INSIDERS, OUTSIDERS, AND INVOLUNTARY UNEMPLOYMENT: SEXUAL HARASSMENT EXACERBATES GENDER INEQUALITY

Daniel L. Chen and Jasmin Sethi*

Abstract Sexual harassment is perceived to be a major impediment to female labor force participation. We use the random assignment of U.S. federal judges setting geographically-local precedent, and the fact that judges' biographies predict decisions in sexual harassment cases, to document the causal impact of forbidding sexual harassment. Consistent with an insider-outsider theory of involuntary unemployment, but in contrast to a mandated benefits theory of employment protections, pro-plaintiff sexual harassment precedent spurred the adoption of sexual harassment human resources policies, encouraged entry of outsiders, and reduced gender inequality in labor supply and wages among the population. These effects were comparable to the effects of the Equal Employment Opportunity Act and greatest in the construction industry, which was heavily affected by sexual harassment litigation.

Keywords: Gender discrimination, microaggression, trauma, safe spaces, prejudice

JEL codes: J81, J83, K31, J31, J71

*Daniel L. Chen, daniel.chen@iast.fr, Toulouse School of Economics, Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France; dchen@law.harvard.edu, LWP, Harvard Law School; Jasmin Sethi, BlackRock, jasminsethi1@gmail.com. First draft: July 2007. Current draft: August 2016. Latest version available at: http://nber.org/~dlchen/papers/Insiders_Outsiders_and_Involuntary_Unemployment.pdf. We thank Frank Dobbin, Erin Kelly, and Andres Sawicki for sharing data. We also thank research assistants and numerous colleagues with helpful comments at economics faculties at LSE, Paris School of Economics, Duke, Lancaster, George Mason, NBER Law and Economics Conference, Stanford SITE Women and the Economy Conference, Econometric Society, Midwest Economic Association Meetings, and Harvard Labor Economics lunch and law faculties at Yale, University of Chicago, Berkeley, Northwestern, University of Virginia, Duke, Fordham, William and Mary, Florida State University, George Mason, UNC-Chapel Hill, Southwestern, SUNY-Buffalo, Thomas Jefferson, ETH Zurich, American Law and Economics Association Meetings, Law and Society Annual Meetings, Federalist Society Workshop, Kauffman Fellows Conference, Harvard Law School Summer Academic Fellows, and Olin Fellows brown bags. Work on this project was conducted while Daniel Chen received financial support from the European Research Council, Swiss National Science Foundation, Ewing Marion Kauffman Foundation, Institute for Humane Studies, Petrie-Flom Center, Templeton Foundation, Agence Nationale de la Recherche, and the Harvard Law School Summer Academic Fellowship. We also acknowledge joint financial support from the John M. Olin Center for Law, Economics, and Business at Harvard Law School.

1 Introduction

The topic of legal precedent and the justice system's effects on the labor market is understudied in labor economics. For labor economists who study discrimination it is worth noting that much of public policy surrounding labor market discrimination is carried out through the court system, in contrast with other areas of public policy, like education policy, welfare policy, tax policy, health care policy, etc. By studying the Courts of Appeals, we study some of the most important cases that establish precedent, after which the vast majority of potential cases must follow suit or settle. We study sexual harassment law, which is primarily court-made, built on top of anti-discrimination statutes that generally predate the concept of sexual harassment.[1] While sexual harassment has received the attention of one economic model (Basu 2003), it receives no mention in two summaries of the empirical literature on discrimination (Bertrand and Duflo 2016; Neumark 2016) despite the presence of sexual harassment law in U.S. courts since the 1980s, the persistence of labor market gender inequality (Blau and Kahn 2006), survey evidence indicating higher job satisfaction resulting from the presence of sexual harassment law (Newman et al. 2003),[2] and recent debates on cyber-harassment,[3] micro-aggression,[4] trauma, trigger warnings, and safe spaces, as these debates share some similarities with the earlier debates on sexual harassment law.[5]

Sexual harassment is perceived to be a major impediment to female labor force participation.[6] Policymakers in both developed and developing countries have taken steps to address this problem. For example, in India and Mexico, female-only trains and buses have been introduced so women

[1] In the early years of sexual harassment law, the dominant understanding of sexual harassment focused on the sexualized nature of the harassment as the problem (bosses groping subordinates, sexualized insults, sleep with me or you're fired, etc). Schultz (1998) made a conceptual argument that sexual harassment is because of gender. So when men on an almost-all-male factory floor sabotage the tools and equipment of the only woman working there in an effort to push her to quit, that's sexual harassment, even if nothing about it is sexualized.

[2] A recent study finds that 30% of students between 7th and 12th grade in the U.S. experienced sexual harassment in the previous school year and 44% experienced in-person harassment. Negative effects such as absenteeism, poor sleep, and stomachaches were reported by 87% of the harassed students (Hill and Kearl 2011).

[3] As the social norms change, courts continue to struggle with exactly how to define sexual harassment and whether to include, for example, cyber-harassment (Franks 2011), as a form of sexual harassment.

[4] The University of California is considering guaranteeing a right to be free from verbal expressions of intolerance and issued a list of suggested phrases to avoid: http://www.ucop.edu/academic-personnel-programs/_files/seminars/Tool_Recognizing_Microaggressions.pdf.

[5] Recently, the University of Oregon's Bias Response Team summarized common practices on college campuses: Students, faculty, and staff who feel threatened, harassed, intimidated, triggered, microaggressed, offended, ignored, under-valued, or objectified because of their race, gender, gender identity, sexuality, disability status, mental health, religion, political affiliation, or size are encouraged to contact the Bias Response Team, whose goal is to eradicate bias on campus, making Oregon a safer place. Bias is defined as "any physical, spoken, or written act" that targets another person, even unintentionally. The team's posters propose examples of bias incidents: statements that qualify include: "Thanks, sweetie," "I don't see color," "Where are you from? But where are you really from?," "I was joking, don't take things so seriously".

[6] Men are also increasingly likely to bring sexual harassment claims. We focus on female sexual harassment plaintiffs for brevity as the majority of lawsuits involve female plaintiffs.

would face less harassment on their way to work. In the U.S., making the work environment friendlier to women has been one of the most dramatic labor market changes in the past half-century. Yet, the consequences of forbidding harassment on female labor force outcomes remain unknown. Forbidding sexual harassment could have exacerbated gender inequality if the primary function of sexual harassment law was to mandate a benefit, imposing costs on the targeted group through lower wages or lower employment (Acemoglu and Angrist 2001), act as a tax on labor demand (Summers 1989), and make women more costly to hire (Epstein 1995), or it could have ameliorated gender inequality by opening job opportunities in previously harassing work environments, as suggested by an insider-outsider theory of involuntary unemployment (Lindbeck and Snower 1988). Under the insider-outsider theory, sexual harassment exacerbates inequality by closing job opportunities in harassing work environments. Previous evidence for the insider-outsider theory uses cross-sectional analyses of firms (Lindbeck and Snower 2001) or lab experiments (Fehr and Fischbacher 2002). Since sexual harassment laws in most countries are not implemented randomly, and sexual harassment laws may have been a consequence of the increase in female labor force participation, establishing the causal consequences of requiring insiders to stop harassing outsiders is challenging.[7]

This paper presents evidence using randomly assigned judges who have regional jurisdiction setting legal precedent for millions of individuals residing in one of 12 U.S. Circuits and the fact that judges' biographies predict their decisions. Between 1982 and 2002, over 250 Circuit cases addressing sexual harassment were decided in the United States, yielding roughly 126 experiments across 21 years and 12 Circuits. While this number may seem small, analysis of state laws usually examine the impact of up to 50 experiments in the U.S. or 34 in the OECD. Because judicial composition of sexual harassment cases is unlikely to be correlated with subsequent labor market outcomes other than through sexual harassment decisions, the random assignment of judges creates exogenous variation in legal precedent that can be used to estimate the causal impact of court-made sexual harassment law on workplace gender inequality.

In the following sections, we present an analysis of data on human resources policies, the Current

[7]Correspondence studies (Bertrand and Mullainathan 2004) are precluded by the difficulty and plausibility of a research design involving the randomization of everyday events that denigrate individuals because they are members of particular groups (Pierce 1970; Rowe 1981). An alternative research design investigates the role of a phenomenon that is closely related to sexual harassment–sexism or prejudice. This is the belief that minorities should appropriately play only certain roles, or should know their "place" (Charles and Guryan 2008; Charles et al. 2010; Chen 2004). Using a correlational approach, the authors find a positive correlation between prejudicial attitudes and labor market inequality. A third kind of research design employs policy variation. For example, the Equal Employment Opportunity Act increased minority employment shares by 0.5 to 1.1 percentage points per year (Chay 1998). Our research design is closest to the third design, except that our policy variation is random.

Population Survey, and data on sexual harassment decisions in Circuit and District Courts collected by other authors and ourselves. Section 2 discusses theoretical and practical aspects of sexual harassment law, emphasizing how labor lawyers and human resources consultants greatly exaggerated the risk of sexual harassment lawsuits after major Circuit decisions to motivate employers to make human resources policy changes to forbid sexual harassment. Forbidding harassment allows outsider women to enter the labor force as they can now compete for jobs previously dominated by men and the insider women who tolerated sexual harassment. In the insider-outsider theory of Lindbeck and Snower (2001), economic rents captured by insiders are dissipated when they can no longer harass the outsiders, and on the margin, some insiders may leave.

It is important to note that the U.S. Federal Courts, and in particular, the U.S. Courts of Appeals (also known as Circuit Courts), have strong impact on sexual harassment law. Only 2% of Circuit cases successfully appeal to the U.S. Supreme Court, so U.S. Courts of Appeals determine the vast majority of decisions that set legal precedent and 98% of their decisions are final. Their decisions are binding precedent within the Circuit, but not outside. They are persuasive precedent on the state courts within the Circuit. Since the Courts of Appeals judges are revealed substantially after the briefs are filed and after litigation costs have been sunk at both the District and Circuit Court level, very few cases settle when the judges are revealed. We also note that the U.S. Courts of Appeal only handles issues of new law–so we should not expect Circuit cases to influence the decision of a subsequent case that appears in the court.[8] The sample precedents that we discuss are indicative of the rather different legal issue presented: whether to use the reasonable person or reasonable woman standard for what constitutes sexual harassment, or, whether the plaintiff has to prove emotional harm in court. To examine the economic impacts of these kinds of legal regulations whose effects can appear with some delay, we present an empirical framework that allows for a delayed effect of the law, and confirm the results are robust to modifying the basic specification.

Section 3 establishes that the random composition of judicial panels was related to sexual harassment Circuit decisions. We assess randomization by examining whether judicial panel characteristics were correlated with pre-trial characteristics (i.e., the facts of the case–who was involved, the egregiousness of the harm, etc.–as determined by the lower District Courts) and determining how similar the string of judicial panel assignments was to a random string. Judge's politics and gender predict sexual harassment decisions (Farhang and Wawro 2004; Epstein 1995; Peresie 2005), but it has been

[8]In any event, when we run the reduced form relationship between previous judicial assignment and future laws, there is little apparent effect.

argued that judges vote more along party lines than along gender lines (Dixon 2010), that U.S. Presidents who appoint nontraditional (e.g., women) candidates take the opportunity to appoint more ideologically extreme individuals (Asmussen 2011), and that female conservatives exhibit prejudice against females (Eisenman 1991). In our setting, female Republicans were 18 percentage points less supportive of sexual harassment plaintiffs, who were predominantly female, while male Democrats were 13 percentage points more supportive of sexual harassment plaintiffs.[9]

Because judicial composition of sexual harassment judicial panels is unlikely to be correlated with subsequent labor market outcomes other than through sexual harassment decisions, the random panels of judges who decide differently from each other, and in a manner predicted by their biographies, creates exogenous variation in appellate precedent forbidding sexual harassment that can be used to estimate the causal impact of court-made sexual harassment law on gender inequality. We also collect data on judicial biographies to exploit the large variation in judicial decisions that is due to random combinations of biographical characteristics, and employ a sparse model, LASSO, to select among the combinatorial possible number of panel characteristics that predict sexual harassment decisions (Belloni et al. 2012).

In Section 4, we consider whether firms subsequently adopt sexual harassment human resources policies and whether females see improvements in labor market outcomes relative to males in Circuits and years with more precedent that is favorable to sexual harassment plaintiffs. Two-stage least squares estimates using this variation imply that forbidding sexual harassment spurred adoption of human resources policies to address sexual harassment, and females were more likely to be employed and report longer working hours and higher earnings relative to males. These outcomes decrease, to some extent, for males.[10]

The presence of sexual harassment human resources policies grew from 15% in 1982 to 96% by 1997 (Dobbin and Kelly 2007), or roughly 5 percentage points per year. Our estimates indicate that sexual harassment law spurred the adoption of sexual harassment human resources policies equivalent to 37% of the yearly increase.[11] Sexual harassment law also increased female employment shares by an amount equivalent to what the Equal Employment Opportunity Act achieved for black

[9] As shorthand, we will refer to judges appointed by Democratic presidents as "Democrats" and to judges appointed by Republican presidents as "Republicans." Democrats were 13 percentage points more supportive of sexual harassment plaintiffs. Because female Republicans were more conservative than female Democrats were liberal, female judges on net were 3 percentage points *less* supportive of sexual harassment plaintiffs.

[10] The point estimates of the individual lags weakly suggest that changes in human resources policies precede changes in labor market inequality, though we cannot reject the possibility that these changes occured simultaneously.

[11] This calculation of equivalence takes into account the typical composition of sexual harassment cases and typical number of cases.

employment shares (Chay 1998).[12] Whether these effects seem large or small depends on one's perspective, but the interpretation is subject to the usual caveats in the literature (Deaton 2010)—causal effects are sufficient, but not necessary conditions for an outcome.[13]

In Section 5, we turn to three additional pieces of evidence for an insider-outsider theory of harassment and involuntary unemployment. First, sexual harassment law should have larger effects in industries with more sexual harassment, and the effects were greatest in the construction industry, which had the highest rate of sexual harassment claims. Second, forbidding harassment should increase female productivity, and sexual harassment law increased earnings per hour of female labor force participants. Third, some females—those who previously obtained insider rents—may see less benefit, especially as they faced increased competition from previously outsider females, and outsider females gained six times more than insider females.[14]

Throughout the paper, we explore the robustness of our empirical design by changing our controls and specifications. Several findings emerge: Social outcomes were not related to Circuit decisions before they were made; results are robust to controls for Circuit-specific time trends, composition of the pool of judges available to be assigned to panels, and state fixed effects; and results are robust to dropping 1 Circuit at a time, varying the lag structure, and collapsing the data to the Circuit-year level. Inferences are also similar using wild bootstrap and simulations that randomly assign legal variation to another Circuit. As Barrios, Diamond, Imbens, and Kolesar (2012) write, "if the covariate of interest is randomly assigned at the cluster level, only accounting for non-zero covariances at the cluster level, and ignoring correlations between clusters, leads to valid standard errors and confidence intervals," which implies similar standard errors across the different methods of accounting for clustering at the Circuit or Circuit-year level. In Section 6, we also present tests of randomization, assess additional identification concerns, and present robustness to accounting for the potential for litigants to pursue an appeal in response to prior years' outcomes. We elaborate on the use of randomly assigned judges in the lower courts (District Courts) from which an appeal

[12] It is also equivalent to 55% of a standard deviation in median male sexism (Charles et al. 2010).

[13] A defendant who shares the same first initial as a judge receives 8% longer sentences, but this effect only explains 0.03% variance (Chen and Prescott 2016). See also Deaton's NYU "Debates in Development" lecture on the topic, where he describes causal effects as Insufficient but Non-redundant parts of a condition which is Unnecessary but Sufficient (INUS).

[14] We infer insider status by contrasting the experience of wage-earning women in the labor force with women in the general population since we do not have a panel of individuals followed before and after decisions. When the population sample is restricted to those *in* the labor force who report non-zero wages, sexual harassment law exacerbated gender inequality in hours worked, earnings, and management status. Wages of all labor force participants increased, and a portion of insider males exited in response to, or because of, the entrance of outsider females. Taking these factors together suggests that outsider females gained 6 times more than insider females.

arises as an additional source of exogenous variation to control for the presence of a case.

In Section 7, we present alternative theories, such as compensating differentials and selection, and argue that they are less consistent with the data. As such, our results provide new evidence on a debate regarding whether labor market gender inequality is due to unobserved physiological differences, labor market choices, or discrimination (Summers 2005). In particular, our results suggest that sometimes labor market choices, like the decision to participate, can be due to harassment. Moreover, they provide evidence against the view that forbidding sexual harassment is only an employment protection that imposed costs on the targeted group through lower wages or lower employment as has been observed for other social policies (Acemoglu and Angrist 2001; Summers 1989; Epstein 1995). Instead, they provide empirical support for models that seek to enrich the role of prejudice (Charles and Guryan 2008; Charles et al. 2010; Chen et al. 2016a) in contributing to the gender wage gap (Card and DiNardo 2002; O'Neill 2003; Black and Strahan 2001). They are also consistent with previous evidence on the effects of anti-discrimination law (Hellerstein et al. 2002; Neumark and Stock 2006; Eberts and Stone 1985; Chay 1998). Our paper presents the advantage of being based on randomization, where identification is more transparent.

Section 8 concludes with some remarks on directions for future research. Social scientists have long speculated on the relationship between the innovation of rights and socio-economic conditions, and scholarship to date has been unsatisfactory in exploring questions of causality vis-à-vis legal precedent.[15] Methods to evaluate the impact of court-made law may help judges who are interested in the broader empirical consequences of their decisions. Judge Richard Posner has lamented that, "[judicial] opinions lack the empirical support that is crucial to sound constitutional adjudication" (Posner 1998); similarly Justice Breyer remarked, "I believe that a[n] interpretive approach that undervalues consequences, by undervaluing related constitutional objectives, exacts a constitutional price that is too high" (Breyer 2006). While a fair number of studies have employed the random assignment of judges to identify the impact of judicial decisions on the subsequent outcomes of litigants (Kling 2006; Maestas, Mullen, and Strand 2013; Dahl, Kostøl, and Mogstad 2014), our analysis is different in scope. We would like to estimate laws' effects that incorporate general equilibrium response to the law, and we capture the general equilibrium effects because our randomization occurs at the national level rather than the individual level.[16]

[15]Socioeconomic conditions can drive judicial decisions or court decisions may affect society (Chen and Yeh 2014b; Chen 2015).

[16]General equilibrium effects can include (but are not restricted to) capital and labor migration across Circuits. Of course, since our unit of analysis is Circuit-year, we are subject to the usual concern that the Circuit-years are, at

2 Background

2.1 Theory Forbidding sexual harassment would have significant labor market consequences in an insider-outsider theory of harassment and involuntary unemployment (Lindbeck and Snower 1988; Schultz 1990). In this model, while harassment is allowed, outsiders are unable to find jobs even though they are prepared to work for less than the prevailing wages of incumbent workers (insiders). The outsiders cannot underbid insiders; if they did and were to become new employees, insiders would withdraw cooperation and make the work experience of these entrants unpleasant. In other words, insiders would harass the entrants, thereby reducing their productivity. Firms, therefore, find it costly to substitute outsiders for insiders. These harassment and labor turnover costs create economic rents, which the insiders capture via wage setting, and as a result, involuntary unemployment arises. Outsiders are unable to find work even though they would be just as profitable to the firm as the insiders, provided they faced identical conditions of employment. The insiders' harassment activities, however, ensure that conditions are not the same for insiders and outsiders.

Applying this insider-outsider theory in its simplest form assumes that females are outsiders and men are insiders. Under these assumptions, forbidding harassment can increase the employment and wages of outsiders. Insider males would no longer be allowed to engage in harassment activities, thereby raising the productivity of females. Firms would be willing to hire females, so their employment and wages would increase. Of course, not all females are outsiders; hence, some females—those who previously obtained the insider rents—may see less benefit in their employment outcomes, especially as they faced increased competition from previously outsider females.

An alternative view suggests that sexual harassment law may have been a tax on the hiring of women, making it more costly to hire women (Epstein 1995). The law, like an unfunded, mandated benefit, may act like a tax on labor demand (Summers 1989). These theoretical effects are described and analogized to those of accommodation mandates (Jolls 2001). Similar to the unfunded mandate in the Americans with Disabilities Act (ADA), all of the direct and indirect costs of sexual harassment law on firms—from establishing internal infrastructures conducive to complaint to marginal costs associated with each female worker who has some probability of filing a complaint or becoming a litigant—may have been passed onto women and lowered their wages or employment relative to

first glance, analyzed as separate labor markets. However, we note that since our randomization occurs nationally and repeatedly, the general equilibrium effects of labor or capital mobility across Circuit-years are incorporated into our estimates, and these would be the estimates a judge may be interested in. The precedents themselves may eventually spill over across circuits but it is not immediate, and citation analysis indicates that the precedential impact is 10 times greater within the Circuit than outside.

men. While the Equal Pay Act of 1963 (see U.S.C. § 206(d)) makes lower wages for women for the same work illegal, women's wages could be constrained through a failure to promote given the same qualifications or through the offering of fringe benefits packages designed to appeal to some workers but not others.

Assuming that the effects of sexual harassment law were predominantly experienced by women, forbidding harassment may further lower female wages by increasing the supply of female labor by making it more pleasant for women who would be willing to work for lower wages in work environments that previously allowed harassment. If employees value the benefit at cost, the resulting equilibrium will result in the same level of employment but with the full cost reflected in lower wages (Basu 2003). If there are wage rigidities, then the cost of the benefit cannot be reflected in wages, and thus unemployment may result (Summers 1989). Some recent empirical work has found that similar social policies regulating labor markets, such as the Americans with Disabilities Act (ADA) and maternity mandates, had detrimental effects on the groups they were intended to protect and undoing, in part, the redistributive goals of these policies (ADA: Acemoglu and Angrist 2001; Employment protection: Autor et al. 2006; Maternity mandates: Gruber 1994). There are, however, reasons to think that employers are less likely to experience sexual harassment law as an unfunded mandated benefit since it may be difficult to know in advance who is going to be a sexual harassment plaintiff and men could bear some of the cost of the mandated benefit as potential harassers. Whether forbidding sexual harassment exacerbates gender inequality is a priori ambiguous and this motivates our empirical investigation.

At a practical level, sexual harassment law–which is primarily court-made–and the imposition of direct and indirect litigation costs has impacted firm behavior in at least three ways. First, for those cases that were actually litigated, the law resulted in direct litigation costs and potentially large damage awards. Second, the potential for litigation likely led many firms to settle with potential plaintiffs, particularly after such plaintiffs filed complaints with the Equal Employment Opportunity Commission (EEOC) and received a right to sue letter from the agency. Third, as has been documented by sociologists and legal ethnographers (Dobbin and Kelly 2007; Edelman 1992; Edelman et al. 1999), labor lawyers and human resources consultants greatly exaggerated the risk of sexual harassment suit after major Circuit decisions in human resources management publications and in law review articles. Initial surveys reported that 90% of women experienced sexual harassment, when later surveys found that only a small fraction actually did. Rapid change in case law added to the uncertainty. Our own discussions with labor lawyers indicate how fearful employers have been of Title VII suits. Furthermore, sexual harassment cases tend to be personal allegations, which could cause risk-averse employees to change behavior in a way exceeding that of their response to gender discrimination, ADA, or maternity mandate lawsuits, whose resolution can turn on evidence that is likely to be more statistical in nature.

Interviews with hundreds of firms from a representative sample of U.S. businesses with 50 or more employees suggest that these labor lawyers and human resources consultants were quite effective in translating Circuit decisions into human resources policy changes (Dobbin and Kelly 2007). Millions of dollars were spent on training programs and establishing grievance procedures aimed to reduce the risk of lawsuit more than to reduce the incidence of harassment. For example, more grievance procedures were established than maternity leaves, even though the law mandated maternity leaves in no uncertain terms and grievance procedures were not part of a bright-line rule until 1998, when the Supreme Court said grievance procedures greatly reduced the liability faced by firms. Many firms invested in formal grievance procedures, in part because of the advice of personnel experts who believed that instituting formal grievance procedures analogous to those for civil rights violations would help defend against damage awards. Some firms mandated training akin to the diversity training implemented in response to the Civil Rights Act of 1964 to educate their workers about sexual harassment. This last approach was controversial since some firms and attorneys feared that such training might make potential plaintiffs more aware of harassment and, therefore, more likely to sue. In the end, it is an empirical question whether firms responded to sexual harassment precedent. We employ a dataset on human resources policies (Dobbin and Kelly 2007) to investigate the timing of adoption of sexual harassment grievance procedures and policies in response to sexual harassment law. The adoption of human resources policies is, of course, only one channel through which sexual harassment law can have labor market consequences. Women could be encouraged to enter the work force even in the absence of firm policy changes if they felt they were more likely to win in the event of a sexual harassment suit.

An important theoretical critique of anti-discrimination law suggests that profit-maximizing firms have their own incentives not to discriminate (Becker 1968; Epstein 1995). Similar arguments could be applied (and rebutted) in the context of sexual harassment law. Profit-maximizing firms should have their own incentive to forbid harassment and retain outsider workers at higher productivity, so forbidding sexual harassment with its accompanying litigation and human resources cost should have no effect and could be inefficient. However, there are at least three reasons why firms might not forbid harassment on their own (Lindbeck and Snower 1988). First, insider employees may be risk averse; forbidding harassment could change the insider profit-sharing scheme and thereby impose additional risk on insider employees, who then suffer a utility loss. The firm may be unable to compensate them for this loss. Second, an insider cooperates with entrants if his gains, a share of the additional profit resulting from his cooperation, exceed his losses in market power as his wage falls towards his reservation wage. However, this only happens if the firm relinquishes a share of gross profit, something that may make it a net loser compared to other firms. Agency and transaction costs may prevent Coasian bargaining between insiders and the firm. Then the firm has no incentive to implement the new contract. Third, there may be additional sources of labor turnover costs preventing firms from simply replacing all the insiders with outsiders. Entry of firms that hire the outsiders may not occur due to setup costs, capital market imperfections, scarcity of entrepreneurial skills, and reduction of product prices. A legal regime equalizes the playing field across all firms when no firm by itself would have the incentive to forbid harassment. As economic rents captured by insiders are dissipated when they can no longer harass the outsiders, so on the margin, some insiders may leave.[17]

2.2 Design of Study The correlation between court-made law and economic outcomes is generally difficult to interpret since the causality may run in both directions and the relationship may reflect omitted variables. Momentous judicial decisions may be caused by rather than be causes of political or socioeconomic changes.[18] In addition, with cross-fertilization across different areas of legal doctrine, if different, but related (i.e., correlated), doctrinal areas have independent effects on employment outcomes, social changes may be misattributed to one legal rule when many legal rules are changing simultaneously. The ideal research design randomly assigns individuals or firms to different legal rules to help resolve uncertainty aboout consequential impacts of law (Abramowicz et al. 2011). Thankfully, judges do not randomize their decisions, but the random assignment of judges provides a close approximation.

[17]Sexual harassment is not limited to the young: 25% of sexual harassment complaints are filed by individuals over 35 (Coles 1986). The definition of insiders and outsiders may be continually evolving, and individuals may feel a duty to exclude outsiders (Chen 2006, 2010; Chen and Schonger 2013). Voice-based snap judgments of perceived masculinity based solely on the introductory sentences of lawyers arguing in front of the Supreme Court of the United States negatively predict outcomes in the Court. From an economic perspective, correlations between malleable advocate characteristics and high-stakes outcomes in the United States Supreme Court should not persist unless law firms believe that that masculine voices are more likely to win or have a preference for masculine-sounding lawyers (Chen et al. 2016b).

[18]The diaries of Supreme Court Justices suggest that social factors influenced their decisions (Klarman 2004).

We exploit a natural experiment where pro-plaintiff sexual harassment precedent varies randomly by Circuit and over time due to the random assignment of judges to Circuit panels. Five aspects of the U.S. judicial system are important for the development of this identification strategy. First, the United States has a common law system where American judges not only apply the law but also make the law: Judge's decisions in current cases become precedent for use in decisions in future cases in the same court and in lower courts of the same jurisdiction.[19] The lower courts are the District Courts, which are the courts of general jurisdiction and hold trials. There are three layers of courts in the U.S. federal judicial system. When trial cases are appealed, they go to appellate courts, referred to as Circuit Courts. Only 2-3% of Circuit cases are appealed again to the U.S. Supreme Court. Therefore, Circuit Courts are quite active in shaping law. They handle the vast majority of cases deciding issues of new law and provide new interpretations or distinctions of pre-existing precedents or statutes.

Second, Circuit Courts are only to hear cases presenting new legal issues (only 10-20% of District Court opinions are appealed). Circuit Courts are credited with continually finding new distinctions with which to expand or contract the space under which an actor would be found liable (Gennaioli and Shleifer 2007). The appendix provides a timeline of the major developments in sexual harassment doctrine. For examples of how doctrinal shifts could make it easier for subsequent sexual harassment plaintiffs to bring and win suit, consider in *Ellison v. Brady,* the replacement of a "reasonable person" standard with a "reasonable woman" standard for determining whether sexual harassment occurred, and in *Harris v. Forklift Systems, Inc.*, the elimination of the requirement to prove psychological harm in courts. These doctrinal developments are typical of the cases in the Circuit Courts. Moreover, cases representing big legal changes occur throughout the time period.

Third, there are twelve Circuit Courts, each in charge of a geographic region (comprising 4-9 states) of the United States, known as a Circuit (See Figure 1). As such, Circuit decisions in one Circuit do not establish precedent that other Circuits must necessarily follow. When Circuits choose to adopt the precedent of another Circuit, it is typically with some delay. For example, a new case bringing the same issue of law must be filed in a District Court, appealed to the Circuit Court, decided upon, and have an opinion issued before the doctrine becomes binding precedent in the new Circuit.

[19] Circuit decisions are binding precedent only in the Circuit of the court delivering the opinion; that is, the District Courts within a Circuit and the Circuit Court itself must follow the precedent set by the Circuit Court's prior decisions. This premise has been checked empirically using a research design with the random assignment of judges (Chen et al. 2014a).

Fourth, judges are randomly assigned in Circuit Courts, three judges to each case. Some judges take a reduced caseload, but all are randomly assigned by a computer algorithm. Section 6 reports tests of randomization. Judges' names are typically not revealed to the litigating parties until after they file their briefs–sometimes only a few days before the hearing, if there is a hearing. Because judges are appointed for life by the U.S. president, and a Circuit can have 20-40 judges in the pool available to be assigned, we observe that the number of possible combinations of judges or combinations of judicial demographic characteristics on a panel is very large. This variation renders a large amount of experimental variation with which to identify causal effects.

Fifth, identification also hinges on whether sexual harassment cases assigned to judges with different background characteristics do in fact have different outcomes. Experience, legal philosophy, or group identity may cause judges to view issues in different ways.[20] A number of papers have documented the effect of judges' demographic background on sexual harassment cases (Farhang and Wawro 2004; Epstein 1995; Peresie 2005). The prior literature, focusing on gender and politics, indicates that judges vote more along party lines than by gender and that female conservatives are more hardline conservative on gender issues (Eisenman 1991; Dixon 2010). Presidents who appoint nontraditional (e.g., women) candidates take the opportunity to appoint more ideologically extreme individuals than they would otherwise (Asmussen 2011). The fact that some judges are more extreme than others allow us to construct a setting akin to a randomized experiment in the establishment of precedent across different regions of the United States.[21]

[20] A large literature documents that judicial decisions are correlated with their biographical characteristics or experience. Some papers show evidence that may have a causal interpretation, for example, that Circuit Court judges behave more partisan before Presidential elections, that wins and losses of sports games affect judicial decisions, that matching first initials with a randomly assigned defendant affects decisions, and that decisions are affected by decisions on recent cases (Berdejó and Chen 2014; Chen and Spamann 2014; Chen and Prescott 2016; Chen 2016; Chen et al. 2015). Further, the explanatory power of these factors persist after employing the best prediction models of judges' decisions (Barry et al. 2016; Chen and Eagel 2016; Chen et al. 2016c,d).

[21] Our research design can be further clarified by the following illustration. Consider a Circuit with a high proportion of judges who are extreme. The designation of extreme is not meant to imply that extreme judges are less reasonable or accurate in their decisions. Rather, extreme judges are those who tend to, for example, vote for plaintiffs more so than the average judge. The empirical strategy does not rely on cases getting more extreme judges in this Circuit as opposed to another Circuit, which could be different for a variety of reasons. Rather, the strategy relies on the fact that, from year to year, the proportion of cases for a particular case category in this Circuit that are assigned extreme judges varies in a random manner. The idiosyncratic variation is not expected beforehand since judicial assignment is not revealed to parties until very late in the appellate process and after each litigant's briefs are filed. In the years when an unexpectedly high number of extreme judges are assigned to panels for a particular case category, the proportion of cases for that case category that will result in pro-plaintiff precedent is also high. Even though we cannot ask any particular Circuit to randomize their decisions, a randomized control trial is in effect created through the random assignment of judges who interpret the facts and the law differently. Random variation in the assignment of Circuit judges is an attractive instrument for a number of reasons. The random assignment of judges is exogenous and unexpected. It varies in both the cross-section and the time-series, so it does not rely on strong assumptions about the comparability of different regions (e.g., Circuits) and years. The enormous variation in legal decisions due to the judicial panel composition also makes the empirical design an ideal setup to study the consequences of law. See Chen and Yeh (2013) for an extended, non-technical discussion of the empirical intuition.

We use this variation to identify the effects of sexual harassment law on human resources policies and gender inequality in the labor market. We should expect to see an effect of Circuit judicial decisions if judges follow precedent and Circuit decisions on the margin make it easier for subsequent plaintiffs to bring and win suit. For example, shifting from a reasonable person standard to a reasonable woman standard in defining sexual harassment would make it easier for subsequent plaintiffs to bring and win suit. We might then expect firms and individuals to respond to Circuit decisions, whether through newspaper publicity, advocates, lawyers, or information consultants greatly exaggerating the risk of suit after major Circuit decisions.[22]

2.3 Testing the Empirical Predictions of the Theory The most robust prediction of the insider-outsider theory, which sets it in contrast with a world where sexual harassment law is solely a mandated benefit whose costs are borne on those whom the law is intended to help, is that outcomes are likely to differ in Circuit-years that issue pro-plaintiff sexual harassment jurisprudence in a manner that favors females (relative to males). To test this, we will simply compare the outcomes in Circuit-years with pro-plaintiff and pro-defendant sexual harassment jurisprudence and perform robustness checks to confirm that the difference seems to be due to sexual harassment jurisprudence.

To link legal precedent with labor force participation, we parameterize $\phi(y) = e^{\rho Law_{ct}} A_{ct}$. Consider a general dynamic growth equation for labor force participation:

$$(1) \quad y_{ct} = A_{ct} + \alpha_1 y_{ct-1} + ... + \alpha_n y_{ct-n} + \rho_0 Law_{ct} + \rho_1 Law_{ct-1} + ... + \rho_n Law_{ct-n} + \varepsilon_{ct}$$

allowing participation to depend on n lags of past participation and adding an error term.

We assume that A_{ct} evolves according to:

$$(2) \quad \Delta A_{ct} = g_c + \gamma_0 Law_{ct} + ... + \gamma_n Law_{ct-n}$$

This allows both current and lagged jurisprudence to affect the growth rate of A.

Substituting ΔA_{ct} into a first differenced version of y_{ct} yields a dynamic panel estimation equation of the form:

$$(3) \quad \Delta y_{ct} = g_c + \alpha_1 \Delta y_{ct-1} + ... + \alpha_n \Delta y_{ct-n} + \gamma_0 Law_{ct} + ... + \gamma_n Law_{ct-n} + \rho_0 \Delta Law_{ct} + \rho_1 \Delta Law_{ct-1} + ... + \rho_n \Delta Law_{ct-n} + \Delta\varepsilon_{ct}$$

The "level effects" of law on output appear through ρ. The "growth effects" of law appear through γ. Rewriting the ΔLaw terms as Law terms yields:

[22] Circuit decisions have been shown to affect stock market outcomes (Araiza et al. 2014). Another study finds that Supreme Court cases from 1999–2013 collectively moved $140 billion in stock prices (Katz et al. 2015).

$$(4) \qquad \Delta y_{ct} = g_c + \alpha_1 \Delta y_{ct-1} + ... + \alpha_n \Delta y_{ct-n} + (\gamma_0 + \rho_0) Law_{ct} + (\gamma_1 + \rho_1 - \rho_0) Law_{ct-1} + ...$$
$$+ (\gamma_n + \rho_n - \rho_{n-1}) Law_{ct-n} - \rho_n Law_{ct-n-1} + \Delta\varepsilon_{ct}$$

Relabeling the coefficients on Law yields:

$$(5) \qquad \Delta y_{ct} = g_c + \alpha_1 \Delta y_{ct-1} + ... + \alpha_n \Delta y_{ct-n} + \sum_{j=0}^{n+1} \beta_j Law_{cj} + \Delta\varepsilon_{ct}$$

To find the growth effect, consider $\Delta y_{ct-j} = \Delta y$ and $Law_{cj} = Law$. Solving yields:

$$(6) \qquad \Delta y_c = \frac{g_c}{1 - \alpha_1 - ... - \alpha_n} + \frac{\sum_{j=0}^{n+1} \beta_j}{1 - \alpha_1 - ... - \alpha_n} Law_c$$

so that the growth effect of jurisprudence is simply $\frac{\sum_{j=0}^{n+1} \beta_j}{1-\alpha_1-...-\alpha_n}$, which is identical to $\frac{\sum_{j=0}^{n} \gamma_j}{1-\alpha_1-...-\alpha_n}$ since the ρ terms all cancel.

As we find variation in Law that is randomly assigned, we focus on $\alpha_j = 0$ for all j and estimate:

$$(7) \qquad g_{ct} = \theta_c + \theta_t + \sum_{n=0}^{L} \beta_{t-n} Law_{ct-n} + \Delta\varepsilon_{ct}$$

where θ_c are Circuit fixed effects, θ_t are time fixed effects, and Law_{ct} is a vector of annual jurisprudence with up to L lags included. This equation captures the growth effect of law, e.g., the effect of law on aspects of the economy, such as human resource policies, that influence labor force participation.

The growth equation allows separate identification of level effects and growth effects through the examination of β_j. In particular, both effects influence the growth rate in the initial period. The difference is that the level effect eventually reverses itself. For example, a jurisprudential shock may affect social activism, but after a few periods, activism returns to normal. By contrast, the growth effect appears during the jurisprudential shock and is not reversed. A failure to innovate in one period leaves the Circuit permanently further behind. The growth effect is identified as the summation of the jurisprudential effects over time. Following the convention in the growth literature, we are interested in the distributed lag effect and test for joint significance of the lags.

Another way to think about our research design is that laws are not likely to have an immediate impact. Firms may need time to adjust to a new legal regime; alternatively, the effects of a law change may fade as expectations adjust. To estimate the delayed effects of the law, we estimate a distributed lag specification. Thus we need to also control for the presence of a Circuit case, and we exploit variation in the presence of a Circuit case that comes from the random assignment of District Court judges. We assume that the demographic characteristics of District judges are correlated with the likelihood of an appeal (since District judges' characteristics are correlated with the likelihood of reversal (Haire, Songer, and Lindquist 2003; Sen 2015; Barondes 2010; Steinbuch 2009), but not with socioeconomic outcomes other than through the courts.[23] Then, our specification should be invariant to the number of lags and leads. The use of leads serves as a check of the identification strategy, namely, whether the assignment of judges to sexual harassment cases may be endogenous to other factors that correlate with socio-economic outcomes.

Our data on labor force outcomes are repeated-cross sections, so we will estimate the growth equation in levels. This means that a level effect is inferred from a persistent set of lags and a growth effect from a set of lags that grow over time. In contrast, convergence in institutions and behavior is inferred from lags that fade over time. To sum up, the simple insider-outsider model of involuntary unemployment contrasts the view of sexual harassment law as a mandated benefit that hurts those intended for help. If forbidding sexual harassment is a mandated benefit, the resulting equilibrium will result in the same level of employment but with the full cost reflected in lower wages, or if there are wage rigidities, unemployment may result. If forbidding sexual harassment prevents insiders from excluding outsiders, labor market outcomes increase along both margins.

2.4 Data Collection Our empirical analysis draws on several sources of data on sexual harassment cases—an established dataset as well as our own data collection of Circuit and District Court cases.

The first dataset is from Boyd et al. (2010), which codes case characteristics, such as the presence of certain fact patterns and legal issues for a subset of Title VII discrimination claims in the Chicago Judges Project data (Sunstein et al. 2006). We use this data to perform a randomization check, namely, if pre-trial case characteristics are correlated with the assignment of the Circuit Court judge.

Our second dataset is composed of our own collection of cases from 1982 to 2002 of all sexual harassment cases brought in an employment context. An earlier draft of the paper used the Chicago Judges Project data, which coded sexual harassment cases from 1995-2002, and found similar results. We follow the data collection and coding method in the Chicago Judges Project and extend the analysis back to 1982 to examine the impact of the totality of sexual harassment cases. The EEOC first issued guidelines with the term "sexual harassment" in 1980. Before 1982, there were very few

[23] To be sure, we are not identifying the causal effects of the presence of a Circuit case. The assignment of a District judge can affect the Circuit panel's decision. We want to identify a portion of the law that is not coming from other social trends or areas of law, and this portion comes from the random assignment of federal judges.

sexual harassment cases, so our search method obtains very few cases in the early years. More than one case per year is needed because our main specification includes year fixed effects. Figure 2 plots the growth in sexual harassment cases, number of pro-plaintiff decisions, and number of pro-defendant decisions during this time period. The coding of the vote follows the coding method used in the Chicago Judges Project database: pro-plaintiff vote if the plaintiff was afforded any relief. Table I indicates that on average, there were 0.996 sexual harassment panels per Circuit-year for a total of 251 cases. A sizeable portion of Circuit-years, 51%, had no sexual harassment panel. Roughly 67% of the decisions were pro-plaintiff.

Our third dataset comprise of District Court cases, collected using a similar methodology as for the Circuit Court collection. We searched Westlaw using "((SEX! +2 DISCRIMINATION) (GENDER +2 DISCRIMINATION)) & (SEX! +2 HARASSMENT)". This resulted in 3,754 cases between 1982 and 2002.

Our fourth dataset is information on judges' characteristics from the Appeals Court Attribute Data, District Court Attribute Data,[24] Federal Judicial Center, and our own data collection.[25] We filled in missing data by searching transcripts of Congressional confirmation hearings and other official or news publications on Lexis. The average Circuit-year had 18.50 judges available for assignment to panels. The expected number of male Democrats per seat was 0.33. We calculate the expectations based on the composition of the Circuit pool of judges available to be assigned in any Circuit-year assuming that all judges have an equal probability of assignment. We weight senior judges according to the frequency that a typical senior judge sits on cases. Additional summary statistics are displayed in Table I.

Our fifth dataset on courts involves administrative data from the Administrative Office of the U.S. Courts (AOC) and PACER filings on District Court cases[26] to obtain judge identities that are missing in the AOC data. We use this data to perform a test related to the identification assumption.

Our data on outcomes comes from two established sources, the Current Population Survey and a dataset of workplaces interviewed on their human resource policies dating back to 1965.

[24] http://www.cas.sc.edu/poli/juri/attributes.html

[25] Variables include: geographic history, education, occupational history, governmental positions, military service, religion, race, gender, and political affiliations. Raw data on religion come from Goldman (1999). Sisk's data are available at http://courseweb.stthomas.edu/gcsisk/religion.study.data/cover.htm. Judges whose religions remained missing or unknown were coded as having no publicly known religious affiliation. As political and social issues divide along religious lines, it is reasonable to hypothesize that judges from different religions come to different conclusions in court cases (Chen and Lind 2007, 2014). One of the coauthors collected these additional religious and personal attributes (Chen and Yeh 2014a).

[26] Sixteen years of Public Access to Court Electronic Records are available on open source sites for 33 Districts.

TABLE I

SUMMARY STATISTICS

Circuit-Year Level	Mean (Standard Deviation)
Number of Judges	18.504 (7.356)
Number of Panels	0.996 (1.471)
Proportion of Circuit-Years with No Panels	51%
Proportion of Pro-Plaintiff Decisions when Circuit-Year has Panels	67%
Expected # of Females per Seat	0.117 (0.081)
Expected # of Democrats per Seat	0.407 (0.121)
Expected # of Female Republicans per Seat	0.035 (0.040)
Expected # of Male Democrats per Seat	0.326 (0.119)
N (Circuit-years)	252

First, we obtained labor market outcomes from the Merged Outgoing Rotation Groups (MORG) Current Population Survey (CPS), which contains individual employment outcomes, including weekly earnings, amount of time worked, employment status, and management status. The primary variable of interest is the distinction between no-employment (including non-labor force participants) vs. part- or full-time employment. Non-labor force participants include discouraged workers. According to the Bureau of Labor Statistics, "Persons who are neither employed nor unemployed are not in the labor force. This category includes retired persons, students, those taking care of children or other family members and others who are neither working nor seeking work." We assume these discouraged workers are outsiders in an insider-outsider theory of involuntary unemployment.

Earnings are adjusted to be in 2000 real terms. In addition, logs are taken of 1+earnings. We use hours worked last week instead of usual weekly hours because usual weekly hours are not consistently available (Autor et al. 2005). We recode the number of hours worked as zero for individuals who are not in the labor force or not employed. We also recode their log earnings as zero. Some studies drop these individuals and find a smaller difference in labor market outcomes between men and women. We include individuals not in the labor force since the insider-outsider theory of involuntary unemployment considers the entire population, not just those who are in the labor force or employed, since discouraged workers may leave the labor market. We later restrict the sample to labor force participants for the analysis of insiders. The designation of labor force participants as insiders does not mean we assume that all labor force participants are harassing non-participants.

Management status is a binary indicator for whether an individual has an administrator, official, public administration, executive, or other management-related occupation. Occupation is available for about 90% of the unemployed and 33% of those not in the labor force, about 10% of which are managerial. Respondents may interpret this question as being about their previous job. The CPS also contains demographic controls, including age, sex, race, marital status, educational attainment, and the geographic location of the individual, which allows us to match the individual's state of residence to the Circuit having legal jurisdiction. We weight our analysis with CPS-provided weights.

We measure the presence of firm-level sexual harassment policies in a national sample of 389 workplaces interviewed in 1997 on the history of human resources practices dating back to 1965. This data was collected by Dobbin and Kelly (2007). They report a response rate of 56% but found no bias in survey response along observable characteristics: establishment size, organization size, subsidiary status, branch/headquarter status, region, and female chief executive. There is no variable indicating whether the firm had offices in multiple Circuits. Estimated effects would be biased towards 0 to the extent that firms make HR decisions at a level that transcends Circuit boundaries.

2.5 Specification Thanks to the randomization built into the U.S. Federal Courts, the basic empirical specification is straightforward. The effect of sexual harassment jurisprudence is obtained by comparing the means of the outcomes of interest after pro-plaintiff precedent and pro-defendant precedent. Note that this difference is not an estimate of the comparison between pro-plaintiff precedent and no precedent. The employer and employee activities can be different than what they would have been if there was no precedent whatsoever. We are mostly interested in the effect of pro-plaintiff, rather than pro-defendant precedent, *when there is a precedent, as this would be the effect of the decision of a judge deliberating on the case.*

Denoting Y_{ict} as the value of the outcome of interest for individual (employer or employee) i and Law_{ct} as a dummy equal to 1 if the decision was pro-plaintiff, this estimate is simply:

$$(8) \quad E[Y_{ict}|Law_{ct} = 1] - E[Y_{ict}|Law_{ct} = 0] \equiv \beta_1.$$

For female labor market outcomes, we expect $\beta_1 < 0$ for at least one dimension of wages or employment if forbidding sexual harassment is a mandated benefit whose costs are borne out on women. However, for an insider-outsider theory, we expect $\beta_1 > 0$ for wages and employment if insiders are prevented from lowering the productivity of outsiders.

Since we are interested in effects over time, we specify a distributed lag and extend our specification to include the presence of a decision, $1\left[M_{ct-n} > 0\right]$. We focus on five years of lags and one lead ($n = -1$ to 5) and vary the lag structure for robustness. M is the number of cases, which is typically 0 or 1 (so typically $Law_{c(t-n)}$ is 1 (100% pro-plaintiff) or 0 (100% pro-defendant)). Since our outcome data is yearly, we take the average law measurements in each Circuit-year. We also considered weighting our estimates by the number of cases in a Circuit-year, where weights are the geometric mean of $M_{c(t-n)} + 1$ over the distributed lag. The statistical significance of the results increases, so we present more conservative estimates without weighting.

$$(9) \quad g_{ct} = \theta_c + \theta_t + \sum_{n=0}^{L} \beta_{1t-n} Law_{ct-n} + \sum_{n=0}^{L} \beta_{2t-n} 1\left[M_{ct-n} > 0\right] + \varepsilon_{ct}$$

Analogizing to coin flips, β_1 captures the effect of the heads-or-tails coin flip (pro-plaintiff vs. pro-defendant precedent), $\beta_1 + \beta_2$ captures the effect of the heads coin flip and the presence of the coin (pro-plaintiff precedent vs. no decision), and β_2 captures the effect of the tails coin flip and the presence of the coin (pro-defendant precedent vs. no decision).

We examine several outcomes: presence of sexual harassment policy, employment status, hours worked, earnings, and management status.[27] We will test whether there are differences in outcomes by gender by running the regression:[28]

$$(10) \qquad Y_{ict} = \theta_c + \theta_t + \textstyle\sum_{n=0}^{L} \beta_{3t-n} Law_{ct-n} + \sum_{n=0}^{L} \beta_{4t-n} 1\left[M_{ct-n} > 0\right] + \sum_{n=0}^{L} \beta_{5t-n} Law_{ct-n} * F_{ict} + \sum_{n=0}^{L} \beta_{6t-n} 1\left[M_{ct-n} > 0\right] * F_{ict} + \eta X_{ict} + \varepsilon_{ict}$$

F_{ict} represents a dummy indicator for being female. We expect $\beta_5 < 0$ if females are adversely affected by sexual harassment law and $\beta_5 > 0$ if females benefit from sexual harassment law. We report the average and the individual coefficients. We consider $n = 0$ as a lag because some statistics refer to calendar year. Most of the effects appear with some slight delay so excluding $n = 0$ in joint significance tests does not affect our results.

In principle, we have 252 experiments (across 21 years and 12 Circuits). With random treatment assignment, adding controls can add precision to the estimates if the controls are strong predictors of the outcomes. We show that our main estimates are robust in Table VI to the inclusion or exclusion of:

- Circuit-fixed effects, θ_c, and time-fixed effects, θ_t;
- Circuit-specific time trends to allow different Circuits to be on different trajectories with respect to outcomes;
- A vector of observable unit characteristics, X_{ict}, such as age, gender, educational attainment, and race, which each enter as dummies with the exception of age;
- and time-varying Circuit-level controls, such as the characteristics of the pool of judges available to be assigned in Circuit c and time $t - n$.

The firm-level analysis uses the same controls as Dobbin and Kelly (2007): number of employees, percent of women at location, sex of top executive named, percent women among named executives, age of establishment, dummies for manufacturing, service, and trucking, regional unemployment rate, and percent female in industry.

As Barrios, Diamond, Imbens, and Kolesar (2012) write, "if the covariate of interest is randomly

[27]Because management status includes previous job, we only examine this outcome when examining the sample of labor force participants.

[28] We do not interact the legal precedent with the gender indicator for our firm-level analysis.

assigned at the cluster level, only accounting for non-zero covariances at the cluster level, and ignoring correlations between clusters, leads to valid standard errors and confidence intervals," so we expect to see similar results whether clustering standard errors at the Circuit or Circuit-year level (Barrios et al. (2012) show that random assignment of treatment addresses serial and spatial correlation across treatment units). Our results are also unaffected using the standard approach with U.S. data, 50 state clusters (see Table VI). Table VI also reports regressions where each datapoint corresponds to 1 experimental subject (21 years x 12 Circuits). We check results using randomization inference that assigns the legal variation to another Circuit and using wild bootstrap. In addition, the leads specification serves as an omnibus check for insufficient adjustment of standard errors. A prior draft clusters standard errors at the Circuit-year level and this draft clusters at the Circuit level, and the results are unchanged.[29]

Court decisions requiring female friendly work environments may have occurred as a consequence of the labor market entry of females, leading to an upward bias in OLS estimates. On the other hand, if labor market entry of females is too low, courts may be more likely to require female friendly work environments, leading to a downward bias in OLS estimates. Therefore, ascertaining a causal effect from judicial decisions to social trends is difficult without idiosyncratic variation in judicial decisions. Since Law_{ct} and ϵ_{ict} may be correlated due to uncontrolled-for social trends or other legal developments that correlate both with Law_{ct} and outcomes Y_{ict}, we develop an instrumental variable for Law_{ct} using judges' biographical characteristics. We use biographical characteristics because the number of sexual harassment cases yields sharp demographic effects but not judge-specific effects, as there are too few cases with which to identify judge-specific effects. In this paper, we do not analyze heterogeneity in terms of, e.g., early or late pro-plaintiff decisions, or the compounding effects of, e.g., pro-plaintiff decisions, and only analyze the average effects.

We drop the subscript n to ease the following exposition. Let N_{ct} be the number of judges assigned to sexual harassment panels who are female Republican. Figures 3A and 3B illustrate the identification strategy. The jagged line displays $\frac{N_{ct}}{M_{ct}}$ and the smooth line displays $\mathbf{E}(\frac{N_{ct}}{M_{ct}})$ in each of the 12 Circuits. The smooth lines indicate the underlying variation in judge-specific characteristics within Circuits over time. The jagged line indicates the random year-to-year variation in female Republicans per seat. We estimate how outcomes respond to idiosyncratic variation in $\frac{N_{ct}}{M_{ct}}$.

More formally, let $p_{ct} = \frac{N_{ct}}{M_{ct}} * \mathbf{1}[M_{ct} > 0]$, i.e., defined to be 0 when $\mathbf{1}[M_{ct} > 0] = 0$. Then: $\mathbf{E}[(p_{ct} -$

[29] We thank our NBER discussant Bentley MacLeod for recommending that we cluster at the Circuit level.

$\mathbf{E}(p_{ct}))\varepsilon_{ict}] = \mathbf{Pr}[M_{ct} > 0]\mathbf{E}[(p_{ct} - \mathbf{E}(p_{ct}))\varepsilon_{ict}|M_{ct} > 0] + \mathbf{Pr}[M_{ct} = 0]\mathbf{E}[(p_{ct} - \mathbf{E}(p_{ct}))\varepsilon_{ict}|M_{ct} = 0] = 0$. Next, $\mathbf{E}[(p_{ct} - \mathbf{E}(p_{ct}))\varepsilon_{ict}] = \mathbf{E}(p_{ct}\varepsilon_{ict}) - \mathbf{E}[\mathbf{E}(p_{ct})\varepsilon_{ict}] = \mathbf{E}(p_{ct}\varepsilon_{ict}) - \mathbf{E}(p_{ct})\mathbf{E}(\varepsilon_{ict}) = \mathbf{E}[p_{ct}\varepsilon_{ict}]$. Thus, p_{ct} and $p_{ct} - \mathbf{E}(p_{ct})$ both serve as valid instruments. Notably, as Table VI shows, the results are unaffected by controlling for $\mathbf{E}(p_{ct})$. This draft presents estimates using the following identification assumption (i.e., moment condition) for causal interpretation: $\mathbf{E}[\frac{N_{ct}}{M_{ct}}\varepsilon_{ict}|\mathbf{E}(\frac{N_{ct}}{M_{ct}}), 1\,[M_{ct} > 0]] = 0$. Previous drafts obtained similar results using $\mathbf{E}[N_{ct}\varepsilon_{ict}|\mathbf{E}(\frac{N_{ct}}{M_{ct}}), \mathbf{1}\,[M_{ct} > 0]\,, M_{ct}] = 0$, which looks at the number of pro-plaintiff decisions controlling for the number of decisions, and $\mathbf{E}[N_{ct}\varepsilon_{ict}|\mathbf{E}(\frac{N_{ct}}{M_{ct}}), \mathbf{1}\,[M_{ct} > 0]\,, Q_{ct}] = 0$, which controls for the size of the court docket and checks if pro-plaintiff vs. pro-defendant decisions had opposite-signed effects. All lags and leads of Law_{ct} are instrumented for in the actual implementation. As standard, we also lag and interact the instruments as we lag and interact Law_{ct}.

For example, if a Circuit-year has a higher fraction of female Republicans assigned, the precedent that year will be that much more pro-defendant. We are interested in the subsequent effects of that precedent on employer and employee outcomes. We are able to do so because the identity of a judge on a case does not directly affect outcomes except through the precedent.[30] It is also worth noting that for our legal domain, pro-plaintiff vs. pro-defendant is the materially relevant legal doctrine. A very interesting feature of the institutional setting, however, is that it is possible to assess this hypothesis in conjunction with another. If there are other aspects of sexual harassment doctrine that are sensitive to judges' biographical characteristics, and if these other aspects of sexual harassment doctrine affect economic outcomes, we should observe correlations between 2SLS residuals and Circuit-year biographical characteristics not used in the first stage. They are not, which suggests that either the pro-plaintiff vs. pro-defendant dimension of these sexual harassment cases is the primary channel through which sexual harassment jurisprudence has an effect, or other aspects of sexual harassment jurisprudence are not polarized along judicial demographic characteristics, which reduces the concerns that we are picking up something other than pro-plaintiff vs. pro-defendant.[31]

Including lags that are important predictors of the outcome improves statistical precision. However, litigants' decisions to appeal may respond to previous years' legal decisions, so controlling for $\mathbf{1}[M_{ct} > 0]$ may bias the coefficient for Law_{ct}; and the bias is more severe for more distant lags while

[30] The court decision is taken as precedent by subsequent courts. Also, judge identity do not predict stock prices at the time of resolution controlling for the manner in which the case was resolved and judge identity do not predict stock prices at the moment that judges are revealed (Badawi and Chen 2014).

[31] Moreover, each Circuit Court decides many thousands of cases per year. The effects we observe are not likely to reflect the effects of other decisions made by other judges as the composition of judicial panels in other legal areas should be uncorrelated. The identity of the judge plausibly affects economic outcomes through the legal precedent alone.

being non-existent for the most advanced lead. We assess whether this potential endogeneity is a significant concern. We describe our instrumental variables strategy for $\mathbf{1}[M_{ct} > 0]$, which relies on the composition of District Court judges assigned to sexual harassment cases. We considered two identification strategies and implemented the first one due to data availability.

In the first strategy, $w_{ct} = \frac{\sum_{d=1}^{J} K_{cdt} * \left(\frac{L_{cdt}}{K_{cdt}}\right)}{\sum_{d=1}^{J} K_{cdt}}$, where K_{cdt} denotes the number of cases filed in District *court* d within Circuit c at time t.[32] L_{cdt} denotes the number of judges with a particular characteristic assigned to cases. The intuition is that assigning District judges who are disproportionately appealed, for whatever reason, leads to the presence of a case in the Circuit, $\mathbf{1}[M_{ct} > 0]$. Note that the terms in the numerator need $K_{cdt} > 0$. An approximation is to define $K_{cdt} * \left(\frac{L_{cdt}}{K_{cdt}}\right)$ as 0 if $K_{cdt} = 0$. Then, the instrument can be constructed if $\sum_{d=1}^{J} K_{cdt} > 0$, i.e., the denominator is non-zero.

In the second definition, $\tilde{w}_{ct} = \sum_{\tilde{d}=1}^{\tilde{J}} K_{\widetilde{cdt}} * \left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right)$, where $\tilde{d}$ denotes District *courthouse* or *court*.[33] A District Court has several courthouses (also referred to as Divisions) and in some District Courts, random assignment is at the courthouse level. The second definition does not rely on the exogeneity of the location of next sexual harassment case, $K_{\widetilde{cdt}}$. It also does not rely on $K_{\widetilde{cdt}} > 0$, though it still relies on $\sum_{\tilde{d}=1}^{\tilde{J}} K_{\widetilde{cdt}} > 0$, i.e., a non-missing value for each Circuit c time t.[34] It is not possible to merge courthouse location for most of the District Court cases.[35] This results in $\tilde{w}_{ct}$ being undefined in over 50% of Circuit-years and thus requiring another dummying out for missing strategy.[36] We opted to use w_{ct}. We restrict our attention to cases authored by District Court judges and exclude recommendations by magistrate judges because litigants cannot directly

---

[32] J goes from 5 to 13 depending on the District.

[33]

Fact The Law of Iterated Expectations (LIE) implies $E\left(K_{\widetilde{cdt}} * \left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right) * \epsilon_{ct}\right) = 0$.

Proof Using LIE, $E\left(K_{\widetilde{cdt}} * \left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right) * \epsilon_{ct}\right) = E\left(E\left[K_{\widetilde{cdt}} * \left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right) * \epsilon_{ct} \middle| K_{c\tilde{d}t}\right]\right)$. Rearranging results in: $E\left(K_{c\tilde{d}t} E\left[\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right) * \epsilon_{ct} \middle| K_{c\tilde{d}t}\right]\right)$. Again by LIE: $E\left[\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right) * \epsilon_{ct} \middle| K_{c\tilde{d}t}\right] = E\left[E\left(\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right) * \epsilon_{ct} \middle| \epsilon_{ct}, K_{c\tilde{d}t}\right) \middle| K_{c\tilde{d}t}\right]$. Rearranging once again: $E\left[\epsilon_{ct} E\left(\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right) \middle| \epsilon_{ct}, K_{c\tilde{d}t}\right) \middle| K_{c\tilde{d}t}\right]$. The expression $\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)$ is the deviation of the ratio of judge assignment characteristics from the mean, so it should be independent of ϵ_{ct} and $K_{c\tilde{d}t}$. Therefore, $E\left(\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right) \middle| \epsilon_{ct}, K_{c\tilde{d}t}\right) = 0$.

[34] It is innocuous to make the approximation $K_{\widetilde{cdt}} * \left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}} - E\left(\frac{L_{\widetilde{cdt}}}{K_{\widetilde{cdt}}}\right)\right)$ as 0 if $K_{\widetilde{cdt}} = 0$.

[35] We tried to link the courthouse information from the AOC via the docket number.

[36] When we use $\tilde{w}_{ct}$, we include a dummy for missing values in $\tilde{w}_{ct}$ and define $\tilde{w}_{ct}$ to be 0 when it would otherwise be missing.

appeal a magistrate judge's recommendation (28 U.S.C. § 636(c)(1)).[37]

We find that the coefficients for Law_{ct} are very similar, so we do not focus our analysis on District Court cases. It would seem to suggest that the willingness to appeal was not significantly affected by prior years' assignment of Circuit judges to sexual harassment cases, at least in a manner that would lead to significant changes in our estimates of the effect of Law_{ct}. In Section 6, we present empirical assessments of randomization and potential issues that arise because our data comprise published opinions: settlement, publication, and strategic use of keywords or citation. Theoretically, settlement is not an issue: judges are revealed after litigants file their briefs in Circuit Courts, sometimes only a few days before the hearing, which gives little opportunity and incentive for settlement upon learning the identity of the panel.[38] Second, unpublished cases do not have precedential authority.[39] Third, we collected additional data to examine how similar the string of actual panel assignments is to a random string.[40]

3 The Effect of Judge Identity on Court Decisions

Judges vote more along party lines than along gender lines (Dixon 2010), and presidents who appoint nontraditional (e.g., women) candidates take the opportunity to appoint more ideologically extreme individuals (Asmussen 2011). Female conservatives exhibit prejudice against females (Eisenman 1991), and female Republicans were 18 percentage points less likely to vote pro-plaintiff ($p < 0.05$). This estimate is reported in Table II Panel A Column 4. Male Democrats were 13 percentage points more likely to vote pro-plaintiff ($p < 0.01$) as can be seen in Column 5. Democrats were also 13 percentage points more likely to vote in favor of sexual harassment plaintiffs ($p < 0.01$) in Column 2. Because female Republicans were more conservative than female Democrats were liberal, female judges on net were 3 percentage points *less* likely to vote in favor of sexual harassment

[37] We also tested the following sensitivity check. With many endogenous variables and many instruments, there is a danger of overfitting with instruments from the wrong year. In robustness checks, we use the contemporaneous instruments to predict $Law_{c(t)}$ and $\mathbf{1}[M_{ct} > 0]$ and use the fitted values as instrumental variables in 2SLS.

To see how, suppose:

$$Y_{ict} = \beta_{10} Law_{c(t)} + \beta_{11} Law_{c(t-1)} + ... + \varepsilon_{ict}$$

Let the first stage be: $L_{c(t)} = Z_0 \Pi_0 + u_0$, where $Z_0 = \left[\ p_{c(t)}\ \right]$ and $L_{c(t-1)} = Z_1 \Pi_1 + u_1$, where $Z_1 = \left[\ p_{c(t-1)}\ \right]$. Set $\hat{X} = [\ \hat{L}_{c(t)} \quad \hat{L}_{c(t-1)} \quad ... \quad \hat{L}_{c(t-j)}\ ]$for $j = 0, 1, ...,$ where $\hat{L}_{c(t-j)} = Z_j \hat{\Pi}_j = Z_j (Z_j' Z_j)^{-1} Z_j' L_{c(t-j)}$.

Observe that $\hat{\beta} = (\frac{\hat{X}'X}{n})^{-1} \frac{\hat{X}'Y}{n} = \beta + (\frac{\hat{X}'X}{n})^{-1} \frac{\hat{X}'\varepsilon}{n}$. Let $\hat{Q} = (\frac{\hat{X}'X}{n})$, then $\sqrt{n}(\hat{\beta} - \beta) = \hat{Q}^{-1} \frac{\hat{X}'\varepsilon}{\sqrt{n}} . \frac{1}{\sqrt{n}} \hat{X}_j' \varepsilon = \frac{1}{\sqrt{n}} \frac{X_j z_j}{n} (\frac{z_j' z_j}{n})^{-1} z_j' \varepsilon = \hat{\Gamma} \sqrt{n} \frac{z_j' \varepsilon}{n}$. Since $\sqrt{n} \frac{z_j' \varepsilon}{n} \rightarrow N(0, \Phi_j)$, so $\sqrt{n}(\hat{\beta} - \beta) \rightarrow N(0, V), V = Q^{-1} \Gamma \Phi \Gamma Q^{-1}$.

[38] Notably, settlement rates were unaffected by the D.C. Circuit announcing their judges earlier (Jordan 2007).

[39] Unpublished cases are routine cases. Judicial ideology has no impact on unpublished cases (Keele et al. 2009) nor the decision to publish (Merritt and Brudney 2001).

[40] To see the random strings test assessed in Chen (2013) as an omnibus test: Suppose female Republicans are more likely to publish cases or use the keywords that generate our sample, then we should expect female Republicans to appear autocorrelated relative to a set of simulated strings.

plaintiffs ($p > 0.1$, Column 1) and were 8 percentage points less likely to vote pro-plaintiff when also controlling for party of appointment ($p < 0.1$, Column 3). We use male Democrat as our instrumental variable to mirror female Republican as instrumental variable. The point estimates change little when both are included (Column 6) and are essentially unaffected with the inclusion of Circuit and year fixed effects in Column 7, the inclusion of the expected number of male Democrats per seat and the expected number of female Republicans per seat in Column 8, and both sets of controls in Column 9.

In Panel B, we examine the verdict at the case-level. An additional female Republican on a three-judge panel reduced the chances of a pro-plaintiff decision by 18 percentage points ($p < 0.05$) in Panel B Column 1.[41] An additional male Democrat on a three-judge panel increased the chances of a pro-plaintiff decision by 11 percentage points ($p < 0.01$) in Column 2, magnitudes that are similar to the judge-level votes. The point estimates and statistical significance change little with the inclusion of both variables (Column 3), fixed effects (Column 4), the expected judge type per seat (Column 5), and both sets of controls (Column 6).

In Panel C, we examine these relationships at the Circuit-year level for the 124 Circuit-years with at least 1 case in Columns 1 to 3. The estimates indicate that an additional female Republican on a three-judge panel decreased the proportion of pro-plaintiff decisions by 34 percentage points in Column 1 ($p < 0.01$) and an additional male Democrat on a three-judge panel increased the proportion of pro-plaintiff decisions by 17 percentage points in Column 2 ($p < 0.01$). The coefficient at the Circuit-year level can be different from the coefficient at the case level due to the fact that cases can bunch up unevenly across Circuit-years. Column 4 adds a dummy indicator for whether there were cases, $1\left[M_{ct} > 0\right]$ and dummies out for missing values.[42] The number of observations increases to the complete time-frame of 252 Circuit-years.[43] The point estimates and F-statistics are completely unaffected by dummying out for missing values, which indicates that the increase in sample size from dummying out for missing values is not driving results.[44] The point estimates remain similar with the inclusion of both variables (Column 6), fixed effects (Column 7), the expected

[41] Female Republicans per seat varies from 0 to 1, so we divide the coefficient by 3.

[42] Dummying out for missing values is a standard approach in the literature. We redefine $Law_{ct} = Law_{ct} * \mathbf{1}\left[M_{ct-n} > 0\right]$, i.e., defined to be 0 when $\mathbf{1}\left[M_{ct-n} > 0\right] = 0$, and regress it onto the instrument $p_{ct} = \frac{N_{ct}}{M_{ct}} * \mathbf{1}\left[M_{ct-n} > 0\right]$, while including $\mathbf{1}\left[M_{ct-n} > 0\right]$ as a control.

[43] Circuit 11 was not founded until 1981, so Circuit 11 has 6 fewer observations than the other Circuits. Circuit 11 was created by splitting it off from Circuit 5; Circuit 5's decisions before this split are considered binding precedent in Circuit 11. We account for this split in our analyses by assigning pre-1981 precedent in Circuit 5 to observations in Circuit 11.

[44] The R-square mechanically increases, but the F-statistic for strength of first stage does not.

TABLE II

First Stage: Relationship Between Pro-Plaintiff Sexual Harassment Courts of Appeals Decisions and Composition of Sexual Harassment Panels, 1982-2002

Panel A: Judge Level	Outcome: Pro-plaintiff Decision								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.0332		-0.0785+						
	(0.0394)		(0.0416)						
Democrat		0.131**	0.143**						
		(0.0358)	(0.0371)						
Female Republican				-0.183*		-0.145+	-0.122+	-0.112+	-0.117
				(0.0757)		(0.0739)	(0.0616)	(0.0612)	(0.0659)
Male Democrat					0.130**	0.121**	0.110**	0.122**	0.111*
					(0.0361)	(0.0362)	(0.0353)	(0.0377)	(0.0373)
Circuit-year controls	N	N	N	N	N	N	Fixed Effects	Expectations	Both
F-statistic	0.71	13.48	7.81	5.87	12.89	7.88	6.06	6.45	5.02
N	752	752	752	752	752	752	752	752	752
R-sq	0.001	0.018	0.021	0.006	0.016	0.019	0.112	0.026	0.112

Panel B: Case Level	Outcome: Pro-Plaintiff Decision					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Republicans	-0.553*		-0.434+	-0.418+	-0.379+	-0.407
per Seat	(0.200)		(0.217)	(0.211)	(0.208)	(0.234)
Male Democrats		0.322**	0.286*	0.285*	0.314**	0.298*
per Seat		(0.0876)	(0.0933)	(0.103)	(0.0967)	(0.110)
Circuit-year controls	N	N	N	Fixed Effects	Expectations	Both
F-statistic of instruments	7.66	13.48	7.91	6.17	7.33	4.61
N	251	251	251	251	251	251
R-sq	0.019	0.032	0.043	0.126	0.046	0.128

Panel C: Circuit-Year Level	Outcome: Proportion of Pro-Plaintiff Decisions								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Republicans	-1.015**		-0.848**	-1.015**		-0.848**	-0.810**	-0.804**	-0.845**
per Seat	(0.224)		(0.256)	(0.224)		(0.256)	(0.217)	(0.221)	(0.199)
Male Democrats		0.513**	0.445*		0.513**	0.445*	0.459*	0.455*	0.494**
per Seat		(0.164)	(0.165)		(0.164)	(0.165)	(0.150)	(0.155)	(0.145)
Circuit-years with no cases	Dropped	Dropped	Dropped	Dummied	Dummied	Dummied	Dummied	Dummied	Dummied
Circuit-year controls	N	N	N	N	N	N	Fixed Effects	FE, Expect	FE, Trends
F-statistic of instruments	20.53	9.84	22.46	20.53	9.84	22.56	63.92	41.23	32.95
N	124	124	124	252	252	252	252	252	252
R-sq	0.067	0.080	0.125	0.605	0.611	0.630	0.669	0.669	0.706

Panel D: Analysis Level Firm HR Policies	Outcome: Proportion of Pro-Plaintiff Decisions						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Republicans	-1.284*		-0.874	-0.871	-0.593	-0.600	-0.528
per Seat	(0.568)		(0.508)	(0.510)	(0.428)	(0.409)	(0.436)
Male Democrats		0.663**	0.535**	0.534**	0.509**	0.523**	0.479**
per Seat		(0.174)	(0.130)	(0.131)	(0.0990)	(0.0943)	(0.102)
Circuit-years with no cases	Dummied	Dummied	Dummied	Dummied	Dummied	Dummied	Dummied
Firm-level controls	N	N	N	Y	Y	Y	Y
Circuit-year controls	N	N	N	N	Fixed Effects	FE, Expect	FE, Trends
F-statistic of instruments	5.12	14.52	10.30	10.15	13.95	16.01	11.09
N	5584	5584	5584	5584	5584	5584	5584
R-sq	0.647	0.663	0.679	0.680	0.733	0.735	0.751

Panel D: Analysis Level Individual MORG CPS	Outcome: Proportion of Pro-Plaintiff Decisions						
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Female Republicans	-1.163**		-0.973**	-0.971**	-0.839**	-0.849**	-0.767**
per Seat	(0.284)		(0.283)	(0.281)	(0.167)	(0.174)	(0.182)
Male Democrats		0.524**	0.441*	0.442*	0.467**	0.466**	0.517**
per Seat		(0.157)	(0.151)	(0.151)	(0.147)	(0.152)	(0.155)
Circuit-years with no cases	Dummied	Dummied	Dummied	Dummied	Dummied	Dummied	Dummied
Individual-level controls	N	N	N	Y	Y	Y	Y
Circuit-year controls	N	N	N	N	Fixed Effects	FE, Expect	FE, Trends
F-statistic of instruments	16.77	11.22	16.96	17.11	50.31	32.32	38.12
N	5429470	5429470	5429470	5418564	5418564	5418564	5418564
R-sq	0.627	0.628	0.652	0.652	0.691	0.691	0.711

Notes: Significant at +10%, *5%, **1%. Heteroskedasticity-robust standard errors are in parentheses. Observations are clustered at the Circuit level. Fixed effects are dummy indicators for Circuit and for year. Expectations are the expected number of Female Republicans and Male Democrats per seat. Trends are Circuit-specific time trends. In Panels C and D, proportions of pro-plaintiff decisions and judicial type per seat during Circuit-years with no cases are defind to be 0.

judge type per seat (Column 8), and Circuit-specific time trends (Column 9).

The F-statistic is 21 for the female Republican instrument and 9.8 for the male Democrat instrument. When both are included, the joint F is 22.5, which is above the weak instruments threshold. The addition of fixed effects, expectations, and Circuit-specific time trends increases the F-statistic on the instruments up to 63.9. The R-square does not change much with the inclusion of controls in Columns 7-9 and does not change at all when expectations (Column 8) are added on top of the fixed effects (Column 7).

Panels D reports the first stage analyzed at the level of our firm-level analysis.[45] The coefficients differ from those in Panel C because the numbers of firms per Circuit-year is not constant. The F-statistic is 10 in Column 3 and increases up to 16 with the inclusion of firm-level controls, fixed effects, and additional Circuit-year controls. Panel E reports the first stage analyzed at the level of our individual-level analysis. The point estimates are more similar to those in Panel C. The F-statistic is 17 in Column 3 and goes up to 50 with the additional controls.

To check whether our linear specifications miss important aspects of the data, Figures 5A and 5B present nonparametric local polynomial estimates of the first stage.[46] The first stage effect is not due to outliers. The results show a monotonically increasing relationship for male Democrats and decreasing one for female Republicans.

Table III presents a falsification test of the instrument and shows that the proportion of pro-plaintiff decisions is not related to the number of male Democrats per seat or the number of female Republicans per seat in the one or two years before and after the true instrument. Typically, Circuit Courts handle cases that present new legal issues, so we should not expect future proportions of pro-plaintiff decisions to respond to past assignment of treatment.

This last finding does not mean that Circuit Court cases have no precedent.[47] Rather, this finding is consistent with only the hardest, most novel legal issues being decided in Circuit Courts. This institutional setting means that $\sum_{n=0}^{\infty} \beta_{1n} = \sum_{n=0}^{\infty} TOT_{ct}^{n} = \sum_{n=0}^{\infty} LATE_{ct}^{n}$. In the $LATE$ setting, compliers are cases that respond to assignment of the judge. These are the cases with enough legal ambiguity such that differences in perspective yield different decisions. Some cases

[45] Bertrand et al. (2004) recommend an analysis at the individual level to be able to control for individual-level covariates.

[46] Estimation proceeds in two steps. In the first step, we regress the proportion of decisions that were progressive on Circuit and year fixed effects and we regress the instrument on the same. Next, we take the residuals from these two regressions and use a nonparametric local polynomial estimator to characterize the relationship between the instrument and progressive decisions.

[47] Chen et al. (2014a) examines cases filed before the Circuit Court decision but resolved after the Circuit decision and finds that lower court judges follow legal precedent.

have no strong legal precedent. In these cases, judges seek guidance (Posner 1998; Breyer 2006; Abramowicz et al. 2011). Always-takers and never-takers are cases without ambiguity.[48] These are the easy cases that do not respond to the assignment of the judge because of strong precedent. Then recall: $TOT \equiv E[Y_{1i} - Y_{0i}|R_i = 1] = E[Y_{1i} - Y_{0i}|R_{1i} > R_{0i}]Pr(R_{1i} > R_{0i}|R_i = 1) + E[Y_{1i} - Y_{0i}|R_i = 1]Pr(R_{1i} = R_{0i} = 1|R_i = 1)$, where R_i indicates whether i received treatment, $R_{1i} > R_{0i}$ indicates whether individual i is a complier and $R_{1i} = R_{0i} = 1$ denotes an always-taker, under the assumption of no defiers. The strong first stage suggests that any bias that results from the presence of non-compliers is likely to be small. The bias from non-monotonicity is given by $\frac{Pr[Defier]}{Pr[Complier]-Pr[Defier]}(\beta^{Complier} - \beta^{Defier})$, which is likely to be small even if there are non-compliers, since the magnitude of the first stage is large (the denominator grows and the numerator shrinks) using biographical characteristics as the instrument.

We also employed LASSO to instrument for Law_{ct}. It is worth noting that a large number of biographical characteristics serve as valid instruments, which results in a weak instruments problem if we used them all.[49] Our approach has the benefit of a surfeit of experimental variation. The statistician must trade-off, however, between the power of the first stage regression with the addition of instruments (Angrist and Imbens 1995; Stock and Yogo 2005). Choosing among a large number of instruments quickly becomes a challenging statistical issue. We use LASSO (Least Absolute Shrinkage and Selection Operator) to address the issue of instrument selection (Belloni et al. 2012). LASSO addresses a problem with OLS, which has low bias but large variance: change the data a bit, and you get different subsets of covariates deemed important. LASSO is a sparse model that automatically sets small estimated coefficients to 0 to reduce model complexity. As such, LASSO is an effective tool in selecting instrumental variables from available judicial biographical characteristics. Since using too many instruments effectively renders the instruments weak, such a selection device is necessary. LASSO minimizes the sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. Because of the nature of this constraint, it tends to produce some coefficients that are exactly 0 and hence gives interpretable models. Intuitively, it

[48] Ambiguity causes polarization in perceptions (Baliga et al. 2013).

[49] The thirty characteristics are: Democrat, male, male Democrat, female Republican, non-White, Black, Jewish, Catholic, No religion, Mainline Protestant, Evangelical, BA received from same state of appointment, BA from a public institution, JD from a public institution, having an LLM or SJD, elevated from District Court, born in the 1910s, 1920s, 1930s, 1940s, 1950s, appointed when president and congress majority were from the same party, ABA score, above median wealth, appointed by president from an opposing party, prior federal judiciary experience, prior law professor, prior government experience, previous assistant U.S. attorney, and previous U.S. attorney. Adding panel-level interactions (e.g., fraction of judge seats assigned to Democrats multiplied by fraction of judge seats assigned to Blacks) yielded a total of 450 possible instruments.

TABLE III

Falsification Test of Instrument: Relationship Between Pro-Plaintiff Sexual Harassment Courts of Appeals Decisions and Composition of Sexual Harassment Panels in Other Years, 1982-2002

Circuit-Year Level	Outcome: Proportion of Pro-Plaintiff Decisions_t			
	(1)	(2)	(3)	(4)
Female Republicans	-0.771*	-0.888**	-0.828**	-0.714*
per Seat_t	(0.249)	(0.226)	(0.178)	(0.249)
Male Democrats	0.442*	0.375*	0.433*	0.460**
per Seat_t	(0.146)	(0.147)	(0.154)	(0.142)
Female Republicans	-0.226	-0.292		
per Seat_{t-1}	(0.320)	(0.374)		
Male Democrats	0.0486	0.00643		
per Seat_{t-1}	(0.120)	(0.107)		
Female Republicans		-0.0758		
per Seat_{t-2}		(0.263)		
Male Democrats		-0.0948		
per Seat_{t-2}		(0.181)		
Female Republicans			-0.301	-0.174
per Seat_{t+1}			(0.173)	(0.203)
Male Democrats			-0.196	-0.0741
per Seat_{t+1}			(0.134)	(0.129)
Female Republicans				-0.0118
per Seat_{t+2}				(0.198)
Male Democrats				0.127
per Seat_{t+2}				(0.113)
Circuit-years with no cases	Dummied	Dummied	Dummied	Dummied
Circuit-year controls	Fixed Effects	Fixed Effects	Fixed Effects	Fixed Effects
N	240	228	240	228
R-sq	0.660	0.665	0.702	0.713

Notes: Significant at +10%, *5%, **1%. Heteroskedasticity-robust standard errors are in parentheses. Observations are clustered at the Circuit level. Fixed effects are dummy indicators for Circuit and for year. Proportions of pro-plaintiff decisions and judicial type per seat during Circuit-years with no cases are defind to be 0.

is a data penalty for having too many covariates. Additional covariates, whose coefficients are not 0, are more likely to violate that constraint. We have a very large number of valid instruments, but not every demographic characteristic matters. LASSO enhances statistical precision when using the random assignment of Circuit and District Court judges. LASSO is theoretically optimal under certain conditions including sparsity described by Belloni et al. (2012). The use of the LASSO-selected instruments also provides a check of over-identification.

The joint F statistic increases to 33 for the Circuit-year level and up to 130 for the analysis-level.[50] The results are unaffected by the inclusion of instruments for $\mathbf{1}[M_{ct} > 0]$. It is worth repeating that including the effects of $\mathbf{1}[M_{ct} > 0]$ renders additional counterfactuals of interest. The focus of the paper is on the difference between pro-plaintiff vs. pro-defendant precedent, but one may also be interested in the difference between pro-plaintiff vs. *no* precedent. The U.S. Federal Courts effectively randomize both the coin flip (pro-plaintiff vs. pro-defendant) and the presence of the coin itself (pro-plaintiff vs. no decision vs. pro-defendant).[51]

4 Estimating the Impact of Sexual Harassment Law on Gender Inequality

4.1 Human Resources Policies We should expect to see an effect if the following three assumptions are met: judges follow precedent; on the margin, pro-plaintiff decisions in Circuit Courts make it easier for subsequent sexual harassment plaintiffs to bring and win suits (see the appendix for a list of major doctrinal developments shifting the legal landscape under which plaintiffs may win); and firms respond to Circuit decisions. While the first two assumptions are less contested, the third bears further examination.

Table IV displays the effect of pro-plaintiff sexual harassment jurisprudence on the adoption of sexual harassment human resources policies. Firms were 5.7 percentage points more likely to have a sexual harassment policy on average in each year in the five years after a pro-plaintiff precedent than after a pro-defendant precedent (the IV estimate is reported in Column 2 and the average lag is reported at the bottom). The most statistically significant effects are found in the first three years after the precedent. This can be seen in Figure 6A, which displays the 95% confidence intervals. The lead effects are always insignificant. Similar findings are found with the LASSO instruments

[50] At the Circuit-year level, the LASSO procedure selected the following three instruments: the interaction between the number of male Democrats per seat and the number of judges born in the 1920s per seat, the interaction between the number of female Republican per seat and the number of judges having an LLM or SJD per seat, and the interaction between the number of female Republican per seat and the number of judges with above median wealth per seat.

[51] A detailed analysis of assumptions behind these alternative conterfactuals and why they might be interesting is presented elsewhere (Chen et al. 2014b; Chen and Yeh 2014b).

TABLE IV

THE EFFECT OF SEXUAL HARASSMENT LAW ON HUMAN RESOURCES SEXUAL HARASSMENT POLICY

	Presence of Sexual Harassment Policy			
	(1)	(2)	(3)	(4)
Proportion Pro-Plaintiff	-0.00108	0.0374	0.0279	0.0353
Circuit Decisions$_{t+1}$	(0.0142)	(0.0252)	(0.0233)	(0.0354)
Proportion Pro-Plaintiff	0.00984	0.0231	-0.0167	0.0282
Circuit Decisions$_t$	(0.0135)	(0.0316)	(0.0443)	(0.0291)
Proportion Pro-Plaintiff	0.0157	0.0408+	0.0455*	0.0223
Circuit Decisions$_{t-1}$	(0.0215)	(0.0244)	(0.0224)	(0.0687)
Proportion Pro-Plaintiff	0.0291	0.0210	0.0978+	-0.0102
Circuit Decisions$_{t-2}$	(0.0188)	(0.0507)	(0.0503)	(0.0542)
Proportion Pro-Plaintiff	0.0616**	0.117**	0.0612	0.112*
Circuit Decisions$_{t-3}$	(0.0153)	(0.0405)	(0.0413)	(0.0510)
Proportion Pro-Plaintiff	0.0494*	0.0542	-0.00609	0.146**
Circuit Decisions$_{t-4}$	(0.0162)	(0.0545)	(0.0497)	(0.0502)
Proportion Pro-Plaintiff	-0.00874	0.0517	0.0533+	0.103
Circuit Decisions$_{t-5}$	(0.0154)	(0.0439)	(0.0290)	(0.0677)
Controls	Y	Y	Y	Y
IV	N	Y	Lasso IV	Lasso IV, District IV
Mean dependant variable	0.543	0.543	0.543	0.646
Average lag	0.029	0.057	0.050	0.075
Joint F lags	9.34	24.09	30.20	13.70
Joint F leads	0.01	2.20	1.44	1.44
N	4014	4014	4014	2617
R-sq	0.260	0.259	0.257	0.121

Notes: Significant at +10%, *5%, **1%. Heteroskedasticity-robust standard errors are in parentheses. Observations are clustered at the Circuit level. These regressions include firm-related controls (number of employees, percent of women at location, sex of top executive named, percent women among named executives, age of establishment, dummies for manufacturing, service, and trucking, regional unemployment rate, and percent female in industry), Circuit fixed-effects, year fixed-effects, Circuit-specific time trends, and a dummy for whether there were no cases in that Circuit-year. The regressions also include all the lags of the dummy for there being no cases, sexual harassment law, and the instruments as well as the interactions of these variables with gender.

(Column 3) and when both the LASSO and District-level instruments are used (Column 4).[52]

From 1982 to 1997, an average of 54% of establishments reported having sexual harassment policies, growing from 15% in 1982 to 96% in 1997. Our baseline point estimate of 5.7 percentage points means that, after taking into account how often there is a case and in which direction a case is typically resolved (half of the Circuit-years had a case and two-thirds of the cases were pro-plaintiff while one-third were pro-defendant), then the impact of sexual harassment law is equivalent to 37% of the yearly increase.

To understand this calculation in more detail, since the typical Circuit-year had sexual harassment panels half of the time and 0.67 proportion of decisions were pro-plaintiff (Table I), multiplying 0.67,

[52]These results are robust to collapsing the data to Circuit-year means and using LASSO and District-level instruments.

0.5, and 5.7 percentage points suggests that, since the genesis of sexual harassment law, its average impact in a typical Circuit-year caused firms to be 1.9 percentage points more likely to have a sexual harassment policy. Assuming a linear 81 percentage point increase in sexual harassment policy during the 16 years, firms were 5.1 percentage points more likely to have a sexual harassment policy in any given Circuit-year. Under these assumptions, pro-plaintiff sexual harassment law appears to have played an important role in the change of human resources policies to address sexual harassment, equivalent to 37% of the yearly change. This back-of-the-envelope calculation is subject to the usual caveats that causal effects are sufficient, but not necessary conditions for an outcome (Deaton 2010).[53]

What we have reported in Table IV are estimates that *condition* on the presence of a Circuit case, that is, we identify the causal effect of a pro-plaintiff decision where the counterfactual is a pro-defendant decision. To examine the *unconditional* effect (where the counterfactual is no decision), i.e., what if sexual harassment law had never existed in the courts whatsoever, we need to consider the average coefficient on the presence of a Circuit case, which is -0.043 for Column 2, -0.038 for Column 3, and -0.027 for Column 4 (not displayed). We focus on Column 4, which uses the random assignment of District judges to construct another set of instruments to isolate a component of the effect of having a Circuit case that is uncorrelated with other social trends and legal developments. Here, a pro-plaintiff precedent increased by 0.075 - 0.027 or 4.8 percentage points the likelihood of having a sexual harassment human resources policy while a pro-defendant precedent decreased by 2.7 percentage points the likelihood of having a sexual harassment human resources policy. This inference is possible because the identification of presence of a case captures the effect of pro-defendant precedent where the counterfactual is no case, whereas, adding the coefficients on pro-plaintiff precedent and on presence of a case yields the effect of pro-plaintiff precedent where the counterfactual is no case. We will refer to the latter as the *unconditional* effect.

The joint F statistic on the coefficient on the presence of appeals is 9.2. Similar back-of-the-envelope calculations imply that the impact of sexual harassment Circuit decisions from a typical Circuit-year is equivalent to 24% of the annual change in human resources policies addressing sexual harassment. In more detail, since 67% of decisions were pro-plaintiff, but only half the Circuit-years saw a Circuit sexual harassment case, we multiply 0.67, 0.50 and 4.8 to obtain a 1.6 percentage point increase in sexual harassment human resources policy due to pro-plaintiff decisions in a typical

[53]To illustrate, Chen et al. (2016e) use a machine learning approach to analyze the effects of court laws and finds that, of 18 factors that predict abortion attitudes, the court variables comprise 25-30% of the sixth factor.

Circuit-year and multiply 0.33, 0.50, and 2.7 to obtain a 0.4 percentage point decrease in sexual harassment human resources policy due to pro-defendant decisions in a typical Circuit-year. The net effect is 1.6 - 0.4, or 1.2 percentage points out of 5.1 percentage points annual increase in sexual harassment human resources policies.

4.2 Labor Force Participation Table V presents the results on labor market outcomes. The odd-numbered columns report OLS estimates while the even-numbered columns report 2SLS. The average interaction lag and the average level effect lag are reported at the bottom. On average, a pro-plaintiff sexual harassment Courts of Appeals decision increased female employment relative to males by 1.6 percentage points in the likelihood of working part-time or full-time in the five years after the decision (Column 2). The level effect reveals that a pro-plaintiff sexual harassment decision reduced the likelihood that males were working part-time or full-time by 1.3 percentage points per year during the five years after the decision. The strongest effects were in the fifth year after the decision, a few years after the strongest effects for sexual harassment human resource policies. Figure 6B plots the coefficients and the 95% confidence intervals. The lags are jointly significant and none of the lead coefficients are statistically significant.

Table VI presents a number of robustness checks on the labor force participation results. Averages of the interaction lags are displayed in Column 1 and joint F tests of the lags are displayed in Column 2. We see that the point estimates and statistical significance of the 2SLS estimates hardly change when we add Circuit specific time trends (row A), remove Circuit and year fixed effects (row B), remove all control variables except gender, presence of Circuit cases, and interaction of gender with presence of a Circuit case (row C), control for the expected number of male Democrats per seat and the expected number of female Republicans per seat (row D), or add state fixed effects (row E). The robustness to these model perturbations is consistent with the first stage estimates being invariant to the inclusion or exclusion of these controls in Table II. The joint F test increases significantly to 16.49 when we do not use CPS weights (row F). The average interaction lag increases from 1.6 percentage points to 2.1 percentage points and has a joint F test of 19.25 when we add a two-year lead (row G); the joint F statistics on the leads are still insignificant. The estimates also change little when we drop one Circuit at a time, though the joint significance varies (row H). Results are similar when we cluster standard errors at the state level (row I), collapse the data to the Circuit-year level (row J), use LASSO IV (row J), and use District IV (row L).

The point estimates are very stable, which is consistent with the use of randomization to estimate

TABLE V

The Effect of Sexual Harassment Law on Gender Inequality

	Employment Status		Hours Worked		Earnings	
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion Pro-Plaintiff	-0.000709	0.00549	0.0690	0.527	0.00981	0.0717
Circuit Decisions$_{t+1}$	(0.00308)	(0.00876)	(0.147)	(0.466)	(0.0251)	(0.0549)
Proportion Pro-Plaintiff	0.00208	0.00163	-0.152	-0.360	-0.00837	-0.0248
Circuit Decisions$_{t+1}$ * Female	(0.00454)	(0.0127)	(0.136)	(0.595)	(0.0285)	(0.0667)
Proportion Pro-Plaintiff	-0.00119	0.00336	0.0655	0.147	0.000603	0.0366
Circuit Decisions$_t$	(0.00300)	(0.00718)	(0.149)	(0.334)	(0.0217)	(0.0424)
Proportion Pro-Plaintiff	0.00207	-0.00309	-0.188	-0.552	-0.0000162	-0.0398
Circuit Decisions$_t$ * Female	(0.00527)	(0.00891)	(0.196)	(0.438)	(0.0284)	(0.0532)
Proportion Pro-Plaintiff	-0.00483	-0.00609	-0.170	-0.300	-0.0234	-0.0166
Circuit Decisions$_{t-1}$	(0.00361)	(0.00694)	(0.152)	(0.405)	(0.0248)	(0.0424)
Proportion Pro-Plaintiff	0.00701	0.00798	0.135	0.105	0.0378	0.0176
Circuit Decisions$_{t-1}$ * Female	(0.00685)	(0.00964)	(0.245)	(0.373)	(0.0382)	(0.0501)
Proportion Pro-Plaintiff	-0.00649	-0.0117	-0.303	-0.739	-0.0456	-0.0912
Circuit Decisions$_{t-2}$	(0.00367)	(0.0135)	(0.176)	(0.681)	(0.0265)	(0.0967)
Proportion Pro-Plaintiff	0.00631	0.00353	0.258	0.539	0.0490	0.0653
Circuit Decisions$_{t-2}$ * Female	(0.00509)	(0.0170)	(0.216)	(0.868)	(0.0293)	(0.113)
Proportion Pro-Plaintiff	-0.00657	-0.0147	-0.293	-0.273	-0.0520	-0.0992
Circuit Decisions$_{t-3}$	(0.00440)	(0.0106)	(0.209)	(0.604)	(0.0322)	(0.0902)
Proportion Pro-Plaintiff	0.00476	0.0182	0.214	0.522	0.0506	0.142
Circuit Decisions$_{t-3}$ * Female	(0.00613)	(0.0140)	(0.278)	(0.646)	(0.0397)	(0.109)
Proportion Pro-Plaintiff	-0.00726	-0.00640	-0.202	-0.117	-0.0463	-0.0552
Circuit Decisions$_{t-4}$	(0.00487)	(0.00513)	(0.216)	(0.259)	(0.0344)	(0.0500)
Proportion Pro-Plaintiff	0.00771	0.00195	0.124	-0.156	0.0559	0.0466
Circuit Decisions$_{t-4}$ * Female	(0.00721)	(0.00887)	(0.229)	(0.354)	(0.0423)	(0.0632)
Proportion Pro-Plaintiff	-0.00694+	-0.0284**	-0.133	-0.723+	-0.0460	-0.166*
Circuit Decisions$_{t-5}$	(0.00373)	(0.0105)	(0.164)	(0.427)	(0.0306)	(0.0689)
Proportion Pro-Plaintiff	0.00803	0.0476*	0.249	1.531+	0.0615	0.293**
Circuit Decisions$_{t-5}$ * Female	(0.00626)	(0.0187)	(0.210)	(0.799)	(0.0369)	(0.113)
Controls	Y	Y	Y	Y	Y	Y
IV	N	Y	N	Y	N	Y
Mean dependant variable - Male	0.813	0.813	34.33	34.33	4.910	4.910
Mean dependant variable - Female	0.646	0.646	22.78	22.78	3.654	3.654
Average interaction lag	0.007	0.016	0.196	0.508	0.051	0.113
Average level effect lag	-0.006	-0.013	-0.220	-0.430	-0.043	-0.086
Joint F of interaction lags		8.53		5.42		7.60
Joint F of interaction leads		0.02		0.37		0.14
Joint F of level effect lags		28.11		5.07		10.86
Joint F of level effect leads		0.39		1.28		1.70
N	3736671	3736671	3608012	3608012	3410738	3410738
R-sq	0.095	0.095	0.131	0.131	0.133	0.133

Notes: Significant at +10%, *5%, **1%. All regressions use MORG CPS and restrict to individuals between the ages of 18 and 65. Heteroskedasticity-robust standard errors are in parentheses. Observations are clustered at the Circuit level. Hours and Log earnings set to 0 for individuals not employed. All estimates are weighted using CPS survey weights. These regressions include individual controls (age, race dummies, educational attainment dummies, and marital status dummy), Circuit fixed-effects, year fixed-effects, a dummy for whether there were no cases in that Circuit-year. The regressions also include interactions between the female indicator and the dummy for there being no cases, sexual harassment law, and the instruments.

TABLE VI

Robustness of IV Estimates

	The Effect of Sexual Harassment Law on Female Employment Share	
	Average of interaction lags	Joint F of interaction lags
	(1)	(2)
A. Add Circuit-Specific Trends	0.016	8.35
B. No Fixed Effects	0.016	8.17
C. No Control Variables (except Presence of Circuit Case, Presence Interacted with Gender, and Gender)	0.017	8.08
D. Control for Expectation	0.016	8.31
E. Add State Fixed Effects	0.016	8.00
F. No CPS Weights	0.013	16.49
G. Add 2-year Lead	0.021	19.25
H. Drop 1 Circuit		
Circuit 1	0.015	6.57
Circuit 2	0.017	14.22
Circuit 3	0.016	13.81
Circuit 4	0.017	17.12
Circuit 5	0.007	37.15
Circuit 6	0.017	6.61
Circuit 7	0.017	8.72
Circuit 8	0.013	6.33
Circuit 9	0.019	5.13
Circuit 10	0.018	34.03
Circuit 11	0.014	17.23
Circuit 12	0.016	8.76
I. Cluster at State Level	0.016	11.88
J. Collapsed to Circuit-Year level	0.017	14.64
K. Collapsed with Lasso IV	0.011	25.47
L. Collapsed with Lasso and District IV	0.013	9.40

causal effects. We can also employ Monte Carlo placebo simulations that randomly assign the laws and panel assignments to different Circuits. The most conservative simulation assigns the complete time series of legal variation for one Circuit to another Circuit (Shoag 2011). The point estimate for the collapsed Circuit-year data is at the 90^{th} percentile in these simulations. Moreover, the use of alternative instruments do not affect the results as we saw in Table IV.

We can compare these estimates with the estimates of the impact of other anti-discrimination laws. For example, the Equal Employment Opportunity Act (EEOA) increased black employment shares by 0.5 to 1.1 points per year (Chay 1998). To understand what the conditional effect of 1.6 percentage points means in terms of its economic significance, multiplying by 0.67, the typical proportion of pro-plaintiff decisions, and 0.5, the proportion of Circuit-years with a Circuit sexual harassment case, suggests that women were 0.55 percentage points more likely to have any employment as compared to men in the typical Circuit-year due to the development of sexual harassment law during this time period. Across all Circuits and years, 81% of men were employed and 65% of women were employed on average. Between 1982 and 2002, the percentage of men with full or any employment stayed relatively constant from 78.8% to 80.5%, but for women, it increased dramatically from 56.7% to 68.6%, or about 0.57 percentage points per year. Since our specification comes from a five-year lag distribution, this suggests that judicial sexual harassment decisions in the previous five years caused 2.75 percentage points impact on gender inequality in employment. Examining the impact on female employment alone, we would add the average level effect lag with the average interaction lag, which results in 0.3 percentage points. The joint F significance of statistical significance is 12.70. Multiplying by 0.67, 0.5, and 5 suggests that Circuit decisions regarding sexual harassment cases in a typical five-year time period caused 0.5 percentage points increase in female employment status, a small but economically significant amount.

Next, to compare with the EEOA, we would need to add the effect of a presence of a Circuit case with the effect of a pro-plaintiff decision. To calculate the unconditional effect of sexual harassment precedent taking into account the effect of the presence of a Circuit case, we use the District IV estimates from row L. The average effect of a pro-plaintiff decision on female employment shares is 0.013. The average coefficient on the dummy indicator for the interaction of gender and presence of a Circuit case is -0.009. The unconditional effect of a pro-plaintiff decision is to decrease female employment shares by 0.4 percentage points. Multiplying 1.3 by 0.67 and 0.5, the proportion of pro-plaintiff decisions and the frequency with which a Circuit case is present, indicates that the conditional effects of pro-plaintiff sexual harassment decisions are on the lower range (0.44) of the effects of EEOA found by Chay (1998); but multiplying 0.4 by those factors indicates that the unconditional effects are one-third of those effects (0.13), or one-quarter of the effects of EEOA. Multiplying by 5 to consider the effect of sexual harassment decisions in a typical five-year time period indicates that sexual harassment law increased female employment shares by 0.67 to 2.1 points per year, suggesting an effect similar to those of EEOA.

4.3 Hours Worked Table V Columns 3 and 4 display the results on hours worked. One pro-plaintiff sexual harassment Courts of Appeals decision increased female hours worked relative to males by 30 minutes worked per week on average in each year during the five years after the decision (Column 4). It also reduced hours worked by males by 26 minutes per week on average in each year during the five years after the decision. The strongest effects are again found in the fifth year after the decision. The joint F of statistical significance is 5.42 on the lag interaction effects and 5.07 on the lag level effects. None of the lead coefficients are statistically significant.

Between 1982 and 2002, gender gap in work hours closed by roughly 16 minutes/week per year, so to understand what 0.51 hours worked means in terms of its economic significance, multiplying by 0.67, the typical proportion of pro-plaintiff decisions, suggests that women were working 0.34 more hours per week compared to men in the typical Circuit-year due to the development of sexual harassment law during this time period. Across all Circuits and years, men worked 34.33 hours per week and women worked 22.78 hours per week. These averages are low because we define hours worked to be 0 if an individual is not employed or in the labor force. In Section 5 we restrict our attention to labor force participants and do not make this imputation. Between 1982 and 2002, the male hours worked per week rose from 32.8 to 33.7 while female hours worked increased from 19.0 to 24.6, or about 0.26 hours worked per week per year. One can also calculate the effects of sexual harassment law conditional on the presence of a case and, in its totality, including the presence of a case, in which the effects are scaled down as in the previous sections.

4.4 Earnings Table V Column 6 reports the impact on the gender wage gap. Pro-plaintiff sexual harassment precedent reduced the wage gap by 0.11 in log real weekly earnings on average in each year during the five years after the decision. It reduced male wages by 0.09 log real weekly earnings on average per year during the five years after the decision. The strongest effects are again found in the fifth year after the decision. The joint F of statistical significance is 7.60 on the lag interaction effects and 10.86 on the lag level effects. None of the lead coefficients are statistically significant.[54]

Our analyses of wages and hours include both insiders and outsiders. We do this since the insider-outsider theory of involuntary employment focuses on labor force participation in the entire population. Analyses of wage gaps sometimes focus on insiders, i.e., the labor force participants who report non-zero wages, and they report comparable effects. Focusing on insider wages, the Equal Employment Opportunity Act narrowed the black-white earnings gap by 0.11 to 0.18 log points per year (Chay 1998) and state laws barring race discrimination increased black men's relative earnings by 0.28% per year (Neumark and Stock 2006). Including outsiders, a standard deviation increase in

[54]Between 1982 and 2002, the male log real weekly earnings rose from 4.75 to 4.95 while female log real weekly earnings increased from 3.11 to 4.02, or about 0.033 decrease in gender wage gap per year.

median male sexism in a state is associated with a 0.031 log point decrease in female wages relative to males (Charles et al. 2010), so sexual harassment law is comparable to the effects of a standard deviation in median male sexism.[55]

To see how our estimates compare with the estimates of the impact of male sexism, we can calculate the unconditional effect of sexual harassment law. A pro-plaintiff decision increased female wages relative to males by 0.11 log points but the presence of an appeal decreased female wages relative to males by 0.05 log points. In other words, a pro-defendant decision had an unconditional effect of decreasing female wages relative to males by 0.05 log points and a pro-plaintiff decision had an unconditional effect of increasing female wages relative to males by 0.06. Multiplying by 0.67 and 0.5, the proportion of pro-plaintiff decisions and the frequency with which a Circuit case is present, indicates that during the development of stricter sexual harassment law, a typical year saw a 0.02 log points increase in female wages relative to males, roughly equivalent to two-thirds of a standard deviation in median male sexism.

5 Evidence for the Insider-Outsider Theory of Involuntary Unemployment

5.1 Disaggregating the Effects of Sexual Harassment Law by Industry Sexual harassment is very difficult to measure objectively.[56] Changing social mores can also make it difficult to measure whether a change in registered sexual harassment complaints is due to an increasing willingness to report, increasing sensitivity to harassment, or increasing harassment. Therefore, we cannot directly measure the impact of sexual harassment law on incidence of sexual harassment.[57] We can, however, disaggregate the effects of sexual harassment law and evaluate its effects specifically in an industry with a reputedly high amount of sexual harassment–the construction industry. Sexual harassment rates per 100,000 women were highest in the construction industry according to complaints filed with the Equal Employment Opportunity Commission (Hersch 2011). The bulk of sexual harassment plaintiffs (38%) are in the blue-collar industry (Juliano and Schwab 2000).

Table VII documents that the ameliorative effects on gender inequality are positive and statistically significant for employment status, hours worked, and earnings in the construction industry but negative for the other industries. Note that we restrict the analysis to workers who report an

[55] When analyses include outsiders, studies make different wage imputations about non-labor force participants. Some studies of the gender wage gap correct for selection by imputing wages above or below the median based on educational attainment for women not working full-time and then estimating median regressions (see, e.g., Neal 2004; Olivetti and Petrongolo 2008; Charles et al. 2010).

[56] In a study using decoy crime victims, Banerjee et al. (2012) find that only 50% of sexual harassment cases are registered by the police, in contrast to 92% of break-ins and more than 64% of motorcycle thefts.

[57] For example, the number of sexual harassment claims filed increased after the 1991 Clarence Thomas hearings.

industry, so the analysis necessarily relies on a less complete picture of the entire population, since labor force participants were more likely to report their industry. The effect on managerial status is negative for females relative to males in both the construction and non-construction industries. Sexual harassment laws hurt the promotion of women to management positions in construction and non-construction industries to a similar degree since dividing by the overall inequality in management status results in a similar negative impact in terms of percentage of overall inequality. This finding is noteworthy for two reasons. First, this suggests that the estimated impacts of sexual harassment law are not simply capturing a general time trend in female labor market outcomes. Second, outsiders entering the labor force are probably not going to be managers in the first instance.

5.2 Insiders and Outsiders Thus far, we have shown positive effects of sexual harassment law on female employment outcomes on average. We now turn to some evidence for the insider-outsider theory of harassment and involuntary unemployment. If some women are insiders, then forbidding sexual harassment should result in less ameliorative impacts for insider women. We do not have a panel of individuals followed before and after legal changes, so in this analysis, we must make the albeit rough approximation that labor force participants are insiders and non-participants outsiders.

Table VIII reports the estimated effects of sexual harassment law for labor force participants. When the analysis is restricted to the labor force, we find that pro-plaintiff sexual harassment decisions *increased* gender inequality by 0.16 hours worked last week, 0.004 in log real weekly earnings, and 0.7 percentage points in the likelihood to be a manager. The bottom of the table reports mean dependent variables for insider males and females that are similar to those found in other studies of labor force participants (Blau and Kahn 2006). The joint F tests of statistical significance are very large, 35, 58, and 30, respectively, for the lag interaction effects and 25, 13, and 6 for the lag level effects. One of the lead coefficients is statistically significant for earnings in Column 1.[58]

Next, we do some arithmetic to assess how much outsider females were gaining relative to insider females, insider males, and outsider males.[59] We want to attribute the effect of sexual harassment law on female outsiders, female outsiders who enter the labor force, and insider females. We want to do the same for male outsiders, male insiders who leave the labor force, and male insiders. (These attributions account only for net flows. If there were significant inflows *and* outflows in the labor

[58]Given the number of F tests for lead coefficients, it may be expected that at least one would be statistically significant.

[59]The mean dependent variables for insider males and females are similar to those found in other studies of labor force participants (Blau and Kahn 2006).

TABLE VII.— The Effect of Sexual Harassment Law on the Construction Industry

	Construction	Non-Construction	Construction	Non-Construction	Construction	Non-Construction	Construction	Non-Construction
	Employment Status		Hours Worked		Earnings		Management	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proportion Pro-Plaintiff	0.00680	0.0118	0.408	0.617+	0.0830	0.0917*	0.00527	0.000504
Circuit Decisions$_{t+1}$	(0.0116)	(0.00777)	(0.595)	(0.348)	(0.0892)	(0.0466)	(0.00601)	(0.00211)
Proportion Pro-Plaintiff	-0.0172	-0.0137	-0.996	-0.725	-0.184	-0.0854	-0.0357+	-0.00120
Circuit Decisions$_{t+1}$ * Female	(0.0307)	(0.0130)	(0.775)	(0.573)	(0.170)	(0.0722)	(0.0197)	(0.00490)
Proportion Pro-Plaintiff	-0.000151	0.00362	-0.386	0.192	-0.00504	0.0405	-0.00638	0.00684*
Circuit Decisions$_t$	(0.00955)	(0.00643)	(0.562)	(0.310)	(0.0775)	(0.0402)	(0.00664)	(0.00344)
Proportion Pro-Plaintiff	-0.0524+	-0.00586	-2.086	-0.653	-0.358	-0.0461	-0.00621	-0.0128*
Circuit Decisions$_t$ * Female	(0.0295)	(0.0138)	(1.276)	(0.584)	(0.223)	(0.0807)	(0.0217)	(0.00500)
Proportion Pro-Plaintiff	-0.0126	0.00516	-1.578	0.129	-0.107	0.0441	0.00272	0.00620**
Circuit Decisions$_{t-1}$	(0.0156)	(0.00788)	(0.995)	(0.381)	(0.129)	(0.0457)	(0.00299)	(0.00219)
Proportion Pro-Plaintiff	-0.000377	-0.0178	0.942	-0.771	-0.0171	-0.114	-0.00183	-0.0104**
Circuit Decisions$_{t-1}$ * Female	(0.0243)	(0.0149)	(1.396)	(0.600)	(0.170)	(0.0823)	(0.0185)	(0.00354)
Proportion Pro-Plaintiff	-0.0186	0.00813	-1.254	-0.0238	-0.153	0.0302	0.000524	0.00288
Circuit Decisions$_{t-2}$	(0.0164)	(0.0114)	(0.831)	(0.533)	(0.122)	(0.0726)	(0.00682)	(0.00275)
Proportion Pro-Plaintiff	-0.0367+	-0.0255	-1.295	-0.444	-0.292	-0.116	-0.0620*	-0.00148
Circuit Decisions$_{t-2}$ * Female	(0.0211)	(0.0204)	(1.348)	(0.904)	(0.193)	(0.119)	(0.0266)	(0.00667)
Proportion Pro-Plaintiff	-0.0187	0.00279	-0.498	0.368	-0.124	0.00775	0.00713	-0.00159
Circuit Decisions$_{t-3}$	(0.0120)	(0.00832)	(0.788)	(0.459)	(0.0931)	(0.0602)	(0.0119)	(0.00260)
Proportion Pro-Plaintiff	0.000774	-0.00979	1.428	-0.598	0.0677	-0.0384	-0.0520*	-0.00337
Circuit Decisions$_{t-3}$ * Female	(0.0418)	(0.0109)	(1.546)	(0.442)	(0.263)	(0.0688)	(0.0239)	(0.00533)
Proportion Pro-Plaintiff	-0.0148	-0.00301	-0.699	0.0529	-0.147+	-0.0245	0.00652	-0.000364
Circuit Decisions$_{t-4}$	(0.00967)	(0.00691)	(0.536)	(0.346)	(0.0798)	(0.0555)	(0.00678)	(0.00253)
Proportion Pro-Plaintiff	-0.00615	0.000147	-1.469	-0.344	-0.0549	0.0142	-0.0215	-0.00236
Circuit Decisions$_{t-4}$ * Female	(0.0361)	(0.00986)	(1.978)	(0.329)	(0.304)	(0.0589)	(0.0352)	(0.00288)
Proportion Pro-Plaintiff	-0.0246	-0.0141	-1.194	-0.0913	-0.129	-0.0693	-0.0111+	0.00384
Circuit Decisions$_{t-5}$	(0.0217)	(0.0114)	(1.104)	(0.379)	(0.166)	(0.0728)	(0.00633)	(0.00266)
Proportion Pro-Plaintiff	0.0907*	0.0175	2.284**	0.210	0.515*	0.0884	0.0537+	-0.0117**
Circuit Decisions$_{t-5}$ * Female	(0.0359)	(0.0176)	(0.873)	(0.732)	(0.223)	(0.110)	(0.0326)	(0.00315)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
IV	Y	Y	Y	Y	Y	Y	Y	Y
Mean dependant variable - Male	0.836	0.892	33.91	38.31	4.977	5.517	0.174	0.129
Mean dependant variable - Female	0.793	0.826	27.45	29.47	4.385	4.775	0.106	0.098
Average interaction lag	0.010	-0.007	0.378	-0.389	0.044	-0.033	-0.017	-0.006
Average level effect lag	-0.018	0.000	-1.045	0.087	-0.132	-0.002	0.001	0.002
Joint F of interaction lags	13.66	2.99	25.83	26.27	10.18	3.14	11.13	32.54
Joint F of interaction leads	0.31	1.13	1.70	1.64	1.14	1.39	3.22	0.07
Joint F of level effect lags	19.27	3.51	49.25	1.45	20.09	2.86	5.94	15.43
Joint F of level effect leads	0.28	2.17	0.45	3.04	0.77	3.68	0.71	0.07
N	210153	2949731	201678	2825198	163297	2666305	210153	2949731
R-sq	0.048	0.044	0.079	0.094	0.071	0.108	0.107	0.059

Notes: Significant at +10%, *5%, **1%. All regressions use MORG CPS and restrict to individuals between the ages of 18 and 65 and who report an industry category. Heteroskedasticity-robust standard errors are in parentheses. Observations are clustered at the Circuit level. Hours and Log earnings set to 0 for individuals not employed. All estimates are weighted using CPS survey weights. These regressions include individual controls (age, race dummies, educational attainment dummies, and marital status dummy), Circuit fixed-effects, year fixed-effects, a dummy for whether there were no cases in that Circuit-year. The regressions also include interactions between the female indicator and the dummy for there being no cases, sexual harassment law, and the instruments as well as the interactions of these variables with gender.

TABLE VIII
The Effect of Sexual Harassment Law on Insiders

	Labor Force Partipants Reporting Non-Zero Wages		
	Earnings	Hours Worked	Management
	(1)	(2)	(3)
Proportion Pro-Plaintiff	0.0256*	0.232	0.000853
Circuit Decisions$_{t+1}$	(0.0101)	(0.244)	(0.00182)
Proportion Pro-Plaintiff	-0.0208	-0.381	-0.00212
Circuit Decisions$_{t+1}$ * Female	(0.0135)	(0.441)	(0.00464)
Proportion Pro-Plaintiff	0.0257	0.0221	0.00545+
Circuit Decisions$_{t}$	(0.0206)	(0.212)	(0.00306)
Proportion Pro-Plaintiff	-0.0243	-0.463	-0.0115*
Circuit Decisions$_{t}$ * Female	(0.0283)	(0.391)	(0.00480)
Proportion Pro-Plaintiff	0.0299*	-0.0781	0.00486*
Circuit Decisions$_{t-1}$	(0.0118)	(0.185)	(0.00233)
Proportion Pro-Plaintiff	-0.0343*	-0.191	-0.00873*
Circuit Decisions$_{t-1}$ * Female	(0.0140)	(0.257)	(0.00403)
Proportion Pro-Plaintiff	-0.00465	-0.306	0.00240
Circuit Decisions$_{t-2}$	(0.0138)	(0.196)	(0.00219)
Proportion Pro-Plaintiff	0.0126	0.485	-0.00326
Circuit Decisions$_{t-2}$ * Female	(0.0175)	(0.419)	(0.00568)
Proportion Pro-Plaintiff	-0.00226	0.343	-0.0000238
Circuit Decisions$_{t-3}$	(0.0222)	(0.245)	(0.00283)
Proportion Pro-Plaintiff	0.00444	-0.328	-0.00636
Circuit Decisions$_{t-3}$ * Female	(0.0184)	(0.287)	(0.00569)
Proportion Pro-Plaintiff	0.00107	0.186	0.00276
Circuit Decisions$_{t-4}$	(0.0169)	(0.164)	(0.00221)
Proportion Pro-Plaintiff	0.00688	-0.389*	-0.00665*
Circuit Decisions$_{t-4}$ * Female	(0.00926)	(0.177)	(0.00285)
Proportion Pro-Plaintiff	0.0150	0.370	0.00358+
Circuit Decisions$_{t-5}$	(0.0291)	(0.363)	(0.00205)
Proportion Pro-Plaintiff	-0.00821	-0.379	-0.00876**
Circuit Decisions$_{t-5}$ * Female	(0.0330)	(0.486)	(0.00260)
Controls	Y	Y	Y
IV	Y	Y	Y
Mean dependant variable - Male	6.298	42.90	0.144
Mean dependant variable - Female	5.854	36.04	0.120
Average interaction lag	-0.004	-0.160	-0.007
Average level effect lag	0.008	0.103	0.003
Joint F of interaction lags	35.16	58.21	30.38
Joint F of interaction leads	2.38	0.75	0.21
Joint F of level effect lags	24.71	13.10	6.25
Joint F of level effect leads	6.47	0.90	0.22
N	2424997	2622664	2755279
R-sq	0.296	0.081	0.057

Notes: Significant at +10%, *5%, **1%. All regressions use MORG CPS and restrict to individuals between the ages of 18 and 65. Heteroskedasticity-robust standard errors are in parentheses. Observations are clustered at the Circuit level. All estimates are weighted using CPS survey weights. These regressions include individual controls (age, race dummies, educational attainment dummies, and marital status dummy), Circuit fixed-effects, year fixed-effects, a dummy for whether there were no cases in that Circuit-year. The regressions also include interactions between the female indicator and the dummy for there being no cases, sexual harassment law, and the instruments as well as the interactions of these variables with gender.

force, insider women would be worse off and outsider women better off than our calculations indicate, so our calculations may be viewed as a lower bound.)

Focusing on the average yearly lag effect over five years, Column 1 in Table VIII indicates that insider men gained by 0.008 log real weekly earnings while insider women gained by 0.004 log real weekly earnings. Column 6 in Table V indicates that insider and outsider men lost 0.086 log real weekly earnings and insider and outsider women gained 0.027 log real weekly earnings. If we approximate the 65% of women with part- or full-time employment to be insiders and similarly for 81% of men, then accounting for net movements into and out of labor force participation (Table V Column 2) yields the following summary for the effects for females and males:

$(0.35 - 0.003) * 0 + (0.003) * 5.9 + (0.65) * 0.004 = 0.020$, which is near 0.027;

$(0.19) * 0 + (0.013) * -6.3 + (0.81 - 0.013) * 0.008 = -0.075$, which is near -0.086;

To explain the first equation, female outsiders who remained outsiders constitute nearly 35%, female outsiders who enter the labor force constitute 0.3%, and insider females are 65%. Male outsiders constitute 19%, insider males who became outsiders constitute 1.3% and the remaining insiders constitute nearly 80%. The numbers 0.004 and 0.008 come from Table VIII Column 1. The numbers 5.9 and 6.3 are the group-mean dependent variables in Table VIII. The numbers +0.3% and -1.3% come from Table V Column 2.

In sum, forbidding sexual harassment caused wages of all labor force participants to increase, and a portion of insider males to exit in response to, or because of, the entrance of outsider females. The resulting gains for females were tilted towards outsider females who gained six times more than insider females; outsider females also gained three times more than remaining insider males. Another indication of the increase in productivity, particularly for females, is that while insider females gained 0.004 log real weekly earnings, they lost in hours worked (a decline of 0.06 hours per week) and also in managerial status (0.4 percentage point decline in likelihood to be a manager).

Taken together, these results suggest that forbidding sexual harassment in large part encouraged the entry of outsider women, who then obtained part- or full-time employment and received wages, and increased the productivity of women in the labor force.[60] Previously insider men lost the most in terms of employment status. We only analyze labor market outcomes, however, and are not doing a welfare calculation along all margins.

[60] Outsider women asked for and received higher wages.

6 Additional Remarks

Our empirical strategy uses the number of extreme judges assigned to sexual harassment cases to approximate a true experiment. This requires that Circuit judges be randomly assigned. At the Circuit-year level, the number of extreme judges per seat needs to be as good as randomly assigned, conditional on having a case. We report three assessments of this assumption.

6.1 Interviews with government officials First, we surveyed a number of courts of appeal and evaluated measures taken by them to ensure that the assignment of judges to panels is random. In one court, two to three weeks before the oral argument, a computer program is used to randomly assign available judges, including any visiting judges, to panels that will hear cases. The program used is an in-house creation. There is a mechanism in the program that ensures the same judges are not sitting together on panels. This is also checked manually, although the clerk could not remember ever having manually to change judicial assignments for this reason. There is no specialization among judges; the cases are "all over the map" in regard to subject matter. Senior judges tell the clerk how often they are willing to sit and hear cases, and they are added to the program for randomized assignment in accordance with their schedules. There is an administrative office that sets the baseline number of cases senior judges must hear per term.

In another court, random assignment of panels occurs before the random assignment of cases. Panels of judges are organized to hear cases on a yearly basis, randomly assigned together by computer program and given dates for hearings. There are "holes" left in some of the panels by the program, and visiting judges are plugged into those spots by the chief judge. This program also ensures that the same judges are not seated together repeatedly. Thus, the judges know at the beginning of the year which days they will be hearing cases and the compositions of the panels on which they will sit.

Once all the briefing is completed, a case is put into a pool of cases "ready to calendar." If a panel of judges has previously looked at a case, it will be sent back to them (for example, if it was remanded to resolve one issue). Otherwise, a different program randomly assigns cases to these pre-established panels and dates. About eight weeks before the scheduled argument, a preliminary calendar is sent out and the judges review it for recusal. If a judge must recuse himself, the case is taken off the calendar and placed back in the pool for reassignment. Senior judges decide how many days and which months they will work, and this information is entered into the program for random assignment. Before the advent of computer programs, one judge did all of the panel assignments by hand, and the clerks randomly assigned the cases by hand. For more information about random assignment of cases at the Circuit level, see Brown Jr. and Lee (2000).[61]

Other variations from random assignment include: en banc cases that are heard by the entire pool of judges (or a significant fraction in Circuit 9). We do not use these cases, which are also relatively infrequent. Judges can also take sick leave or go on vacation, but this is determined far in advance. Not accounting for vacation, sick leave, senior status, en banc, remand, and recusal can lead to the inference that judges are not randomly assigned. Our identification strategy assumes that these kinds of deviations from random assignment are ignorable. Even a gold-standard random process — the roll of a die — has a deterministic element. If known with precision, the force and torque applied to the die, the subtle air currents, the hardness of the surface, etc., might allow us (or a physicist) to determine with certainty the outcome of these "random" rolls. Despite this obvious non-randomness, we would still have faith in the outcome of a trial with treatment assignments based on die rolls because we are certain that the factors affecting the assignment have no impact on the outcome of interest and hence are ignorable.

6.2 Orthogonality with pre-trial characteristics As a second randomization assessment, we use data from Boyd et al. (2010), which codes some case characteristics for a subset of 415 gender discrimination cases in the Chicago Judges Projects data (Sunstein et al. 2006). We regress case characteristics on male Democratic (female Republican) judges per seat and find that most characteristics are not correlated with the judicial panel composition. Table IX shows that of 19 case characteristics, one is correlated with male Democrats per seat and one is correlated with female Republicans per seat. Both correlations are statistically significant at the 10% level. Given the number of tests, it may be expected that 10% would be statistically significant at the 10% level. For the 79 cases that we coded from 1982 to 1995, we also noted whether the plaintiff was the victim of sexual harassment and which party appealed. The last two rows of Table IX show that these case characteristics are also uncorrelated with judicial panel composition. Judicial characteristics do appear orthogonal to case characteristics as determined before the assignment of Circuit judges.

6.3 Omnibus Test for Circuit Courts: Random Strings In our third randomization assessment, we examine deviations from random assignment by seeing whether the sequence of proportions of judges is similar to a random process. Figures 3A and 3B suggest visually that panel composition is not serially correlated. We formally investigate this by:

[61] See also, http://law.du.edu/images/uploads/neutral-assignment/Neutral_assignment_links.pdf.

TABLE IX

RANDOMIZATION CHECK: ORTHOGONALITY WITH CASE CHARACTERISTICS AS DETERMINED BY LOWER COURT

Case Characteristics as Determined by Lower Court	Male Democrats per seat (1)	Female Republicans per seat (2)
Direction of Lower Court Decision	0.0115	-0.171
	(0.0856)	(0.187)
P claims employer acted in retaliation	-0.102	0.184
	(0.0936)	(0.205)
All plaintiffs are female	0.0126	-0.0920
	(0.0747)	(0.164)
Title IX claim	0.0415	-0.0558
	(0.0252)	(0.0553)
Section 1983 claim	0.0533	-0.0474
	(0.0500)	(0.110)
Constructive discharge from employment	0.00764	0.0726
	(0.0559)	(0.122)
Procedural issues dominate	0.0167	0.163
	(0.0586)	(0.128)
P suing under state law	0.0677	-0.283
	(0.0830)	(0.181)
P claims illegally denied promotion	-0.0591	-0.0465
	(0.0755)	(0.165)
P claims illegally not being hired	-0.0909+	0.105
	(0.0529)	(0.116)
P claims illegally fired	0.0460	-0.159
	(0.0961)	(0.210)
P claims unequal pay	-0.0235	-0.0868
	(0.0675)	(0.148)
P sued under 14th Amendment	0.0606	-0.167+
	(0.0429)	(0.0938)
P sued under 1st Amendment	0.0574	-0.0503
	(0.0353)	(0.0775)
Damages major point of contention	0.0765	0.166
	(0.0669)	(0.147)
Contains Section 1981 claim	0.0295	-0.0818
	(0.0585)	(0.128)
Contains age discrimination claim	0.0368	-0.241
	(0.0695)	(0.152)
Contains pregnancy discrimination claim	0.0232	0.0911
	(0.0484)	(0.106)
Contains emotional distress claim	-0.0781	0.0432
	(0.0530)	(0.116)
P not victim of harassment	-0.0312	-0.338
	(0.131)	(0.340)
P is appellant	-0.109	-0.349
	(0.208)	(0.541)

Notes: Significant at +10%, *5%, **1%. Heteroskedasticity-robust standard errors are in parentheses. Each coefficient represents a separate regression of a distinct case characteristic on male Democrats (female Republicans) per seat. “P” refers to plaintiff.

1. Propose a statistic that can be computed from the sequence of numbers of female Republicans (male Democrats) per seat within a Circuit.
2. Compute the statistic for the actual sequence, s^*.
3. Compute the statistic for each of 1,000 bootstrap samples from the actual sequence, i.e., s_1, s_2, s_3 . . .s_n. Since there were changes in the expected number of female Republicans (male Democrats) per seat over time, we treat our bootstrap samples as a vector of realized random variables, with the probability based on the expectation during the Circuit-year.
4. Compute the empirical p-value, p_i by determining where s^* fits into s_1, s_2, s_3 . . .s_n.
5. Repeat steps 1-4 and calculate p_i for each unit.

We use the following statistics:

Autocorrelation: We see if the value in the j$^{\text{th}}$ case depends on the outcome in the j-1$^{\text{th}}$case. This statistic can detect whether judicial assignments are "clustered," meaning a higher than expected number of back-to-back seat assignments to a particular type of judge. This test tells us whether certain judges sought out sexual harassment cases, perhaps in sequence.

Mean-Reversion: We test whether there is any form of mean reversion in the sequence, meaning that the assignment in the n$^{\text{th}}$ case is correlated with the assignment in previous $n-1$ cases. This test tells us whether judges or their assignors were attempting to equilibrate their presence, considering whether a judge was "due" for a sexual harassment case.

Longest-Run: We test whether there are abnormally long "runs" of certain types of judges per seat. This test tells us whether certain Circuits may have assigned certain judges with sexual harassment cases during certain time periods, for example, to achieve specialization.

Number of Runs: Instead of simulating 1000 random strings, we compute the exact statistic for number of runs. This test captures violations of randomization at the case level rather than Circuit-year. In power calculations, this test has less Type II error compared to the other tests.[62]

With a truly random process, the collection of all unit p-values should be uniformly distributed. (Imagine that you generate summary statistics for 1000 random strings. The 1001$^{\text{th}}$ random string should have a summary statistic that is equally likely to be anywhere from 1 to 1000.) A visual examination suggests that the empirical distributions for our p-values approach the CDF of a uniform distribution (Figure 4), which we formally test using a Kolmogorov-Smirnov test statistic from an empirical null (Table X). Our p-values are uniformly distributed for all six tests except

[62]These random strings tests also have lower Type II error compared to regression.

TABLE X

Randomization Check: P-values

	Female Republican	Male Democrat
Auto-correlation	0.34*	0.24
Mean Reversion	0.32	0.16
Longest Run	0.22	0.23

Kolmogorov-Smirnov Test

Values of Dn for Various P

n = 9		n = 12	
Prob(sqr(n)*Dn<b)	b/sqr(n)	Prob(sqr(n)*Dn<b)	b/sqr(n)
0.01	0	0.01	0.009
0.05	0.0373	0.05	0.0345
0.1	0.0619	0.1	0.0553
0.25	0.1091	0.25	0.09
0.5	0.1804	0.5	0.1574
0.75	0.2608	0.75	0.2275
0.9	0.3392	0.9	0.2958
0.95	0.3874	0.95	0.3381
0.99	0.4795	0.99	0.4288

autocorrelation for female Democrats, which is largely due to Circuit 6. In our robustness checks, we drop one Circuit at a time.

We also checked one or two years before the true instrument, that judicial decision-making is not correlated with future judicial assignment.

6.4 Randomization for District Judges District Courts assign one judge to a case randomly or rotationally (Taha 2009; Bird 1975).[63] For example, one District told us that random assignment occurs within 24 hours of a case filing, which is handled in the order of its arrival. Waldfogel (1995) reports that one District Court uses three separate randomization wheels and each wheel

[63]Cases being returned on remand from the Circuit court are not randomly assigned. We do not use remanded cases in our dataset.

corresponds to the anticipated case length.[64] Related[65] cases, if filed within a few weeks, may be consolidated.[66] Consolidation only occurs for relatively high-frequency case types, which does not include free speech jurisprudence. For the handful of District cases that do overlap such that they are consolidated, we assume the decisions about case relatedness occur in a manner exogenous to judge assignment.

District Courts judges are revealed much earlier than Circuit Court judges. Ideally, we would use docket filings in the Administrative Office of the U.S. Courts pertaining to free speech cases, but judges are omitted for most cases prior to 2000, so we must use published District opinions to construct our District IV. We buttress the assumption that settlement, publication, and strategic use of keywords or citations are exogenous: 1) in District Courts, judges are much more constrained and ideology has been found to play hardly any role. Judicial ideology does not predict settlement rates (Ashenfelter et al. 1995; Nielsen et al. 2010), settlement fees (Fitzpatrick 2010), publication choice (Taha 2004), or decisions in published or unpublished cases (Keele et al. 2009);[67] 2) we

[64] The ideal construction of $\tilde{w}_{ct}$ takes a weighted sum across wheels of deviation from expectations, $E\left(\frac{L_{c\tilde{d}t}}{K_{c\tilde{d}t}}\right)$, separately for senior and non-senior judges. Senior judges can elect not to be assigned to certain wheels. Another District Court uses, instead of wheels, thirteen computer generated decks of cards–one deck for each case category and an identical number of cards (two or five) for each active judge (http://www.mnd.uscourts.gov/cmecf/Order-for-Assignment-of-Cases.pdf). The decks refill when the majority of the deck has been exhausted. Senior judges can request to be assigned to certain decks. Even within a deck, senior judges can ex ante request a "bye" for specialized case types. Within each District Court are several courthouses (also referred to as Divisions). The appropriate Division is determined by where the parties are located and where the cause of action arose. Some Divisions get their own deck of cards. Taha (2009) reports that in 29 Districts, a case may be assigned to any judge in that District, while in the others, the cases are assigned to a geographic Division within the District and randomly assigned to one of the judges in that Division.

However, since $E\left(\frac{L_{c\tilde{d}t}}{K_{c\tilde{d}t}}\right)$ is uncomputable for senior judges since we would need to know the senior "byes" in every District courthouse, we drop senior District judges for calculating $\tilde{w}_{ct}$; we also drop visiting (judges routinely visit other courts to assist with caseload) and magistrate judges (they assist District Court judges but do not have life tenure and we do not have their biographical data) for similar reasons, collectively resulting in less than 10% sample loss. Non-ideological cases are referred to magistrate judges (Nash 2015), so omitting them will not matter. Identification is unaffected by dropping judges even if they are in the same wheel. Some courts spin separate random wheels for District judges and for magistrate judges. In some Districts, parties can decline assignment to a magistrate judge within a certain time period and request another random draw. This will not affect identification because it happens before the random assignment that we use. In some Districts, when the federal government is a litigant on the case, the U.S. attorney can pick the wheel.

In sum, conditional on case type, there is random assignment at the court or courthouse level, and we must only calculate the yearly expected composition of judges in District courthouses, $E\left(\frac{L_{c\tilde{d}t}}{K_{c\tilde{d}t}}\right)$, and we drop judges whose $E\left(\frac{L_{c\tilde{d}t}}{K_{c\tilde{d}t}}\right)$ is unknowable.

[65] Related means that one decision will substantially resolve all cases.

[66] Waldfogel (1995) reports that plaintiffs can argue the case is related to another pending case and, if the judge agrees, the cases will be consolidated. A clerk reported 8% of filed cases were accepted as related in 1991 in SDNY. In another District Court, if a clerk identifies and two judges agree that a new civil case is related to another open civil case, they will be consolidated in the interests of justice or judicial economy. The clerk brings the possible connection to the attention of the judge of the new case, who then confers with the judge of the earlier case to determine whether they are in fact related cases.

[67] This is consistent with the District judge identity only affecting outcomes through the presence of an appeal but not through the District Court decision, but this is not necessary for the empirical analysis.

examine these issues directly:[68] we test whether District Court judicial biographical characteristics in *filed* cases jointly predict publication.[69] We are able to conduct this test because we link PACER filing data, which has judge identity, to AOC data, which has information on publication.[70] We assume that remaining deviations from random assignment, like vacation days, are ignorable.

7 Alternative Theories

In this section, we consider several alternative theories to explain our findings.

Tax on Hiring Men: An alternative view of sexual harassment law is simply that it is a tax on the hiring of men, who are potential harassers.[71] A less sympathetic view is that sexual harassment law mandates a transfer from females to males. However, if sexual harassment law was essentially a tax on the hiring of men, we should have observed insider male outcomes to fall relative to insider female outcomes, but they did not.

Machismo: The insider-outsider theory suggests intentional harassment, but the productivity of females could have been lower simply due to their unfamiliarity with the machismo culture prevalent before females were hired in substantial numbers. However, if this were the case, wages should not be related to the elimination of machismo, but we find that both male and female insider wages increased.

Compensating Differentials: Compensating differentials do not appear sufficient to explain the findings. If insider women were compensated for having to face sexual harassment, their wages would decline with as sexual harassment law developed; instead, their wages increased. Moreover, a pure compensating differentials story would not explain why outsider women joined the work force. Wages inside the labor force would have adjusted downwards for the decrease in sexual harassment and outsider women on the margin would have been indifferent to entry.

Change in the Composition of the Female Labor Force: Perhaps in the absence of sexual harassment law, firms chose to hire less productive women and after forbidding sexual harassment, firms hired more productive women. However, this does not easily explain the exit of insider males nor why insider male wages increased more than the wages of insider women.

---

[68] The random strings test is ineffective because some Districts use rotational assignment or random drawing of judges from card decks without replacement.

[69] We use LASSO to select biographical characteristics and no characteristic was chosen.

[70] We obtained all freely available PACER (Public Access to Court Electronic Records) data on District cases from 32 districts for 1980 to 2008 for a total of 359,595 non-duplicated cases. This data contains the name of the District where the case was filed, the filing and termination date (missing for 10% of cases), the assigned docket number, and the name of the District or magistrate judge presiding on the case. We merge the names of the judges into the Administrative Office of the U.S. Courts (AOC) database.

[71] The vast majority of sexual harassment cases were filed by women regarding sexual harassment by men.

Mandated Benefit: While the mandated benefits view is not an alternative theory for the results, explaining why maternity mandates and the ADA had different effects from sexual harassment law is worth mentioning. First, unlike maternity mandates, sexual harassment directly improves the productivity of women by making a better work environment. Second, unlike ADA, it is far more difficult to determine precisely which women are likely to impose the costs of sexual harassment on a firm. Disabilities are often visible to employers and, therefore, the unfunded mandate of accommodations may have led to calculated decisions to not hire particular disabled workers whereas employers could not as easily make the same calculated decisions vis-a-vis women. Third, unlike the cost of complying with the ADA or the federal requirement of providing maternity mandates, the cost of compliance with sexual harassment law could be reduced by not hiring either the group being harassed or the group doing the harassing. Men were hired less as well as women hired more, unlike what happened with ADA. Fourth, the costs of sexual harassment law are quite high, although exact figures for all the direct and indirect costs are difficult to obtain. Some labor lawyers observe that the ADA further required large fixed costs upfront in physical infrastructure, whereas the fear of a Title VII suit is always looming.

8 Conclusion

Making the workplace friendlier to women has been one of the most dramatic labor market changes in the past half-century, yet, the consequences of forbidding sexual harassment on female labor force outcomes remain unknown. The fact that female labor market outcomes improve more than insider male labor market outcomes decline is consistent with the idea that there was some degree of rent capture by insiders and suggests that firms were not profit-maximizing and forbidding harassment on their own in the absence of sexual harassment law. It is also consistent with the persistence of beliefs that women should play only certain roles, or should know their “place”, and that this persistence accounts for a substantial portion of inequality.

Interpreting anti-discrimination law to forbid sexual harassment has been a key contribution of academic scholarship in employment law. Unlike other employment laws, sexual harassment law is generally considered "good" social policy and has not come under fire for its potential negative consequences in the way that other employment protections, such as ADA and maternity mandates, have. Yet, economic theory, at first glance, suggests that the potential effects of forbidding sexual harassment may be similar to those of other employment mandates. It may exacerbate gender inequality overall because it could be viewed as a tax on the hiring of women. We identify the impact of court-made sexual harassment precedent on gender inequality by using the fact that federal judges are randomly assigned to Circuit cases along with the fact that gender and party of appointment of judges affect sexual harassment decisions. We find that sexual harassment law does not appear to exacerbate gender inequality. Pro-plaintiff precedent increases female wages and employment relative to that of men. However, when restricted to people previously in the work force, sexual harassment law has less ameliorative effects for females. These findings are more consistent with an insider-outsider theory of involuntary unemployment–where insiders harass outsiders in order to capture economic rents and forbidding harassment increases entry of outsiders, raising their employment and then wages–than with compensating wage differentials or mandated benefits models of sexual harassment.

Methodologically, the empirical framework developed here provides causal estimates of court precedent holding all else equal including unobserved factors. It overcomes the basic issues of omitted variables and reverse causality. Furthermore, it has the advantages that the exclusion restriction is likely to hold, the LATE interpretation of the IV estimates are policy relevant, the general equilibrium effects are those which we would want to include, and the impulse response function is well-identified. We hope it proves fruitful for policy-makers and judges interested in assessing the impact of court-made law as well as for scholars and theorists interested in evaluating theories of behavioral responses to the law, exploring heterogeneity in terms of early or late decisions, or exploiting variation in the sequence of decisions.

References

Abramowicz, Michael, Ian Ayres, and Yair Listokin, 2011, Randomizing Law, *University of Pennsylvania Law Review* 159, 929–1005.

Acemoglu, Daron, and Joshua D. Angrist, 2001, Consequences of Employment Protection? The Case of the Americans with Disabilities Act, *The Journal of Political Economy* 109, 915–957.

Angrist, Joshua D., and Guido W. Imbens, 1995, Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity, *Journal of the American Statistical Association* 90, 431–442.

Araiza, Alberto G., Daniel L. Chen, and Susan Yeh, 2014, Does Appellate Precedent Matter? Stock Price Responses to Appellate Court Decisions on FCC Actions, in Yun chien Chang, ed., *Empirical Legal Analysis: Assessing the Performance of Legal Institutions*, volume 19 of *Economics of Legal Relationships* (Routledge).

Ashenfelter, Orley, Theodore Eisenberg, and Stewart J. Schwab, 1995, Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes, *Journal of Legal Studies* 24, 257–281.

Asmussen, Nicole, 2011, Female and Minority Judicial Nominees: President's Delight and Senators' Dismay?, *Legislative Studies Quarterly* 36, 591–619.

Autor, David H., John J. Donohue, and Stewart J. Schwab, 2006, The Costs of Wrongful-Discharge Laws, *The Review of Economics and Statistics* 88, 211–231.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, 2005, Rising Wage Inequality: The Role of Composition and Prices, NBER Working Papers 11628, National Bureau of Economic Research.

Badawi, Adam B., and Daniel L. Chen, 2014, The Shareholder Wealth Effects of Delaware Litigation, Working paper, ETH Zurich, Mimeo.

Baliga, Sandeep, Eran Hanany, and Peter Klibanoff, 2013, Polarization and ambiguity, *The American Economic Review* 103, 3071–3083.

Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh, 2012, Can Institutions be Reformed from Within? Evidence from a Randomized Experiment with the Rajasthan Police, NBER Working Papers 17912, National Bureau of Economic Research.

Barondes, Royce De Rohan, 2010, Federal District Judge Gender and Reversals, in *5th Annual Conference on Empirical Legal Studies Paper*, Working Paper.

Barrios, Thomas, Rebecca Diamond, Guido W. Imbens, and Michal Kolesár, 2012, Clustering, Spatial Correlations and Randomization Inference, *Journal of the American Statistical Association* 107, 578–591.

Barry, Nora, Laura Buchanan, Evelina Bakhturina, and Daniel L. Chen, 2016, Events Unrelated to Crime Predict Criminal Sentence Length, Technical report.

Basu, Kaushik, 2003, The Economics and Law of Sexual Harassment in the Workplace, *The Journal of Economic Perspectives* 17, 141–157.

Becker, Gary S., 1968, Crime and Punishment: An Economic Approach, *Journal of Political Economy* 76, 169–217.

Belloni, Alex, Daniel L. Chen, Victor Chernozhukov, and Chris Hansen, 2012, Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain, *Econometrica* 80, 2369–2429.

Berdejó, Carlos, and Daniel L. Chen, 2014, Priming Ideology? Electoral Cycles without Electoral Incentives among U.S. Judges, Working paper, ETH Zurich.

Bertrand, Marianne, and Esther Duflo, 2016, Field Experiments on Discrimination, Technical report, National Bureau of Economic Research.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, 2004, How Much Should We Trust Differences-In-Differences Estimates?, *The Quarterly Journal of Economics* 119, 249–275.

Bertrand, Marianne, and Sendhil Mullainathan, 2004, Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination, *The American Economic Review* 94, 991–1013.

Bird, Susan Willett, 1975, The Assignment of Cases to Federal District Court Judges, *Stanford Law Review* 27, 475–487.

Black, Sandra E., and Philip E. Strahan, 2001, The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry, *The American Economic Review* 91, 814–831.

Blau, Francine D., and Lawrence M. Kahn, 2006, The U.S. Gender Pay Gap in the 1990s: Slowing Convergence, *Industrial and Labor Relations Review* 60, 45–66.

Boyd, Christina, Lee Epstein, and Andrew D. Martin, 2010, Untangling the Causal Effects of Sex on Judging, *American Journal of Political Science* 54, 389–411.

Breyer, Stephen, 2006, *Active Liberty: Interpreting Our Democratic Constitution* (Vintage Books).

Brown Jr., Robert J., and Allison Herren Lee, 2000, Neutral Assignment of Judges at the Court of Appeals, *Texas Law Review* 78, 1037–1116.

Card, David E., and John Enrico DiNardo, 2002, Technology and U.S. Wage Enequality: A Brief Look, *Economic Review* 87, 45–62.

Charles, Kerwin K., Jonathan Guryan, and Jessica Pan, 2010, Sexism and Women's Labor Market Outcomes, Working paper, University of Chicago.

Charles, Kerwin Kofi, and Jonathan Guryan, 2008, Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination, *The Journal of Political Economy* 116, 773–809.

Chay, Kenneth Y., 1998, The Impact of Federal Civil Rights Policy on Black Economic Progress: Evidence from the

Equal Employment Opportunity Act of 1972, *Industrial and Labor Relations Review* 51, 608–632.
Chen, Daniel, Yosh Halberstam, and Alan Yu, 2016a, Covering: Mutable Characteristics and Perceptions of (Masculine) Voice in the U.S. Supreme Court, Technical report, Department of Economics-Toulouse School of Economics.
Chen, Daniel, Yosh Halberstam, and Alan Yu, 2016b, Covering: Mutable Characteristics and Perceptions of (Masculine) Voice in the U.S. Supreme Court, Technical report, Department of Economics-Toulouse School of Economics.
Chen, Daniel L., 2004, Gender Violence and the Price of Virginity: Theory and Evidence of Incomplete Marriage Contracts, Working paper, University of Chicago, Mimeo.
Chen, Daniel L., 2006, Islamic Resurgence and Social Violence During the Indonesian Financial Crisis, in Mark Gradstein, and Kai A. Konrad, eds., *Institutions and Norms in Economic Development*, chapter 8, 179–199 (MIT Press).
Chen, Daniel L., 2010, Club Goods and Group Identity: Evidence from Islamic Resurgence during the Indonesian Financial Crisis, *The Journal of Political Economy* 118, 300–354.
Chen, Daniel L., 2013, The Deterrent Effect of the Death Penalty? Evidence from British Commutations During World War I, Working paper, ETH Zurich.
Chen, Daniel L., 2015, Can Markets Stimulate Rights? On the Alienability of Legal Claims, *RAND Journal of Economics* 46, 23–65.
Chen, Daniel L., 2016, Priming Ideology: Why Presidential Elections Affect U.S. Courts of Appeals Judges, Technical report.
Chen, Daniel L., Xing Cui, Lanyu Shang, and Jing Zhang, 2016c, What Matters: Agreement Between U.S. Courts of Appeals Judges, Technical report.
Chen, Daniel L., Matt Dunn, Rafael Garcia Cano Da Costa, Ben Jakubowki, and Levent Sagun, 2016d, Early Predictability of Asylum Court Decisions, Technical report.
Chen, Daniel L., and Jess Eagel, 2016, Can Machine Learning Help Predict the Outcome of Asylum Adjudications?, Technical report.
Chen, Daniel L., Jens Frankenreiter, and Susan Yeh, 2014a, Measuring the Effects of Legal Precedent in U.S. Federal Courts, Working paper, ETH Zurich.
Chen, Daniel L., Kristen Kwan, Maria Maass, , and Luisa Ortiz, 2016e, Law and Norms: Using Machine Learning to Predict Attitudes Towards Women and Abortion, Technical report.
Chen, Daniel L., Vardges Levonyan, and Susan Yeh, 2014b, Do Policies Affect Preferences? Evidence from Random Variation in Abortion Jurisprudence, Working paper, ETH Zurich.
Chen, Daniel L., and Jo Lind, 2007, Religion, Welfare Politics, and Church-State Separation, *Journal of Ecumenical Studies* 42, 42–52.
Chen, Daniel L., and Jo Thori Lind, 2014, The Political Economy of Beliefs: Why Fiscal and Social Conservatives and Liberals Come Hand-in-Hand, Working paper.
Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue, 2015, Decision-Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires, Working paper, ETH Zurich.
Chen, Daniel L., and James J. Prescott, 2016, Implicit Egoism in Sentencing Decisions: First Letter Name Effects with Randomly Assigned Defendants .
Chen, Daniel L., and Martin Schonger, 2013, Social Preferences or Sacred Values? Theory and Evidence of Deontological Motivations, Working paper, ETH Zurich, Mimeo.
Chen, Daniel L., and Holger Spamann, 2014, This Morning's Breakfast, Last Night's Game: Detecting Extraneous Factors in Judging, Working paper, ETH Zurich.
Chen, Daniel L., and Susan Yeh, 2013, Distinguishing Between Custom and Law: Empirical Examples of Endogeneity in Property and First Amendment Precedents, *William & Mary Bill of Rights Journal* 21, 1081–1105.
Chen, Daniel L., and Susan Yeh, 2014a, Growth Under the Shadow of Expropriation? The Economic Impacts of Eminent Domain, Working paper, ETH Zurich and George Mason University.
Chen, Daniel L., and Susan Yeh, 2014b, How Do Rights Revolutions Occur? Free Speech and the First Amendment, Working paper, ETH Zurich.
Coles, Frances S., 1986, Forced to Quit: Sexual Harassment Complaints and Agency Response, *Sex Roles* 14, 81–95.
Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad, 2014, Family Welfare Cultures, *Quarterly Journal of Economics* 129, 1711–1752.
Deaton, Angus, 2010, Instruments, Randomization, and Learning about Development, *Journal of Economic Literature* 48, 424–455.
Dixon, Rosalind, 2010, Female Justices, Feminism, and the Politics of Judicial Appointment: A re-examination, *Yale Journal of Law and Feminism* 21, 297–338.
Dobbin, Frank, and Erin L. Kelly, 2007, How to Stop Harassment: Professional Construction of Legal Compliance in Organizations1, *American Journal of Sociology* 112, 1203–1243.
Eberts, Randall W., and Joe A. Stone, 1985, Male-Female Differences in Promotions: EEO in Public Education, *The Journal of Human Resources* 20, 504–521.
Edelman, Lauren B., 1992, Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law, *American Journal of Sociology* 97, 1531–1576.
Edelman, Lauren B., Christopher Uggen, and Howard S. Erlanger, 1999, The Endogeneity of Legal Regulation:

Grievance Procedures as Rational Myth, *American Journal of Sociology* 105, 406–54.

Eisenman, Russell, 1991, Gender and Racial Prejudice of Conservative College Women, *Psychological Reports* 68, 450–450.

Epstein, Richard A., 1995, *Forbidden Grounds: The Case Against Employment Discrimination Laws* (Harvard University Press).

Farhang, Sean, and Gregory Wawro, 2004, Institutional Dynamics on the U.S. Court of Appeals: Minority Representation Under Panel Decision Making, *Journal of Law, Economics, and Organization* 20, 299–330.

Fehr, Ernst, and Urs Fischbacher, 2002, Why Social Preferences Matter - The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives, *The Economic Journal* 112, C1–C33.

Fitzpatrick, Brian T, 2010, An Empirical Study of Class Action Settlements and Their Fee Awards, *Journal of Empirical Legal Studies* 7, 811–846.

Franks, Mary Anne, 2011, Unwilling Avatars: Idealism and Discrimination in Cyberspace, *Columbia Journal of Gender and Law* 20, 224–261.

Gennaioli, Nicola, and Andrei Shleifer, 2007, The Evolution of Common Law, *The Journal of Political Economy* 115, 43–68.

Goldman, Sheldon, 1999, *Picking Federal Judges: Lower Court Selection from Roosevelt Through Reagan* (Yale University Press).

Gruber, Jonathan, 1994, The Incidence of Mandated Maternity Benefits, *The American Economic Review* 84, 622–641.

Haire, Susan B., Donald R. Songer, and Stefanie A. Lindquist, 2003, Appellate Court Supervision in the Federal Judiciary: A Hierarchical Perspective, *Law & Society Review* 37, 143–168.

Hellerstein, Judith K., David Neumark, and Kenneth R. Troske, 2002, Market Forces and Sex Discrimination, *The Journal of Human Resources* 37, 353–380.

Hersch, Joni, 2011, Compensating Differentials for Sexual Harassment, *The American Economic Review Papers and Proccedings* 101, 630–634.

Hill, Catherine, and Holly Kearl, 2011, *Crossing the Line: Sexual Harassment at School* (American Association of University Women).

Jolls, Christine, 2001, Antidiscrimination and Accommodation, *Harvard Law Review* 115, 642–699.

Jordan, Samuel P., 2007, Early Panel Announcement, Settlement, and Adjudication, *Brigham Young University Law Review* 2007, 55–107.

Juliano, Ann, and Stewart J. Schwab, 2000, The Sweep of Sexual Harassment Cases, *Cornell Law Review* 86, 548–592.

Katz, Daniel Martin, Michael James Bommarito, Tyler Soellinger, and James Ming Chen, 2015, Law on the Market? Evaluating the Securities Market Impact of Supreme Court Decisions, *SSRN (August 24, 2015)* .

Keele, Denise M., Robert W. Malmsheimer, Donald W. Floyd, and Lianjun Zhang, 2009, An Analysis of Ideological Effects in Published Versus Unpublished Judicial Opinions, *Journal of Empirical Legal Studies* 6, 213–239.

Klarman, Michael, 2004, *From Jim Crow to Civil Rights: The Supreme Court and the Struggle for Racial Equality* (Oxford University Press, New York).

Kling, Jeffrey R., 2006, Incarceration Length, Employment, and Earnings, *The American Economic Review* 96, 863–876.

Lindbeck, Assar, and Dennis J. Snower, 1988, Cooperation, Harassment, and Involuntary Unemployment: An Insider-Outsider Approach, *The American Economic Review* 78, 167–188.

Lindbeck, Assar, and Dennis J. Snower, 2001, Insiders versus Outsiders, *The Journal of Economic Perspectives* 15, 165–188.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand, 2013, Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt, *American Economic Review* 103, 1797–1829.

Merritt, Deborah Jones, and James J. Brudney, 2001, Stalking Secret Law: What Predicts Publication in the United States Courts of Appeals, *Vanderbilt Law Review* 54, 69–121.

Nash, Jonathan R., 2015, Examining Federal District Judges' Referrals to Magistrate Judges, in *2015 Annual Meeting of the International Society of New Institutional Economics (ISNIE)*, number 1-44.

Neal, Derek, 2004, The Measured Black - White Wage Gap among Women Is Too Small, *Journal of Political Economy* 112, S1–S28.

Neumark, David, 2016, Experimental Research on Labor Market Discrimination, Technical report, National Bureau of Economic Research.

Neumark, David, and Wendy A. Stock, 2006, The Labor Market Effects of Sex and Race Discrimination Laws, *Economic Inquiry* 44, 385–419.

Newman, Meredith A., Robert A. Jackson, and Douglas D. Baker, 2003, Sexual Harassment in the Federal Workplace, *Public Administration Review* 63, 472–483.

Nielsen, Laiura B., Robert L. Nelson, and Ryon Lancaster, 2010, Individual Justice or Collective Legal Mobilization? Employment Discrimination Litigation in the Post Civil Rights United States, *Journal of Empirical Legal Studies* 7, 175–201.

Olivetti, Claudia, and Barbara Petrongolo, 2008, Unequal Pay or Unequal Employment?: A Cross-Country Analysis

of Gender Gaps, *Journal of Labor Economics* 26, 621–654.
O'Neill, June, 2003, The Gender Gap in Wages, circa 2000, *The American Economic Review* 93, 309–314.
Peresie, Jennifer L., 2005, Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts, *The Yale Law Journal* 114, 1759–1790.
Pierce, Chester, 1970, Offensive mechanisms, *The black seventies* 265–282.
Posner, Richard A., 1998, Against Constitutional Theory, *New York University Law Review* 73, 1–22.
Rowe, Mary, 1981, The minutiae of discrimination: The need for support, *Outsiders on the inside: Women and organizations* 155–71.
Schultz, Vicki, 1990, Telling Stories about Women and Work: Judicial Interpretations of Sex Segregation in the Workplace in Title VII Cases Raising the Lack of Interest Argument, *Harvard Law Review* 103, 1749–1843.
Schultz, Vicki, 1998, Reconceptualizing sexual harassment, *The Yale Law Journal* 107, 1683–1805.
Sen, Maya, 2015, Is Justice Really Blind? Race and Appellate Review in U.S. Courts, *Journal of Legal Studies* 44, (In Press).
Shoag, Daniel, 2011, The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns, Working paper, Harvard University, Mimeo.
Steinbuch, Robert, 2009, An Empirical Analysis of Reversal Rates in the Eighth Circuit During 2008, *Loyola of Los Angeles Law Review* 43, 51–19.
Stock, James H., and Motohiro Yogo, 2005, Testing for Weak Instruments in Linear IV Regression, in Donald W.K. Andrews, and James H. Stock, eds., *Identification and Inference for Economic Models: Essays in Honor of Thomas Rothenberg*, 80–108 (Cambridge University Press, Cambridge, MA).
Summers, Lawrence H., 1989, Some Simple Economics of Mandated Benefits, *The American Economic Review* 79, 177–183.
Summers, Lawrence H., 2005, Remarks at NBER Conference on Diversifying the Science and Engineering Workforce, in *National Bureau of Economic Research Conference on Diversifying the Science and Engineering Workforce. Cambridge, MA. Retrieved March*, volume 9.
Sunstein, Cass R., David Schkade, Lisa M. Ellman, and Andres Sawicki, 2006, *Are Judges Political?: An Empirical Analysis of the Federal Judiciary* (Brookings Institution Press).
Taha, Ahmed E., 2004, Publish or Paris? Evidence of How Judges Allocate Their Time, *American Law and Economics Review* 6, 1–27.
Taha, Ahmed E., 2009, Judge Shopping: Testing Whether Judges' Political Orientations Affect Case Filings, *University of Cincinnati Law Review* 20, 101–135.
Waldfogel, Joel, 1995, The Selection Hypothesis and the Relationship between Trial and Plaintiff Victory, *The Journal of Political Economy* 103, 229–260.

Figure 1

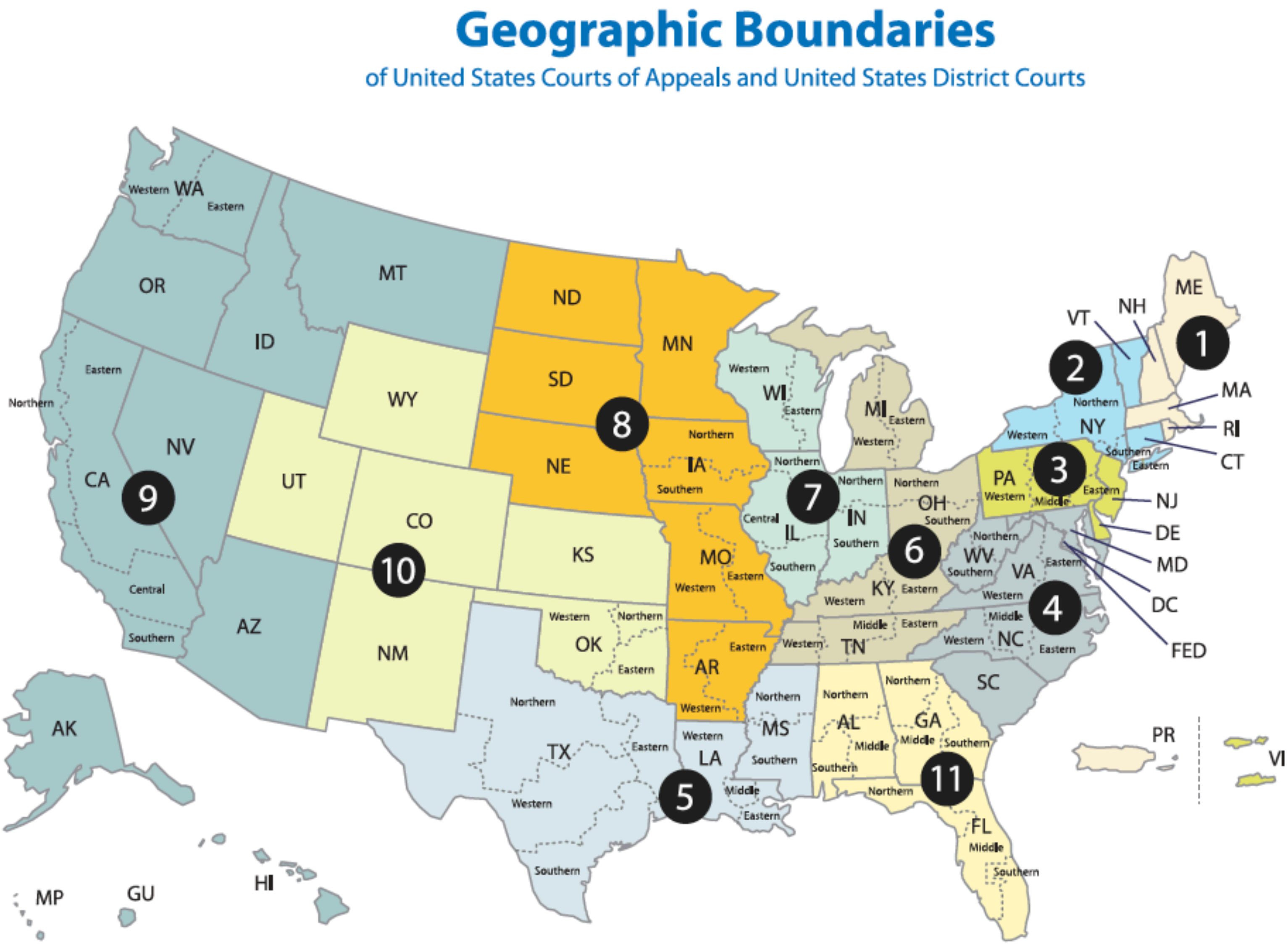

Geographic Boundaries
of United States Courts of Appeals and United States District Courts
WA
Western
Eastern
OR
MT
ID
ND
MN
SD
WY
NE
IA
Northern
Southern
WI
Western
Eastern
MI
Eastern
Western
CA
Northern
Eastern
Central
Southern
NV
UT
CO
AZ
NM
KS
MO
Western
Eastern
OK
Western
Northern
Eastern
AR
Eastern
Western
TX
Northern
Eastern
Western
Southern
LA
Western
Middle
Eastern
MS
Northern
Southern
IL
Northern
Central
Southern
IN
Northern
Southern
OH
Northern
Southern
KY
Western
Eastern
TN
Western
Middle
Eastern
AL
Northern
Middle
Southern
GA
Northern
Middle
Southern
FL
Northern
Middle
Southern
SC
NC
Western
Middle
Eastern
VA
Western
Eastern
WV
Northern
Southern
PA
Western
Middle
Eastern
NY
Western
Northern
Southern
Eastern
ME
NH
VT
MA
RI
CT
NJ
DE
MD
DC
FED
AK
HI
GU
MP
PR
VI
1
2
3
4
5
6
7
8
9
10
11

Figure 2: Sexual Harassment Decisions, 1982-2002

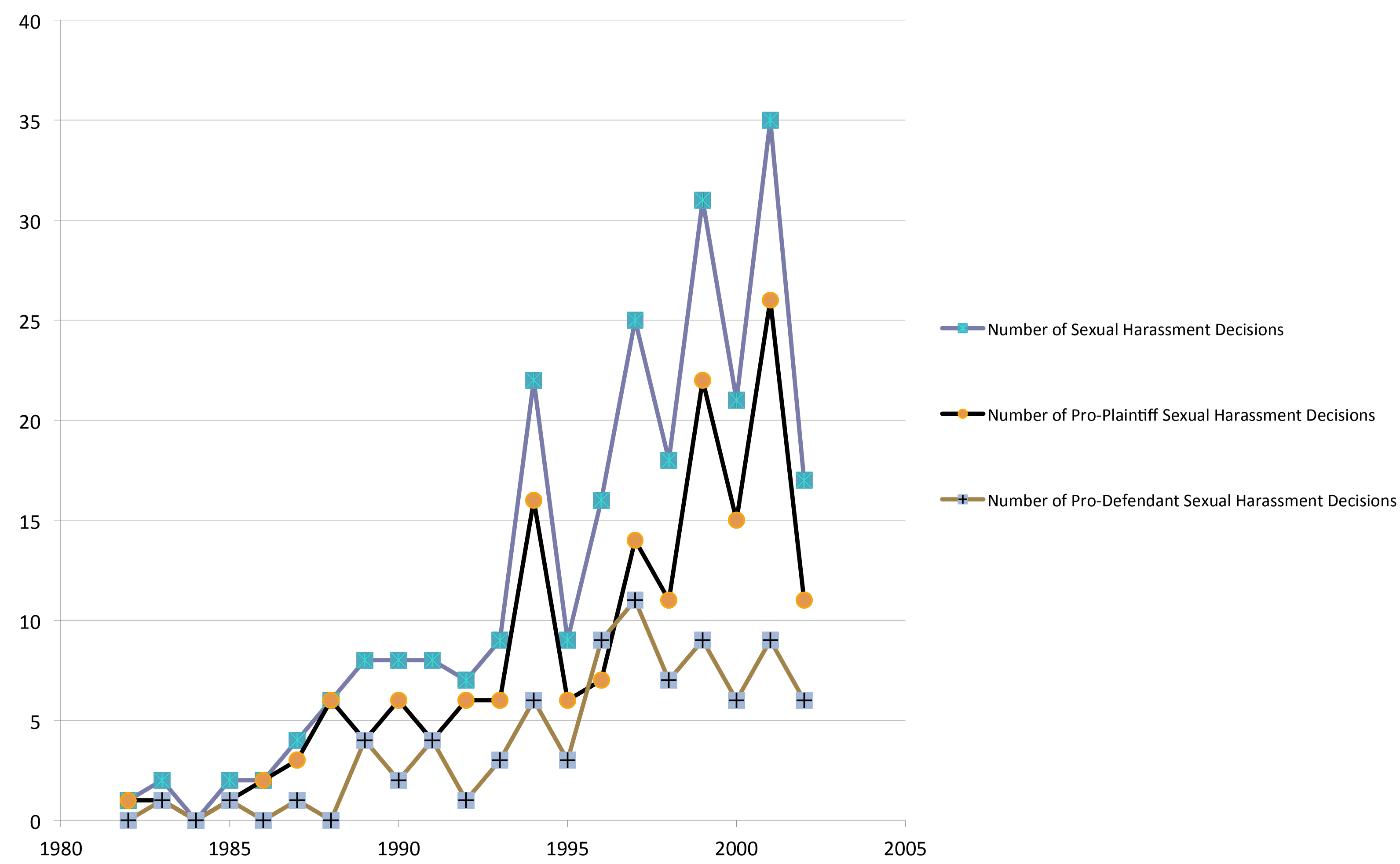


Figure 3A

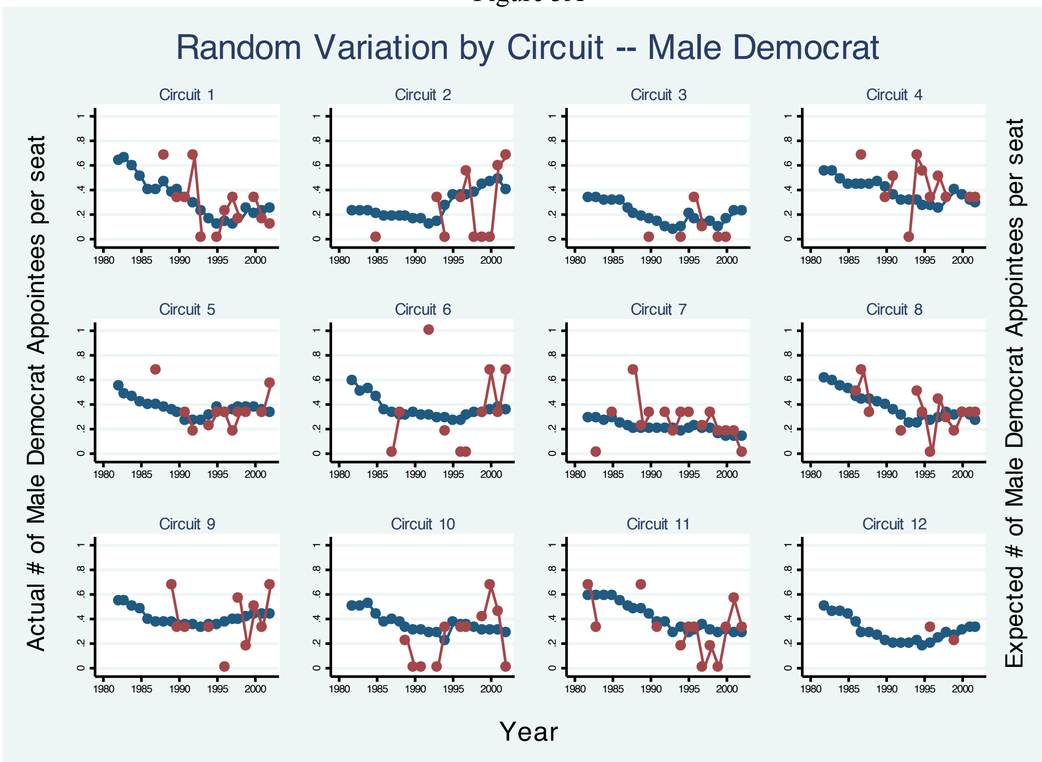


Figure 3B

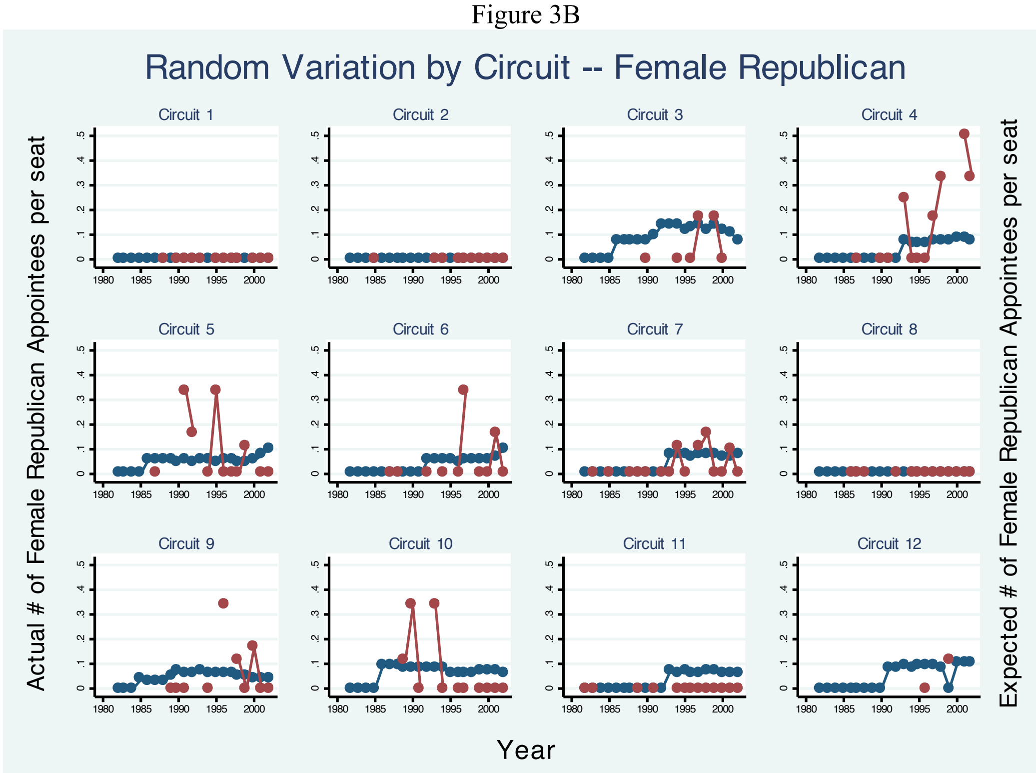


Figure 4: Randomization Check

P-Values of Male Democrat strings

P-Values of Female Republican strings

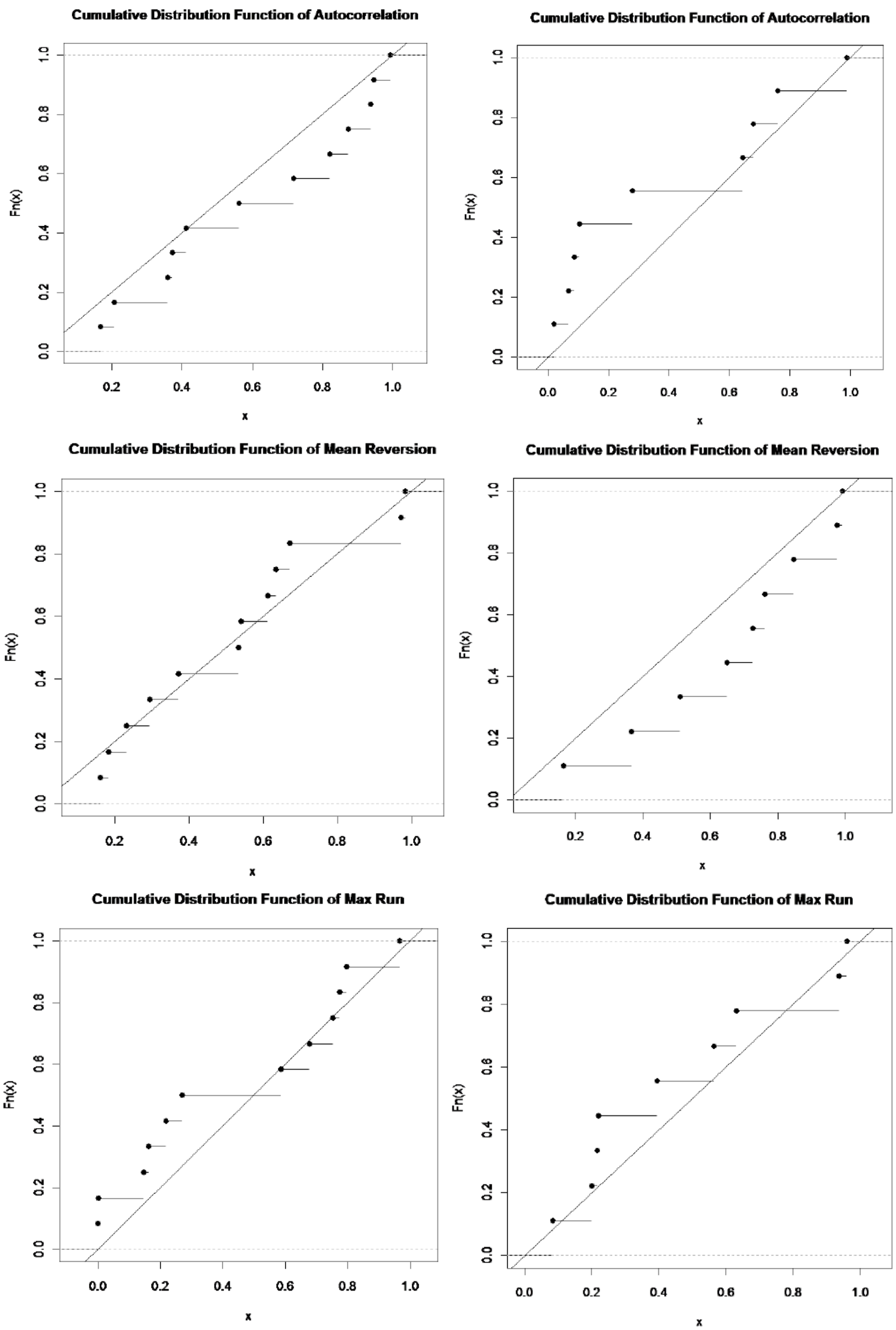
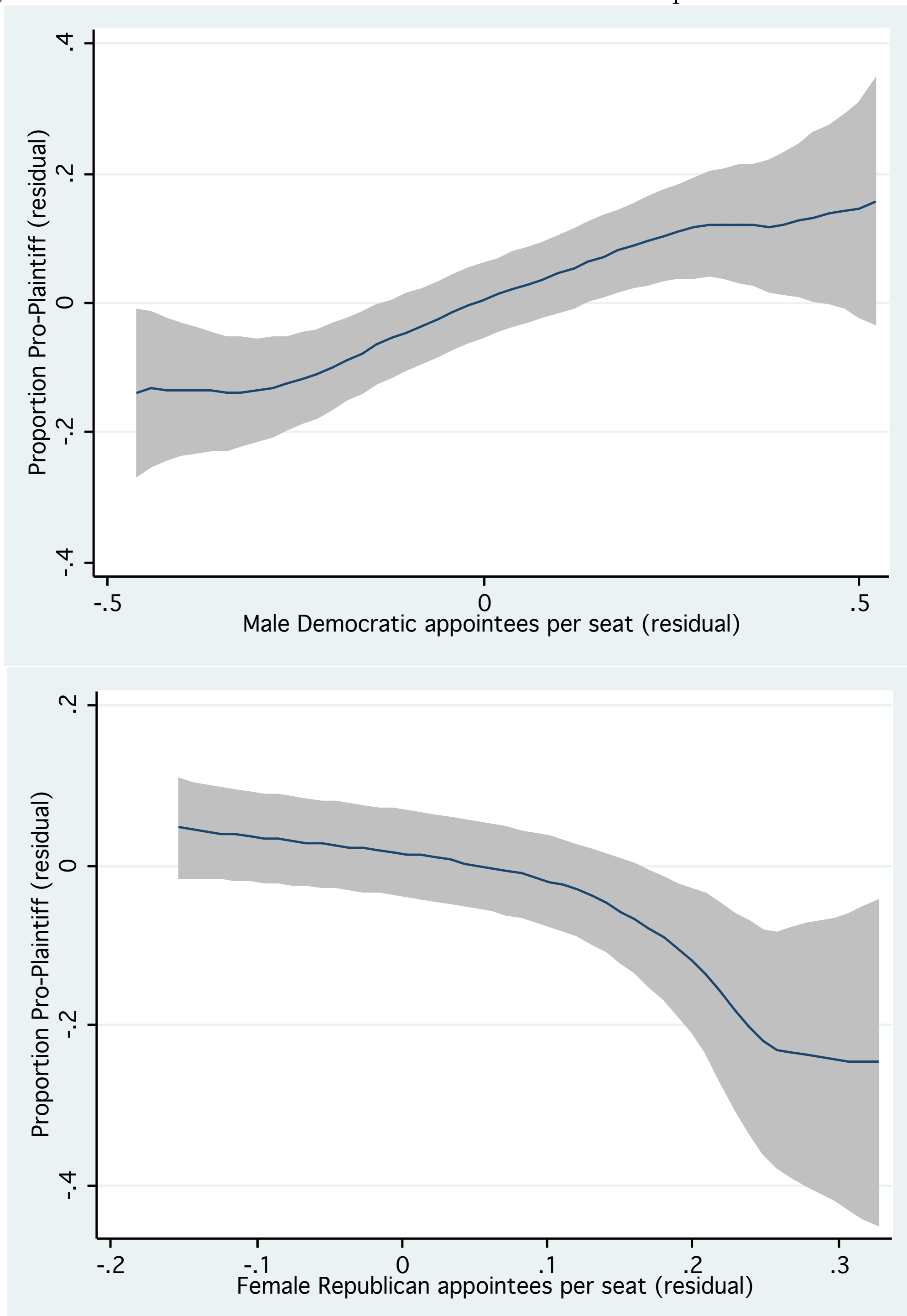
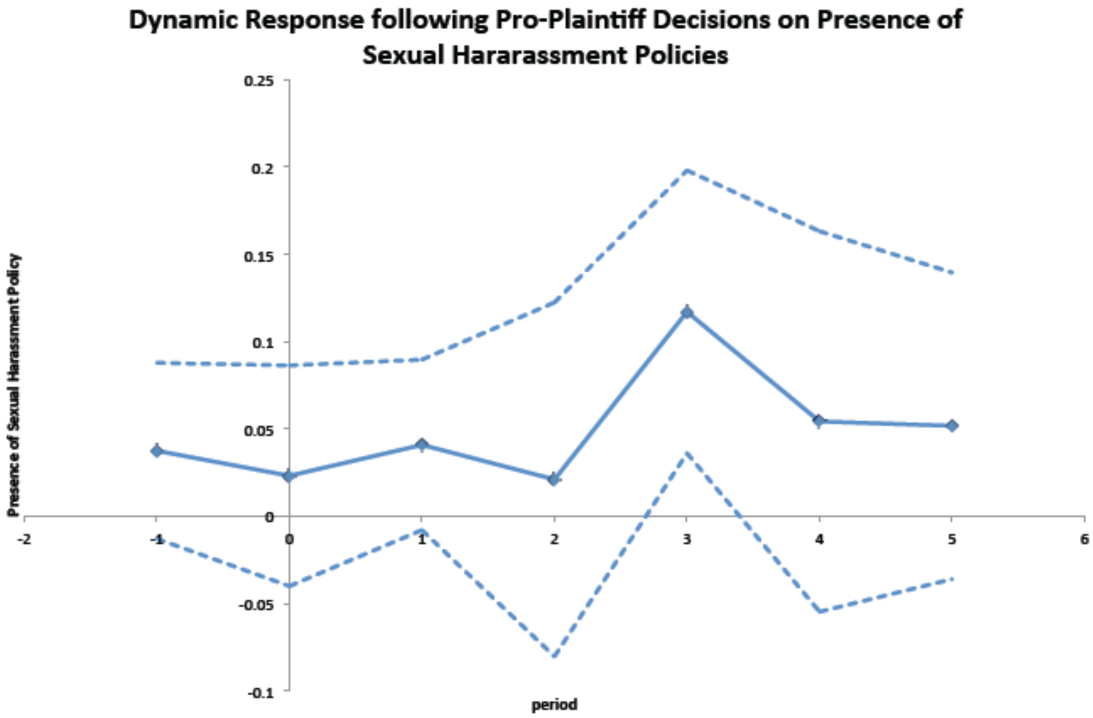


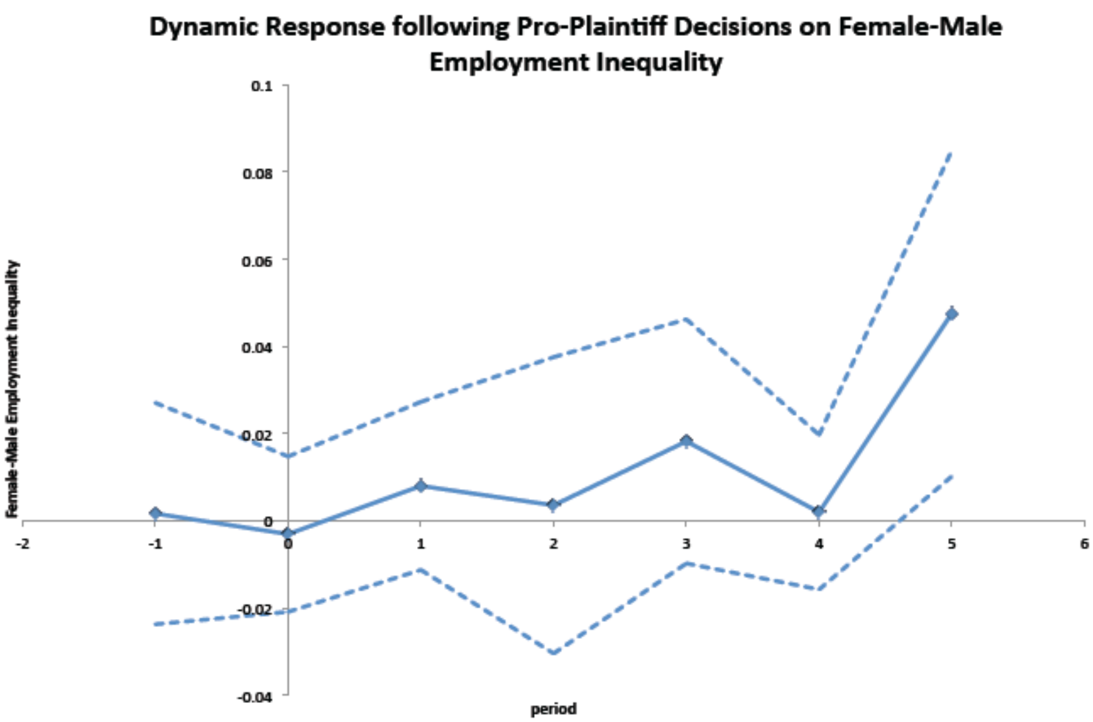
Figure 5A: Pro-Plaintiff Sexual Harassment Decisions and Composition of Judicial Panel



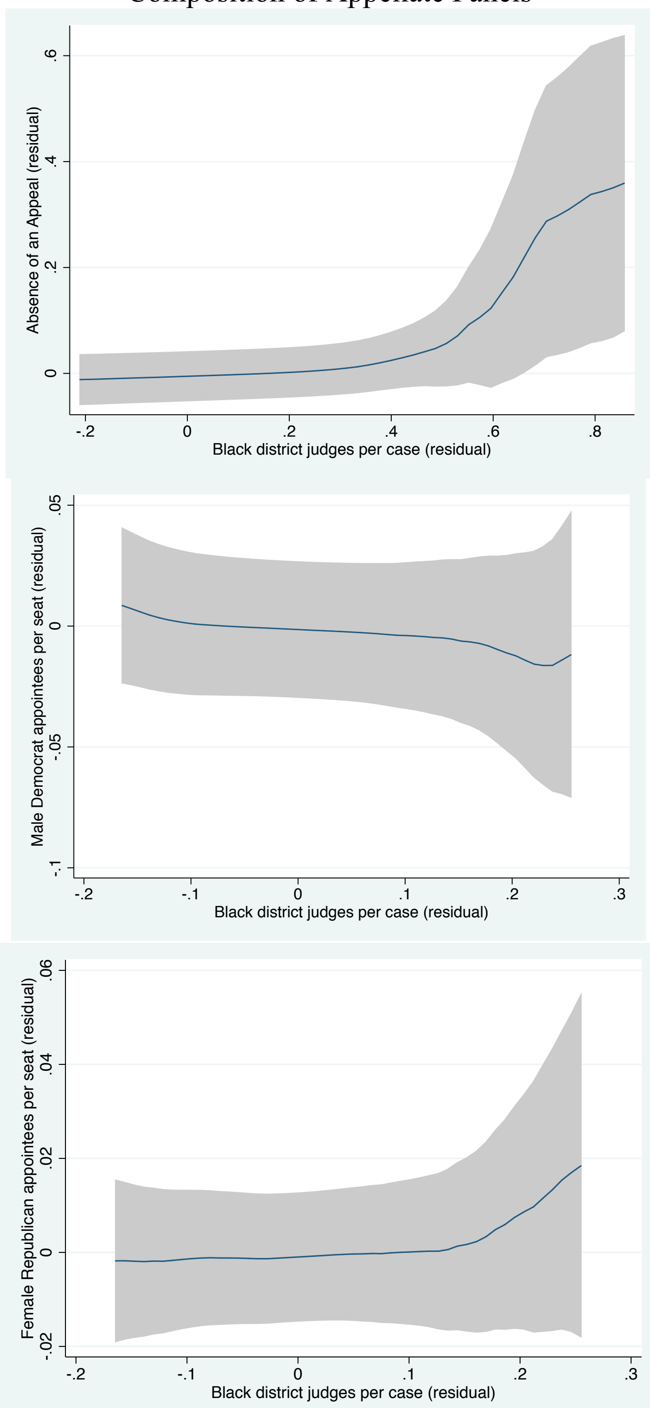
Nonparametric local polynomial estimates are computed using an Epanechnikov kernel. The bandwidth is 0.14 for the male Democrat graph and 0.09 for the female Republican graph. Shaded area indicates 90 percent confidence bands. The residuals are calculated removing circuit and year fixed effects.

Figure 6: Sexual Harassment Law and Labor Markets





Appendix Figure 1: Composition of District Panels, Absence of Appellate Panels, and Composition of Appellate Panels



Nonparametric local polynomial estimates are computed using an Epanechnikov kernel. The bandwidth is 0.24 for the absence of appellate panels graph and 0.11 and 0.09 for the male Democrat and female Republican graph. Shaded area indicates 90 percent confidence bands. The residuals are calculated removing circuit and year fixed effects.

Appendix

A *Major Doctrinal Developments in Sexual Harassment Law*

1964 – **Title VII** – prohibits sex discrimination in employment.

1976 – **Williams v. Saxbe** – Court recognized sexual harassment as a form of sex discrimination when sexual advances by male supervisor towards female employee, if proven, would be deemed an artificial barrier to employment placed before one gender and not another.

1977 – **Barnes v. Costle** – U.S. Court of Appeals for the Second District ruled that retaliation against a female employee for rejecting sexual advances of her boss is a violation of Title VII's prohibition against sex discrimination.

1980 – **EEOC** issues guidelines forbidding "sexual harassment" as a form of sex discrimination.

1985 – **McKinney v. Dole** - U.S. Court of Appeals for the DC Circuit ruled that physical violence, even if it is not overtly sexual, can be sexual harassment if the unwelcome conduct is based on the victim's gender.

1986 – **Meritor Savings Bank, FSB v. Vinson** – The Supreme Court first recognized "sexual harassment" as a violation of Title VII and established the standards for analyzing whether the conduct was welcome and levels of employer liability.

1988 – **Hall v. Gus Construction** - U.S. Court of Appeals for the Eighth District finds that when male construction workers "hazed" three female colleagues, even if the conduct was not specifically sexual in nature, was gender based harassment.

1991 – **Ellison v. Brady** – Changed analysis of conduct from reasonable person to reasonable women test when determining whether actionable sexual harassment occurred.

1991 – **Civil Rights Act of 1991** provides for jury trials and for increased damages in Title VII sexual harassment suits.

1993 – **Harris v. Forklift Systems, Inc** – plaintiff may bring sexual harassment claim without necessarily showing psychological harm. In addition to Meritor, the factors when analyzing whether sexual harassment occurred include: (i) Frequency of conduct; (ii) Severity; (iii) Whether the conduct is physically threatening or humiliating; (iv) Or is a mere offensive utterance; (v) And whether the conduct unreasonably interferes with employees work performance; (vi) No Single Factor is Required but Totality of the Circumstances Test.

1998 – **Faragher v. City of Boca Raton** - Supreme Court decision that establishes that an employer is subject to vicarious liability for hostile environment created by a supervisor unless the employer can demonstrate that it exercised reasonable care to prevent and correct promptly any sexually harassing behavior and that the employee unreasonably failed to take advantage of any preventative or corrective opportunities provided by the employer.

1998 – **Burlington Industries, Inc v. Ellerth** - Companion Supreme Court decision to Faragher that further elaborates that the employer's "Faragher" defense to vicarious liability is not available if the employee suffers a tangible job consequence as result of supervisor's actions.

APPENDIX TABLE I

First Stage Relationship Between Absence of Circuit Cases and Judicial Composition of District Court Cases, 1986-2002

Circuit-Year Level	Outcome: Absence of Circuit Cases$_t$			
	(1)	(2)	(3)	(4)
Black judges per District case$_t$	0.470*	0.485**	0.419*	0.459**
	(0.176)	(0.138)	(0.145)	(0.144)
Black judges per District case$_{t+1}$		-0.0492	-0.0248	
		(0.369)	(0.357)	
Black judges per District case$_{t+2}$			-0.0192	
			(0.374)	
Circuit-year controls	N	N	N	Fixed Effects
N	203	190	177	203
R-sq	0.019	0.020	0.014	0.372

Notes: Significant at +10%, *5%, **1%. Heteroskedasticity-robust standard errors are in parentheses. Observations are clustered at the Circuit level. Fixed effects are dummy indicators for Circuit and for year.

B District IV

We report a first-stage correlation between the presence of a Circuit case and the proportion of District cases in a Circuit that received a Black judge, with an F-statistic of 7. With the inclusion of Circuit and year fixed effects, the point estimates are identical and the F-statistic goes up to 10 as can be seen from comparing Columns 1 and 4 in Appendix Table I. The average Circuit-year had 8% of District cases ruled by a Black judge, which decreased the probability the Circuit-year had an appeal by 46% (Appendix Table I and Appendix Figure 1). Sen (2015) also reports that ethnicity of the District judge, in particular, being Black, predicts differences in how their opinions are treated.

Columns 2 and 3 of Appendix Table I show that the proportion of District cases heard by a Black judge is not related to the presence of a Circuit case in the previous one or two years. The proportion of District cases in a Circuit that received a Black judge is uncorrelated with the number of male Democrats (female Republicans) per seat in Circuit cases as indicated in the second and third figures of Appendix Figure 1. For 1982-1985, most Circuits did not have any District-level sexual harassment cases so when the District IV is employed, those years are dropped from the sample. Since we may expect a lag between District and Circuit Court rulings, we also considered the relationship between the presence of a Circuit case and the previous year's assignment of District judges to sexual harassment cases. Regardless of the source of variation, the two-stage least squares estimates are very similar, which is consistent with the 2SLS results being unaffected by the inclusion of District instruments.