defendants we 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 ethnic, rather than racial sentencing differentials. The main document Hispanic-White ethnic sentencing differential is simply masked in this specification within the white-race impacts.[31]

A.4 Evidence in Support of the Identifying Assumptions

A.4.1 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 A7 first addresses this concern by extending specification (3) to additionally control for the defendant's time in the CJS

[31] In this specification, 68% of defendants are White, 28% are Black, and the other groups (American Indian, Asian/Pacific Islander, Multi-Racial, Other Race) each do not constitute more than 2% of defendants.

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. As shown in Table A7, the earlier results are 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 A8 where we use OLS and survival models to estimate the time between last offense and sentencing date for each defendant, and then test whether this changes significantly, by ethnicity, 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 ethnicity (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 just for specific offense types (Columns 2a-4c).

A.4.2 Time Confounders

The second identifying assumption is that there are no ethnicity-time effects in ethnic sentencing differentials that naturally occur around 9-11 each year. We use the $MCFS$ data on cases from earlier years (1999 onwards) to estimate placebo 9-11 impacts by ethnicity.[32] The results are shown in Table A9 and confirm that there are no natural ethnicity-time effects around 9-11 along either sentencing margin. Column 1 shows 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 2: the implied DiD impact in 2001 is to reduce judicial departures for them by 5.5pp.

A candidate time confounder 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 (3) 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 A2, the graphs the estimated impact on Hispanics for non-

[32] 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.

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, 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 G.W.Bush was appointing 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 A3 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 A10 we re-estimate our baseline results from Table 2 allowing for the post 9-11 impacts on each ethnic 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.

Table 1: Pre 9-11 Ethnic Sentencing Differentials, Judicial Decisions

Sample: Federal Cases up for Sentencing between 10/1/1998 and 09/10/2001
Standard errors in parentheses clustered by ethnicity-district
Selection on unobservables (SoU) bounds in brackets

	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)
[Bounds: $\delta_{B(0)}$, $\delta_{B(1)}$]		[-.008, .007]		[-.692, -.649]		[-8.70, 3.88]
τ required for coefficient of 0		.551		-10.9		.315
Hispanic	.133**	.010	.156	-.768***	1.72	4.08***
	(.062)	(.015)	(.424)	(.100)	(4.17)	(.611)
[Bounds: $\delta_{H(0)}$, $\delta_{H(1)}$]		[-.035, .01]		[-1.05, -.768]		[4.08, 4.85]
τ required for coefficient of 0		.223		-1.65		-5.18
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
Unadjusted R-squared		.245		.219		.744
R^{max}=min(1, 1.3 x unadjusted R-squared)		.318		.285		.967
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 ethnicity-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 ethnicity (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. In parentheses we report bounds on the OLS estimate accounting for selection on unobservables using the Oster [2017] method: the bounds are set assuming the coefficient of proportionality is zero or one. Below the bounds we report the coefficient of proportionality that is required for the implied point estimate to be zero.

Table 2: Judicial 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	-.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 Ethnic 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 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.

Table 3: Citizenship and Offense Type

Dependent Variable: Downward Departure Granted by Federal Judge
Standard errors in parentheses clustered by ethnicity-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 ethnicity-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 covera ll 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. 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 4: Pre 9-11 Ethnic Sentencing Differentials, Prosecutorial Decision Making

Sample: Federal Cases up for Sentencing between 10/1/1998 and 09/10/2001
Standard errors in parentheses clustered by ethnicity-district
Selection on unobservables (SoU) bounds in brackets

	Non-zero Statutory Minimum			Statutory Minimum			Substantial Assistance Departure	
	(1) Unconditional	(2) Conditional	(3) Conditional	(4) Unconditional	(5) Conditional	(6) Conditional	(7) Unconditional	(8) Conditional
Black	.233***	.168***	.051***	29.0***	21.6***	7.81***	-.002	-.025***
	(.022)	(.015)	(.006)	(2.66)	(1.69)	(.869)	(.019)	(.006)
[Bounds: $\delta_{B(0)}$, $\delta_{B(1)}$]		[.142, .168]	[-.002, .051]		[18.6, 21.6]	[2.12, 7.81]		[-.034, -.025]
τ required for coefficient of 0		2.52	.970		2.62	1.33		-2.64
Hispanic	.054	.126***	.056***	4.30	13.9***	7.37***	-.115***	-.090***
	(.036)	(.020)	(.009)	(4.03)	(2.14)	(.982)	(.018)	(.008)
[Bounds: $\delta_{H(0)}$, $\delta_{H(1)}$]		[.126, .155]	[.056, .068]		[13.9, 17.9]	[7.37, 9.70]		[-.090, -.081]
τ required for coefficient of 0		-1.22	-4.32		-1.20	-2.58		4.86
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
Unadjusted R-squared		.148	.499		.136	.369		.143
R^{max}=min(1, 1.3 x unadjusted R-squared)		.192	.648		.177	.480		.186
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 ethnicity-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 ethnicity (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. In parentheses we report bounds on the OLS estimate accounting for selection on unobservables using the Oster [2017] method: the bounds are set assuming the coefficient of proportionality is zero or one. Below the bounds we report the coefficient of proportionality that is required for the implied point estimate to be zero.

Table 5: Prosecutorial Decision Making around 9-11

Standard errors in parentheses clustered by ethnicity-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)	(6) NE Sample
Initial charges post 9-11*Hispanic	.075*	10.7**	9.33**	1.81	-.037	.016
	(.042)	(5.34)	(4.65)	(2.65)	(.042)	(.012)
Initial charges post 9-11*Black	-.010	.684	-5.39	.846	-.053	.007
	(.048)	(7.50)	(7.36)	(3.66)	(.048)	(.014)
Initial charges post 9-11	-.033	-5.96	-8.29**	-.873	.035	-.004
	(.033)	(3.90)	(3.94)	(2.34)	(.037)	(.010)
Offender, Legal and District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Offense Type Codes	No	No	No	Final	Final	Final
Guideline Cell Dummies	No	No	No	Yes	No	No
p-value: [Post*B = Post*H]	.046	.172	.030	.755	.654	.382
Adjusted R-squared	.171	.147	.190	.797	.180	.155
Observations	3,612	3,600	3,612	3,612	3,612	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. 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 Columns 5 and 6 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 6: Pre-sentence Reports

OLS regression estimates; standard errors in parentheses clustered by ethnicity-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 ethnicity (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 ethnicity, 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 7: Juhn-Murphy-Pierce Decompositions of Hispanic-White Differentials

	Cohort 1: Judge Decisions	Cohort 2: Prosecutor Decisions
	(1) Downwards Departure	**(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 + σ-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: σ-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.

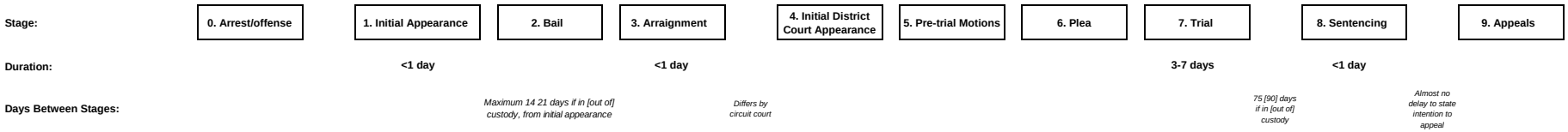
Table 8: Judges and Ingroup Bias

Dependent Variable in Columns 1-4: Coefficient on post 9-11 x Hispanic x District dummy, from NE sample
Dependent Variable in Column 5: Coefficient on Hispanic x District dummy, from full sample
Observations weighted by district share of Hispanics in 2001, robust standard errors in parentheses

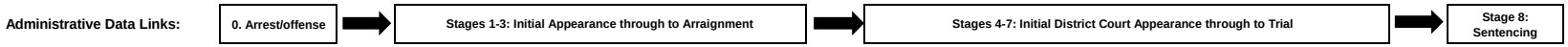
	Natural Experiment Sample				Pre 9-11 Sample
	(1) Ethnicity	(2) Other Judge Characteristics	(3) District Population	(4) Effect Size	(5) Hispanic Coefficient, Effect Size
District Proportion Hispanic Judges	.225***	.204**	.554***	.032***	.063**
	(.073)	(.101)	(.207)	(.012)	(.031)
District Proportion Black Judges	.272	.338	.097	.008	-.014
	(.217)	(.222)	(.207)	(.018)	(.020)
District Proportion Senior Status Judges		-.066	.027	.004	-.007
		(.076)	(.090)	(.014)	(.027)
District Proportion Male Judges		-.022	-.143	-.017	-.057**
		(.095)	(.093)	(.011)	(.028)
District Mean Judge Age		.006*	.004	.015	.051*
		(.003)	(.003)	(.014)	(.029)
District Proportion Democratic President Appointees		.180**	.137**	.025**	-.002
		(.076)	(.066)	(.012)	(.020)
District Proportion of Post-Period Window with Bush-Appointed US Attorney		.026	-.046	-.017	-
		(.027)	(.033)	(.013)	-
District Proportion Black (2000)			.275**	.032**	-.047**
			(.127)	(.015)	(.021)
District Proportion Hispanic (2000)			-.337*	-.034*	-.090*
			(.184)	(.019)	(.048)
Change in District Proportion Black (1990 - 2000)			-2.59**	-.027**	-.004
			(1.06)	(.011)	(.017)
Change in District Proportion Hispanic (1990 - 2000)			-.100	-.002	.057*
			(.519)	(.011)	(.033)
Mean of Dependent Variable			-.016		.041
Adjusted R-squared	.105	.172	.287	.287	.366
Observations	88	88	88	88	90

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The results in Columns 1 to 4 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. The results in Column 5 is based on the full sample (those that come up for sentencing from 10/1/1998 to 09/30/2003). 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. For Columns 1-4, 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. In Column 5, the dependent variable is the coefficient on Hispanic*District from a difference-in-difference regression for the full sample period with a full set of controls, and where the dependent variable is whether a downward 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 in Columns 1-4 we restricted 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. We perform an analogous sample cut of judges relevant for the full sample in Column 5. 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 Columns 4 and 5, effect sizes on all covariates are reported.

Figure 1: Federal CJS Timeline



Linkage Rates



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%

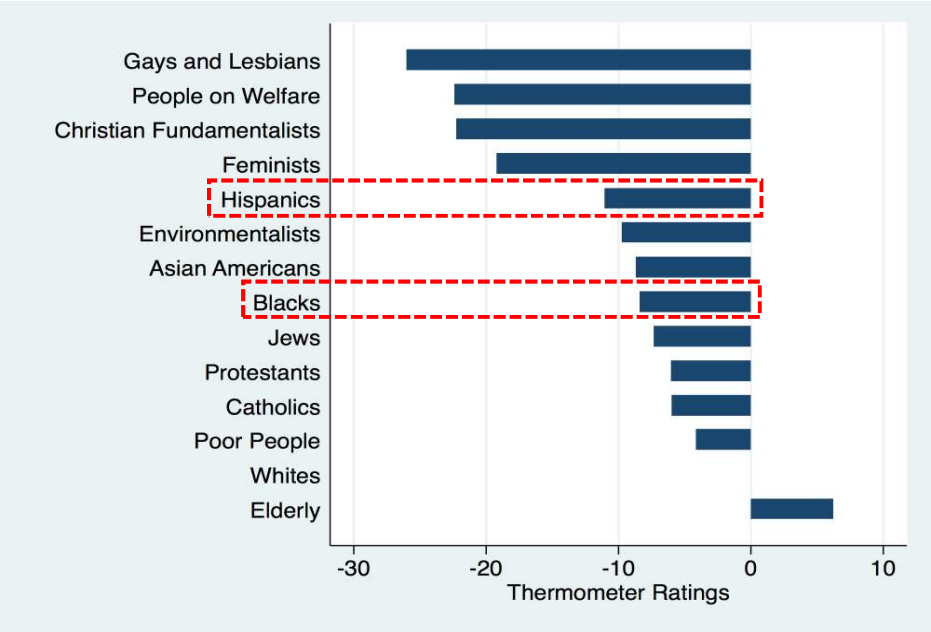
Panel B. Left-to-Right Linkage Rates

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%

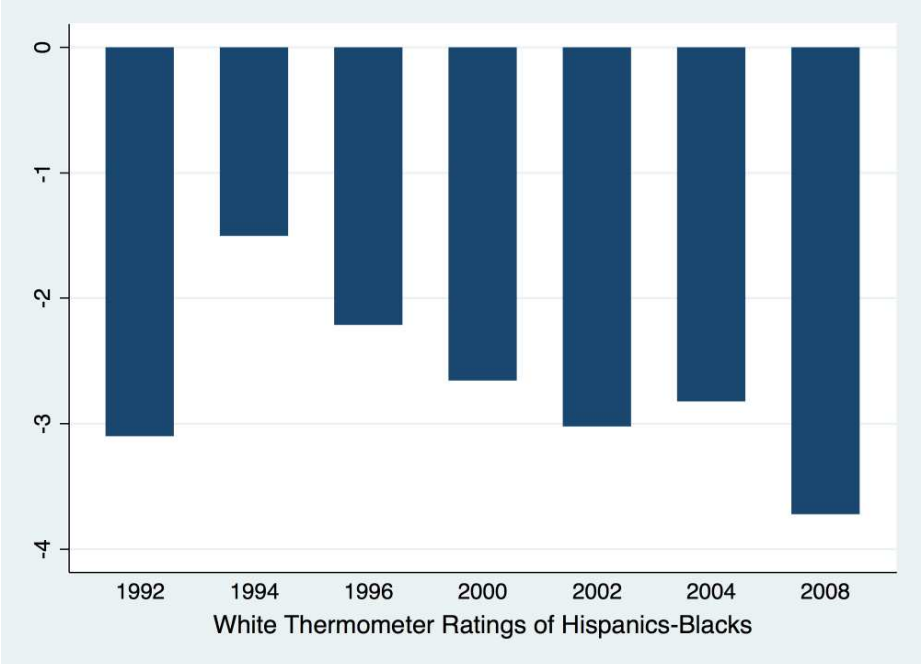
Notes: We use the Monitoring of Federal Criminal Sentences (MFCS) data set for our analysis. This comprises information gathered from four linked administrative data sources. As described in Rehavi and Starr [2014], the four linked data sets are: (i) US Marshals Service (USMS) data, that covers the arrest/offense 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) US Sentencing Commission (USSC) data, 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 2A: Pre 9-11 Sentiments Towards Hispanics

A. NES 2000 (Normalized by White-White Thermometer Rating)



B. NES Time Series: 1992-2008

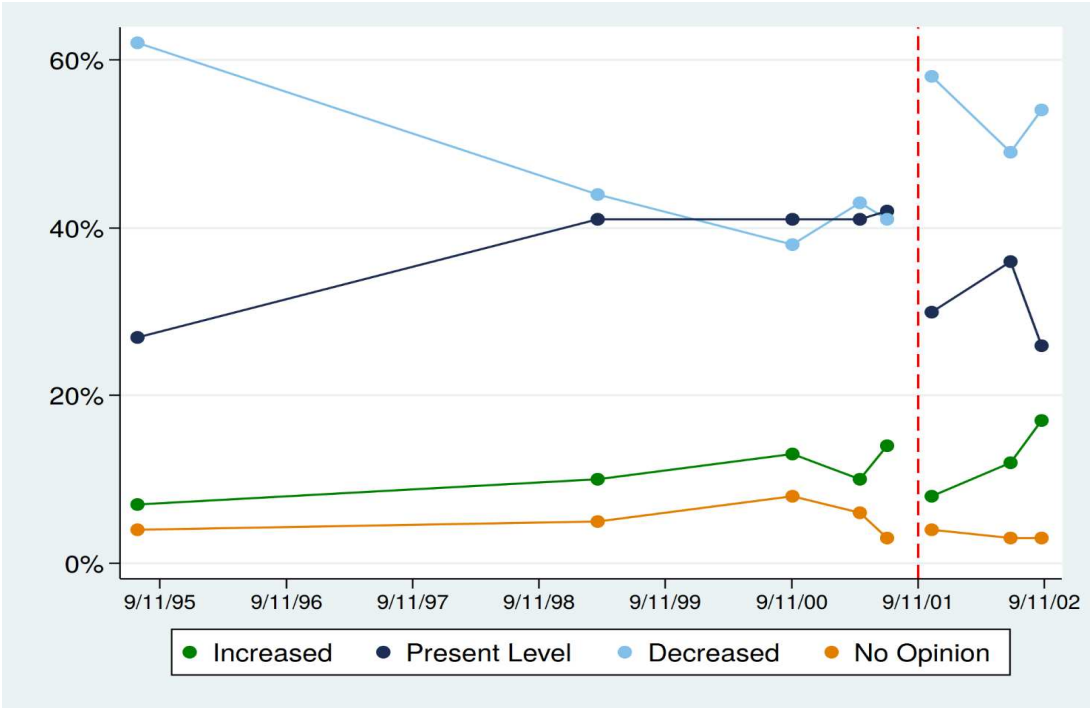


Notes: The graphs in panels A and B are constructed from the National Election Survey, and are based on the thermometer ratings of White respondents only. Respondents were asked about their feelings towards many groups in American society, and to represent these opinions on a "feeling thermometer". Respondents 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). In Panel A 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. In Panel B, the dependent variable is the difference between white respondents weighted mean rating of Hispanics minus that of Blacks, for each of the relevant survey years.

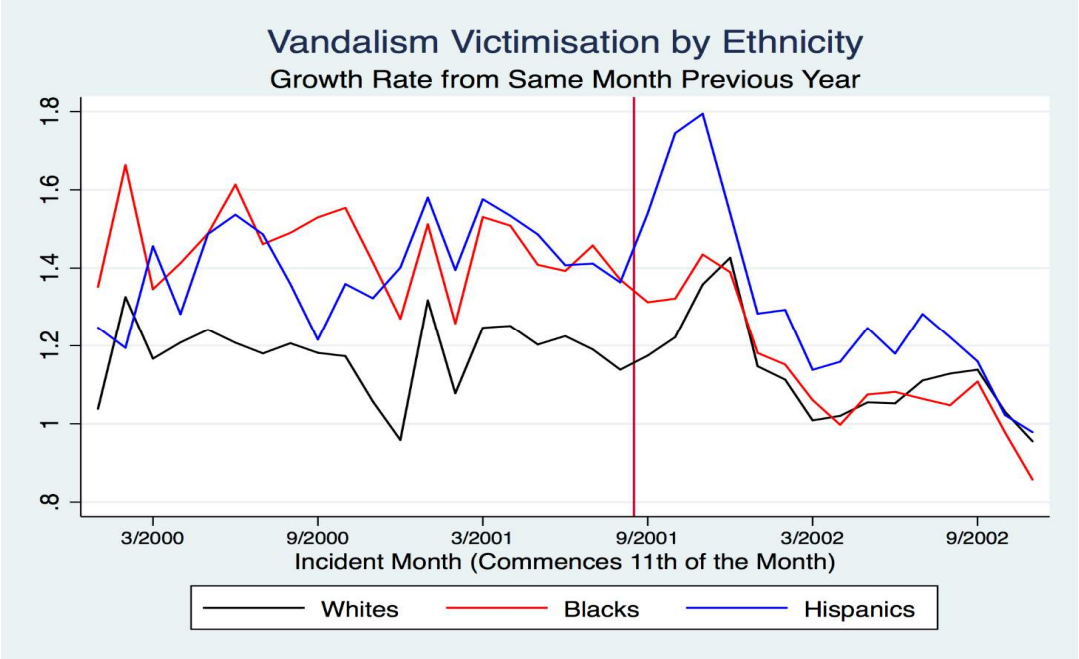
Figure 2B: Sentiments Towards Hispanics Around 9-11

A. Gallup Poll on Immigration

Q: Should Immigration be Kept at Its Present Level, Increased or Decreased?



B. Victimization

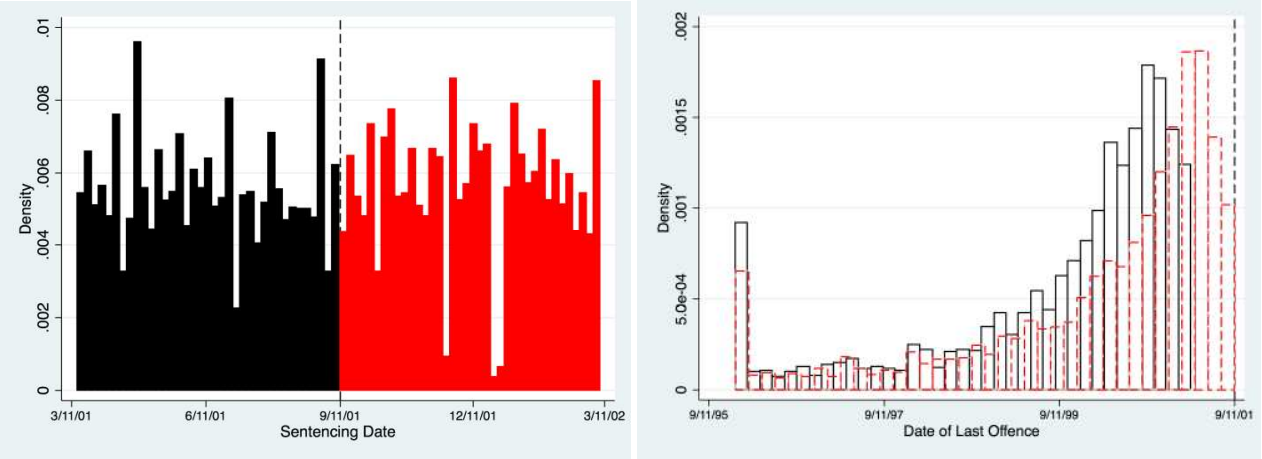


Notes: Panel A is based on a Gallup Poll that asks respondents, "Thinking more about immigration - that is, people who come from other countries to live here in the United States, in your view, should immigration be kept at its present level, increased or decreased?". The data was accessed via http://www.gallup.com/poll/1660/immigration.aspx. 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.

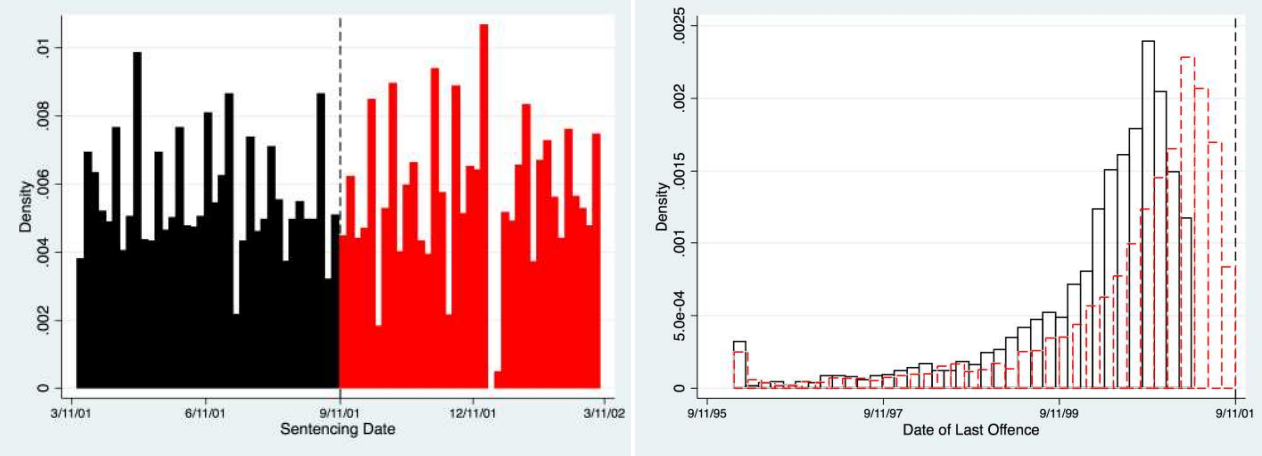
Figure 3: Sentencing and Last Offense Dates, by Ethnicity

A. Sentencing Date **B. Date of Last Offense**

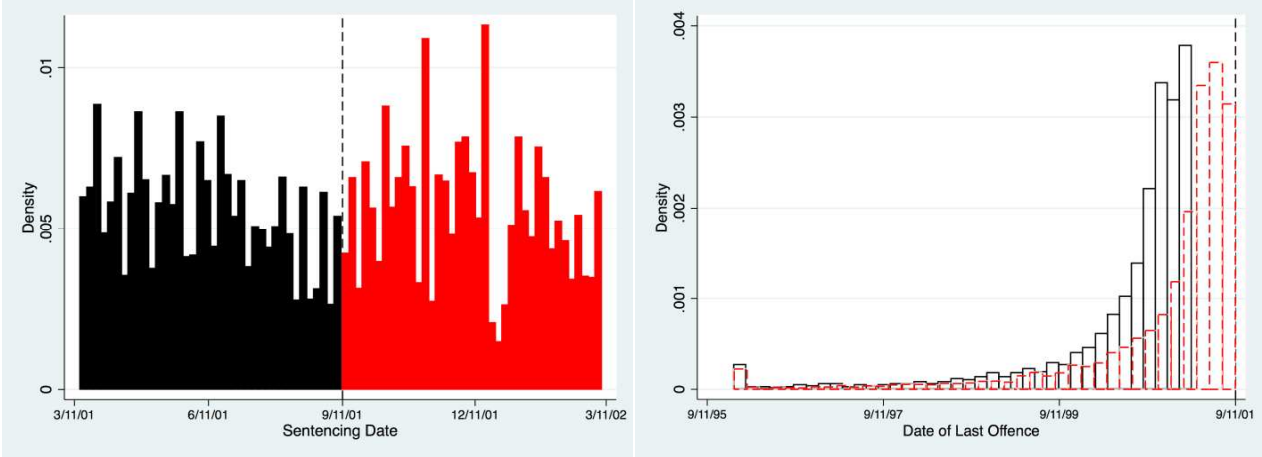
Whites



Blacks



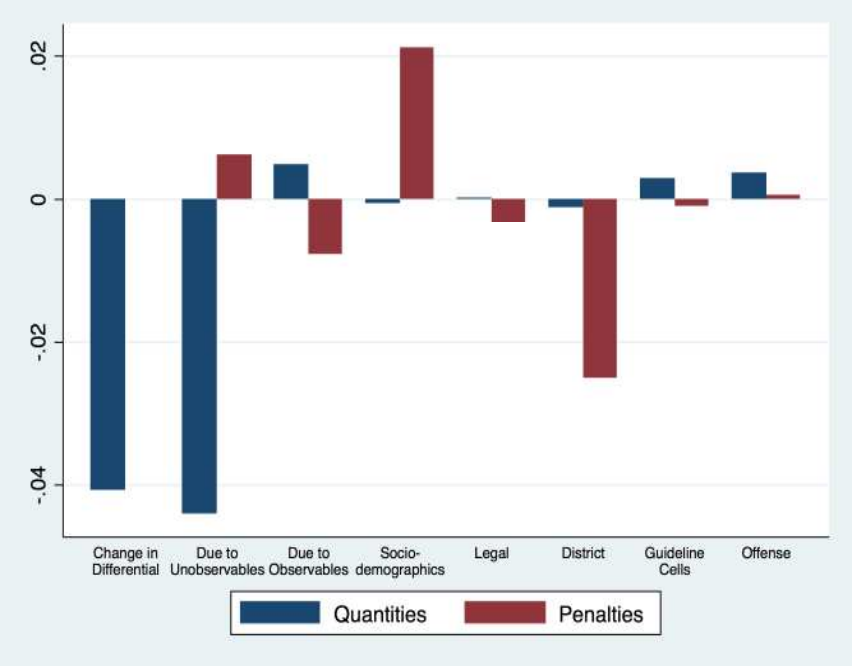
Hispanics



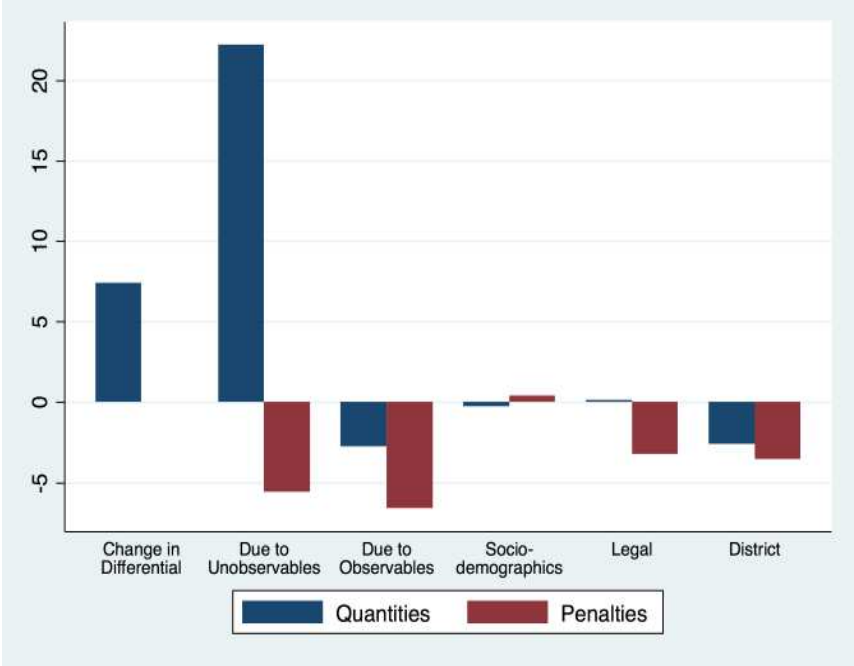
Notes: The left hand side figures show the distribution of dates of sentencing date, for each ethnicity: 9/11 is indicated by the vertical dashed line. The right hand side figures show the distribution of the dates of last offenses, by ethnicity. The first bar corresponds to a last offense date on or 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

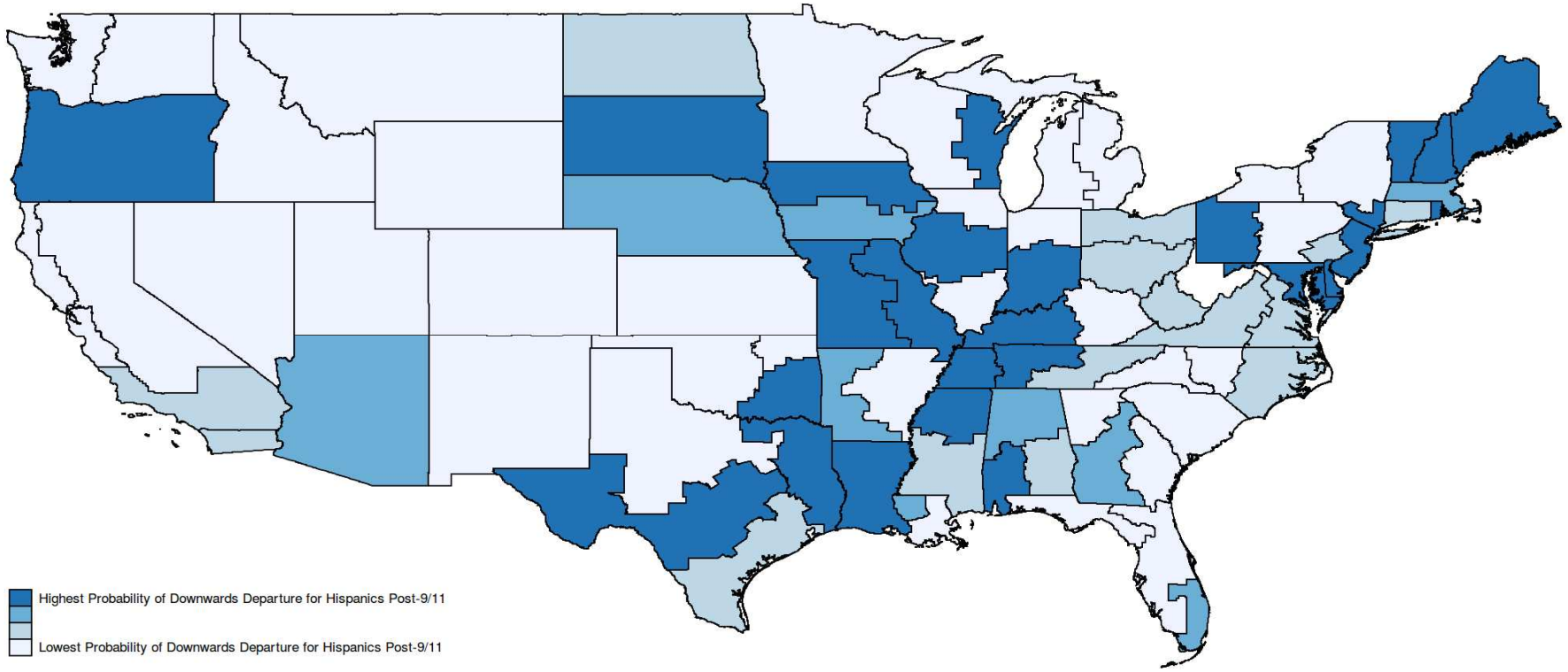


Cohort 2: Prosecutor Decisions
Statutory Minimum Sentence Length



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.

Figure 5: Spatial Patterns of Hispanic-White Sentencing Differentials



Notes: 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.

Table A1: Sentencing Guideline Cells (in months imprisonment)

		Criminal History Category (Criminal History Points)					
		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

Offense Level

Source: Chapter 5, 2001 Federal Sentencing Guidelines Manual [http://www.ussc.gov/sites/default/files/pdf/guidelines-manual/2001/manual/CHAP5.pdf]

Table A2: Detailed Federal CJS Timeline

Stage		Who is	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.

Source: http://gan.fd.org/pdfs/NDGA%20Timeline.pdf, accessed March 7th 2016.

Table A3a: Descriptives for the Pre 9-11 Sample

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

	White			Black			Hispanic			Total		
	Raw Sample	Working Sample	p-value	Raw Sample	Working Sample	p-value	Raw Sample	Working Sample	p-value	Raw Sample	Working Sample	p-value
Sample Size	41959	40694		35745	34649		61289	55552		138993	130895	
Number Dependents	1.171	1.171	[.999]	1.678	1.68	[.967]	1.823	1.828	[.961]	1.579	1.578	[.997]
	(1.442)	(1.441)		(1.842)	(1.84)		(1.798)	(1.796)		(1.73)	(1.728)	
Marital Status:												
Single	.294	.295	[.9]	.505	.506	[.881]	.288	.301	[.594]	.345	.354	[.683]
	(.456)	(.456)		(.5)	(.5)		(.453)	(.459)		(.475)	(.478)	
Married	.369	.371	[.888]	.209	.21	[.889]	.332	.351	[.446]	.312	.32	[.557]
	(.482)	(.483)		(.406)	(.407)		(.471)	(.477)		(.463)	(.466)	
Cohabiting	.077	.078	[.894]	.136	.138	[.873]	.142	.15	[.634]	.121	.124	[.74]
	(.267)	(.268)		(.343)	(.344)		(.349)	(.357)		(.326)	(.33)	
Divorced	.172	.173	[.857]	.066	.066	[.909]	.051	.053	[.746]	.091	.094	[.79]
	(.377)	(.379)		(.248)	(.249)		(.219)	(.223)		(.288)	(.292)	
Widowed	.006	.006	[.805]	.003	.003	[.987]	.002	.002	[.623]	.004	.004	[.73]
	(.079)	(.08)		(.055)	(.055)		(.048)	(.049)		(.061)	(.062)	
Separated	.05	.051	[.904]	.051	.051	[.949]	.044	.046	[.699]	.048	.049	[.677]
	(.219)	(.219)		(.219)	(.22)		(.205)	(.21)		(.213)	(.215)	
Education Level:												
Less than High School	.25	.252	[.896]	.4	.402	[.876]	.581	.615	[.447]	.435	.446	[.759]
	(.433)	(.434)		(.49)	(.49)		(.493)	(.487)		(.496)	(.497)	
High School Graduate	.371	.373	[.846]	.362	.364	[.827]	.156	.162	[.772]	.274	.281	[.792]
	(.483)	(.484)		(.481)	(.481)		(.363)	(.369)		(.446)	(.45)	
Some College	.226	.228	[.84]	.184	.185	[.883]	.071	.072	[.93]	.147	.15	[.841]
	(.418)	(.419)		(.387)	(.388)		(.256)	(.258)		(.354)	(.357)	
College Graduate	.131	.132	[.966]	.037	.038	[.984]	.019	.019	[.999]	.058	.059	[.894]
	(.338)	(.338)		(.19)	(.19)		(.138)	(.138)		(.233)	(.236)	
Age	38.802	38.848	[.912]	31.674	31.683	[.975]	31.999	31.987	[.982]	34.011	34.061	[.932]
	(12.195)	(12.178)		(9.284)	(9.267)		(9.168)	(9.197)		(10.714)	(10.745)	
Defense Counsel:												
Privately Retained	.173	.175	[.966]	.078	.079	[.957]	.072	.073	[.974]	.104	.106	[.914]
	(.379)	(.38)		(.268)	(.27)		(.259)	(.26)		(.306)	(.308)	
Court Appointed	.173	.174	[.975]	.179	.181	[.969]	.303	.298	[.969]	.232	.229	[.954]
	(.378)	(.379)		(.384)	(.385)		(.459)	(.457)		(.422)	(.42)	
Federal Public Defender	.111	.112	[.981]	.132	.131	[.97]	.246	.248	[.961]	.176	.175	[.965]
	(.315)	(.315)		(.339)	(.338)		(.431)	(.432)		(.381)	(.38)	
Self-represented	.004	.004	[.830]	.003	.003	[.931]	.001	0	[.712]	.002	.002	[.852]
	(.062)	(.06)		(.052)	(.051)		(.023)	(.021)		(.046)	(.044)	
Rights waived	.004	.003	[.798]	.004	.003	[.843]	.001	.001	[.882]	.002	.002	[.787]
	(.059)	(.056)		(.06)	(.057)		(.031)	(.03)		(.049)	(.047)	
Other Arrangements	0	0	[.968]	.001	.001	[.969]	0	0	[.905]	0	0	[.922]
	(.021)	(.021)		(.022)	(.023)		(.019)	(.019)		(.02)	(.021)	
Criminal History Score	2.155	2.162	[.922]	2.973	2.987	[.830]	2.402	2.395	[.952]	2.475	2.479	[.954]
	(1.629)	(1.632)		(1.823)	(1.823)		(1.699)	(1.687)		(1.74)	(1.737)	
Offense Severity	17.59	17.636	[.883]	22.255	22.328	[.865]	18.808	18.822	[.987]	19.336	19.381	[.928]
	(8.414)	(8.405)		(9.605)	(9.576)		(7.968)	(7.869)		(8.75)	(8.711)	

Notes: The full sample refers to all Federal cases that come up for sentencing from 10/1/1998 to 09/10/2001, the pre 9-11 period. For each ethnicity (and the sample as a whole), we show the descriptive statistic for all these cases (the "Raw Sample" Columns), and for those cases used in the main analysis where there is non-missing information for key covariates (the "Working Sample" Columns). Specifically, observations were dropped from the raw sample if the following variables were missing: district, race/ethnicity, criminal history, offense severity, sentence length or offense type. Means and standard deviations (in parentheses) 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 ethnicity-district.

Table A3b: Descriptives for the Natural Experiment Sample

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

	White			Black			Hispanic			Total		
	Raw Sample	Working Sample	p-value	Raw Sample	Working Sample	p-value	Raw Sample	Working Sample	p-value	Raw Sample	Working Sample	p-value
Sample Size	14226	12994		12054	10876		18212	16358		44492	40228	
Number Dependents	1.102	1.101	[.977]	1.685	1.689	[.938]	1.857	1.839	[.851]	1.561	1.555	[.934]
	(1.42)	(1.415)		(1.823)	(1.831)		(1.792)	(1.776)		(1.72)	(1.713)	
Marital Status:												
Single	0.332	0.336	[0.650]	0.53	0.534	[0.762]	0.319	0.328	[0.678]	0.380	0.387	[0.738]
	(0.471)	(0.473)		(0.499)	(0.499)		(0.466)	(0.47)		(0.485)	(0.487)	
Married	0.359	0.356	[0.828]	0.208	0.206	[0.793]	0.344	0.355	[0.547]	0.312	0.315	[0.819]
	(0.48)	(0.479)		(0.406)	(0.404)		(0.475)	(0.478)		(0.463)	(0.464)	
Cohabiting	0.078	0.077	[0.966]	0.128	0.128	[0.970]	0.153	0.155	[0.877]	0.122	0.123	[0.935]
	(0.268)	(0.267)		(0.334)	(0.334)		(0.36)	(0.362)		(0.327)	(0.328)	
Divorced	0.159	0.159	[0.975]	0.061	0.061	[0.953]	0.052	0.052	[0.959]	0.089	0.089	[0.962]
	(0.365)	(0.366)		(0.239)	(0.239)		(0.223)	(0.222)		(0.284)	(0.285)	
Widowed	0.005	0.005	[0.905]	0.003	0.003	[0.591]	0.002	0.002	[0.881]	0.003	0.003	[0.734]
	(0.068)	(0.067)		(0.054)	(0.051)		(0.046)	(0.046)		(0.056)	(0.055)	
Separated	0.048	0.048	[0.960]	0.048	0.048	[0.950]	0.047	0.047	[0.983]	0.048	0.047	[0.981]
	(0.214)	(0.214)		(0.214)	(0.213)		(0.211)	(0.211)		(0.213)	(0.213)	
Education Level:												
Less than High School	0.262	0.265	[0.810]	0.403	0.405	[0.852]	0.613	0.634	[0.529]	0.444	0.453	[0.799]
	(0.440)	(0.441)		(0.49)	(0.491)		(0.487)	(0.482)		(0.497)	(0.498)	
High School Graduate	0.383	0.384	[0.950]	0.372	0.371	[0.933]	0.182	0.183	[0.938]	0.298	0.299	[0.960]
	(0.486)	(0.486)		(0.483)	(0.483)		(0.386)	(0.387)		(0.457)	(0.458)	
Some College	0.22	0.22	[0.982]	0.181	0.181	[0.940]	0.076	0.074	[0.884]	0.15	0.15	[0.991]
	(0.414)	(0.414)		(0.385)	(0.385)		(0.264)	(0.262)		(0.357)	(0.357)	
College Graduate	0.124	0.122	[0.830]	0.038	0.037	[0.892]	0.022	0.02	[0.778]	0.059	0.057	[0.874]
	(0.33)	(0.327)		(0.191)	(0.189)		(0.146)	(0.14)		(0.235)	(0.233)	
Age	38.439	38.206	[0.541]	31.797	31.718	[0.787]	32.325	32.162	[0.723]	34.143	33.997	[0.769]
	(12.167)	(12.093)		(9.251)	(9.242)		(9.285)	(9.234)		(10.707)	(10.655)	
Defense Counsel:												
Privately Retained	0.165	0.166	[0.973]	0.079	0.08	[0.949]	0.082	0.079	[0.887]	0.108	0.107	[0.985]
	(0.371)	(0.372)		(0.27)	(0.272)		(0.275)	(0.27)		(0.31)	(0.31)	
Court Appointed	0.170	0.170	[1]	0.159	0.16	[0.963]	0.275	0.274	[0.994]	0.21	0.21	[0.998]
	(0.376)	(0.376)		(0.365)	(0.367)		(0.446)	(0.446)		(0.407)	(0.407)	
Federal Public Defender	0.132	0.134	[0.905]	0.152	0.154	[0.952]	0.256	0.267	[0.852]	0.188	0.194	[0.869]
	(0.338)	(0.341)		(0.359)	(0.361)		(0.437)	(0.442)		(0.391)	(0.395)	
Self-represented	0.004	0.003	[0.576]	0.003	0.002	[0.646]	0.000	0.000	[0.718]	0.002	0.002	[0.487]
	(0.061)	(0.054)		(0.056)	(0.047)		(0.02)	(0.017)		(0.047)	(0.041)	
Rights waived	0.001	0.001	[0.819]	0.002	0.002	[0.987]	0.001	0.001	[0.930]	0.001	0.001	[0.951]
	(0.032)	(0.034)		(0.047)	(0.047)		(0.029)	(0.029)		(0.036)	(0.036)	
Other Arrangements	0.000	0.000	[0.948]	0.000	0.000	[0.942]	-	-	-	0.000	0.000	[0.924]
	(0.012)	(0.012)		(0.018)	(0.019)		-	-		(0.012)	(0.012)	
Criminal History Score	2.209	2.214	[0.944]	3.067	3.061	[0.934]	2.377	2.39	[0.900]	2.511	2.514	[0.961]
	(1.66)	(1.657)		(1.832)	(1.822)		(1.673)	(1.667)		(1.748)	(1.741)	
Offense Severity	18.107	17.81	[0.311]	22.119	21.705	[0.304]	19.307	18.982	[0.671]	19.687	19.34	[0.395]
	(8.358)	(8.205)		(9.209)	(9.025)		(7.971)	(7.695)		(8.594)	(8.376)	

Notes: The natural experiment 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, 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. For each ethnicity (and the sample as a whole), we show the descriptive statistic for all these cases (the "Raw Sample" columns), and for those cases used in the main analysis where there is non-missing information for key covariates (the "Working Sample" Columns). Specifically, observations were dropped from the raw sample if the following variables were missing: district, race/ethnicity, criminal history, offense severity, sentence length, offense type or date of final offense. We further restrict the sample to cases in which: (i) guilt pleas are filed (that is so for 96% of defendants); (ii) three or fewer offenses were committed because for offenses in the 2002 tax year (those that come up for sentencing from 01/10/2001 through to 30/09/2002), in the MCFS data we only observe the date of offense for the first three offenses. Means and standard deviations (in parentheses) 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 ethnicity-district.

Table A4: Robustness Checks on Ethnic Sentencing Differentials Around 9-11

Dependent Variable: Downward Departure Granted (0/1)

Standard errors in parentheses, clustered by ethnicity-district unless otherwise stated

	(1) Cluster on sentence week x ethnicity	(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 ethnicity in Column 1, and by ethnicity-district in Column 2. 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 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 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 A5: Ethnic Sentencing Differentials Around 9-11, by Ethnicity

Dependent Variable: Downward Departure Granted (0/1)
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 A6: Racial Sentencing Differentials Around 9-11

Dependent Variable: Downward Departure Granted (0/1)
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 A7: Time in the Federal CJS

Dependent Variable: Downward Departure Granted (0/1)
Standard errors in parentheses clustered by ethnicity-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 ethnicity-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 ethnicity (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 A8: Time Between Dates of Last offense and Sentencing

OLS and survival regression estimates; standard errors in parentheses, clustered by ethnicity-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 ethnicity (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 A9: Placebo

Dependent Variable: Downward Departure Granted (0/1)
Standard errors in parentheses clustered by ethnicity-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 ethnicity-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 ethnicity (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 A10: Bush Appointed US Attorneys

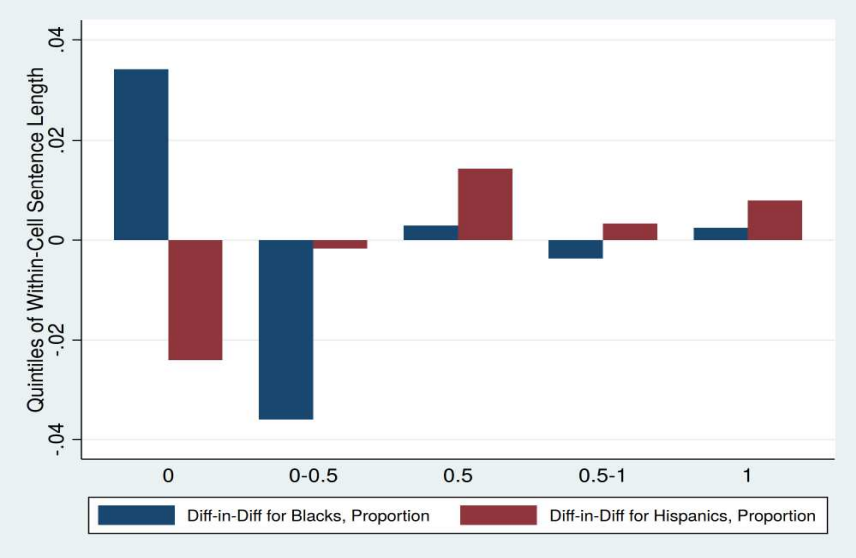
Dependent Variable: Downward Departure Granted (0/1)
Standard errors in parentheses clustered by ethnicity-district
Deviations from mean at district level

	(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

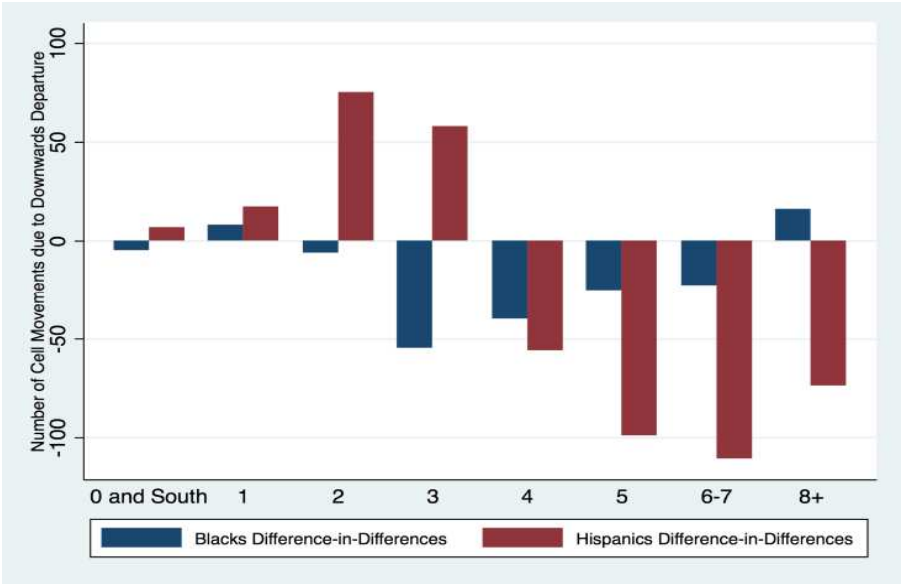
Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. OLS regression estimates are shown in all Columns. Standard errors are reported in parentheses, where these are clustered by ethnicity-district. The sample of 40,228 Federal cases is used (those that come up for sentencing in a six month 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.

Figure A1: Judicial Decision Making, Other Channels

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



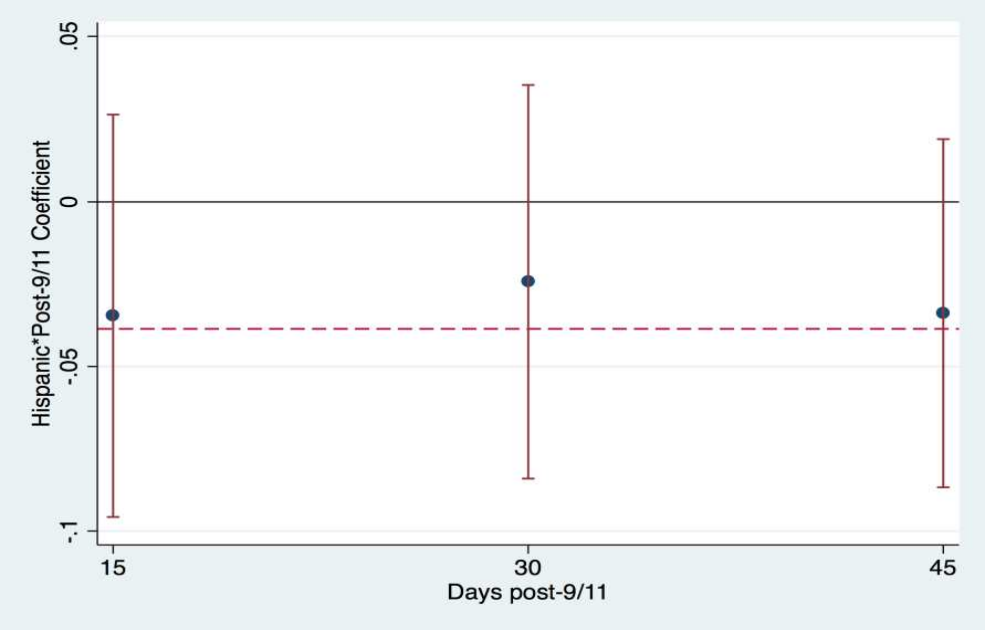
B. Cell Movements | Downward Departure = 1



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 Figure 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.

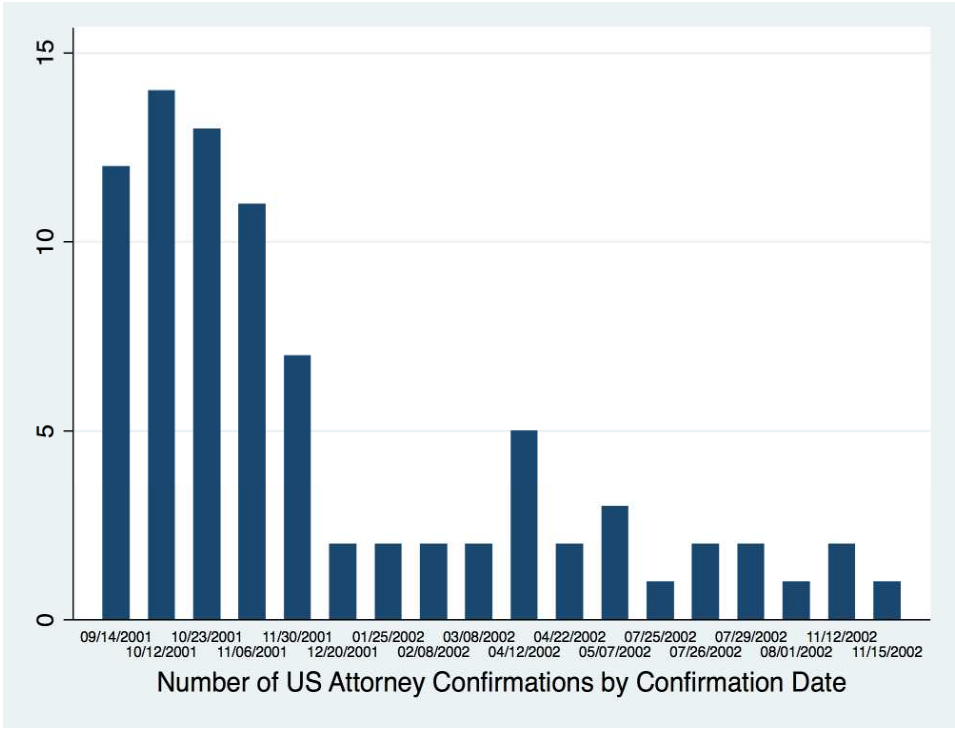
Figure A2: Patriot Act

A. 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 dummies.

Figure A3: 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.