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

Brendon McConnell Imran Rasul

May 2020

Abstract

We investigate whether increased animosity towards Muslims post 9-11 had spillover effects on Black and Hispanic individuals in the federal criminal justice system. Using linked administrative data tracking defendants from arrest through to sentencing, we find that after 9-11, sentence and pre-sentence outcomes for Hispanic defendants significantly worsened. Outcomes for Black defendants were unchanged. The findings are consistent with judges and prosecutors displaying social preferences characterized by contagious animosity from Muslims to Hispanics. Our findings provide among the first field evidence of contagious animosity, so that social preferences across outgroups are interlinked and malleable.

*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, Neale Mahoney, Alan Manning, Olivier Marie, Ioana Marinescu, Michael Mueller-Smith, Aviv Nevo, 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

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

We advance this literature using three novel pillars of analysis to identify and measure the decisions of judges and prosecutors that determine differential outcomes by race/ethnicity. The building blocks underlying our analysis are: modifying the notion of ingroup and outgroup bias in societies comprising multiple groups/identities, using a novel research design built around this notion, and exploiting linked administrative records tracking defendants through all stages of the federal criminal justice system (CJS).

A vast literature examines the biological and evolutionary roots of ingroup bias [Tajfel *et al.* 1971]. Individuals are assumed to have some social preference over the payoffs to their ingroup, and their outgroup, where they favor their ingroup more strongly. As with individual preferences, the standard view is that such social preferences are stable and immutable.[1] However, there has been increasing attention on alternative formulations that suggest such social preferences are malleable. A nascent body of laboratory evidence shows agents can display contagious altruism: under this view, positive altruism towards an outgroup fosters altruism towards the ingroup. A second scenario is one of parochial altruism: under this view, greater rivalry between groups fosters more cooperation within the ingroup.[2]

We apply these notions to US society, where individuals can have one of many identities. There is thus one ingroup and multiple outgroups so social preferences are defined over all these groups. We then ask does increased animosity towards one outgroup drive social preferences towards another outgroup. The answer is no if social preferences across outgroups are independent. On the other hand, there can be *contagious animosity* across outgroups so hostility towards one outgroup drives hostility towards others. Alternatively, there might be *parochial animosity* so hostility

[1]Social psychologists have documented dimensions such as race, ethnicity, religiosity and political affiliation, as all being salient across contexts, in driving ingroup biases. In economics, ingroup biases have been studied in laboratory settings and show to emerge even in artificially created groups [Shayo 2009, Bertrand and Duflo 2016].

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

towards one outgroup increases altruism towards other outgroups. While the study of ingroup-outgroup biases goes back decades, to the best of our knowledge, there has been little examination of spillover effects across outgroups [Bertrand and Duflo 2016]. The notion is important because it implies a malleability of outgroup biases, and that anti-discrimination policies against one outgroup can have positive or negative externalities on other outgroups.

We use the ideas of contagious/parochial animosity to construct a research design to examine racial/ethnic sentencing differentials in the federal CJS: a high stakes and professional economic environment. This is a setting in which defendants are of multiple identities (by race, ethnicity, citizenship etc.) and the vast majority of federal judges and prosecutors during our study period are White, so we view them as the ingroup. We consider 9-11 as an exogenously timed event that heightened the salience of insider-outsider differences in US society, and specifically, increased animosity towards Muslims [Human Rights Watch 2002, Davis 2007, Woods 2011]. We use this exogenously timed shock towards one outgroup to measure spillovers on sentencing outcomes in the CJS for other outgroups, namely for Black and Hispanic defendants.

A priori, not all outgroups would be equally impacted through spillovers induced by the structure of social preferences. In particular, there are reasons why Hispanic defendants are closer to Muslims in social construct than other outgroups. Drawing on work in sociology, we provide a detailed account of how Islamophobia and immigration have become gradually intertwined in American consciousness since the mid 1990s, but were most forcefully framed together in the aftermath of 9-11 [Romero and Zarrugh 2018]. Three channels are identified linking Islamophobia and Hispanics: (i) political rhetoric; (ii) policy framing; (iii) restructured institutions.

We examine the impact of 9-11 on sentencing gaps across races/ethnicities using the Federal Justice Statistics Resource Center (*FJSRC*) data combined with the Monitoring of Federal Criminal Sentences (*MFCS*) data set. This covers the universe of all male defendants up for sentencing from 1998 to 2003, so either side of 9-11 and totaling 230,000 federal criminal cases. It is nationally representative, covering cases from all 90 mainland US Districts, defendants of all ages, and all types of criminal offense. Such large and representative samples allow for both Black-White and Hispanic-White differentials to be studied. Moreover, the *FJSRC* comprises four linked administrative data sources covering the time from a defendant's initial arrest and offense charge, and all subsequent stages of their processing through the federal CJS. This linked administrative dataset thus allows pre-sentencing differential treatment arising from the behavior of prosecutors or legal counsel to be studied alongside the behavior of judges at sentencing. Furthermore, it enables us to pin down whether judges and prosecutors display similar kinds of social preference structures across outgroups, and to address long-standing challenges for empirical work on the CJS that is typically based on sentencing data only [Klepper *et al.* 1983].

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

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

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

Our core results are as follows. We first confirm that relative to Whites, Blacks and Hispanics sentenced pre 9-11 receive significantly longer prison sentences. For Hispanics sentenced post 9-11, sentencing differentials become further exacerbated through a specific channel: they become 13.5% less likely to receive a downward departure than Whites. The implied increase in sentence length for Hispanics is .736 months, corresponding to 18% of the conditional pre 9-11 differential in sentence length. Placing a monetary value on this increased incarceration suggests the spillover effects from heightened animosity towards Muslims post 9-11 towards Hispanics, led to an increase of $1547 in incarceration costs per Hispanic defendant. This maps to a large increase in total costs for the federal CJS given the modal defendant in the study period is Hispanic.

We further develop an approach to identify the marginal defendants most likely to be impacted by changes in judges' propensity to downward depart. We find that among marginal defendants, 9-11 led to a increased Hispanic-White sentence differential of just over two months, corresponding to 50% of the conditional pre 9-11 differential in sentence length. The magnitude of this is comparable to sentencing differentials across groups that opened up after sentencing guidelines were abolished altogether in 2005 [Yang 2015].

Black-White sentencing differentials around 9-11 are unaffected along all sentencing margins, and as far as the data allows, we find the post 9-11 impacts to be statistically similar for Hispanic citizens and Hispanic non-citizens. Overall, the results are consistent with judges displaying contagious animosity from Muslims to Hispanics, while their social preferences are independent between Muslim and Black defendants, and we find no evidence that 9-11 leads to greater altruism within the majority ingroup.

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

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

As with judges, the results on prosecutors' decisions are consistent with them displaying contagious animosity from Muslims to Hispanics and their social preferences being independent between Muslim and Black defendants. More precisely, Hispanic defendants initially charged post 9-11 are 7.5pp more likely to receive an initial offense that carries a statutory minimum, and their statutory minimum sentence is 10.7 months longer. These impacts correspond to: (i) 60% of the pre 9-11 Hispanic-White gap in the the likelihood of an initial offense charge with a mandatory minimum; (ii) 77% of the pre 9-11 Hispanic-White gap in the statutory minimum sentence length. Indeed, these causal responses to 9-11 lead the Hispanic-White differential on each margin to become as large as the pre 9-11 Black-White differential.[3]

[3]On prosecutorial biases, Rehavi and Starr [2014] use related linked administrative data from the Federal CJS to show that prosecutor's initial offense charges account for half the Black-White sentencing gap. They do so for

Having established a causal spillover of 9-11 on Hispanic outcomes in the federal CJS, our final set of results probe the data to narrow the interpretation of these widening Hispanic-White differentials. As best as the data allows, we explore whether the results can be explained through statistical discrimination (say through higher expected recidivism rates of Hispanics post 9-11).

We first present a Juhn *et al.* [1993] decomposition of sentencing differentials between those that come up for sentencing post 9-11, where Hispanics are significantly less likely to receive a downwards departures from judges. The decomposition shows that only negligible amounts of the unconditional DD in outcome can be attributed to either differences in their observables relative to Whites, or the sentencing penalties of such observables. This helps to rule out explanations for the increased Hispanic-White differential based on the harshness with which certain offense types are dealt with post 9-11, offender characteristics including those that might perhaps closely predict recidivism such as the guideline cell they are assigned to, or explanations related to effort or allocation of legal counsel to defendants post 9-11. Overall, the decomposition suggests explanations for why Hispanic-White sentencing differentials worsen post 9-11 based on statistical discrimination do not easily fit the evidence.

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

Both strategies suggest 9-11 had spillover effects on Hispanics through decisions made by judges, with them having social preferences displaying contagious animosity from Muslim to Hispanic outgroups, but independence between Muslim and Black outgroups.

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

We provide among the first field evidence based on a quasi-experimental research design of the existence of contagious animosity. We do so in the high stakes and professional environment of the federal CJS. Earlier work on sentencing differentials in other parts of the CJS has explicitly

the period 2006-8, after sentencing guidelines have been abolished.

or implicitly framed the issue in terms of ingroup and outgroup biases [Bushway and Piehl 2001, Shayo and Zussman 2011, Abrams *et al.* 2012, Anwar *et al.* 2012, Rehavi and Starr 2014]. By allowing for multiple outgroups and developing the notion of contagious/parochial altruism, our work has the important implication that in multi-group societies, effective anti-discrimination policies targeting one group can have positive externalities onto other minority groups. Our analysis also helps address an appeal made in recent overviews of the economics of discrimination literature on the need to better bridge to the psychology literature on the origins of discriminatory behavior [Charles and Guryan 2011, Bertrand and Duflo 2016].

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

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

2 The Federal Criminal Justice System

Criminal cases are filed in federal court if prosecuted by a federal agency or related to federal law. In 2000 the three most frequent criminal offenses were for drug trafficking (40%), immigration (22%), and fraud (9%). This is a high stakes setting: cases heard in federal courts tend to be more serious than those in state courts. 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.[4]

[4]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 prosecutors hold greater sway in such decisions given the greater resources at their disposal [Jeffries and Gleeson 1995]. The sorting of cases into systems is therefore an executive branch decision: judges and defense counsel have no formal role. The DD research design we use to estimate sentencing differentials eliminates cross sectional differences between defendants, by race, being sent to trial in the federal system. Glaeser *et al.* [2000] provide a theoretical and empirical analysis of the sorting of cases into state and federal systems. The difference in severity across courts is not driven by the

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

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

2.1 Timeline

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

Two other aspects of the timeline are of note. First, a magistrate judge handles the first stages of a defendant's passage through the CJS. At arraignment, the magistrate will issue a scheduling order and which district court judge will actually preside over the case. With the exception of pre-trial motions hearings which are heard by the magistrate, the district court judge presides over the rest of the case (Stage 6 onwards). Second, the recommended guideline cell is determined between

composition of offenses: within offense type there is considerably harsher sentencing in federal courts, reflecting the greater seriousness of such crimes.

trial and sentencing (Stages 7 and 8): this is when the pre-sentence report (PSR) is drafted by the (neutral) Probation Office, the defendant's legal counsel and prosecutors. 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). It reports a determined criminal history score and the offense severity and thus the recommended guideline cell.

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

2.2 Linked Administrative Data

The *FJSRC* dataset comprises four linked administrative data sources covering the arrest/offense stage before an individual enters the federal CJS (Stage 0), and all subsequent stages shown in Figure 1. For sentencing stage 8, we use the *MFCS* data (that can be linked to earlier data sets in the *FJSRC*).[5] We focus on male defendants. Our sample covers $230,000$ federal criminal cases up for sentencing from October 1998 to September 2003 across nearly all US districts. The Appendix provides further data details. To estimate Black-White and Hispanic-White sentencing differentials, we use two variables available at sentencing Stage 8. In the first, defendants are coded as Hispanic (41%) or non-Hispanic (59%). A separate race code then identifies defendants as white-race (71%), black-race (29%), other-race ($< .1\%$). We code Whites as white-race non-Hispanic; Blacks as black-race non-Hispanic; Hispanics as white- or black-race Hispanics. This implies 31% of defendants are White, 26% are Black and 43% are Hispanic.

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

[5]As explained in the Appendix, the *MFCS* data is superior to the USSC data in the *FJSRC* (even though it also originates from the USSC) because it contains exact sentence dates, and dates of last offense.

[6]These include kidnaping/hostage taking, sexual abuse, assault, bank robbery (including arson), drugs: trafficking, drugs: communication, drugs: simple possession, firearms: use (including burglary/breaking and auto theft), larceny, fraud, embezzlement, forgery/counterfeiting, bribery, tax offenses, money laundering, racketeering (including gambling/lottery), civil rights offenses, immigration, pornography/prostitution, offenses in prisons,

cell recommended to the judge in the pre-sentence report. This effectively proxies all case-specific factors the prosecution and legal counsel deem judges should factor into sentencing. However, the data does not identify the cell the defendant was then placed into if downward departed: we only observe the sentence length, that as Figure A1 makes clear, might correspond to many different cells. We later detail the algorithm we use to provide an indication of the number of cells moved conditional on being downward departed.

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

2.3 Federal Sentencing Guidelines

Federal sentencing guidelines were introduced in the Sentencing Reform Act of 1984 by the US Sentencing Commission (USSC). The goal was to alleviate sentencing disparities through *determinate* sentencing, limiting the discretion judges had over sentencing. Parole boards were also abolished so that actual incarceration length became a fixed threshold of 85% of determined sentences.

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

Table A1 shows the full set of guideline cells, mapping each combination of offense severity (1

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

to 43) and criminal history (1 to 13, grouped into 6 bins) into a sentencing range. There are 43 x 6 = 258 guideline cells. These include those in Zone A, where the guidelines include zero sentence length, and cells in Zone D where the guidelines impose a life sentence.

Between trial/conviction and sentencing (Stages 7 and 8), the pre-sentence report is drafted by prosecutors, legal counsel and an independent probation officer. This recommends a guideline cell. However, the guidelines still provide judges discretion to *downwards depart* from the recommended guideline cell, and move in a Northerly direction in the guideline cell Table A1. A judge can do so if they find mitigating circumstances of a kind not adequately taken into consideration by the USSC in formulating the sentencing guidelines. These circumstances include diminished capacity or rehabilitation after the offense but prior to sentencing, family responsibilities or prior good works. Downward departures may also be warranted if "*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 their reason(s) for downward departing.

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

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

3 Descriptives, 9-11, Research Design

3.1 Pre 9-11 Sentencing Differentials

We examine pre 9-11 sentencing differentials along two margins of judicial decision making: (i) if a downward departure is granted; (ii) the sentence length (in months).

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

[7]The guideline cells were in operation until 2005. The Supreme Court's 2005 decision in US v. Booker found the mandatory application of guidelines to be unconstitutional. The guidelines are now considered advisory.

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

Columns 2 and 4 show that conditioning on covariates, there are large changes in the Black- and Hispanic-dummy coefficient estimates on each margin. This is expected given defendants in each group differ on observables. However, even conditional on covariates including the recommended guideline cell, we see that statistically significant Black-White and Hispanic-White sentencing differentials remain. For example, Black and Hispanic defendants have significantly longer sentence lengths. A natural benchmark we use for the later analysis on any spillover impacts of 9-11 on outgroups, is the pre 9-11 conditional sentencing gap, that is around 4 months for both outgroups relative to Whites, or around 10% of the White sentence length.

3.2 Linking Muslim and Hispanic Outgroups

We aim to understand whether judges and prosecutors display social preferences characterized by contagious or parochial animosity across outgroups. We do so by exploiting 9-11 as an exogenously timed increase in animosity towards one outgroup: Muslims. 9-11 certainly increased animosity towards Muslims [Human Rights Watch 2002, Davis 2007, Woods 2011], and reduced their rates of assimilation [Gould and Klor 2016]. Not all outgroups would be impacted by any resulting contagious/parochial animosity but there are reasons why Hispanics are closer to Muslims in social construct than other outgroups. To understand the link between 9-11 and Hispanics, we draw on work in sociology by Romero and Zarrugh [2018]. They provide a detailed account of how Islamophobia and immigration have become gradually intertwined in American consciousness since the mid 1990s, but were most forcefully framed together in the aftermath of 9-11. They build an evidence base for this thesis by analyzing government reports, media accounts, non-governmental evaluations, statements by politicians, and other secondary sources. They argue that Islamophobia – or the extreme and irrational fear of Muslims and Islam – was deployed against Hispanics to garner political support, and justify increased surveillance and immigration enforcement. Romero and Zarrugh [2018] 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. On policy, immigration and terrorism issues have slowly become intertwined since the 1995 Oklahoma bombings. Two prominent legislative Acts linked immigration and terrorism pre 9-11: the Illegal Immigration Reform and Responsibility Act, and the Anti-Terrorism and Effective Death Penalty Act. Both became law in 1996, linking terrorism and immigration and broadening the set of federal criminal cases subject to deportation. Post 9-11 the Patriot Act came into effect 45 days later, further increasing the link between terrorism and immigration through its near exclusive focus on immigration offenses. On institutions, the formation of the Department of Homeland Security (DHS) represented the first time terrorism and immigration agencies had been merged. The DHS merged 22 federal agencies, and as such the culture of the joint bureaucracy changed.

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

To provide quantitative evidence on impacts on Hispanics in the immediate post 9-11 period, Panel A of Figure 2 shows time series evidence from a Gallup Poll on immigration: this highlights a marked shift against immigration post 9-11. Panel B shows vandalism victimization rates, by race/ethnicity. The data show a spike in Hispanics reporting being victims of vandalism post 9-11, with the growth rates in victimization rates only slowly returning back to trend. Other studies show 9-11 worsened labor market outcomes for Hispanics [Orrenius and Zavodny 2009].[8]

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

3.3 Research Design

To isolate the impact 9-11 had on sentencing outcomes, we compare outcomes between: (i) defendants who committed their last offense before 9-11 and were sentenced *before* 9-11; (ii) to defendants who *also* committed their last offense before 9-11, but were sentenced *after* 9-11. We construct a second difference in outcomes across race/ethnicity to estimate a DD impact of 9-11 on criminal sentencing. Our working sample is based on a ±180 day sentencing window around 9-11 2001, where *all* defendants have committed their offense prior to 9-11, and hence entered the federal CJS timeline in Figure 1, but some were sufficiently far advanced along so as to come up

[8] Legewie [2013] documents worsening attitudes towards immigrants in response to terrorist events in a range of countries; Hopkins [2010] uses panel data around 9-11 to show that it had a profound short run impact on attitudes towards immigrants.

for sentencing pre 9-11, while others had only just entered the timeline prior to 9-11 and so ended up being sentenced post 9-11. To maintain comparability of both groups we restrict the sample further so that for those defendants sentenced before 9-11, their last offense was committed at least 180 days before 9-11.[9]

The working sample covers 40,228 cases: 32% of defendants are White, 27% are Black, and 41% are Hispanic. Table 2 shows the characteristics of each group of defendants, for cases up for sentencing pre and post 9-11. The samples are well balanced on these defendant and legal characteristics, and the difference-in-differences in characteristics are nearly all not different from zero. Where there is imbalance, the magnitudes are small. Given 9-11 was unanticipated, our evidence is based on a sample of defendants and offenses that are representative of caseloads in the federal CJS more broadly.

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

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

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

where s_{iet} is the sentencing outcome for individual i of outgroup e sentenced on day t based on a ± 180 sentencing day window around 9-11, $Post_t$ is a dummy equal to one if the defendant comes up for sentencing post 9-11, and all covariates (X_i, L_i, OFF_{if}, G_{ig}, D_{id}) are as described earlier. ε_{iet} is clustered by federal district. Our data does not contain judge identifiers, so we do not control for judge fixed effects.

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

3.4 Identifying Assumptions and Interpreting ϕ_e

Three assumptions underpin ϕ_e identifying a causal effect of 9-11 on sentencing outcomes for outgroup e. First, the time a defendant spends in the CJS between when they commit their last offense and when they come up for sentencing should not be differentially impacted by 9-11 across groups. This concern is ameliorated by there being 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 such queue jumping. We further address the concern using survival analysis to predict the time a defendant spends in the CJS between the date of last offense and sentencing by group. Second, we require there to be no race/ethnicity-time effects in sentencing differentials that naturally occur around 9-11 each year. We assess this using placebo checks using data from earlier years, and also extend our pre-period to allow us to check for differential time trends across groups. Finally, we require there to be no missing covariates that determine sentencing outcomes, vary across groups *and* change post 9-11 2001 (but not in placebo years). We address this issue by estimating bounds on the key difference in differences terms accounting for selection on unobservables.

Under these assumptions, ϕ_e still need not be interpretable as reflecting contagious/parochial animosity: it might reflect that judges anticipate changes in behavior of defendants post 9-11, with these expectations differing across outgroups. For example, 9-11 might have altered labor market outcomes for minorities and this can affect recidivism rates differentially across groups [Orrenius and Zavodny 2009]; alternatively, judges might anticipate post 9-11 the police will reallocate resources in a way that differentially changes future detection probabilities by race/ethnicity. Taken together, such channels represent different forms of statistical discrimination, where stereotyping of defendants by group can lead to differential outcomes by race/ethnicity post 9-11.[10]

We use two strategies to narrow the interpretation: (i) decomposition analysis to show how much of the differential is attributable to changing sentencing penalties on observables; (ii) correlating sentencing differentials to judge characteristics, including race/ethnicity, in the spirit of rank order tests used to distinguish statistical discrimination from animus in the literature using police arrest data [Anwar and Fang 2006, Park 2017].

[10] 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*).

4 Judges and Sentencing Outcomes

4.1 Downward Departures

Table 3 presents estimates of (1) for downward departures, the key margin of judicial discretion at sentencing. Column 1 shows that Hispanic-White sentencing gaps open up post 9-11: relative to Whites, the likelihood Hispanics receive a downward departure falls significantly by 3.8pp. We see no such impact on Black defendants, on whom the post 9-11 impact for downward departures is a precisely estimated zero (and as shown at the foot of the Column, this is significantly different to the post 9-11 impact on Hispanics, $p = .041$). Recall that as shown in Table 1, no Hispanic-White differential in rates of downward departure existed pre 9-11. This Hispanic-White sentencing differential only opens up post 9-11. If 9-11 sparked a rise in animosity towards Muslims, this pattern of results across outgroups is consistent with judges displaying contagious animosity from Muslims to Hispanics, while their social preferences are independent between Muslims and Blacks.

Judges have to provide an explanation for downward departures: Columns 2 to 5 code these into broad categories. The differential impact on Hispanics is driven by judges being less likely to downwards depart due to: (i) a belief that the criminal history of the defendant is overrepresented; (ii) other reasons. For the first type of downward departure, the post 9-11 impact on Hispanics is significantly different from that on Blacks ($p = .036$). There is no statistically significant shift in downward departures related either to general mitigating circumstances, and no precisely estimated impact on downward departures related to plea bargains.

A greater Hispanic-White sentencing differential post 9-11 could be due to either contagious animosity where anti-Muslim sentiment hurts Hispanics, or parochial animosity where anti-Muslim sentiment increases ingroup altruism towards Whites. The evidence rules out the latter interpretation because: (i) the post 9-11 indicator on the likelihood of downward departure (for Whites) is a precisely estimated zero; (ii) we find statistically significant differences in the impacts between Hispanic and Black defendants, again suggesting the results are not driven by increased altruism towards the White ingroup.

We can convert the impacts on the propensity to downward depart into an implied change in expected sentence length as follows. To do so, we calibrate sentence length impacts assuming the *only* channel through which 9-11 impacts sentence length is through the likelihood of downward departure, and so hold constant other channels such as: (i) the number of guideline cells shifted conditional on downward departure; (ii) sentence length within guideline cell conditional on no departure. We return to these other channels below.

For the current exercise we denote the probability of being assigned to guideline cell g as p_g, the probability of being downward departed as p_d, and the expected sentence conditional on being sentenced within the range of guideline cell g as $E[s|g]$. The implied change in expected sentence length is,

$$\sum_g p_g \Delta p_d \{E[s|g-4] - E[s|g]\}, \quad (2)$$

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

To monetize these sentencing impacts we note: (i) the marginal annual cost per year of imprisoning a male prisoner is $\$29,000$ [Congressional Research Service 2013]; (ii) in the federal system, the elasticity of incarceration with respect to sentence $\simeq .87$ [Rehavi and Starr 2014]. Combining these with our implied sentence impact suggests that 9-11 lead to an increase of \$1547 in incarceration costs per Hispanic defendant, mapping to a large increase in total costs of the federal CJS given that 40% of all defendants are Hispanic.[12]

The analysis conditions on the offense type the defendant is charged with. This replicates earlier work in economics on sentencing outcomes, so conditional on all information available to judges at the point they make their key decision. An alternative approach, following Rehavi and Starr [2014], is to only condition on observables determined at the point a defendant enters the federal CJS. The justification for doing so is that prosecutors might manipulate the offense level, say through selective fact-finding, and perhaps in anticipation of judge's behavior [Schanzenbach and Tiller 2007, Cohen and Yang 2019]. To address this issue we exploit information from the arrest stage of the criminal time line (Stage 0): for the 67% of cases that can be linked back to the arrest stage we condition on over 400 codes corresponding to the precise offense the defendant was originally arrested for (rather than conditioning on the 31 offense type codes or 258 guideline

[11] 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 ($p = .041$, $.058$ respectively).

[12] Mueller-Smith [2016] estimates the total social cost generated by one year of incarceration to be between $\$56,000$ and $\$66,000$. An alternative benchmark is how sentencing differentials in the federal CJS have been impacted by institutional reforms. For example, sentencing guidelines were abolished in 2005 following the Supreme Court's decision in US v. Booker. There is mixed evidence on what impact this abolition had on sentencing differentials. Fischman and Schanzenbach [2012] report no effects, while Yang [2015] uses individual matched judge and defendant data, finds Black sentences rise by two months as a result. Hence the magnitude of our main effect arising from contagious animosity corresponds to just over one third of this.

cells based on prosecutor decisions during the timeline). Column 6 shows that conditional on original arrest codes, the Hispanic-White differential post 9-11 on downward departures remains significant, and is larger in absolute value at $-.046$pp. This impact remains statistically different than any post 9-11 impact on Black defendants ($p = .079$) and the implied sentence length impact is .889 months, nearly 30% of the conditional pre 9-11 Hispanic-White sentence differential.

4.2 Sentence Length

We next consider sentence length as the outcome s_{iet}. The calibration exercise in (2) assumed the *only* channel through which 9-11 impacts sentence length is through the likelihood of downward departure, holding constant other channels such as: (i) the number of guideline cells shifted conditional on downward departure (that from Table 1 we see applies to 17% of defendants); (ii) sentence length within guideline cell conditional on no departure (that applies to the remaining 83% of defendants). Measuring an overall Hispanic-White sentence differential is complicated by the fact that a small share of defendants are impacted through downward departures, and channels (i) and (ii) above might move in opposite directions.

Notwithstanding this issue, to begin with, Table 4 shows impacts on overall sentence length (in months) from estimating (1). Column 1 shows $\widehat{\phi}_H$ not to be statistically different from zero. In Column 2 we remove defendants with a life sentence (as these are all top coded at $s_{iet} = 470$ months). The point estimate of $\widehat{\phi}_H$ then becomes positive, but is still not different from zero. To make the results less sensitive to impacts on the tails of the distribution of sentence lengths caused through channels (i) and (ii) above, Column 3 shows estimates from a quantile regression at the median sentence length, following the approach of Firpo *et al.* [2007]. The point estimate of the Hispanic-White sentencing differential rises to $\widehat{\phi}_H = .715$, closely matching the calibrated sentence length of .736 (that assumed no impacts within cell or in cell movements). We reject the null that the differential effects of 9-11 on sentence lengths for Hispanics and Blacks are the same [$p = .062$].

To build a more complete picture of the sentence impacts of 9-11 that also sheds light on channels (i) and (ii), we next define a sentence adjustment for defendant i initially assigned to guideline cell G: $sa_{iG} = s_i - \min(s_{iG})$. Negative values of sa_{iG} represent a final sentence below the guideline cell range (that arises from a downward departure), $sa_{iG} = 0$ represents the sentence being at the lower bound of the guideline cell (that is a natural focal point for sentence length, with 33% of sentences being at this bound pre 9-11), and positive values represent a higher sentence within the guideline cell (that could also be due to a binding statutory minimum sentence length requirement). We then estimate specifications analogous to (1) where the outcome variable is prob($sa_{iG} \leq \tau$) where $\tau = -1, -2, .. - 24$, prob($sa_{iG} = 0$), and prob($sa_{iG} \geq \tau$) where $\tau = 1, 2, .., 12$. The asymmetry reflects that downward sentence adjustments of up to two years are far more common that upwards sentence adjustments beyond 12 months of the guideline cell minimum. We note that excluding life sentences, the average width of a guideline cell is 15 months.

The resulting sequence of difference-in-difference estimates is shown in Figure 4. The top panels show the estimated Hispanic-White differential for each sentence adjustment, and the corresponding 95% confidence intervals. The left hand Figure does so unconditionally; the right hand Figure controls for the full set of covariates in (1).

For sentence adjustments below the minimum of the guideline cell ($\tau < 0$) we see that: (i) Hispanic defendants are significantly less likely to have sentence adjustments between -9 and -1 months (the range is slightly larger when we do not conditional on covariates); (ii) there is no significant differential impact of 9-11 on sentence adjustments below this level ($\tau \leq -10$). This result suggests the *marginal* Hispanic defendant less likely to be downwards departed post 9-11 is in a sentencing adjustment band just below the minimum of their original guideline cell. Defendants further away to begin with from this minimum are inframarginal, and are not differentially impacted by 9-11.

The right hand side of each panel provides an indication of where the marginal Hispanic defendant is then shifted to: for sentence adjustment at or above the minimum of the guideline cell ($\tau \geq 0$) we see an increased mass of defendants precisely at the minimum of the guideline cell ($\tau = 0$), with declining impacts for conditional sentence adjustments of one month and above.

The lower panels of Figure 4 repeat the analysis for Black-White sentencing adjustment differentials. Both the unconditional and conditional estimates are smaller in magnitude, and not ever statistically different from zero.

As a final step of analysis, we focus in on the resulting impacts on sentence lengths from these changes in sentence adjustments. Our approach is to try and identify those defendants that in the counterfactual absent 9-11, would have been most likely to be downward departed, and then measure their sentence differential post 9-11 against this counterfactual. We proceed as follows. First, we use the entire pre 9-11 sample (back to October 1998) to estimate the likelihood of a downward departure using the same covariates as in (1) but allowing for more detailed categorizations of age and the number of dependents (because the sentencing guidelines make explicit that downward departures can occur partly based on family responsibilities or prior good works). We estimate this prediction model using a probit specification, and do so separately by outgroup e. We then take our baseline working sample of defendants up for sentencing in the window around 9-11, and group defendants into percentile bands of their predicted probability of downward departure, $\widehat{p}_{DD}$, based on the pre 9-11 models. In each subsample, we keep observations if the predicted probability exceeds any given percentile value, so moving from the fifth to the ninetieth percentile we progressively keep fewer observations. Based on each of these sub-samples, we run our standard difference-in-difference specification where the dependent variable is sentence length. Finally, we plot the difference-in-difference for these percentile subsamples of $\widehat{p}_{DD}$ along with their corresponding 95% confidence interval, and overlaid with the histogram of $\widehat{p}_{DD}$.

The results are shown in Panel A of Figure 5. We see that for defendants between the 70th and 85th percentiles of the predicted probability of downward departure, there is a significant increase in sentence lengths. The magnitude of this effect is just over two months. Consistent with the results on sentence adjustments we see that defendants with the highest predicted probability of being downward (over the 90th percentile of $\widehat{p}_{DD}$) have no change in the sentence outcome – as Figure 4 showed, they are not the marginal defendant differentially impacted by 9-11. Second, we see that the majority of defendants – those below the 70th percentile or above the 90th percentile of $\widehat{p}_{DD}$ – have no significant impact on their sentence length, and this is line with 83% of them are not being subject to downwards departures (Table 1). This is what mutes the overall impact on sentence lengths shown in Table 4.

Panel B shows the findings if the first stage prediction model for the likelihood to be downward departed includes additional interactions between the number of children and the six broad categories of criminal history shown in Table A1.

How large is a two month impact on sentence length? It corresponds to 50% of the conditional Hispanic-White sentencing gap pre 9-11 shown in Table 1. It is also comparable in magnitude to the sentencing impacts documented in Yang [2015], who studied racial sentencing differentials once sentencing guidelines were struck down in 2005. She finds that increasing judicial discretion in sentence lengths increased average sentence lengths for Black defendants relative to Whites by two months. Hence our findings suggest the impact on sentence lengths arising through social preference structures and contagious animosity around 9-11 being transmitted from Muslims to Hispanics, are around the same magnitude as that arising from an institutional change in sentencing policy on Black defendants.

4.3 Citizenship and Offense Type

There are two obvious reasons why Hispanic-White sentencing differentials might become exacerbated after 9-11, while Black-White differentials remain unchanged, and that have nothing to do with contagious animosity across outgroups. The first is that Hispanics constitute the majority of non-US citizen defendants. Punishments for non-citizens, such as deportation, differ from those available for citizens and residents/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.

71% of defendants are citizens, 43% of Hispanic defendants are citizens, and 91% of non-citizens are Hispanic. Given this close alignment between race and citizenship status, it is hard to cleanly separate the two but we do so to the extent the data allows. Column 1 of Table 5 allows impacts to vary between Hispanics citizens (US citizen, resident/legal alien) and Hispanic non-citizens (illegal aliens, non-US citizen, status unknown). For both groups of Hispanic, those that are sentenced post 9-11 are significantly less likely to be downward departed. For Hispanic citizens the impact is a 2.8pp reduction in the likelihood of a downwards departure, corresponding to an implied higher sentence length of 17% of the pre 9-11 Hispanic citizen-White differential. For Hispanic non-citizens the impact is a 4.4pp reduction in downwards departure, an implied sentence length increase mapping to 16% of the pre 9-11 Hispanic non-citizen-White sentencing differential. There is no statistical difference between the two impacts ($p = .278$).

A second reason why Hispanic-White sentencing differentials might increase post 9-11 is that they are more likely to be charged with immigration offenses. If such offenses are more severely punished post 9-11, ϕ_H might just pick up that Hispanics are charged with immigration offenses at a greater rate than others. To address the issue, the remaining Columns of Table 5 split the sample by offense type (drug, immigration, other), while still allowing the impact of ethnicity to vary between Hispanic citizens and Hispanic non-citizens. For immigration offenses the vast majority of defendants in the federal system are Hispanic (either citizens or non-citizens). Hence when examining those offenses we restrict the sample further to Hispanics only.

Across offense types, we find no significant differences between impacts of 9-11 on Hispanic citizens and non-citizens: (i) Hispanic non-citizens are significantly less likely to receive downward departures for drug offenses (Column 2) but this effect is not different from that for Hispanics citizens ($p = .210$); (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 if we focus exclusively on border states (Columns 3 and 4); (iii) the lower likelihood of downward departures post 9-11 is largely driven by the impact on Hispanic citizens for other offenses (Column 5), but again this is not different from that for Hispanics citizens ($p = .722$): these constitute around 40% of all offenses and often relate to firearms.[13]

In the online Appendix, Table A3 shows these results by offense types to continue to hold when we use the original arrest codes from the start of criminal time line (Stage 0): we find no robust

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

evidence that sentencing differentials for drug, immigration or other offenses change differentially post 9-11 between Hispanic citizens and Hispanic non-citizen defendants.

4.4 Robustness and Support for Identifying Assumptions

Online Appendix Tables A4 to A6 conduct a battery of robustness checks on our core finding from Table 3. These show the result to be robust to: (i) alternative levels of clustering standard errors; (ii) excluding cases where perhaps because of prosecutor's decision making over the initial offense charges filed (Stage 3 in Figure 1), statutory minima or maxima bind partially over the range set by the guideline cell [Rehavi and Starr 2014]; (iii) estimating (1) separately for each group. We also combine information on Hispanic origins and race to examine whether our findings pick up ethnic, rather than racial, sentencing differentials.

In each set of robustness checks, we find the results hold irrespective of whether we control for final offense codes or initial arrest codes.

The Online Appendix also provides evidence in support of the three identifying assumptions required to interpret ϕ_e as measuring a causal impact: (i) Table A7 shows the main results to be robust to controlling for time of offense (and irrespective of whether we use final offense codes or initial arrest codes), (ii) Table A8 uses survival analysis to show the time a defendant spends between their last offense and when they come up for sentencing is not differentially impacted by 9-11 across groups.

We next address the concern there are race/ethnicity-time effects in sentencing differentials that naturally occur around 9-11 each year. We do so using four pieces of evidence. First, we use data from earlier years to construct placebo 9-11 effects. As Table A9 shows, the impact for Hispanics on downward departures only occurs post 9-11 in 2001. Again, this result is robust to controlling for either final offense codes or initial arrest codes. Second, we check for pre-trends by considering all offenses committed prior to 9-11 (even if the defendant has been sentenced pre 9-11 and exited the system). We thus define the pre-period as starting from October 1998. In this extended sample we can control for linear time trends in rates of downward departure, that can vary by group. Table A10 shows our core result remains robust: there remains a significant fall in the likelihood of Hispanic defendants being downward departed post 9-11 (Column 3). The magnitude of the effect is $-.042$ ($se = .012$) that is near identical to be baseline estimate of $-.038$ ($se = .010$). This is over and above the long run upward trend in the likelihood of Hispanics being downward departed shown (and the magnitude of this trend is slight (.002)).

Third, we address concerns impacts are driven by the Patriot Act, that was enacted 45 days after 9-11. To shed light on the matter we estimate a dynamic specific analogous to (1) that estimates impacts in 15-day windows post 9-11. As we earlier showed immigration offenses do not drive the main result, Figure A2 documents how impacts on judicial departures for Hispanics appear post 9-11 and pre- and post-Patriot Act, for offenses unrelated to the Patriot Act. We find that the point estimates are of similar magnitude to the main estimate from (1) and relatively stable over each of these 15-day windows, including those before the Patriot Act was introduced.

Fourth, we collect data on the date of confirmation of Bush-appointed US Attorneys (shown in Figure A3), to establish in Table A11 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 Attorney, that might otherwise signal a change in how the CJS views the trade-off between justice and social protection. Again, this is robust to controlling either for final offense codes or initial arrest codes.

The final identifying assumption required is that there are no missing covariates that determine sentencing outcomes, vary across groups *and* change post 9-11 2001 (but not in placebo years). We address this following Altonji *et al.* [2005] and Oster [2019] to estimate *bounds* on the treatment effect of $Outgroup_e$ accounting for selection on unobservables. The results in Table A12 show these bounds on $\widehat{\phi}_e$ are tight. For them to include zero requires unobserved factors changing for Hispanics post 9-11 that are orders of magnitude more predictive of sentencing outcomes than the covariates in (1), including the full set of guideline cell dummies.

5 Prosecutors and Pre-sentencing Outcomes

Prosecutors represent a second crucial actor determining defendant outcomes. We extend our research design to examine the pre-sentence prosecutorial decision making. This enables us to provide insight on whether prosecutors, who around 9-11 were overwhelmingly White, display behaviors towards outgroups consistent with the results found for judges.[14]

Prosecutors decide the initial offense charge filed against defendants (Stage 3 in Figure 1). In the federal criminal code, definitions of crimes often overlap, providing prosecutors discretion over initial charges. These charges are crucial because they determine: (i) if statutory minima/maxima sentences bind and take precedence over guideline cell sentence ranges; (ii) outside options in plea bargaining (defendants might plead to a lesser charge to avoid being charged with an offense with a mandatory minimum) [Yang 2016].[15]

In Table 6, we use the pre 9-11 sample to first document, by outgroup: (i) the frequency with

[14] A recent study of state prosecutors by the *Women Donors Network* found that: (i) 95% of elected prosectors are Whites; (ii) the majority of states have no elected Black prosecutors. A summary of the findings are available at http://wholeads.us/justice/wp-content/themes/phase2/pdf/key-findings.pdf (accessed May 13th 2016).

[15] 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).

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).[16] 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). We next condition on a rich set of codes corresponding to the original offense the defendant was arrested for. The result in Column 3 shows that doing so, there remain significant Black-White and Hispanic-White differences in the likelihood of non-zero statutory minimum offense charge being given.

Columns 4 to 6 document these differences translate into a similar pattern of differentials pre 9-11 for statutory minimum sentence lengths. Blacks receive charges carrying minimum sentences that are conditionally 22 months longer than Whites, falling to 7.8 months in cases linked to arrest offense codes. For Hispanics, prosecutors set initial charges with associated statutory minimums that are 14 months longer (or 63% higher) than for Whites, falling to 7.4 months in cases that can be linked to arrest offense codes.

We next use our research design to examine whether 9-11, that increased animosity towards Muslims, had spillover effects on other outgroups in the federal CJS through prosecutors' decisions. 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; (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. We estimate a specification analogous to (1) but where the outcomes are: (i) whether the defendant receives an initial charge with a non-zero statutory minimum sentence; (ii) the length of statutory minimum sentence associated with their

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

initial offense. We do not condition on final offense type or the later determined guideline cell.[17]

The results are in Table 7: (i) Hispanic defendants initially charged post 9-11 are 7.4pp more likely to receive an initial offense that carries a statutory minimum corresponding to a 22% increase over the pre 9-11 period (an impact statistically different from Blacks, $p = .032$); (ii) their statutory minimum sentence is 10.7 months longer; (iii) there is no evidence that 9-11 impacts prosecutors' initial offense charges filed against Black defendants along either margin ($\widehat{\phi}_B = 0$ in Columns 1 and 2). The magnitude of these responses to 9-11 correspond to: (i) 60% of the pre 9-11 Hispanic-White gap in the the likelihood of an initial offense charge with a mandatory minimum; (ii) 77% of the pre 9-11 Hispanic-White gap in the statutory minimum sentence length. Indeed, these impacts of 9-11 leaves the overall post 9-11 Hispanic-White differential on each margin to be at least as large as the Black-White differential.

This pattern of results closely mirrors those found earlier for judges: they are consistent with the structure of social preferences across outgroups for prosecutors being such that there is contagious animosity from Muslims to Hispanics, while their social preferences are independent between Muslims and Blacks.

In the Appendix we consider two further dimensions of prosecutor behavior: (i) granting of substantial assistance departures (that can occur at the plea stage of the timeline); (ii) drafting of the pre-sentence report (that occurs between trial and sentencing). On (i) we find no differential impacts on the likelihood prosecutors grant substantial assistance departures: this helps rule out that the increase in statutory minimum sentence lengths driven by initial offense charges is later undone through defendant cooperation in plea bargains. On (ii) for both outgroups we see no change in the minimum sentence in the guideline cell defendants are placed in. Hence prosecutor-legal counsel interactions at the pre-sentence report stage between trial and sentencing are *not* a major source of differential treatment of defendants by outgroup post 9-11. This suggests increased Hispanic-White sentencing gaps post 9-11 are not due to diminished effort on the part of legal counsel of Hispanic defendants.

6 Interpretation

We have documented an impact of 9-11 on outcomes for a major (non-Muslim) minority group in the high stakes and professional environment of the federal CJS. One interpretation is that the changes in behavior of ingroup judges and prosecutors are driven by their social preference

[17] 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.

structures over outgroups. In particular, their behavior can be rationalized by them having contagious animosity from Muslims to Hispanics, while social preferences are independent between Muslims and Blacks. We now probe the data further using two very different approaches to rule out alternative interpretations of $\widehat{\phi}_e$.

6.1 Decomposition Analysis

We first present a decomposition of sentencing differentials to understand whether they are being driven by changes in observables, or sentencing penalties for those observables. We focus on defendants that come up for judicial sentencing just around, among whom we have documented that Hispanics are significantly less likely to be downward departed (Table 3). We use the Juhn *et al.* [1993] decomposition. This is implemented by first considering the following sentencing equation for White defendant i sentenced in period T: $s^W_{iT} = X'_{iT}\beta^W_T + u^W_T\theta_{iT} = X'_{iT}\beta^W_T + \varepsilon^W_{iT}$, where β^W_T are sentence penalties for Whites, and ε^W_{iT} is a residual for White defendant i in period T. The explicit assumption is that the residuals and covariates are independent [Fortin *et al.* 2011]. The Hispanic-White sentencing differential in period T is then, $\Delta s_T = s^H_T - s^W_T = \Delta X_T\beta^W_T + \Delta\varepsilon_T$. Given our DD research design we take a *second* difference over pre- to post 9-11 time periods ($T = 0$ to $T = 1$):[18]

$$\Delta s_1 - \Delta s_0 = (\Delta X_1 - \Delta X_0)\beta^W_0 + \Delta X_1(\beta^W_1 - \beta^W_0) + (\Delta\varepsilon_1 - \Delta\varepsilon_0). \tag{3}$$

The unconditional DD in the likelihood of downward departure to be explained is $\Delta s_1 - \Delta s_0 = -.041$. The $(\Delta X_1 - \Delta X_0)\,\beta^W_0$ component, or X-effect, measures the contribution to the DD in sentencing gaps of observables. The $\Delta X_1(\beta^W_1 - \beta^W_0)$ component, or β-effect, measures changes in sentencing penalties pre- and post 9-11 for observables.[19]

Figure 6 shows the X- and β-effects for specific covariates, where the y-axis shows the implied sentencing differential that can be attributed to each X- and β- effect. As expected, this shows that each X-effect, on quantities, is small. This is because of our research design, and this result is essentially analogous to what was shown in Table 2 that defendant observables are balanced

[18] While it is well understood that such decompositions do not represent formal tests for statistical discrimination [Charles and Guryan 2011], in our setting the usual concerns related to decomposition analysis for studying discrimination are partly ameliorated because: (i) the DD set-up provides common support in the cross-section of covariates across groups; (ii) the inclusion of guideline cell dummies allows us to capture many case-specific factors driving outcomes.

[19] To check the validity of basing the JMP decomposition off a linear probability model, we have also conducted cross-sectional decompositions in the pre- and post 9-11 periods separately, using 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.

pre- and post 9-11 by group. A more interesting pattern of changing penalties across covariates emerges, with the penalties on some covariates rising and others falling. Due to the alternating signs of the effects, only 7% of the unconditional DD is overall attributable to observables either through the X-effects or the β-effects.

For example, penalties related to education, being married and having children all rise, suggesting that post-911 Hispanics would have been *more* likely to be downward departed than Whites. On covariates related to offense types, we note the X- and β-effects never explain more than 17% of the observed sentencing gap between Hispanics and Whites, while differences in defense counsel types do not explain more than 9% of the overall gap.

Taken together, these findings help rule out explanations for the results based on the harshness with which certain offense types are dealt with post 9-11, offender characteristics including those that might perhaps closely predict recidivism such as the guideline cell they are assigned to, or explanations related to effort or allocation of legal counsel to defendants post 9-11. All this suggests explanations for why Hispanic-White sentencing differentials worsen post 9-11 based on statistical discrimination alone, are not easily reconcilable with the evidence. This is also fits with evidence that recidivism rates did not change across groups pre- and post 9-11 [BJS 2014, 2018].[20]

However, the one covariate that can potentially explain the observed sentencing gap is the federal district of the case: the X-effect is again small and only corresponds to 3% of the unconditional DD, but the β-effect can explain 60% of the gap ($-.025$ of the actual gap, $-.041$). We therefore next examine one important source of spatial variation that might be being reflected in increasing penalties in the decomposition: judge characteristics.

6.2 Judge Characteristics

In federal court data, judge identifiers are typically unavailable (or only a subset of cases can be linked) because these cases are considered more serious and often of national importance.[21] To make progress on how judge characteristics correlate to the change in sentencing differentials, we have coded the characteristics of federal judge's by district, sourced from the *Biographical*

[20] BJS [2014] reports recidivism rates by race for two cohorts of defendants: those released in 1994 and those released in 2004. This suggests: (i) three-year recidivism rates of all groups have risen over time; (ii) there has been no great differential increases across groups over time in recidivism rates. BJS [2018] reports recidivism rates by race over a 9-year follow up period for defendants released in 2005: this shows Hispanics have higher one-year recidivism rates than Whites, but 9-years post release recidivism rates are found to be almost equal between Whites and Hispanics, but are higher for Black defendants. In sum, this evidence does not strongly suggest that post 9-11, recidivism rates among Hispanics rose more than for other groups.

[21] An important relevant exception is Yang [2015], who links individual judge data to federal cases to examine how racial sentencing differentials are impacted once sentencing guidelines were struck down in *United States vs Booker* in 2005.

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

Similar to Guryan and Charles [2011], we proceed in two steps. First, we estimate (1) allowing for a full set of interactions between each federal district d and ($Hispanic_e \times Post_t$) to estimate the coefficient of interest: $\phi_{H,d}$. We do so for the likelihood of a downward departure. Figure A4 shows the spatial pattern of changes in sentencing differentials, plotting $\widehat{\phi}_{H,d}$ for each district. Second, we regress $\widehat{\phi}_{H,d}$ against $\mathbf{J}_d$ and other district characteristics, where observations are weighted by the share of defendants in district d in the working sample that are Hispanic. Observations are weighted because the underlying regression from which each $\widehat{\phi}_{H,d}$ is estimated is based on individual observations, and these vary by district. Robust standard errors are reported.

The weighted mean share of Hispanic (Black) judges in a district is 14% (7%). We note that 16 out of 90 districts (18%) have at least one Hispanic judge, the weighted mean share of Hispanic judges is 13.4%, the median share is 16% and the share conditional on there being at least one Hispanic judge is 19%. Hispanic judges are more likely to be in districts with more Hispanic defendants: the correlation between the share of Hispanic judges and Hispanic defendants in districts is .78 (when districts are weighted by the share of Hispanic defendants). 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.

Table 8 shows the second stage results. In Column 1 we only control for judge race/ethnicity. We find that in districts where there are a higher proportion of Hispanic judges, the change in the Hispanic-White sentencing differential, $\widehat{\phi}_{H,d}$, is significantly smaller. Column 2 shows this is robust to controlling for the seniority, gender, age and appointment characteristics of federal district judges, as well as the share of the post 9-11 window the district spends under a Bush-appointed US Attorney. This suggests the Hispanic ethnicity of judges is not merely picking up them being Democrat appointees, and consistent with the evidence in Schanzenbach [2005] and Harris and Sen [2019], the presence of Democratic appointed judges has an independent correlation with changes in the Hispanic-White sentencing differential.[22]

Column 3 controls for the population shares of ethnic groups in the district, as well the change (1990 to 2000) in proportions for each group. This *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

[22]Our results are consistent with Cohen and Yang [2019], where they use individual judge data to show how Republican judges give harsher sentences to Black defendants.

Hispanic judges does not appear to proxy population characteristics of where the case is heard.

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

The fact that judge ethnicity correlates to the change in the Hispanic-White sentencing differential is *prima facie* evidence against the results being explained by statistical discrimination: if so, then *all* judges, irrespective of their own characteristics should use defendant ethnicity as a marker for unobservable traits in determining sentencing outcomes. This is in the spirit of rank order tests used to distinguish statistical discrimination from animus in the literature using data on police arrests or on individual judges [Anwar and Fang 2006, Park 2017].[23] This interpretation is further reinforced by noting that there more experienced judges are uncorrelated with smaller changes in sentencing differentials (measured either through the senior status of judges or their age). This is counter to the Altonji and Pierret [2001] test of statistical discrimination exploiting the fact that with experience, decision makers learn the true characteristics of agents and become less reliant on proxies such as race/ethnicity.

7 Conclusions

Ingroup bias is a central aspect of human behavior where 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]. We extend this notion to contexts in which social preferences are defined over multiple outgroups. We use a quasi-experimental research design around 9-11 to shed new light on the structure of social preferences across outgroups. Our research design allows us to investigate whether increased animosity towards Muslims in the aftermath of 9-11 had spillover effects on Black and Hispanic individuals in the context of the high stakes and professional environment of the federal criminal justice system.

Our core finding is that as 9-11 sparked a rise in animosity towards Muslims, Hispanic defendants experience worsening sentence and pre-sentence outcomes, in line with judges and prosecu-

[23] Such hit-rate tests for racial bias in the context of arrest data have been devised to deal with the non-random selection of individuals into police stops. In our setting, such concerns over the infra-marginality problem of detecting bias are weaker because there is random matching of defendants to judges in the Federal CJS.

tors having social preferences characterized by *contagious animosity* from Muslims to Hispanics. In contrast, the social preferences of judges and prosecutors are independent between Muslim and Black defendants. We underpin a causal interpretation of these findings by providing evidence in favor of the identifying assumptions underlying our research design, and we narrow down the interpretation of the results by ruling out that they are driven by citizenship, or by statistical discrimination against Hispanic defendants. 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]. We do so with two important caveats: (i) we have exploited a particularly traumatic event that could have triggered a strong emotional response, even in this high stakes setting, in line with nascent well-identified causal evidence on emotions driving judicial decisions [Shayo and Zussman 2011, Chen *et al.* 2016, Philippe and Ouss 2018]; (ii) our research design does not allow us to estimate whether the impacts persist beyond the short-run window of cases in our sample.

Our findings provide among the first field evidence of contagious animosity, that social preferences across outgroups are malleable. This adds to a nascent body of work examining the structure of social preferences, that has so far typically been based on self-reported or observational data collected in post-conflict environments [Bauer *et al.* 2016]. An important implication of our findings is that anti-discrimination policies towards one outgroup can have externalities on other outgroups. On policy implications, our results suggest appointing more Hispanic judges to federal district courts or as federal prosecutors, might go some way towards reducing Hispanic-White sentencing differentials.

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

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

Second, there are many potential outgroups one could consider, over which there is a rich set of social preferences structures to identify. There is no reason to expect contagious animosity/altruism to characterize all pairs. More broadly, there can be circumstances in which individuals have multiple identities, and other circumstances in which individuals can endogenously choose an identity in anticipation of the kinds of interlinked social preference structures we have documented. This opens up a wide array of research questions at the intersection of the formation of social preferences and the economics of identity.

8 References

Abrams, David, Marianne Bertrand and Sendhil Mullainathan 2012. Do Judges Vary in their Treatment of Race? *Journal of Legal Studies* 41:347-83.

Alexander, .Richard.D 1987. *The Biology of Moral Systems*. New York: Aldine De Gruyter.

Allport, Gordon.W 1954. *The Nature of Prejudice*. Cambridge. MA: Perseus Books.

Altonji, Joseph.G, Todd .E.Elder and Christopher.R.Taber 2005. Selection on Observed and Unobserved Variables:Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy* 113:151-84.

Altonji, Joseph.G and Charles.R.Pierret 2001. Employer Learning and Statistical Discrimination. *Quarterly Journal of Economics* 116:313-50.

Anwar, Shamena and Hanming Fang 2006. An Alternative Test of Racial Prejudice in Motor Vehicle Searches:Theory and Evidence. *American Economic Review* 96:127-51.

Anwar, Shamena, Patrick Bayer and Randi Hjalmarsson 2012. The Impact of Jury Race in Criminal Trials. *Quarterly Journal of Economics* 127:1017-55.

Bauer, Michal, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts 2016. Can War Foster Cooperation? *Journal of Economic Perspectives* 30:249-74.

Bayer, Patrick, Randi Hjalmarsson and David Pozen 2009. Building Criminal Capital Behind Bars:Peer Effects in Juvenile Corrections. *Quarterly Journal of Economics* 124:105-47.

Bertrand, Marianne and Esther Duflo 2016. Field Experiments on Discrimination, forthcoming in A.Banerjee and E.Duflo (eds.). *Handbook of Field Experiments*.

BJS 2014. *Recidivism of Prisoners Released in 30 States in 2005:Patterns from 2005 to 2010.* NCJ 244205.

BJS 2018. *Update on Prisoner Recidivism:A 9-Year Follow-up Period 2005-2014.* . NCJ 250975.

Boyd, R, H.Gintis, S.Bowles and P.J.Richerson 2003. The Evolution of Altruistic Punishment. *Proceedings of the National Academy of Sciences* 100:3531-5.

Bushway, Shawn.D and Anne.M.Piehl 2001. Judging Judicial Discretion: Legal Factors and Racial Discrimination in Sentencing. *Law and Society Review* 35:733-64.

Charles, Kerwin.K and Jonathan Guryan 2011. Studying Discrimination: Fundamental Challenges and Recent Progress. *Annual Review of Economics* 3:479-511.

Chen, Daniel.L, Tobias.J.Moskowitz and Kelly Shue 2016. Decision Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires. *Quarterly Journal of Economics* 131:1181-42.

Cohen, Alma and Crystal Yang 2019. Judicial Politics and Sentencing Decisions. *American Economic Journal:Economic Policy.* 11:160-91.

Congressional Research Service 2013. *The Federal Prison Population Buildup:Overview, Policy Changes, Issues and Options.* Report 7-5700, Washington DC:CRS.

Davis, Darren 2007. *Negative Liberty: Public Opinion and the Terrorist Attacks on America.* Russell Sage Foundation.

Depew, Briggs, Ozkan Eren and Naci Mocan 2017. Judges, Juveniles and In-group Bias. *Journal of Law and Economics* 60:209-39.

Eifert, Ben, Edward Miguel and Daniel.N.Posner 2010. Political Competition and Ethnic Identification in Africa. *American Journal of Political Science* 54:494-510.

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.

Firpo, Sergio, Nicole.M.Fortin and Thomas Lemieux 2007. Unconditional Quantile Regressions. *Econometrica* 77:953-73.

Fischman, Joshua.B and Max.M.Schanzenbach 2012. Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums. *Journal of Empirical Legal Studies* 9:729-64.

Fortin, Nicole, Thomas Lemieux and Sergio Firpo 2011. Decomposition Methods in Economics,in O.Ashenfelter and D.Card (eds.). *Handbook of Labor Economics* Vol. 4A, Elsevier.

Fowler, James.H and Nicholas.A.Christakis 2010. Cooperative Behavior Cascades in Human Social Networks. *Proceedings of the National Academy of Sciences* 107:5334-8.

Glaeser, Edward.L, Daniel.P.Kessler and Anne.M.Piehl 2000. What Do Prosecutors Maximize? An Analysis of the Federalization of Drug Crimes. *American Law and Economics Review* 2:259-90.

Gould, Eric.D and Esteban.F.Klor 2016. The Long-run Effect of 9/11: Terrorism, Backlash, and the Assimilation of Muslim Immigrants in the West. *Economic Journal* 126:2064-114.

Harris, Allison.P and Maya Sen 2019. Bias and Judging. *Annual Review of Political Science* 22:241-59.

Hopkins, Daniel.J 2010. Politicized Places:Explaining Where and When Immigrants Provoke Local Opposition. *American Political Science Review* 104:40-60.

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.

Jeffries, John.C Jr and John Gleeson 1995. The Federalization of Organized Crime:Advantages of Federal Prosecution. *Hastings Journal* 46:1095-134.

Jordan, Jillian J., David G.Rand, Samuel Arbesman, James H.Fowler and Nicholas A.Christakis 2013. Contagion of Cooperation in Static and Fluid Social Networks. *PloS One* 8, e66199.

Juhn, Chinhui, Kevin.M.Murphy and Brooks Pierce 1993. Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy* 101:410-42.

Klepper, Steven, Daniel Nagin and Luke-Lon 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.

Legewie, Joscha 2013. Terrorist Events and Attitudes Towards Immigrants: A Natural Experiment. *American Journal of Sociology* 118:1195-245.

Liptak, Adam 2003. For Jailed Immigrants, a Presumption of Guilt. *New York Times*, June 3.

McGovern, Virginia, Stephen Demuth and Joseph E.Jacoby 2009. Racial and Ethnic Recidism Risks. *Prison Journal* 89:309-27.

Mueller-Smith, Michael 2016. The Criminal and Labor Market Impacts of Incarceration, mimeo, University of Michigan.

Mustard, David.B 2001. Racial, Ethnic and Gender Disparities in Sentencing: Evidence from the US Federal Courts. *Journal of Law and Economics* 44:285-314.

Orrenius, Pia.M and Madeline Zavodny 2009. The Effects of Tougher Enforcement on the Job Prospects of Recent Latin American Immigrants. *Journal of Policy Analysis and Management* 28:239-57.

Oster, Emily 2019. Unobservable Selection and Coefficient Stability: Theory and Validation.

Journal of Business Economics and Statistics 37:187-204.

Park, Kyung.H 2017. Do Judges Have Tastes for Discrimination? Evidence from Criminal Courts. *Review of Economics and Statistics* 99:810-23.

Philippe, Arnuad and Aurélie Ouss 2018. No Hatred or Malice, Fear or Affection: Media and Sentencing. *Journal of Political Economy* 126:2134-78.

Rehav, Marit.M and Sonja.B.Starr 2014. Racial Disparity in Federal Criminal Sentences. *Journal of Political Economy* 122:1320-54.

Romero, Luis.A and Armino Zarrugh 2018. Islamophobia and the Making of Latinos as Terrorist Threats. *Ethnic and Racial Studies* 41:2235-54.

Schanzenbach, Max.M 2005. Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics. *Journal of Legal Studies* 34:57-92.

Schanzenbach, Max.M and Emerson.H.Tiller 2007. Strategic Judging Under the US Sentecning Guidelines: Positive Political Theory and Evidence. *Journal of Law, Economics and Organization* 23:24-56.

Shayo, Moses 2009. A Model of Social Identity with an Application to Political Economy: Nation, Class and Redistribution. *American Political Science Review* 103:147-74.

Shayo, Moses and Asaf Zussman 2011. Judicial Ingroup Bias in the Shadow of Terrorism. *Quarterly Journal of Economics* 126:1447-84.

Starr, Sonja.B and Marit.M.Rehavi 2013. Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of *Booker*. *Yale Law Journal* 123:2-80.

Suri, Siddharth and Duncan J.Watts 2011. Cooperation and Contagion in Web-based, Networked Public Goods Experiments. *PloS One*, 6, e16836.

Tajfel, Henri, M.G.Billig, R.P.Bundy and Claude Flament 1971. Social Categorization and Intergroup Behavior. *European Journal of Social Psychology* 1:149-78.

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.].

Woods, Joshua 2011. The 9/11 Effect: Toward a Social Science of the Terrorist Threat. *Social Science Journal* 48:213-33.

Yang, Crystal.S 2015. Free At Last? Judicial Discretion and Racial Disparities in Federal Sentencing. *Journal of Legal Studies* 44:75-111.

Yang, Crystal.S 2016. Resource Constraints and the Criminal Justice System:Evidence from Judicial Vacancies. *American Economic Journal: Economic Policy* 8:289-332.

This document is best read along with FINAL_FIGURES_ONLY.xlsx, which shows the intended layout of the multi-part figures.

1.) Title:
Figure 1: Federal CJS Timeline

Notes:
Notes: We use the Federal Justice Statistics Resource Center (FJSRC) data and the Monitoring of Federal Criminal Sentences (MFCS) data set for our analysis. The FJSRC comprises information gathered from four linked administrative data sources, and we replace the fourth stage USSC data that covers sentencing Stage 8, with the MFCS data (because it is superior to the USSC data that is part of the FJSRC). As described in Rehavi and Starr [2014], the 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) at Stage 8, we then use the Monitoring of Federal Criminal Sentences (MFCS) data set.

2.) Title:
Figure 2: Pre and Post 9-11 Sentiments
Subtitle A Level 1:
A: Sentiments Towards Immigrants Around 9-11
Subtitle A Level 2:
Gallup Poll on Immigration
Subtitle A Level 3:
Q: Should Immigration be Kept at Its Present Level, Increased or Decreased?
Subtitle B Level 1:
B: Crime Rates Around 9-11
Subtitle B Level 2:
Vandalism Victimization
Subtitle B Level 3:
Growth Rate from Same Month in Previous Year
Notes:
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.

3.) Title:

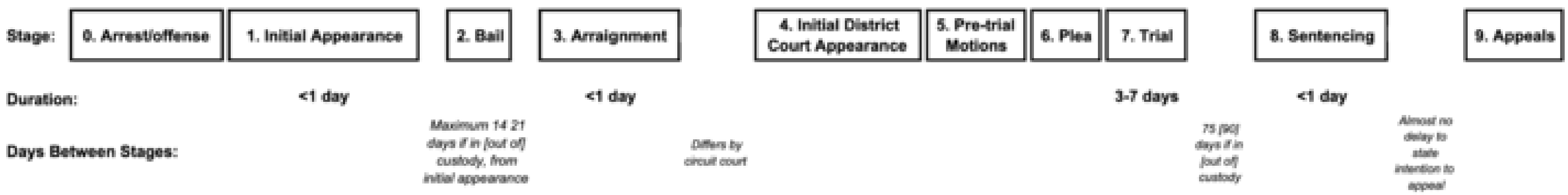
Figure 3: Sentencing and Last Offense Dates
Left hand side subtitle:
A. Sentencing Date
Right hand side subtitle:
B. Date of Last Offense
Row panel 1 subtitle:
Whites
Row panel 2 subtitle:
Blacks
Row panel 3 subtitle:
Hispanics
Notes:
Notes: The left hand side figures show the distribution of dates of sentencing date, for each group: 9/11 is indicated by the vertical dashed line. The right hand side figures show the distribution of the dates of last offenses, by group. The first bar corresponds to a last offense date on or before 1st January 1996. The overlaid histograms are for those sentenced pre- and post-9/11. For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001.

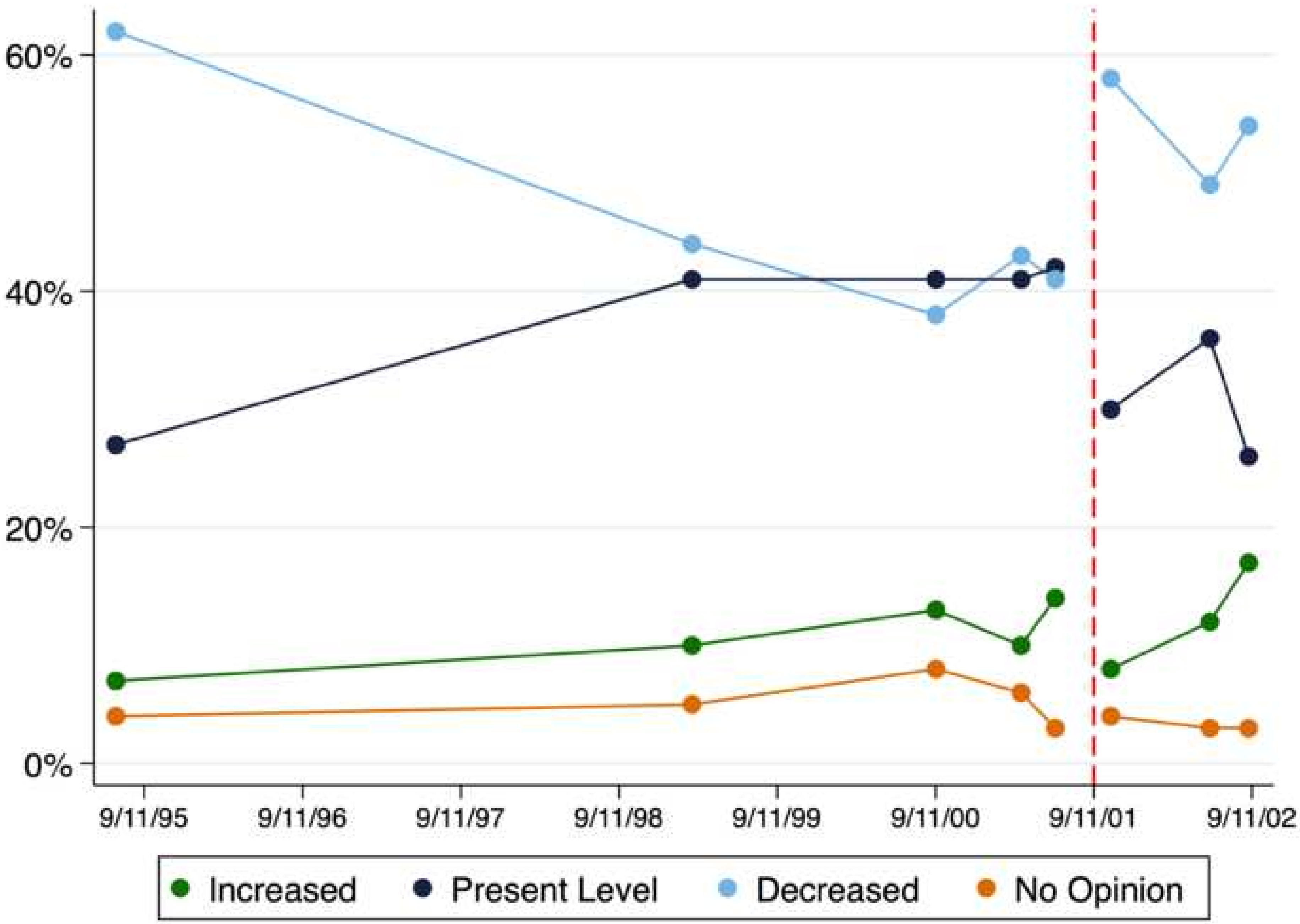
4.) Title:
Figure 4: Sentencing Adjustments
Left hand side subtitle:
Unconditional
Right hand side subtitle:
Conditional
Row panel 1 subtitle:
Hispanic
Row panel 2 subtitle:
Black
Notes:
Notes: The figures show estimates from different regressions, where each coefficient and corresponding 95% confidence interval comes from a separate regression. 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. Standard error or clustered by district. The regressions are based on the difference between an individual's sentence length, and the minimum sentence length recommendation in their allocated guideline cell. Based on this difference, we create a series of dummy variables, which are the dependent variables in the figures above. The first set take a value of 1 (0 otherwise) if the difference in sentence length-guideline cell minimum is less than or equal to a negative integer in the range -24 to -1 (The estimates based on these dependent variables are represented by solid black circles above). We treat zero separately, creating a dummy if sentence length equals the guideline cell minimum (corresponding estimates for this dependent variable are represented by hollow black squares above). Finally we create a set of dummy variables that take a value of 1 (0 otherwise) if sentence length-guideline cell minimum is greater than or equal to a positive integer in the range 1 to 12 (Estimates for which are represented by solid red triangles). We then run a separate OLS regression based on each of these dependent variables, and estimate difference-in-differences models, both without and without a set of additional control variables. In the unconditional models we condition on defendant group (White, Black, Hispanic), whether the case comes up post 9-11, and interactions between the two. In the conditional models we include

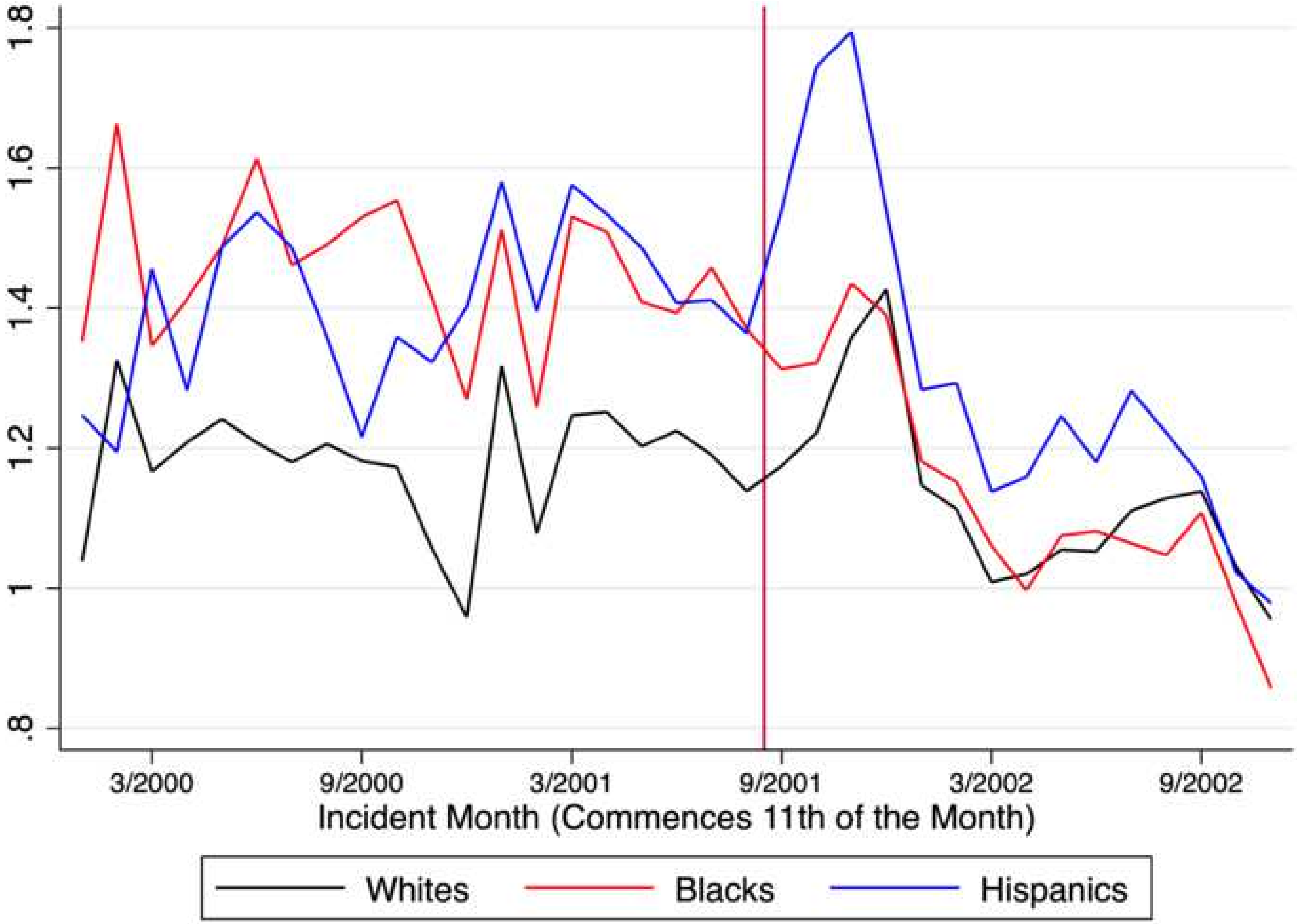
the regular set of ocntrols. Estimates for the two difference-in differences terms post 9-11 x Hispanic and post 9-11 x Black are presented above.

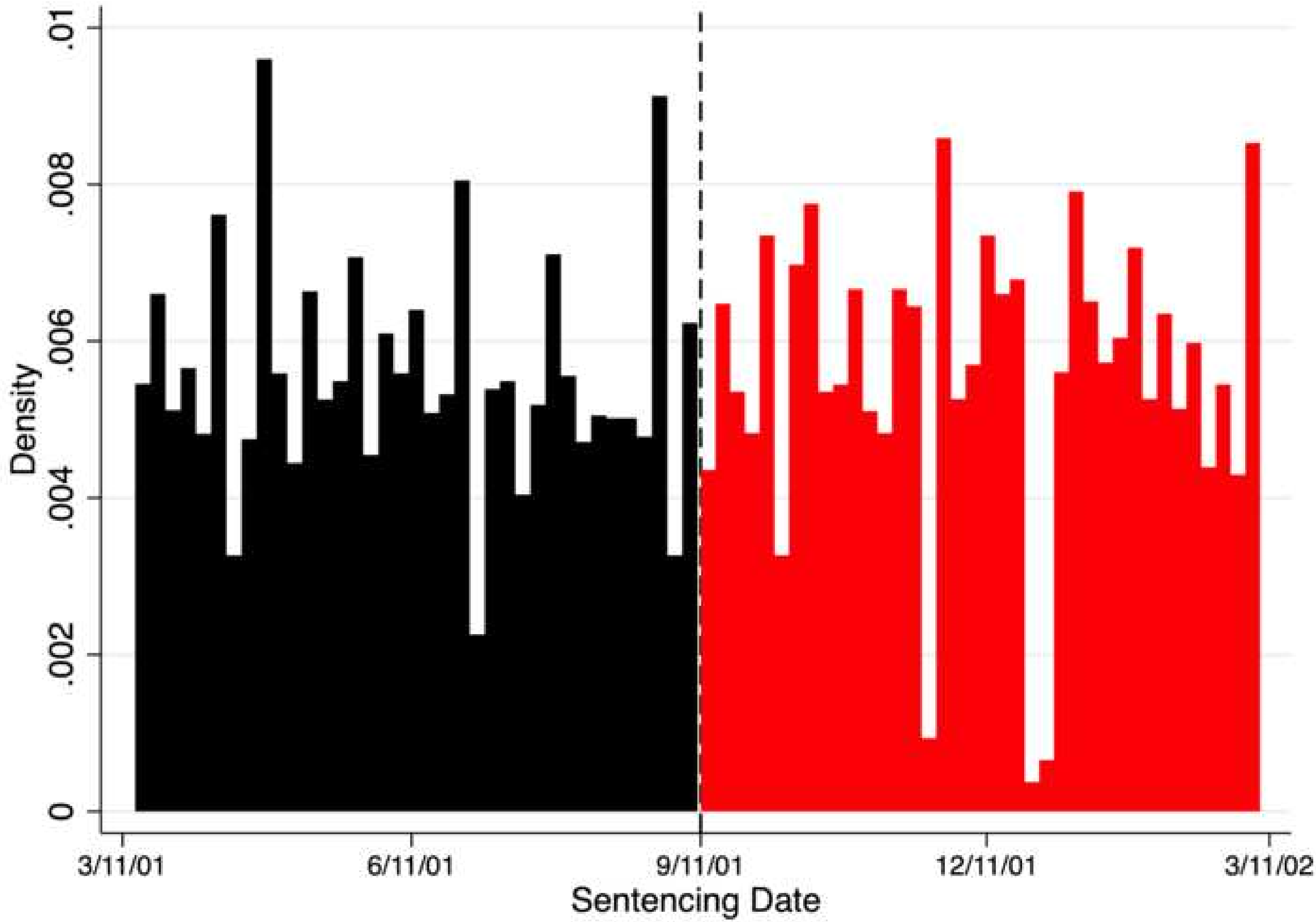
5.) Title:
Figure 5: Predicted Impact on Sentence Length (months)
Left hand side subtitle:
A. Baseline Controls
Right hand side subtitle:
B. Extended Set of Interactions
Notes:
Notes: Each Panel shows estimates where each coefficient and corresponding confidence interval comes from a separate difference-in-differences regression. The regressions are based on different sub-samples of the baseline sample of 40,116 Federal case. The subsamples are created as follows. We begin with an expanded sample of all non-life sentence Federal cases that come up for sentencing between 10/1/1998 and 180 days after 9/11/2001. For cases sentenced pre 9-11, we run a probit regression by ethnicity where the dependent variable is a dummy for downwards departure. Figure A is based on our regular set of controls. Figure B is based on the same set of controls but additionally controls for a set of dummies based on an interaction between number of dependants and criminal history category. We use this first stage regression to predict the probability of a downwards departure for the full, expanded sample (i.e. including post 9-11) and then restrict the sample to the 180 day window around 9/11/2001. We use this predicted probability to create the sub-samples on which the sentence length regressions are based. We calculate the percentiles of the predicted probability of downwards departure for values from 5 to 90 in increments of 5. We subsequently keep observations if the predicted probability exceeds this percentile value. Based on each sub-samples, we run a difference-in-differences regression, where the dependent variable is sentence length, and the regular set of control variables are included. Point estimates and corresponding 95% confidence intervals are shown for the post 9-11 x Hispanic term. On the right hand y-axis we show the value of the predicted probability at each percentile cut-off. In each Figure, the dashed line represents the diff-in-diff estimate based on our working sample around the 9-11 window, excluding defendants with life sentences.

6.) Title:
Figure 6: Decomposition of Hispanic-White Differentials in Downward Departures
Notes:
Notes: The graph shows key results from a Juhn-Murphy-Pierce [1993] decomposition, using a non-parametric procedure. This decomposes the unconditional difference-in-difference for downwards departures between Hispanics and Whites, 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 controls in this decomposition are Offender characteristics, defense counsel type, primary offense type dummies and a dummy for multiple offenses, guideline cell dummies, and Federal district dummies. For the Juhn-Murphy-Pierce decomposition, Whites are chosen as the reference group.

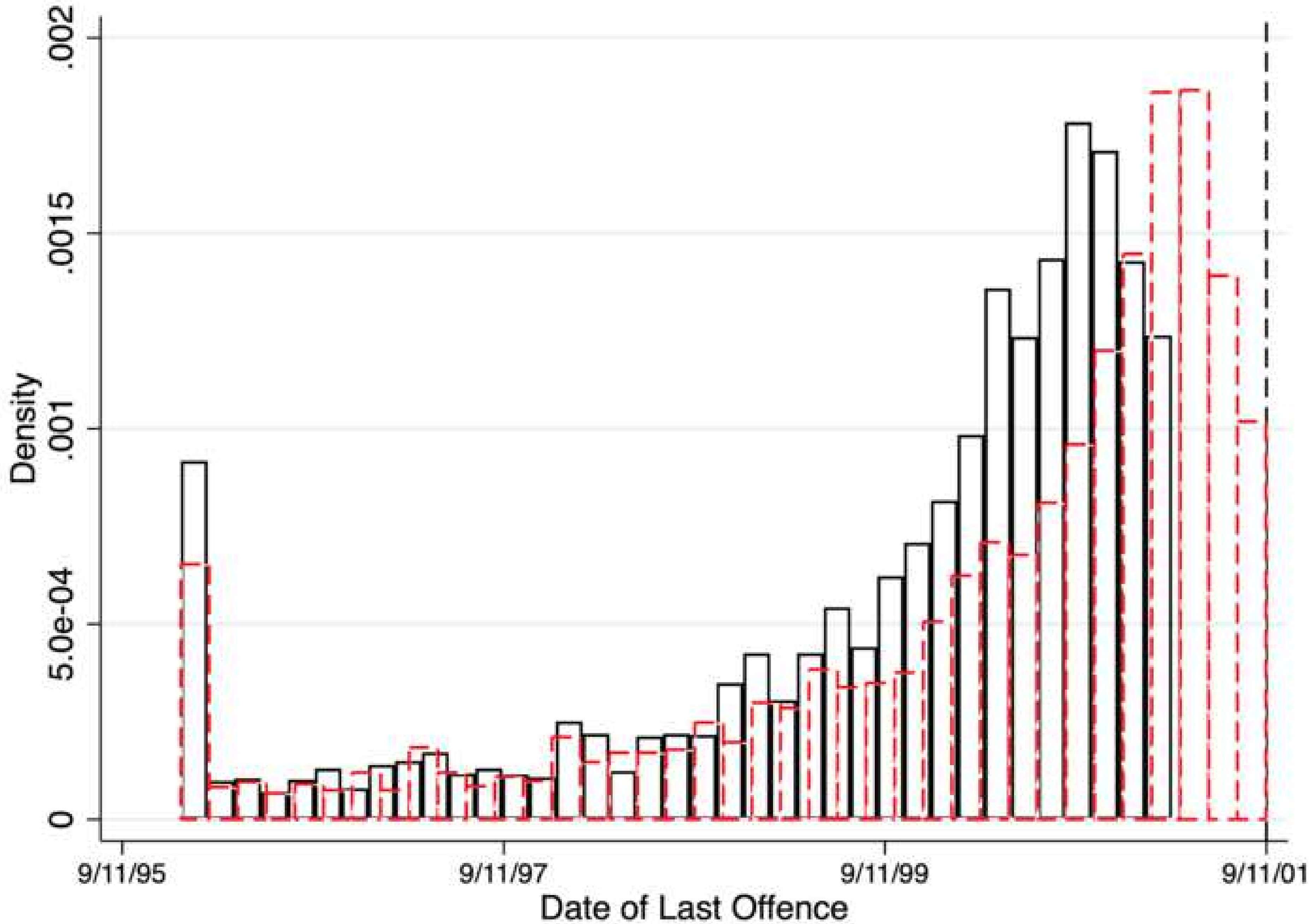

Stage:
0. Arrest/offense
1. Initial Appearance
2. Bail
3. Arraignment
4. Initial District Court Appearance
5. Pre-trial Motions
6. Plea
7. Trial
8. Sentencing
9. Appeals
Duration:
<1 day
<1 day
3-7 days
<1 day
Days Between Stages:
Maximum 14 21 days if in [out of] custody, from initial appearance
Differs by circuit court
75 [90] days if in [out of] custody
Almost no delay to state intention to appeal

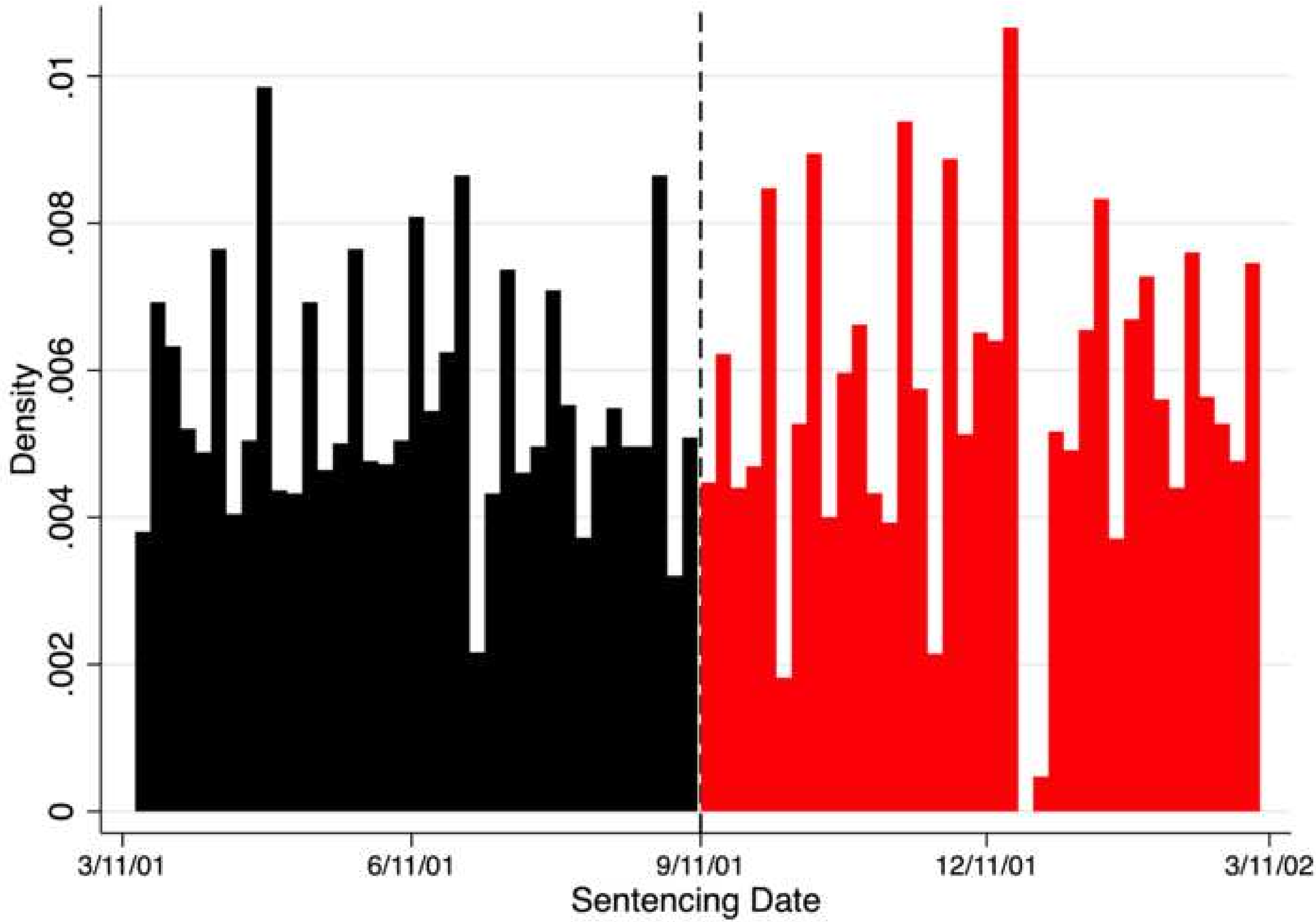

60%
40%
20%
0%
9/11/95
9/11/96
9/11/97
9/11/98
9/11/99
9/11/00
9/11/01
9/11/02
Increased
Present Level
Decreased
No Opinion

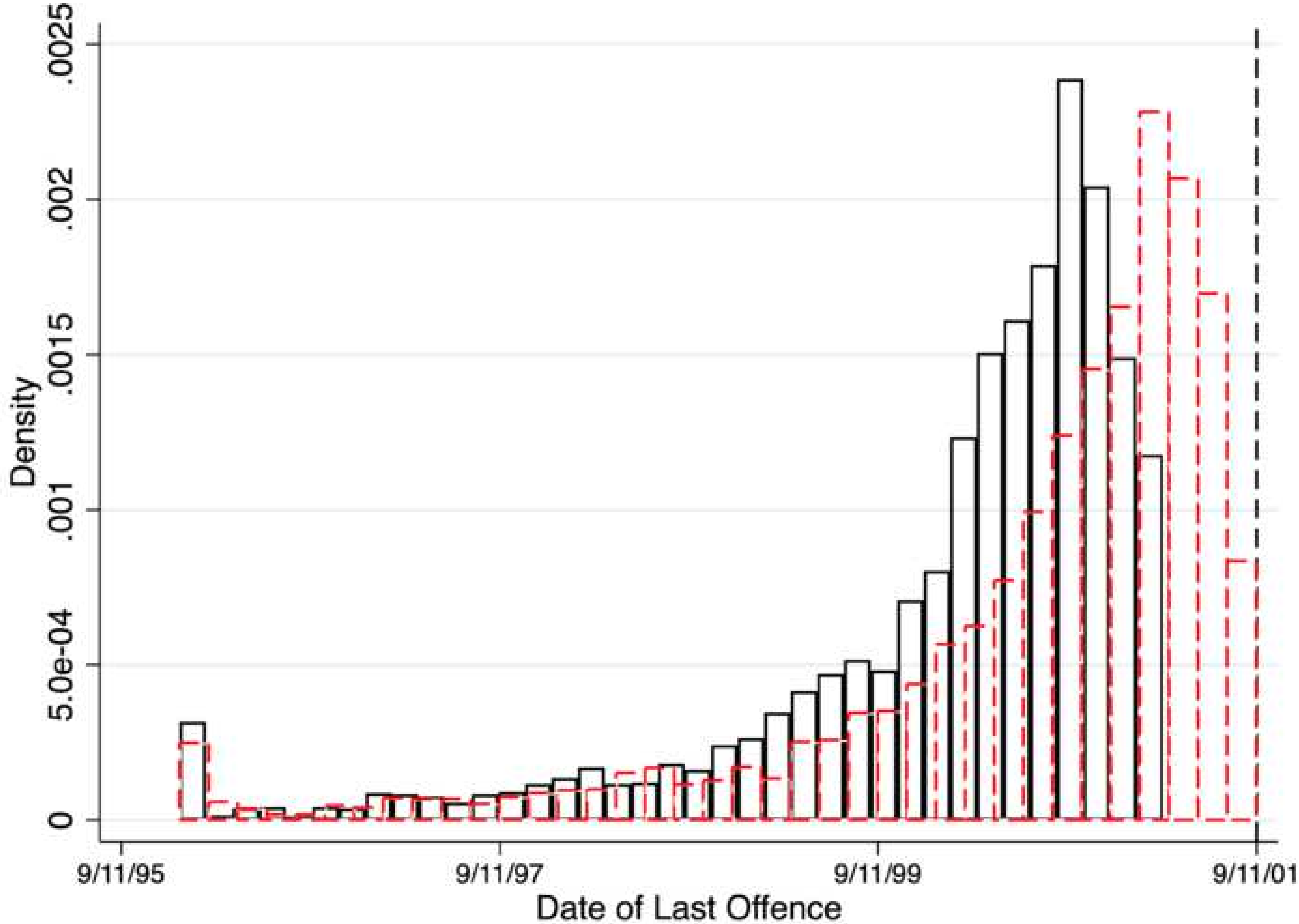

1.8
1.6
1.4
1.2
1
.8
3/2000
9/2000
3/2001
9/2001
3/2002
9/2002
Incident Month (Commences 11th of the Month)
Whites
Blacks
Hispanics



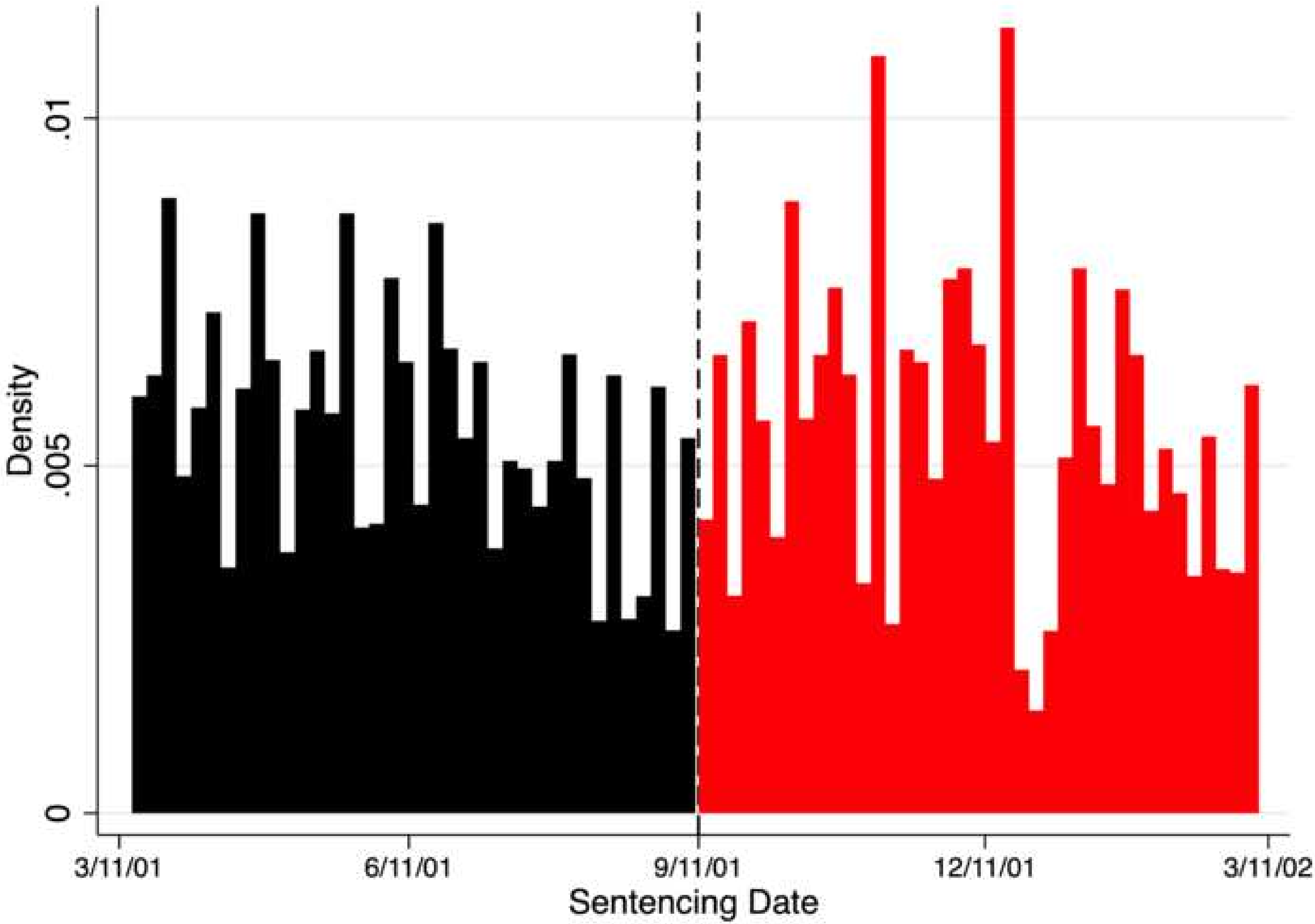
Density
.01
.008
.006
.004
.002
0
3/11/01
6/11/01
9/11/01
12/11/01
3/11/02
Sentencing Date

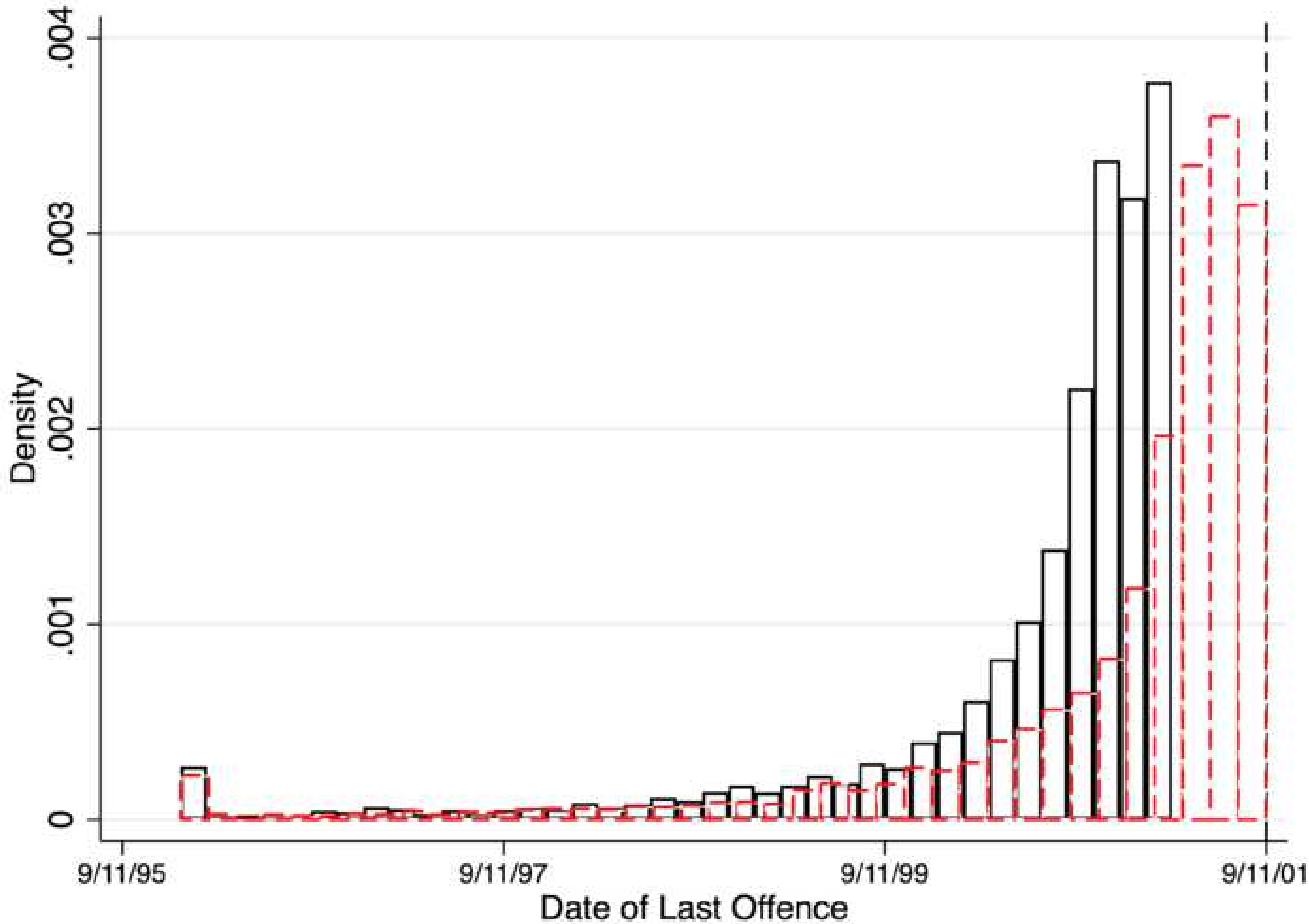

.002
.0015
.001
5.0e-04
0
Density
9/11/95
9/11/97
9/11/99
9/11/01
Date of Last Offence


Density
.01
.008
.006
.004
.002
0
3/11/01
6/11/01
9/11/01
12/11/01
3/11/02
Sentencing Date

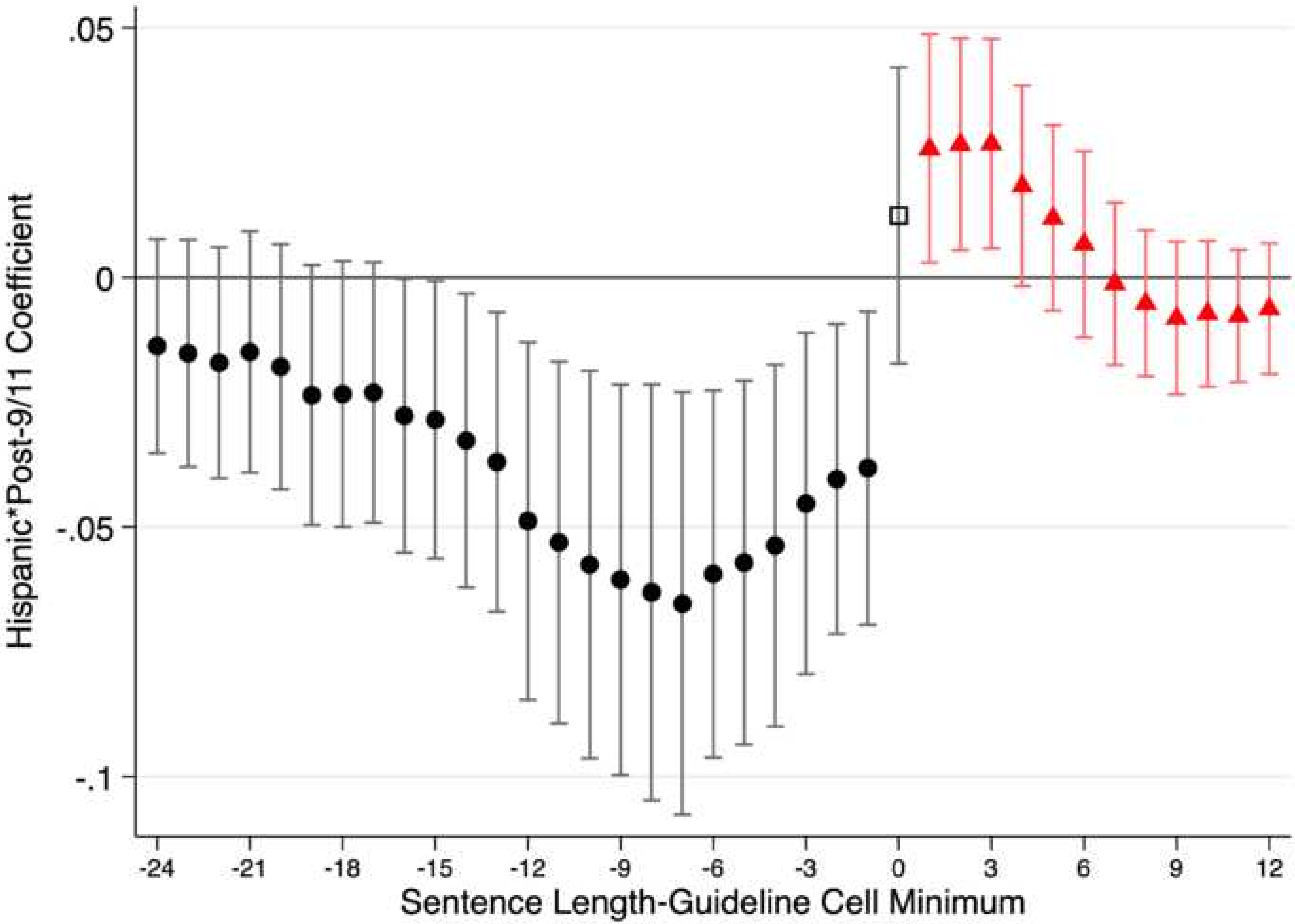


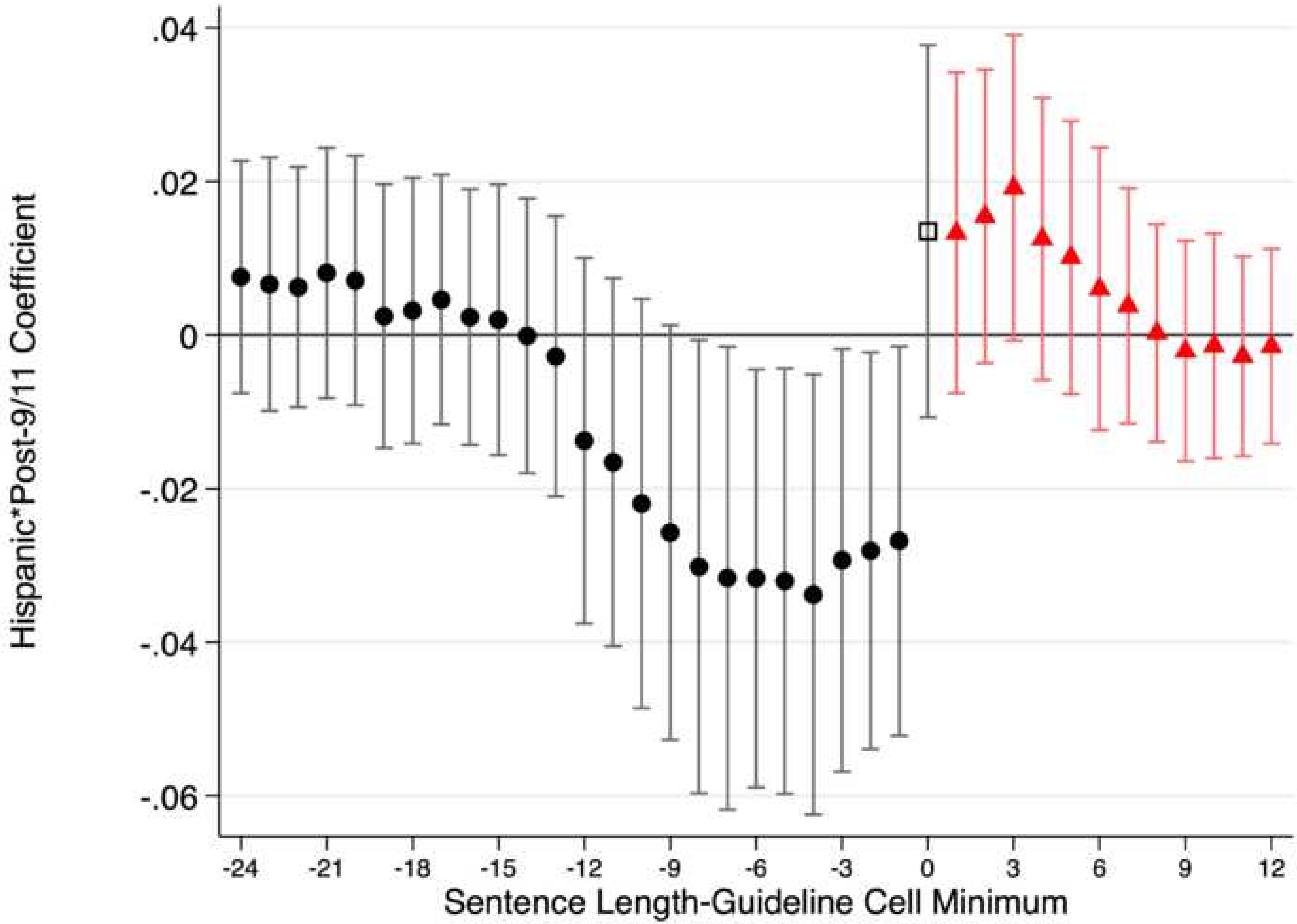
Density
0
5.0e-04
.001
.0015
.002
.0025
9/11/95
9/11/97
9/11/99
9/11/01
Date of Last Offence


Density
.01
.005
0
3/11/01
6/11/01
9/11/01
12/11/01
3/11/02
Sentencing Date

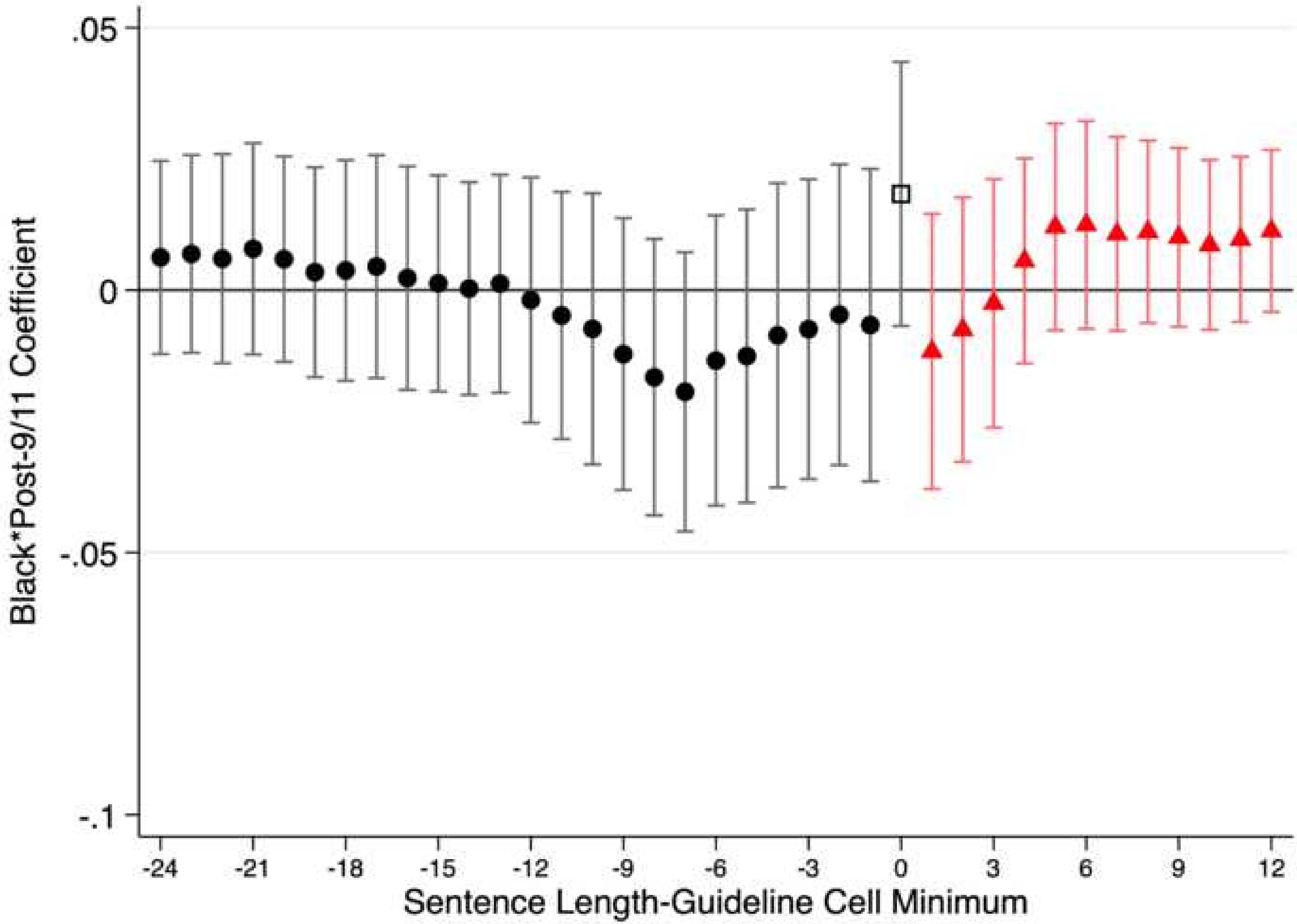


Density
0
.001
.002
.003
.004
9/11/95
9/11/97
9/11/99
9/11/01
Date of Last Offence

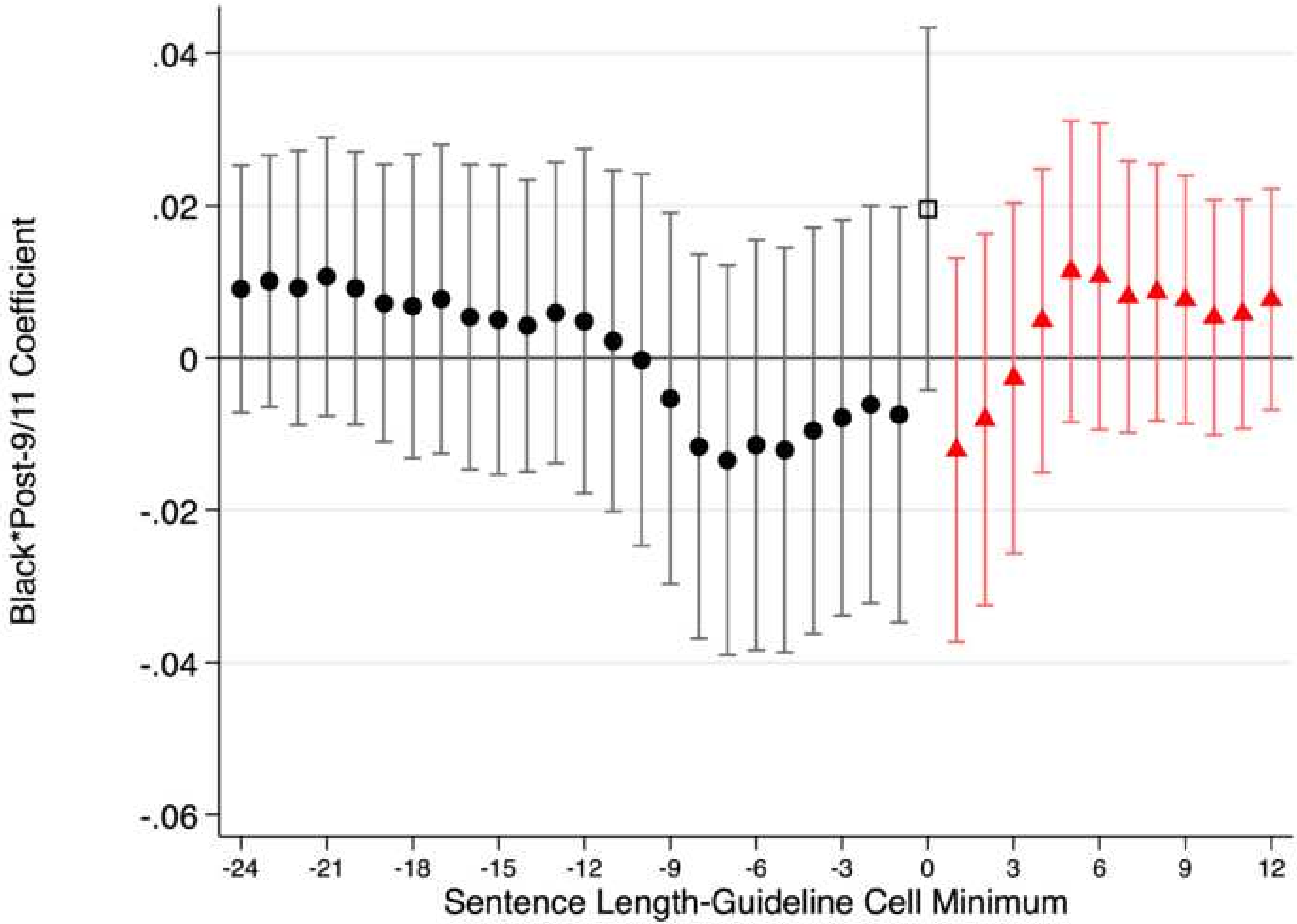

.05
0
-.05
-.1
Hispanic*Post-9/11 Coefficient
-24
-21
-18
-15
-12
-9
-6
-3
0
3
6
9
12
Sentence Length-Guideline Cell Minimum

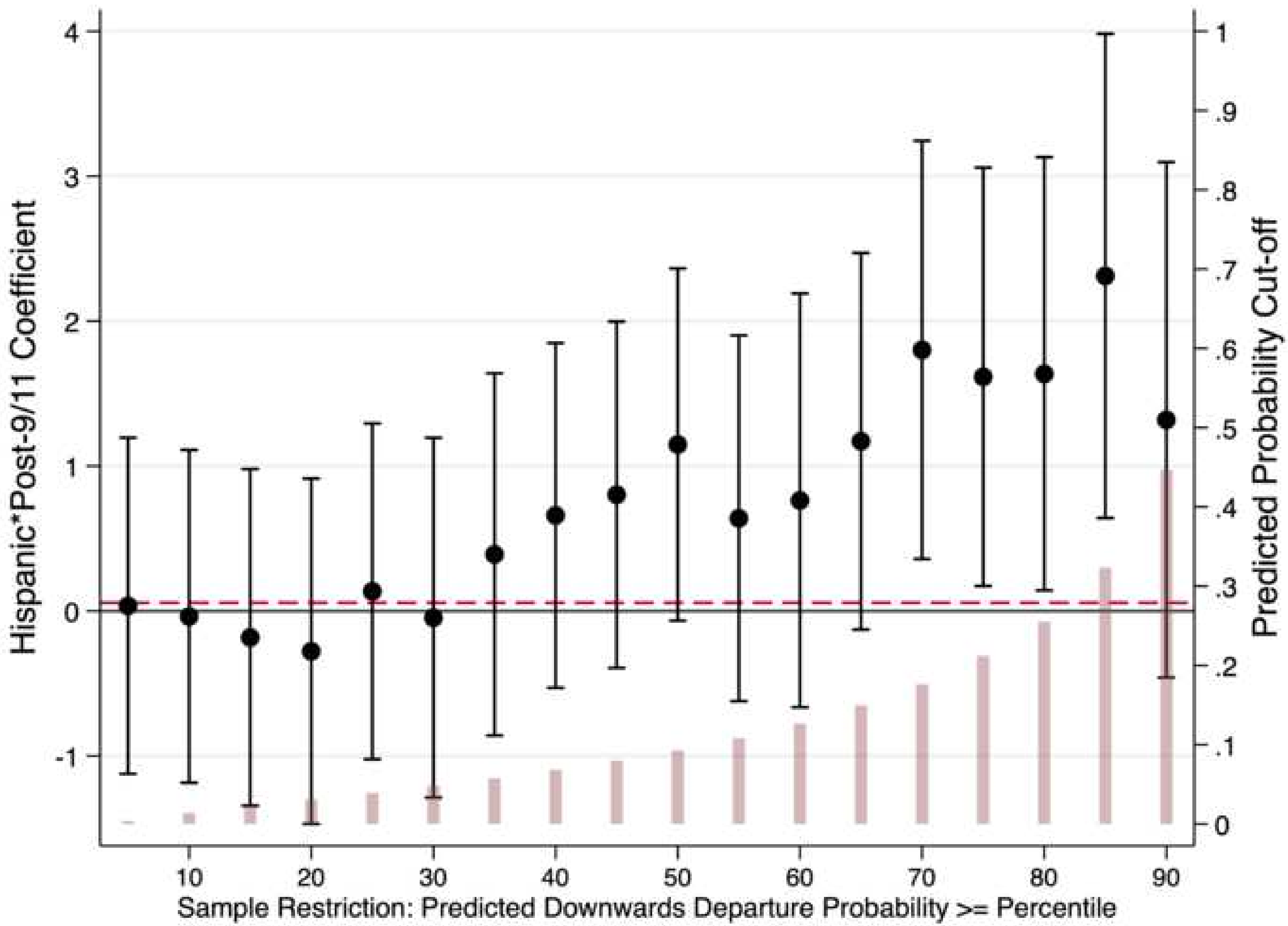


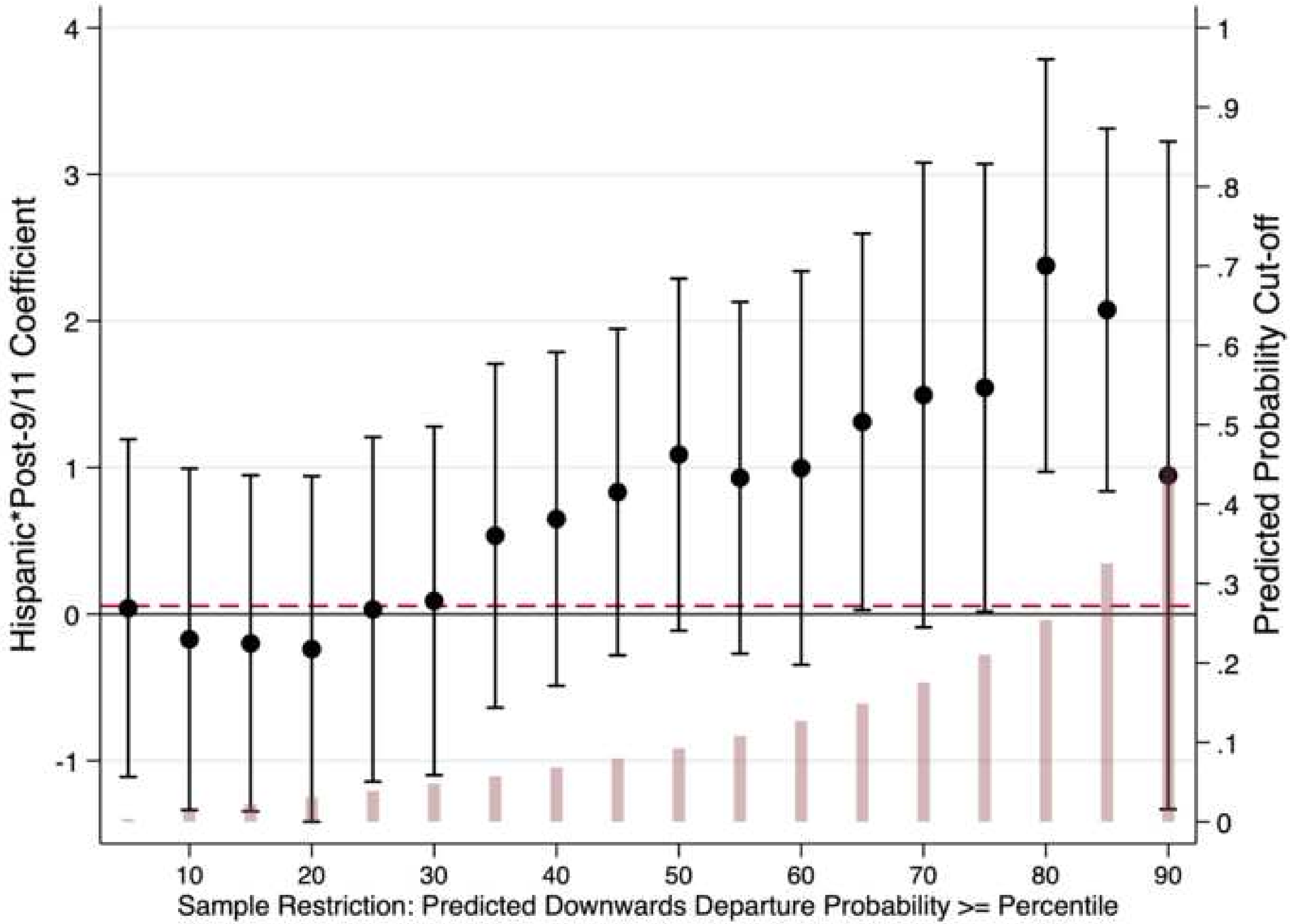
.04
.02
0
-.02
-.04
-.06
Hispanic*Post-9/11 Coefficient
-24
-21
-18
-15
-12
-9
-6
-3
0
3
6
9
12
Sentence Length-Guideline Cell Minimum



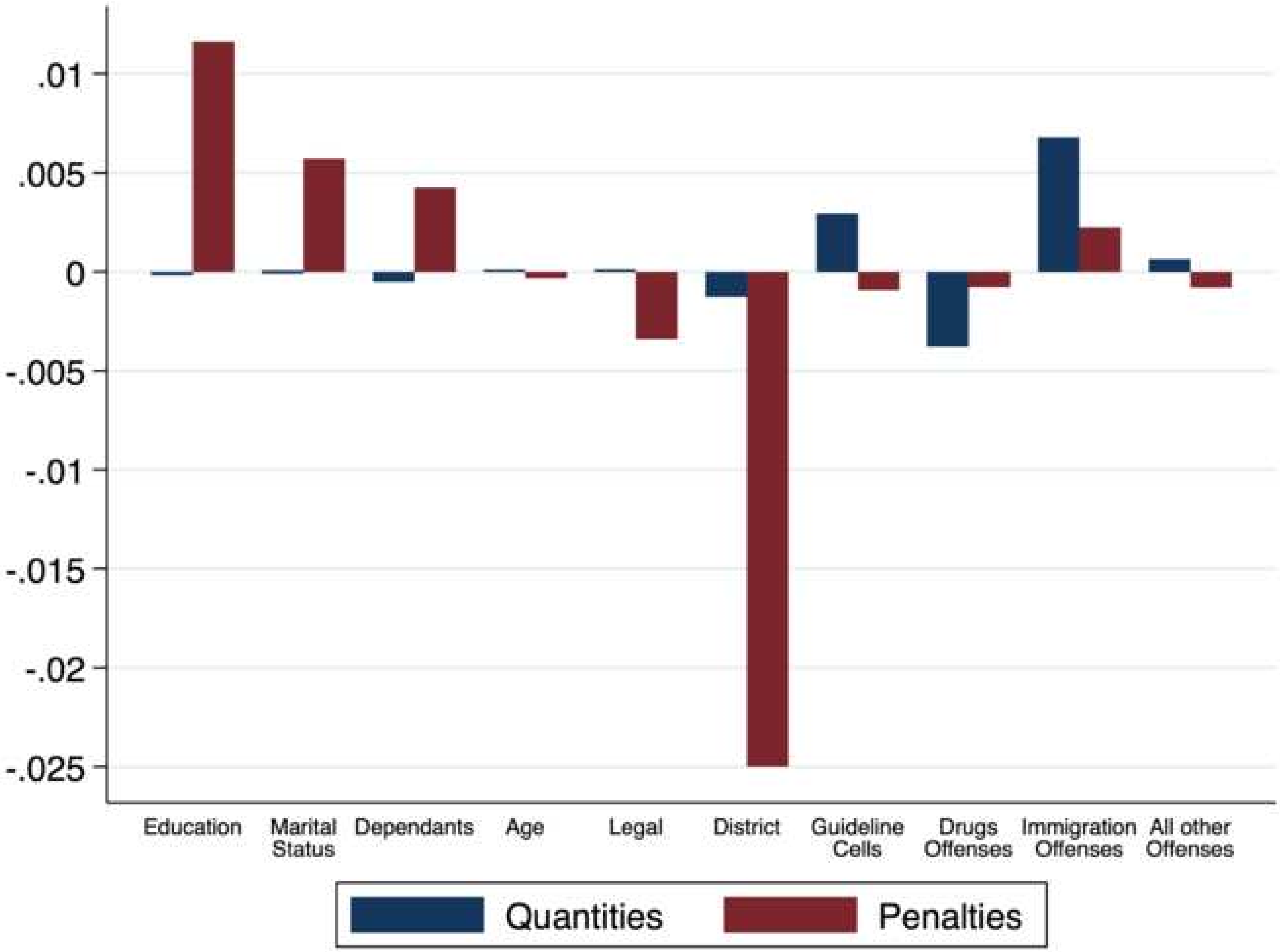
.05
0
-.05
-.1
Black*Post-9/11 Coefficient
-24
-21
-18
-15
-12
-9
-6
-3
0
3
6
9
12
Sentence Length-Guideline Cell Minimum


.04
.02
0
-.02
-.04
-.06
Black*Post-9/11 Coefficient
-24
-21
-18
-15
-12
-9
-6
-3
0
3
6
9
12
Sentence Length-Guideline Cell Minimum


Hispanic*Post-9/11 Coefficient
Predicted Probability Cut-off
Sample Restriction: Predicted Downwards Departure Probability >= Percentile
4
3
2
1
0
-1
1
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
10
20
30
40
50
60
70
80
90



Hispanic*Post-9/11 Coefficient
4
3
2
1
0
-1
Predicted Probability Cut-off
1
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
10
20
30
40
50
60
70
80
90
Sample Restriction: Predicted Downwards Departure Probability >= Percentile


.01
.005
0
-.005
-.01
-.015
-.02
-.025
Education
Marital Status
Dependants
Age
Legal
District
Guideline Cells
Drugs Offenses
Immigration Offenses
All other Offenses
Quantities
Penalties

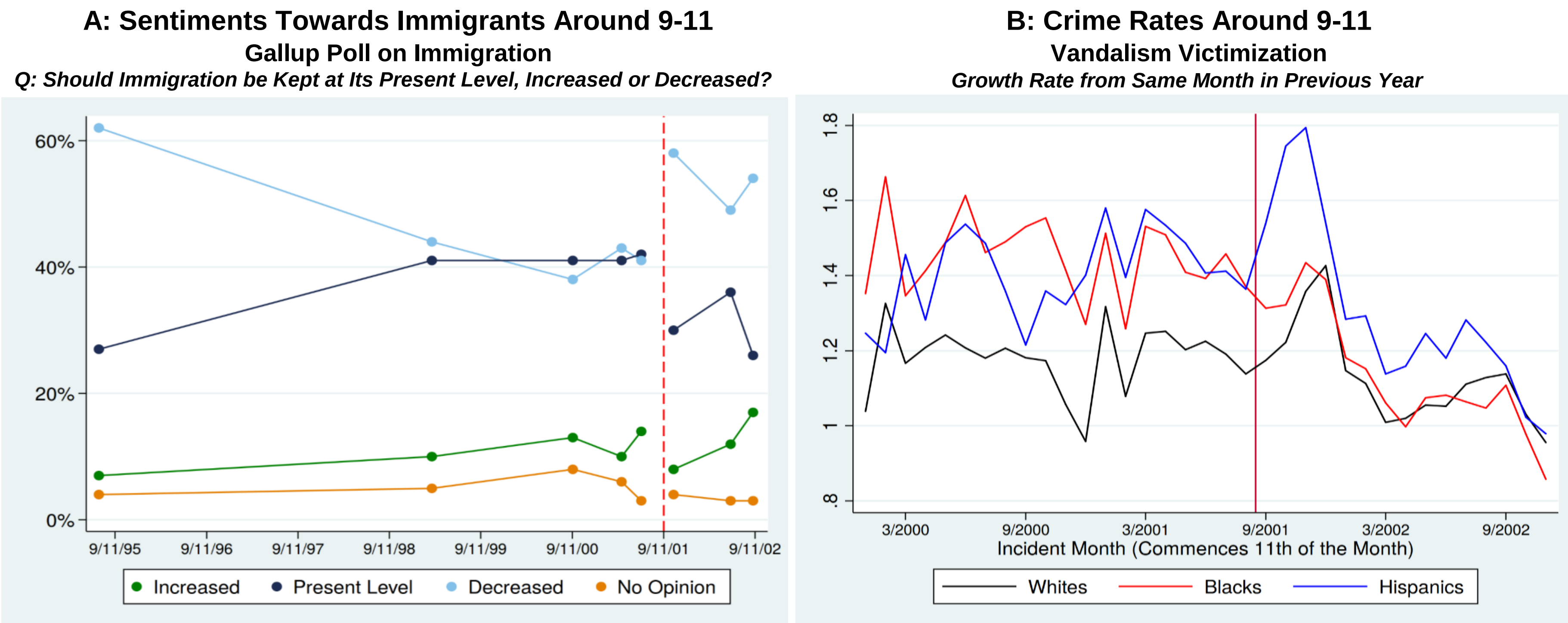
Table

Figure 1: Federal CJS Timeline

Stage:	**0. Arrest/offense**	**1. Initial Appearance**	**2. Bail**	**3. Arraignment**		**4. Initial District Court Appearance**	**5. Pre-trial Motions**	**6. Plea**	**7. Trial**		**8. Sentencing**		**9. Appeals**
Duration:		**<1 day**		**<1 day**					**3-7 days**		**<1 day**		
Days Between Stages:			*Maximum 14 21 days if in [out of] custody, from initial appearance*		*Differs by circuit court*					*75 [90] days if in [out of] custody*		*Almost no delay to state intention to appeal*	

Notes: We use the Federal Justice Statistics Resource Center (FJSRC) data and the Monitoring of Federal Criminal Sentences (MFCS) data set for our analysis. The FJSRC comprises information gathered from four linked administrative data sources, and we replace the fourth stage USSC data that covers sentencing Stage 8, with the MFCS data (because it is superior to the USSC data that is part of the FJSRC). As described in Rehavi and Starr [2014], the 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) at Stage 8, we then use the Monitoring of Federal Criminal Sentences (MFCS) data set.

Figure 2: Pre and Post 9-11 Sentiments

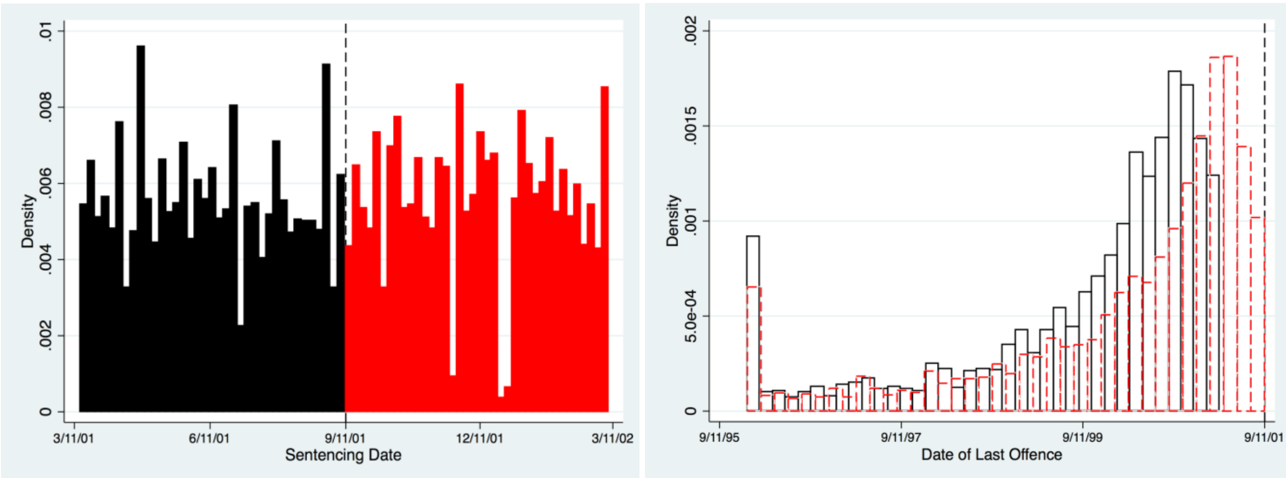


Notes: Panel A is based on a Gallup Poll that asks respondents, "Thinking more about immigration - that is, people who come from other countries to live here in the United States, in your view, should immigration be kept at its present level, increased or decreased?". The data was accessed via http://www.gallup.com/poll/1660/immigration.aspx.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

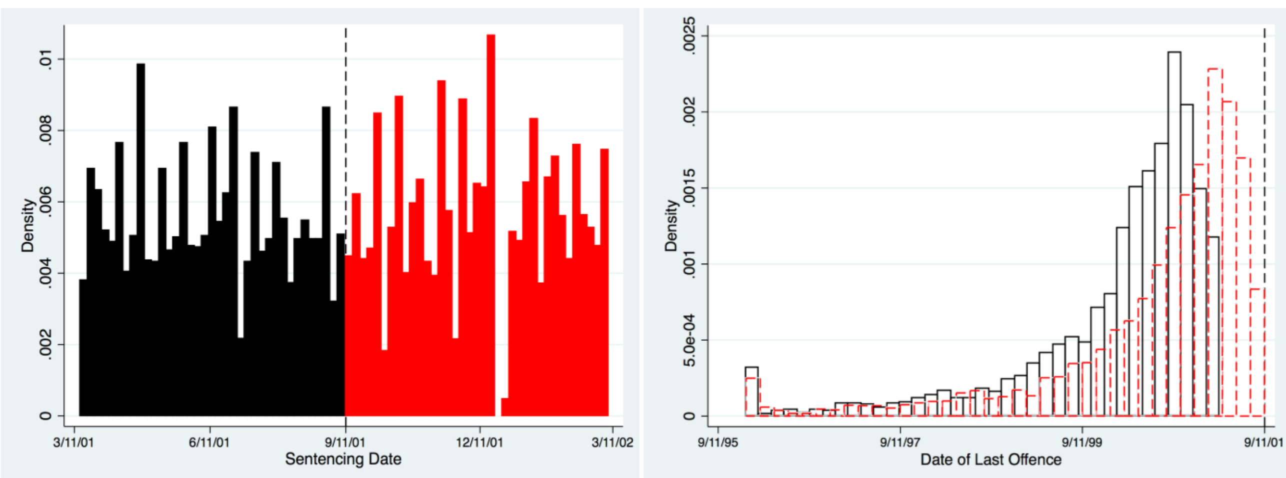
Figure 3: Sentencing and Last Offense Dates

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

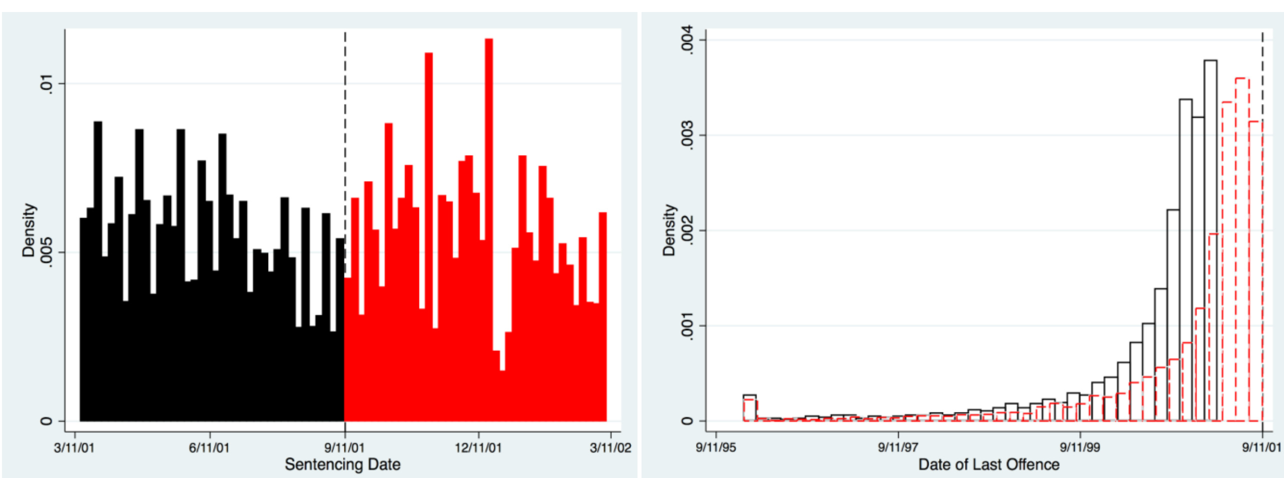
Whites



Blacks

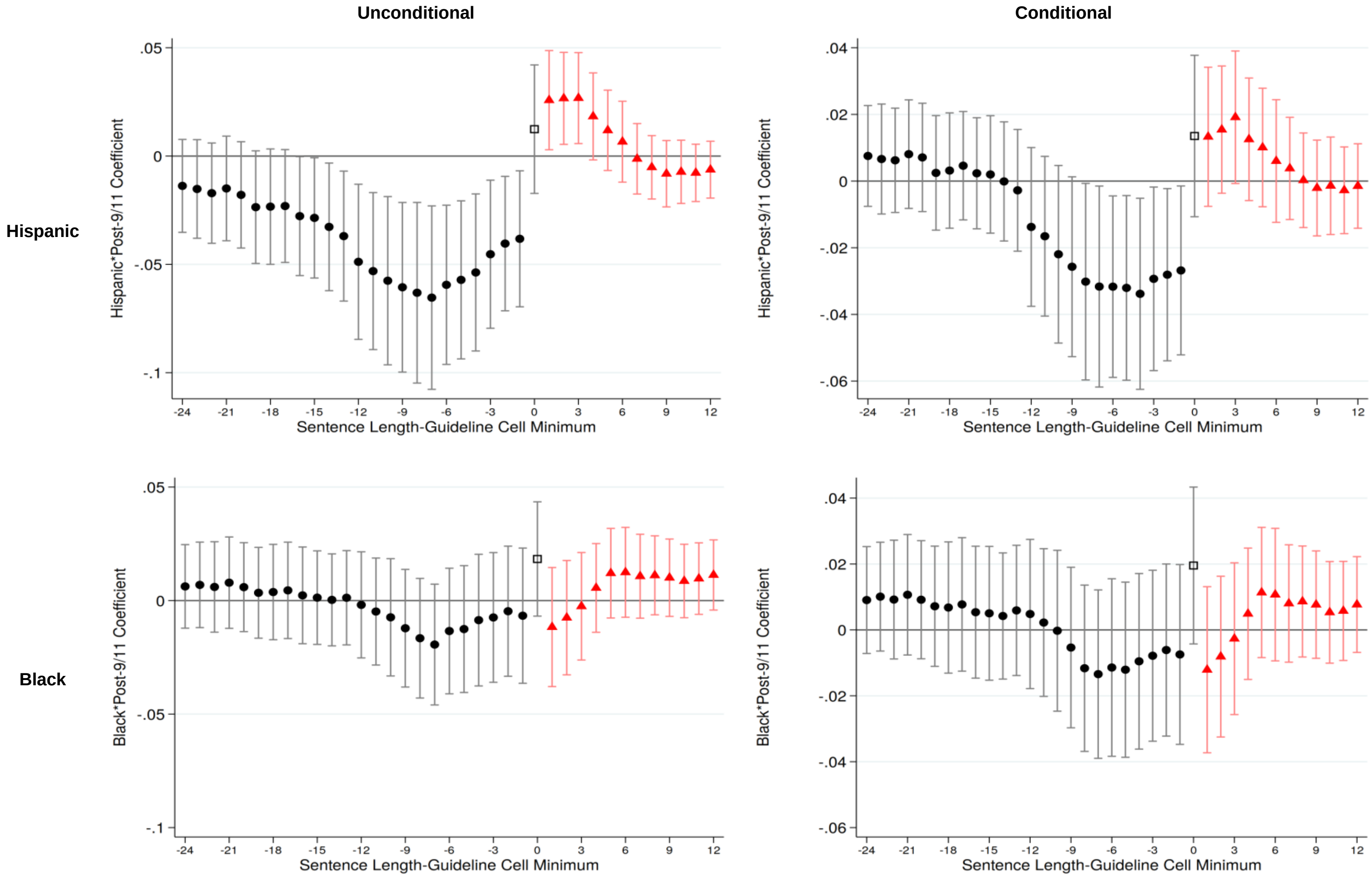


Hispanics



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

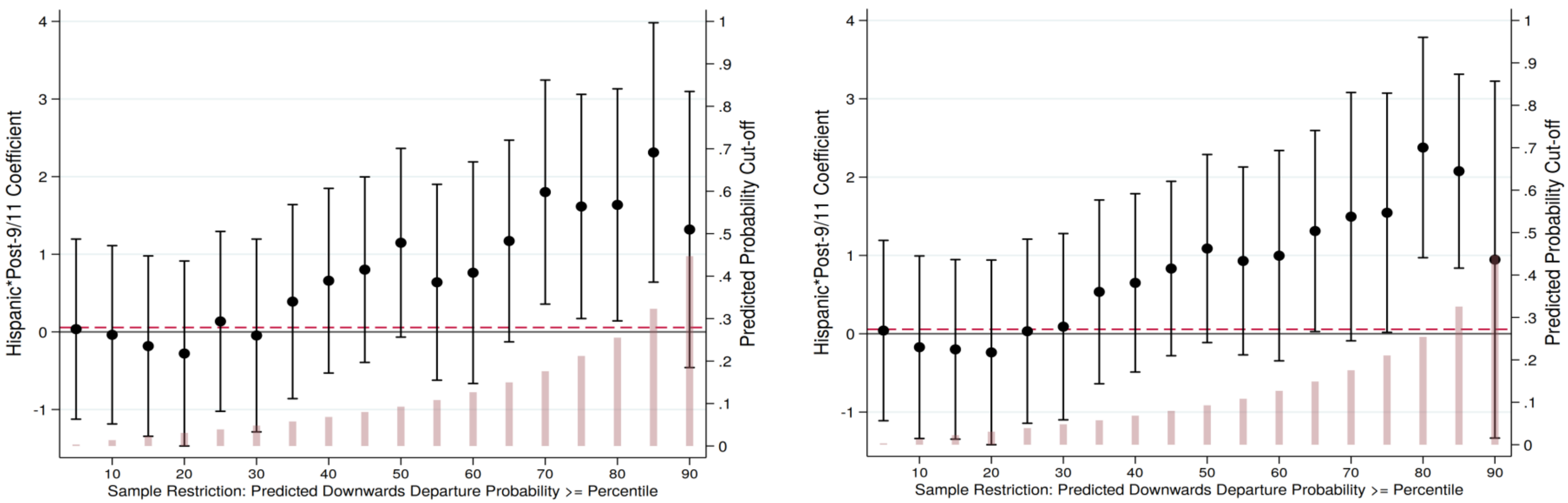
Figure 4: Sentencing Adjustments



Notes: The figures show estimates from different regressions, where each coefficient and corresponding 95% confidence interval comes from a separate regression. 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. Standard error or clustered by district. The regressions are based on the difference between an individual's sentence length, and the minimum sentence length recommendation in their allocated guideline cell. Based on this difference, we create a series of dummy variables, which are the dependent variables in the figures above. The first set take a value of 1 (0 otherwise) if the difference in sentence length-guideline cell minimum is less than or equal to a negative integer in the range -24 to -1 (The estimates based on these dependent variables are represented by solid black circles above). We treat zero separately, creating a dummy if sentence length equals the guideline cell minimum (corresponding estimates for this dependent variable are represented by hollow black squares above). Finally we create a set of dummy variables that take a value of 1 (0 otherwise) if sentence length-guideline cell minimum is greater than or equal to a positive integer in the range 1 to 12 (Estimates for which are represented by solid red triangles). We then run a separate OLS regression based on each of these dependent variables, and estimate difference-in-differences models, both without and without a set of additional control variables. In the unconditional models we condition on defendant group (White, Black, Hispanic), whether the case comes up post 9-11, and interactions between the two. In the conditional models we include the regular set of ocntrols. Estimates for the two difference-in differences terms post 9-11 x Hispanic and post 9-11 x Black are presented above.

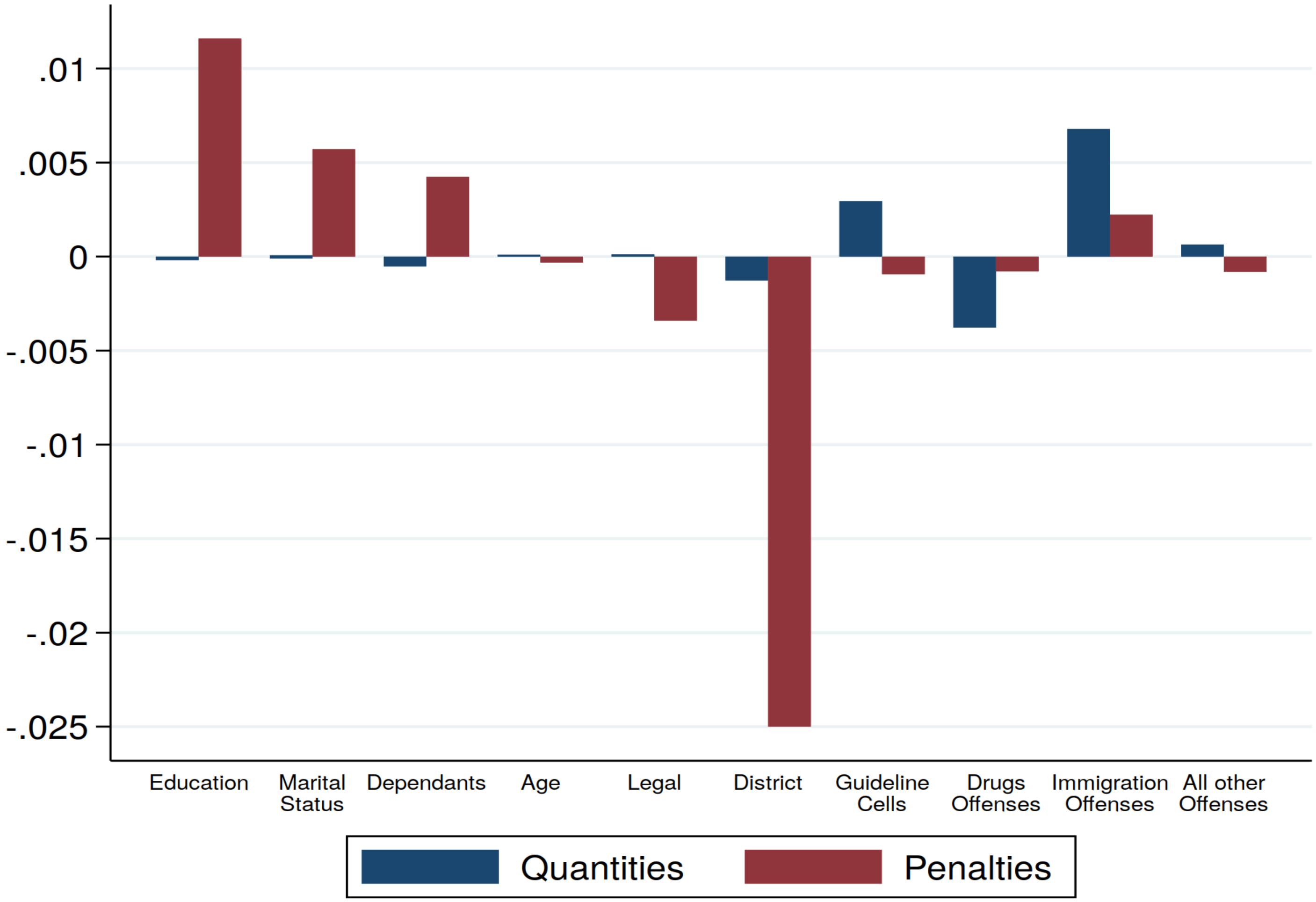
Figure 5: Predicted Impact on Sentence Length (months)

A. Baseline Controls **B. Extended Set of Interactions**



Notes: Each Panel shows estimates where each coefficient and corresponding confidence interval comes from a separate difference-in-differences regression. The regressions are based on different sub-samples of the baseline sample of 40,116 Federal case. The subsamples are created as follows. We begin with an expanded sample of all non-life sentence Federal cases that come up for sentencing between 10/1/1998 and 180 days after 9/11/2001. For cases sentenced pre 9-11, we run a probit regression by ethnicity where the dependent variable is a dummy for downwards departure. Figure A is based on our regular set of controls. Figure B is based on the same set of controls but additionally controls for a set of dummies based on an interaction between number of dependants and criminal history category. We use this first stage regression to predict the probability of a downwards departure for the full, expanded sample (i.e. including post 9-11) and then restrict the sample to the 180 day window around 9/11/2001. We use this predicted probability to create the sub-samples on which the sentence length regressions are based. We calculate the percentiles of the predicted probability of downwards departure for values from 5 to 90 in increments of 5. We subsequently keep observations if the predicted probability exceeds this percentile value. Based on each sub-samples, we run a difference-in-differences regression, where the dependent variable is sentence length, and the regular set of control variables are included. Point estimates and corresponding 95% confidence intervals are shown for the post 9-11 x Hispanic term. On the right hand y-axis we show the value of the predicted probability at each percentile cut-off. In each Figure, the dashed line represents the diff-in-diff estimate based on our working sample around the 9-11 window, excluding defendants with life sentences.

Figure 6: Decomposition of Hispanic-White Differentials in Downward Departures



Notes: The graph shows key results from a Juhn-Murphy-Pierce [1993] decomposition, using a non-parametric procedure. This decomposes the unconditional difference-in-difference for downwards departures between Hispanics and Whites, 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 controls in this decomposition are Offender characteristics, defense counsel type, primary offense type dummies and a dummy for multiple offenses, guideline cell dummies, and Federal district dummies. For the Juhn-Murphy-Pierce decomposition, Whites are chosen as the reference group.

Table A3: Citizenship and Offense Type, by Initial Arrest Codes

Dependent Variable: Downward Departure Granted by Federal Judge

Standard errors in parentheses clustered by district

	(1) Drug Offenses	(2) Immigration Offenses: Hispanics Only	(3) Immigration Offenses: Hispanics Only, Border States	(4) All Other Offenses
Sentenced post 9-11*Hispanic Citizen	-.007	-.079	-.060	-.042*
	(.014)	(.049)	(.062)	(.023)
Sentenced post 9-11*Hispanic Non-Citizen	-.076*	.026	-.003	.005
	(.044)	(.045)	(.057)	(.031)
Sentenced post 9-11*Black	-.005			-.026*
	(.014)			(.015)
Sentenced post 9-11	-.008			.013
	(.013)			(.010)
Offender, Legal and District Controls	Yes	Yes	Yes	Yes
Offense Type Codes	Arrest	Arrest	Arrest	Arrest
Guideline Cells	No	No	No	No
Implied Sentence Length Impact (H, Citizen)	.384 [4.1%]	1.033	.714	.457 [-49.3%]
Implied Sentence Length Impact (H, Non-citizen)	1.80 [23.1%]	1.02 [29.2%]	1.20 [31.0%]	-.295 [12.3%]
p-value: [Post*H Citizen= Post*H Non Citizen]	.102	.259	.640	.309
Adjusted R-squared	.333	.262	.220	.080
Observations	11,871	4,534	3,478	9,837

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 sample of 26,676 Federal cases is used (those that come up for sentencing in a six month window either side of 9/11/2001), for which defendant citizenship is not missing and that can be linked back to arrest (Stage 0). 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 1 to 4 are restricted to drug, immigration and other offenses respectively, where for immigration offenses, only Hispanic defendants are included and Column 3 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 regular set of controls is included. At the foot of each Column, the percentage reported in square brackets is the percentage of the pre 9-11 differential the implied sentence length impact corresponds to. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Hispanic Citizen and post 9-11 x Hispanic Non Citizen dummy interactions are equal against a two sided alternative.

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

Dependent Variable: Downward Departure Granted by Federal Judge

Standard errors in parentheses – see footnotes for clustering details

	(1) Cluster on district x group	(2) Cluster on district x group	(3) Cluster on sentence week x group	(4) Cluster on sentence week x group	(5) Excluding Cases Where Statutory Minima or Maxima Bind Partially	(6) Excluding Cases Where Statutory Minima or Maxima Bind Partially
Sentenced post 9-11*Hispanic	-.038***	-.046**	-.038***	-.046***	-.041***	-.052***
	(.013)	(.019)	(.011)	(.015)	(.008)	(.015)
Sentenced post 9-11*Black	-.013	-.013	-.013	-.013	-.016**	-.015
	(.008)	(.011)	(.008)	(.010)	(.008)	(.011)
Sentenced post 9-11	.006	.003	.006	.003	.009	.006
	(.007)	(.009)	(.006)	(.008)	(.007)	(.009)
Offender, Legal and District Contı	Yes	Yes	Yes	Yes	Yes	Yes
Offense Type Codes	Final	Arrest	Final	Arrest	Final	Arrest
Guideline Cells	Yes	No	Yes	No	Yes	No
p-value: [Post*B = Post*H]	.042	.064	.022	.0194	.017	.030
Adjusted R-squared	.256	.257	.256	.257	.275	.266
Observations	40,228	26,852	40,228	26,852	32,430	21,844

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 district x group in Columns 1 and 2, sentence week x group in Columns 3 and 4, and clustered by district in Columns 5 and 6. In Columns 1 and 3 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 Columns 2 and 4, we restrict the sample to those cases that can be linked back to arrest (Stage 0). In Column 5 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. Column 6 further restricts the sample used in Column 5 to those cases that can be linked back to arrest (Stage 0). The dependent variable is a dummy for whether the case receives a downwards departure. In all Columns we condition on defendant group (White, Black, Hispanic), whether the case comes up post 9-11, and interactions between the two, and the regular set of controls. In odd-numbered Columns we control for the primary offense type and a dummy for multiple offenses and the guideline cell. In even-numbered Columns 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 A5: Sentencing Differentials Around 9-11, by Group

Dependent Variable: Downward Departure Granted by Federal Judge

Standard errors in parentheses clustered by district

	(1) White	(2) Black	(3) Hispanic	(4) White	(5) Black	(6) Hispanic
Sentenced post 9-11	.004	-.008	-.030***	.003	-.011	-.042**
	(.006)	(.005)	(.011)	(.009)	(.007)	(.017)
Difference with Whites		-.011	-.034***		-.014	-.045***
		(.008)	(.010)		(.011)	(.015)
Difference with Blacks			-.023*			-.031*
			(.012)			(.018)
Offender, Legal and District Controls	Yes	Yes	Yes	Yes	Yes	Yes
Offense Type Codes	Final	Final	Final	Arrest	Arrest	Arrest
Guideline Cells	Yes	Yes	Yes	No	No	No
Adjusted R-squared	.151	.074	.313	.151	.071	.288
Observations	12,994	10,876	16,358	8,415	6,976	11,461

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 Columns 4-6 we restrict the sample to those cases that can be linked back to arrest (Stage 0). In Column 1 and 4 only criminal cases involving White defendants are used. In Column 2 and 5 only criminal cases involving Black defendants are used. In Column 3 and 6 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 regular set of controls. In Columns 1-3 we control for the primary offense type and a dummy for multiple offenses and the guideline cell. In Columns 4-6 we instead control for arrest offense codes, but not guideline cells. In Columns 2 and 5 we report the difference between the coefficient estimate between Blacks and Whites (and the corresponding standard error). In Columns 3 and 6 we report the differences between the coefficient estimate between Hispanics and Whites, and Hispanics and Blacks (and the corresponding standard error).

Table A6: Sentencing Differentials Around 9-11, by Ethnicity

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

	(1) Downward Departure	(2) Downward Departure
Sentenced post 9-11*Black	.009	.020
	(.010)	(.014)
Sentenced post 9-11*American Indian	-.037	-.031
	(.023)	(.029)
Sentenced post 9-11*Asian/Pacific Islander	.034	.051
	(.024)	(.033)
Sentenced post 9-11*Multi-Racial	.004	-.073
	(.095)	(.122)
Sentenced post 9-11*Other Race	-.118	-.014
	(.147)	(.142)
Sentenced post 9-11	-.016*	-.026**
	(.009)	(.013)
Offender, Legal and District Controls	Yes	Yes
Offense Type Codes	Final	Arrest
Guideline Cells	Yes	No
Adjusted R-squared	.254	.257
Observations	40,858	27,228

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 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 regular set of controls. In Column 1 we control for the primary offense type and a dummy for multiple offenses and the guideline cell. In Column 2 we instead control for arrest offense codes, but not guideline cells.

Table A7: Time in the Federal CJS

Dependent Variable: Downward Departure Granted by Federal Judge

Standard errors in parentheses clustered by district

	(1) Include Dummies for 20 Groupings of Time Between Last Offense and Sentence	(2) Include Dummies for 20 Groupings of Time Between Last Offense and Sentence	(3) Include Dummies for 20 Groupings of Last Offense Date	(4) Include Dummies for 20 Groupings of Last Offense Date
Sentenced post 9-11*Hispanic	-.035***	-.043**	-.042***	-.047***
	(.010)	(.016)	(.009)	(.014)
Sentenced post 9-11*Black	-.013	-.014	-.014*	-.015
	(.008)	(.011)	(.008)	(.011)
Sentenced post 9-11	.006	.005	-.002	-.003
	(.007)	(.009)	(.008)	(.011)
Offender, Legal and District Controls	Yes	Yes	Yes	Yes
Offense Type Codes	Final	Arrest	Final	Arrest
Guideline Cells	Yes	No	Yes	No
p-value: [Post*B = Post*H]	.0841	.132	.0126	.0366
Adjusted R-squared	.261	.260	.257	.257
Observations	40,228	26,852	40,228	26,852

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. In all Columns we condition on defendant group (White, Black, Hispanic), whether the defendant is sentenced after 9-11 and interactions between this treatment dummies and offender ethnicity, and the regular set of controls. In Columns 1 and 2 we additionally include dummies to group the days between last offense and sentencing date into 20 bins, and in Column 3 and 4 we instead additionally include dummies to group the date of last offense into 20 bins. In odd-numbered Columns we control for the primary offense type and a dummy for multiple offenses and the guideline cell. In even-numbered Columns 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 A8: Time Between Dates of Last Offense and Sentencing

OLS and survival regression estimates; standard errors in parentheses, clustered by 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	8.064	-.036	.033*	12.326	-.085**	.039*	64.367*	-.078	.035	13.115	.036	.002
	(11.796)	(.028)	(.019)	(15.931)	(.042)	(.021)	(38.666)	(.091)	(.052)	(27.684)	(.069)	(.036)
Sentenced post 9-11*Black	13.895	-.021	.022	13.912	-.009	.005	84.703	-.033	.047	15.278	-.033	.034
	(13.749)	(.026)	(.019)	(18.878)	(.044)	(.029)	(68.225)	(.211)	(.106)	(20.019)	(.039)	(.025)
Sentenced post 9-11	5.955	-.024	.007	5.496	-.033	.008	-61.443*	.072	-.006	9.576	-.046*	.019
	(11.144)	(.020)	(.015)	(15.568)	(.042)	(.021)	(35.317)	(.076)	(.055)	(14.706)	(.027)	(.019)
Controls (incl. guideline cell)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value: [Post*B = Post*H]	.588	.579	.535	.909	.022	.174	.716	.805	.888	.933	.317	.333
Observations	40,228	40,228	40,228	18,370	18,370	18,370	6,790	6,790	6,790	15,068	15,068	15,068

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample of cases refers to those 40,228 cases for which sentencing occurs within a 6-month window of 9/11/2001. For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001, and if sentenced before 9/11/2001, the last offense was committed at least 180 days prior to 9/11/2001. In Columns 1a-1c, the full natural experiment sample is used. In Columns 2a-2c (3a-3c) (4a-4c) the sample is restricted to drug (immigration) (other) offenses. The dependent variable is the number of days between the date of the last offense and the sentencing date. In Columns 1a, 2a, 3a and 4a an OLS model is estimated. In Columns 1b, 2b, 3b and 4b a Cox proportional hazard model is estimated so that a negative coefficient means a lower hazard rate, and thus a longer duration. In Columns 1c, 2c, 3c and 4c a log-logistic model with a frailty parameter is estimated. In this model a positive coefficient implies a longer duration. In all Columns we condition on defendant group (White, Black, Hispanic), whether the defendant is sentenced after 9-11 and interactions between this treatment dummies and offender ethnicity, and the regular set of controls. 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 Checks on 9-11 Impacts

Dependent Variable: Downward Departure Granted by Federal Judge

Standard errors in parentheses clustered by district

	(1) Downward Departure	(2) Downward Departure
Sentenced post 9-11*Hispanic*2001	-.047***	-.049**
	(.014)	(.021)
Sentenced post 9-11*Hispanic	.008	.002
	(.006)	(.008)
Sentenced post 9-11*Black*2001	-.016*	-.017
	(.009)	(.012)
Sentenced post 9-11*Black	.002	.001
	(.004)	(.005)
Sentenced post 9-11*2001	.008	.012
	(.008)	(.010)
Sentenced post 9-11	-.003	-.008
	(.004)	(.005)
DDD Impact: POST*H*2001 - POST*H	-.055***	-.051*
	(.019)	(.027)
Confidence Interval	[-.093, -.016]	[-.105, .004]
Offender, Legal and District Controls	Yes	Yes
Offense Type Codes	Final	Arrest
Guideline Cells	Yes	Yes
Adjusted R-squared	.243	.242
Observations	114,642	70,368

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 sample of cases used are those 114,642 cases for which sentencing occurs within a 6-month window of 9/11 in years 1998 to 2001. For those defendants sentenced after 9/11 each year, the last offense was committed prior to 9/11 that year, and if sentenced before 9/11 each year, the last offense was committed at least 180 days prior to 9/11 that year. The dependent variable is a dummy for whether the case receives a downwards departure. We condition on defendant group (White, Black, Hispanic) whether the case comes up post 9-11, and interactions between the two, and three way interactions between a post 9/11 dummy, a dummy for the 2001 Natural Experiment period, and ethnicity. Throughout the regular controls are included. 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: Pre-Trends

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

	All Offenses NE Sample		All Offenses Full Pre-Sample		All Non-Drugs Offenses Full Pre-Sample		All Drug Offenses Full Pre-Sample		All Non-Marijuana Drug Offenses Full Pre-Sample		Marijuana Drug Offenses Full Pre-Sample	
	(1) Offense Codes	(2) Initial Arrest Codes	(3) Offense Codes	(4) Initial Arrest Codes	(5) Offense Codes	(6) Initial Arrest Codes	(7) Offense Codes	(8) Initial Arrest Codes	(9) Offense Codes	(10) Initial Arrest Codes	(11) Offense Codes	(12) Initial Arrest Codes
Sentenced post 9-11*Hispanic	-.038***	-.046***	-.042***	-.050***	-.038***	-.046**	-.037**	-.049**	-.014	-.022	-.032	-.044
	(.010)	(.016)	(.012)	(.018)	(.011)	(.018)	(.017)	(.023)	(.016)	(.019)	(.039)	(.043)
Sentenced post 9-11*Black	-.013	-.013	-.008	-.017	-.009	-.012	-.003	-.018	-.017	-.035**	.048**	-.009
	(.008)	(.011)	(.008)	(.011)	(.010)	(.014)	(.013)	(.015)	(.011)	(.013)	(.023)	(.041)
Sentenced post 9-11	.006	.003	.001	.002	.003	.003	-.004	-.006	.007	.010	-.044**	-.040*
	(.007)	(.009)	(.006)	(.009)	(.009)	(.012)	(.011)	(.014)	(.009)	(.012)	(.018)	(.023)
Hispanic	.022**	.039**	-.018	-.035*	-.038***	-.030*	-.002	-.023	.013	-.016	-.006	-.010
	(.009)	(.015)	(.016)	(.020)	(.014)	(.016)	(.026)	(.030)	(.011)	(.020)	(.051)	(.040)
Black	-.002	.005	-.005	-.003	-.010	-.001	.010	.012	.007	-.002	.014	-.014
	(.006)	(.009)	(.006)	(.011)	(.009)	(.014)	(.013)	(.013)	(.009)	(.015)	(.018)	(.033)
Linear Trend*Hispanic			.002**	.003***	.002**	.002***	.002*	.003***	.000	.001*	.003**	.004***
			(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.002)
Linear Trend*Black			.000	.001**	-.000	.000	.000	.001***	.000	.001***	.000	.002
			(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.001)	(.001)
Linear Trend			.000	.000	.000	.001*	.000	.001	.000	.000	.001	.001
			(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.001)	(.001)
Offender, Legal and District Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Offense Type Codes	Final	Arrest	Final	Arrest	Final	Arrest	Final	Arrest	Final	Arrest	Final	Arrest
Guideline Cells	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
p-value: [Post*B = Post*H]	.041	.079	.013	.139	.001	.0819	.149	.281	.873	.546	.076	.615
Adjusted R-squared	.256	.257	.255	.250	.261	.225	.307	.348	.128	.149	.430	.449
Observations	40,228	26,852	139,096	77,979	76,765	44,345	62,331	33,634	42,075	21,209	20,256	12,425

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 district. The dependent variable throughout is a dummy for whether the case receives a downwards departure. In Column 1, the sample of 40,228 Federal cases (which we call the Natural Experiment or NE sample) 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 the NE sample is restricted to those cases that can be linked back to arrest (Stage 0). In Columns 3-10 we use a different sample – labelled as the Full Pre-Sample– those cases sentenced from 1/10/1998 up until 180 days post 9/11/2001. For those defendants sentenced after 9/11/2001, the last offense was committed prior to 9/11/2001. For those sentenced before, we do not impose a restriction on date of last offense. In even numbered Columns, the sample immediately to the left is restricted to those cases that can be linked back to arrest (Stage 0). The samples (and respective columns) are as follows: all offenses (Columns 3 and 4), all offense except drug offenses (Columns 5 and 6), drug offenses (Columns 7 and 8), all non-marijuana drug offenses (Columns 9 and 10) and marijuana drug offenses (Columns 11 and 12). In all Columns we condition on defendant group (White, Black, Hispanic), whether the case comes up post 9-11, and interactions between the two, and the regular set of controls. In Columns 1,3, 5, 7 and 9 to 7 we control for the primary offense type and a dummy for multiple offenses and guideline cell. In evenly numbered columns we instead control for arrest offense codes, but not guideline cells. Finally in Columns 3-12 we also include ethnic-specific time trends. 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 A11: Bush Appointed US Attorneys

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

	(1) Downward Departure	(2) Downward Departure
Sentenced post 9-11*Hispanic	-.039***	-.049***
	(.010)	(.014)
Sentenced post 9-11*Hispanic*Post-period share under Bush US Attorney	.005	-.002
	(.027)	(.032)
Sentenced post 9-11*Black	-.012	-.012
	(.008)	(.012)
Sentenced post 9-11*Black*Post-period share under Bush US Attorney	.015	.006
	(.020)	(.024)
Sentenced post 9-11	.004	-.000
	(.007)	(.010)
Sentenced post 9-11*Post-period share under Bush US Attorney	-.030	-.043
	(.022)	(.030)
Offender, Legal and District Controls	Yes	Yes
Offense Type Codes	Final	Arrest
Guideline Cells	Yes	No
Implied Sentence Length Impact (H)	.820	1.03
% of Pre 9-11 Ethnic Differential	20.1%	34.5%
p-value: [Post*B = Post*H]	.014	.025
Adjusted R-squared	.257	.257
Observations	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 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 regular set of controls. In Column 1 we control for the primary offense type and a dummy for multiple offenses and the guideline cell. In Column 2 we instead control for arrest offense codes, but not guideline cells. The share of time the district spends in the post period with a Bush appointed US Attorney is measured in deviation from mean. The p-value at the foot of each Column is on the null that the coefficients on the post 9-11 x Black and post 9-11 x Hispanic dummy interactions are equal against a two sided alternative.

Table A12: Treatment Effect Bounds on Judge's Decision Making Around 9-11

Dependent Variable: Downward Departure Granted by Federal Judge
Standard errors in parentheses clustered by 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***
	(.010)	(.003)	(.007)	(.007)	(.007)	(.016)
[Bounds: $\delta_{H(0)}$, $\delta_{H(1)}$]	[-.038, -.036]	[-.013, -.013]	[-.011, -.010]	[-.001, -.001]	[-.013, -.012]	[-.048, -.046]
τ required for coefficient of 0	24.704	-98.256	14.609	-5.009	15.11	-20.731
Sentenced post 9-11*Black	-.013	-.005	.002	-.003	-.007	-.013
	(.008)	(.004)	(.003)	(.003)	(.005)	(.011)
[Bounds: $\delta_{H(0)}$, $\delta_{H(1)}$]	[-.013, -.012]	[-.006, -.005]	[.002, .003]	[-.003, -.003]	[-.007, -.007]	[-.013, -.013]
τ required for coefficient of 0	20.69	-19.444	-1.542	-1200	-25.561	33.198
Sentenced post 9-11	.006	.003	-.000	.001	.002	.003
	(.007)	(.002)	(.002)	(.004)	(.004)	(.009)
[Bounds: $\delta_{H(0)}$, $\delta_{H(1)}$]	[.006, .006]	[.003, .003]	[-.001, -.000]	[.001, .002]	[.002, .002]	[.003, .004]
τ required for coefficient of 0	-17.536	-8.832	-.794	-24.56	-6.311	-3.188
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]	.0411	.0359	.123	.757	.351	.0788
Implied Sentence Length Impact (H)	.736					.889
% of Pre 9-11 Ethnic Differential	18%					29.8%
Unadjusted R-squared	.264	.052	.296	.077	.143	.267
R^{max}=min(1, 1.3 x unadjusted R-squared)	.343	.067	.385	.1	.186	.347
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 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 regular set of controls. In Columns 1 to 5 we control for the primary offense type and a dummy for multiple offenses. 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 A13: Prosecutors' Substantial Assistance Departure Around 9-11

Standard errors in parentheses clustered by district

	(1) Substantial Assistance Departure Granted	(2) Substantial Assistance Departure Granted
Initial charges post 9-11*Hispanic	-.037	-.019
	(.044)	(.062)
Initial charges post 9-11*Black	-.053	-.069
	(.054)	(.076)
Initial charges post 9-11	.035	.057
	(.042)	(.059)
Offender, Legal and District Controls	Yes	Yes
Offense Type Codes	Final	Arrest
Guideline Cell Dummies	No	No
p-value: [Post*B = Post*H]	.673	.159
Adjusted R-squared	.180	.171
Observations	3,612	1,758

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 district. 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 2 we restrict the sample to those cases that can be linked back to arrest (Stage 0). The dependent variable is a dummy for whether the case receives a substantial assistance downwards departure at sentencing. In both Columns the regular set of controls are included. In Column 1 we control for the primary offense type and a dummy for multiple offenses. In Column 2 we instead control for arrest offense codes. 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 A14: Pre-sentence Reports

OLS regression estimates

Standard errors in parentheses clustered by 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
	(.038)	(.224)	(1.72)
Convicted and Sentenced after 9-11 [T2]*Black	.036	-.040	2.02
	(.053)	(.217)	(1.96)
Convicted and Sentenced after 9-11 [T2]	.048	.391***	2.57**
	(.036)	(.135)	(1.29)
Offender, Legal and District Controls	Yes	Yes	Yes
Offense Type Codes	Final	Final	Final
Mean	2.514	19.340	57.6
Standard Deviation	1.741	8.376	64.2
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. 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 in Column 1 (2) is the criminal history score (offense severity score) reported in the pre-sentence report, and in Column 3 it is the lowest sentence in the recommended guideline cell. In all Columns we condition on defendant group (White, Black, Hispanic), whether the defendant is convicted before 9-11 but sentenced after 9-11 [treatment group T1], whether the defendant is convicted and Sentenced after 9-11 [treatment group T2], and interactions between the two treatment dummies and offender group, and the following additional controls: on offender characteristics, we control for dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this non-missing age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a non-missing dependents dummy; on legal controls, we control for a dummy whether information on the defense counsel is missing, and a non-missing dummy interacted with the type of defense counsel (privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements); the primary offense type and a dummy for multiple offenses, 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.

Figure A1: Linkage Rates Across Administrative Data Sets

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

Panel A. Right-to-Left Linkage Rates

Ethnicity	Offense Type	Dyadic Linkage Rate from Stages 1-3 Back to Stage 0	Dyadic Linkage Rate from Stages 4-7 Back to Stages 1-3	Dyadic Linkage Rate from Stage 8 Back to Stages 4-7
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%

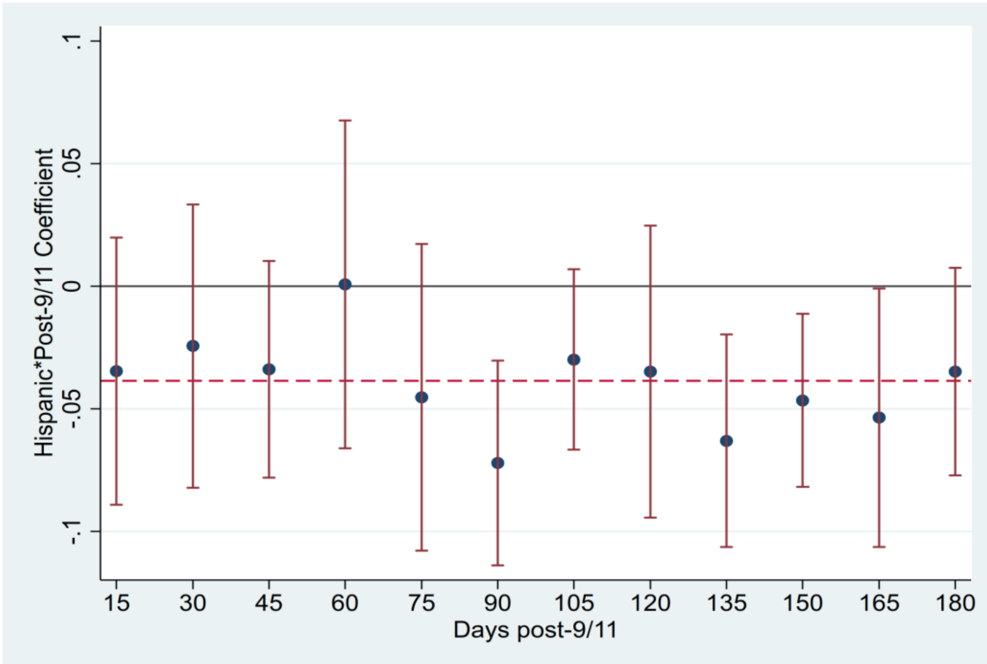
Panel B. Left-to-Right Linkage Rates

Race	Offense Type	Dyadic Linkage Rate from Stage 0 Forward to Stages 1-3	Dyadic Linkage Rate from Stages 1-3 Forward to Stages 4-7	Dyadic Linkage Rate from Stages 4-7 Forward to Stage 8
All	All	38.2%	95.6%	84.3%
White, Black	All	37.8%, 39.3%	95.6%, 95.6%	83.7%, 86.0%
White, Black	Drug	55.1%, 53.8%	86.2%, 87.7%	86.2%, 87.7%
White, Black	Immigration	34.1%, 44.5%	81.7%, 76.2%	81.7%, 76.2%

Notes: We use the Federal Justice Statistics Resource Center (FJSRC) data and the Monitoring of Federal Criminal Sentences (MFCS) data set for our analysis. The FJSRC comprises information gathered from four linked administrative data sources, and we replace the fourth stage USSC data that covers sentencing Stage 8, with the MFCS data (because it is superior to the USSC data that is part of the FJSRC). As described in Rehavi and Starr [2014], the 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) at Stage 8, we then use the Monitoring of Federal Criminal Sentences (MFCS) data set.

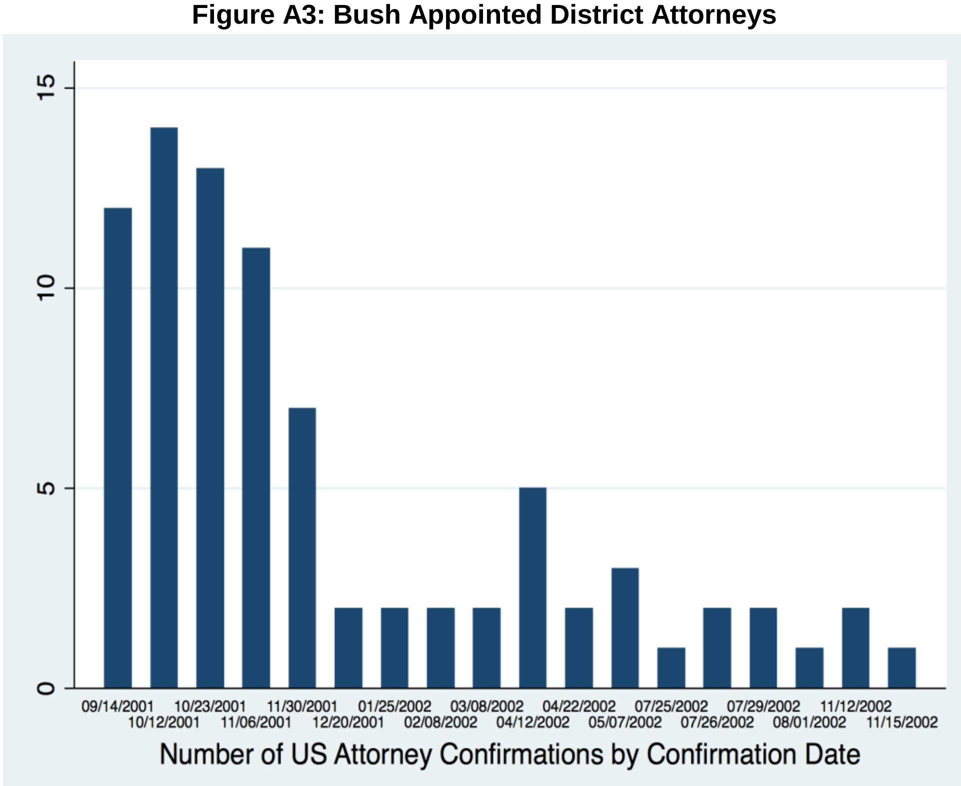
Figure A2: Patriot Act

Hispanics: Non-PA Offenses, Downwards Departure



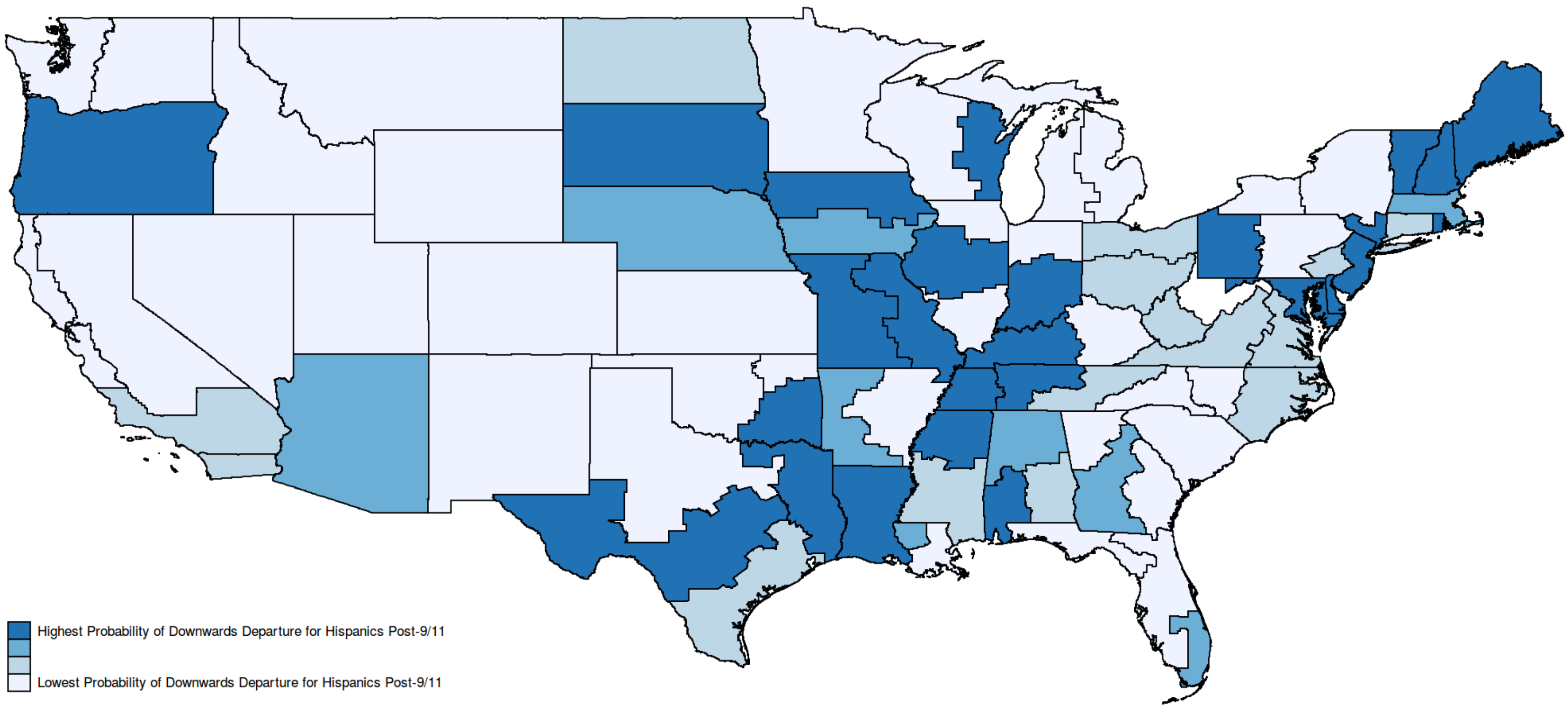
Notes: Figure A3 is based on a restricted version of the Natural Experiment sample. This sub-sample excludes Patriot Act related offenses (Money Laundering and Immigration), resulting in a sample of 32,930 cases (those that come up for sentencing in a six month window either side of 9/11/2001). For those 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. The graphs display output 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 all the 12 post-9/11 periods (and their associated standard error). The first three estimates correspond to before the Patriot Act came into effect; the remaining nine, after. The dashed line shows the corresponding estimate for the NE sample assuming a homogenous post impact. The regression coefficients for the Hispanic*post-9/11 terms are shown. In the regression we condition on the regular set of controls.

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.

Figure A4: Spatial Pattern of Hispanic-White Sentencing Differentials



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