Making Background Salient: The Effects of Open Access Criminal Databases on Offender Behavior*

Erika Forsberg† Hans Grönqvist‡ Susan Niknami§ Mårten Palme¶

Preliminary and Incomplete Draft - August 8, 2024

Abstract

Technological advances have led to the creation of large online criminal databases easily accessible to the general public. However, the impact of this accessibility on the offenders in these databases remains unclear. We investigate this by studying the launch of Sweden's first online criminal database, which allows for anonymous and free searches for individuals charged with a crime. The platform transformed the process of criminal background checks in Sweden, drawing 32 million checks in the first year alone. Leveraging administrative rules that restricted the identification of individuals charged before specific dates, we estimate the effects by comparing outcomes of exposed and non-exposed individuals. We find significant adverse effects of exposure on earnings but not on employment or criminal recidivism. However, there are significantly stronger detrimental effects on both labor market outcomes and recidivism in defendant subgroups; those with at least a high school degree, acquitted individuals, and those either charged with a misdemeanor or living in places with a relatively low concentration of ex-criminals. In these groups we also find adverse effects on educational performance among the children of the defendants. When investigating the sources underlying these results, we find that exposed workers are more likely to work in firms that pays a lower earnings premia. Our results suggest that stigma may be an important mechanism explaining responses to criminal justice interactions. Our findings inform the ongoing debate on what type of information should be disclosed in background checks.

Keywords: *Background checks*; *Informal sanction*; *Stigma*; *Mark of a criminal record*; *Spillovers*; *Recidivism*;
JEL Codes: K42

*This paper has benefited from discussions with Anna Bindler, Eric Chyn, Will Dobbie, Thomas Giebe, Jeff Grogger, Rasmus Landersø, Jens Ludwig, Enrico Moretti, Peter Nilsson, Sam Norris, Aurelie Ouss, Jonathan Tebes, Crystal Yang and seminar participants at the LSE/Chicago conference on the economics of crime and justice 2024, CESifo Area Conference on Labor Economics, Uppsala Labor Group, IFAU, IFN and Linnaeus University. Adam Birgersson and Gabriel Nielsen provided excellent research assistance. We thank Marcos Demetry for generously sharing his data. We are also grateful to the company behind the database for providing some of the data used in the analysis and for graciously taking the time answering our questions. Funding from Handelsbankens forskningsstiftelser and the Swedish Research Council is gratefully acknowledged.

†Uppsala University and IFAU. erika.forsberg@ifau.uu.se
‡Linnaeus University and IFN. hans.gronqvist@lnu.se
§Stockholm University. susan.niknami@sofi.su.se
¶Stockholm University. marten.palme@ne.su.se

1 Introduction

Criminal background information is arguably more readily available to the general public today than ever before. Large online databases containing information on criminal offenders, accessible through state agencies or private background check firms, have substantially reduced the cost for citizens, employers, and landlords to access criminal records.[1] The effect of these databases on the outcomes of the offenders is heavily debated. On the one hand, it is possible that being included in criminal databases could improve the outcomes of offenders through specific deterrence (Anker, Doleac and Landersø, 2021; Chalfin and McCrary, 2017). On the other hand, the information could send a negative signal that harm rehabilitation efforts. The potential for this is supported by results in quasi-experimental studies showing that arrests and incarceration reduce formal sector employment (Mueller-Smith, 2015; Dobbie, Goldin and Yang, 2018; Dobbie et al., 2018; Grenet, Grönqvist and Niknami, 2024). Because of this concern, many states in the US have implemented reforms to conceal criminal background information, e.g. through expungement (Prescott and Starr, 2020; Agan et al., 2024; Selbin, McCrary and Epstein, 2018), sealing (Mueller-Smith and Schnepel, 2021), or "Banning-the-Box" (Doleac and Hansen, 2020; Rose, 2021). While the literature has mostly focused on how employers respond to this type of information,[2] the behavioral adjustments of the offenders actually included in these databases are much less clear. Most empirical studies also struggle with the fact that having a criminal record is not random and therefore likely correlates with unobserved risk factors.

We address these issues by studying the effects of accessing criminal records within the context of the unexpected launch of Sweden's first online criminal database. The platform, introduced in January 2014, transformed the way criminal background checks were conducted by enabling the general public to anonymously search by name for individuals charged with a crime. The criminal background information is sourced from records mandated by law to be provided by courts. Unlike most criminal databases, which offer limited background information on specific crime types (e.g., sex crimes) only for some groups of offenders and target certain users (e.g., employers or landlords), this database covers all criminal charges in Swedish courts and is accessible to everyone without the need for a user profile or any costs. Search results map name to current age and place of residence and reveal whether an individual has faced a criminal charge. For a nominal fee (approximately USD 5), users can obtain the complete verdict, detailing the crime type, conviction year, sentence, and any additional circumstances of the crime. The platform rapidly garnered widespread use. By 2019, when competing platforms were finally launched, searches for criminal records on the site had neared 100 million in total, averaging around 10 searches per person in Sweden.

To identify the causal effects of revealing criminal background information, we exploit a legislative rule which stipulates that courts should not disclose individual information on criminal

[1] This development has been particularly stark in the United States where it is estimated that the three largest background screening companies alone (out of approximately 2,000) in 2019 conducted 56 million searches NCLC, 2019).

[2] See e.g. Pager (2003); Holzer, Raphael and Stoll (2006); Agan and Starr (2018); Cullen, Dobbie and Hoffman (2024); Agan et al. (2024); Doleac and Hansen (2020); Rose (2021); Finlay (2008).

charges dating back more than five years since the inquiry. Consequently, this legislation created a time window beyond which the company behind the platform was unable to access criminal records information. By analyzing data provided by the company we were able to determine the specific cutoff month for each court, which depend on the date when the company applied for and was later granted access to the records. We then utilize administrative data containing information on the complete criminal histories of individuals. These data enable us to distinguish between individuals charged after the specified cutoff date, resulting in their criminal records being exposed online, and non-exposed defendants who were charged prior to the cutoff date. The data also contains a wide array of labor market, educational, and demographic characteristics, which allows us to investigate behavioral responses of the defendants.

We start by verifying empirically that individuals on both sides of the cutoff date are similar in their baseline characteristics and that the pre-reform trends for exposed and non-exposed offenders are parallel. Moreover, we show that there is no evidence of bunching of cases or changes in offender characteristics near the cutoff. These findings confirm that neither courts nor offenders were able to foresee and respond to the future arrival of the platform.

We then proceed to the results from our difference-in-differences model which accounts for correlated unobservables by contrasting the change in outcomes for exposed versus non-exposed defendants. Since all individuals in our sample were charged with a crime and exposure is defined before the platform was introduced our empirical design also addresses potential concerns of general deterrence effects. The results suggest that exposure significantly decrease log earnings by 9.3 percent. This effect size corresponds to about one quarter of that from event study estimates of the effect of a criminal charge on earnings in the period before the platform was introduced. We find no significant effect on labor supply at the extensive margin or criminal recidivism. There are, however, significantly stronger detrimental effects on both labor market outcomes and recidivism in defendant subgroups: those with at least a high school degree, acquitted individuals, native born, and those either charged with a misdemeanor or living in places with a relatively low concentration of ex-criminals. These are groups of the population with arguably a relatively higher opportunity cost of engaging in crime. In many of these groups we also find adverse effects on educational performance among the children of the defendants.

When trying to identify the sources of the earnings losses we find small and mostly insignificant exposure effects on behaviors such as enrolling in labor market training programs, residential mobility, family formation, and sorting across industries based on the share of ex-offenders working in a given industry. However, we find a significant differential effect in defendant sorting with respect to the firm earnings premium estimated using the methods suggested by Abowd, Kramarz and Margolis (1999). Our results suggest that workers in the subgroups most strongly affected by the mark of a criminal record are more likely to work in firms that pay a lower earnings premia. This observation aligns with job ladder models, which suggest that workers continuously search for better job opportunities throughout their careers, gradually transitioning to higher-paying employers. In these models, pay disparities among firms reflect differences in how the surplus is distributed between workers and employers (Burdett and Mortensen, 1998; Haltiwanger et al., 2018; Lachowska, Mas and Woodbury, 2020). Within this framework,

affected workers suffer earnings losses due to the impact of moving down the job ladder to employers that offer lower wages.

Our results are robust to a battery of specification checks, including changing the bandwidth for inclusion in our sample, and we also show that there is no significant effect when assigning court-specific placebo cutoff dates one year before and one year after the actual cutoff date.

Our study provides some of the very first pieces of evidence regarding the causal effect of open access to online criminal background information on the behavior of offenders. One of few exceptions is the study by Finlay (2008) who uses cross-state cross-time variation to study the labor market effects of employer access to state-provided online information on incarceration history for individuals with moderate to long prison sentences. The results show adverse employment consequences in states with open records policies.[3] Our study differs in that the database includes the universe of individuals charged with a crime regardless their sentence and type of crime committed. This is important since, for instance, the ciminal records of felony conviction may have different consequences for defendants, relative to criminal records of misdemeanor charges. We are also able to document the effects for a wide range of outcomes as well as probe underlying behavioral mechanisms.

Our results also add to the literature on the impact of criminal background checks from the employers' perspective.[4] This body of literature either studies the callback rate to fictitious job applications signaling criminal background or examines policies intended to influence the hiring practices of employers. A common finding in the former line of research is that employers are averse to calling back or hiring job applicants with criminal records (Pager, 2003; Holzer, Raphael and Stoll, 2006; Agan and Starr, 2018; Uggen et al., 2014; Cullen, Dobbie and Hoffman, 2024).[5] The latter line of research primarily focuses on Ban-the-Box (BTB) legislation, which restricts inquiries about criminal histories on job applications until later in the hiring process. It often relies on state-level variations in these laws to address correlated unobservables. Doleac and Hansen (2020) demonstrate that these policies decrease the probability of employment by 3.4 percentage points for young, low-skilled black men, suggesting that in the absence of information about an applicant's criminal history, employers statistically discriminate against demographic groups with higher proportions of ex-offenders. Rose (2021) finds that a 2013 Seattle BTB law had minimal effects on ex-offenders' labor market outcomes, indicating that employers may delay background checks until the final stages of the hiring process or that

[3]Finlay (2008) studies a sample of individuals included in the National Longitudinal Survey of Youth (NLSY), in which 369 offenders had received prison sentences for a minimum of one year.

[4]These studies are typically motivated by the fact that individuals with a criminal record are more likely to be unemployed. For instance, Bushway et al. (2022) show that 46 % of unemployed men in the United States have been convicted. Other studies include e.g. Western (2002) and Freeman (1992).

[5]For instance, Pager (2003) demonstrates that disclosing a criminal record to employers who do not explicitly request such information has minimal impact on hiring decisions. Agan and Starr (2018) reveal a significant six-fold increase in callback rates for white-to-black applicants following the implementation of Ban-the-Box legislation in New Jersey and New York. Cullen, Dobbie and Hoffman (2024) conduct experimental research to assess strategies for increasing the demand for workers with criminal records. Their study involves presenting hiring managers with hypothetical scenarios that influence whether workers with criminal records can accept job offers in the future. Results indicate that initially, 39% of employers are open to hiring individuals with criminal records. However, this figure rises to over 50% when managers are offered incentives such as crime insurance, a single performance review, or a limited background check covering only the past year.

ex-offenders may gravitate towards jobs where clean records are not obligatory.[6] While the responses of employers to criminal background information represents a potentially important mechanism, the setting we consider allows us to investigate the effects of criminal background information that is much more widely available also to the general public.

A few recent studies examine policies that not necessarily only target employers. Agan et al. (2024) study the labor market impacts of retroactively reducing felonies to misdemeanors in San Joaquin County, CA where criminal justice agencies reclassified certain theft and drug offenses from felonies to misdemeanors, without requiring input or action from affected individuals. The results reveal no discernible benefits among individuals whose records were reduced proactively.[7] Mueller-Smith and Schnepel (2021) study diversion programs in Harris County, Texas, providing first-time felony defendants with an opportunity to circumvent a criminal record either by reclassifying drug and property offenses or by diverting low-risk defendants. Their findings suggest that these programs significantly reduce recidivism rates and improve quarterly earnings by about 50 percent. We add to this literature by studying an intervention that not only affected certain groups of felony offenders but the universe of individuals charged with a crime. We also contribute by using data that allows us to shed light on the sources underlying the labor market penalty of having a criminal record.

Importantly, our results are connected to an extensive literature that examines the effects of criminal justice sanction strictness on specific deterrence, often finding mixed evidence.[8] However, while the formal sanctions studied in this literature (e.g. incarceration or community service) are the criminal justice system's primary tool to discourage deviant behavior, many scholars also argue that informal social sanctions attached to the stigma of being marked as a criminal may potentially also act as a crime deterrent (Chalfin and McCrary, 2017; Zimring, Hawkins and Vorenberg, 1973; Nagin and Pogarsky, 2003; Weibull and Villa, 2005; Funk, 2004).[9] In this case, conveying socially valuable information on criminal background may actually be an efficient low-cost strategy for society to combat crime.[10] On the other hand, the criminal labeling theory, one of the most fundamental models in criminology, predicts that exposing criminal

[6]Furthermore, Shoagii and Veugeriii (2016) estimate the effects of BTB by considering impacts on residents of high-crime versus low-crime neighborhoods. The results show positive effects on employment in high-crime neighborhoods, which the authors rationalize by minority men may benefit from the law overall, despite negative impacts on some sub groups as suggested by Doleac and Hansen (2020). Jackson and Zhao (2017) use unemployment insurance records to study a 2010 BTB reform in Massachusetts in a difference-in-differences framework. Their results suggest that BTB lowered ex-offender's employment by 2.4 percentage points and quarterly earnings by USD 300.

[7]Agan et al. (2024) also conduct a field experiment notifying a subset of individuals about their proactive reduction also showing null results, implying that lack of awareness is unlikely to explain the findings.

[8]Specific studies include but are not limited to: Agan, Doleac and Harvey (2023); Abrams (2012); Aizer and Doyle Jr (2015); Barbarino and Mastrobuoni (2014); Bhuller et al. (2020); Buonanno and Raphael (2013); Dobbie, Goldin and Yang (2018); Dobbie et al. (2018); Drago, Galbiati and Vertova (2009); Garin et al. (2023); Helland and Tabarrok (2007); Hjalmarsson and Lindquist (2022); Kessler and Levitt (1999); Kling (2006); Kuziemko (2013); Lee and McCrary (2017); Mueller-Smith (2015); Owens (2009); Grönqvist, Grenet and Niknami (2022); Ouss et al. (2023); Rose and Yotam (2021); Raphael (2014).

[9]In fact, in the Becker (1968) economic model of crime, informal sanction costs alone can fully predict criminal behavior if individuals are more sensitive to changes in the probability of apprehension, thus acquiring a criminal record, than to changes in the sanction regime (Chalfin and McCrary, 2017; Nagin, 2013).

[10]See e.g. Mungan (2016); Nagin (2013); Polinsky and Shavell (2007); Kahan and Posner (1999); Rasmusen (1996).

background will amplify offending behavior. According to this view, interactions with the criminal justice system labels the offender a criminal, which leads him to internalize stigmatizing attitudes and conform to a deviant identity (e.g. Becker, 1968; Lemert, 1967).[11] However, since a conviction almost always involves at least one sanction, it has proven almost impossible to separate informal sanction costs from the imposed formal ones (Chalfin and McCrary, 2017).[12] This is important as it is fully possible that these mechanisms operate in different direction.

A few recent studies provide quasi-experimental evidence from contexts that arguably approaches distinguishing between these effects. Agan, Doleac and Harvey (2023) estimate the causal effects of misdemeanor prosecution on defendants' subsequent criminal justice involvement. The analysis exploits the quasi-random assignment of nonviolent misdemeanor cases to District Attorneys in Suffolk County, MA, who vary in their leniency of prosecuting decisions. The results show that nonprosecution of a nonviolent misdemeanor offense leads to a 53 percent reduction in the likelihood of a new criminal complaint. These effects are largest for defendants without prior criminal records, suggesting that averting criminal record acquisition is an important mechanism. Kamat, Norris and Pecenoc (2024) use a random-judge design and show that misdemeanor convictions cause an increase in the number of new offenses committed over the following five years while incarceration on more serious felony charges has no effect after release. The study by Mueller-Smith and Schnepel (2021) on diversion programs arguably also approaches distinguishing between these effects. While these studies compellingly empirically rule out many alternative explanations the authors are also clear that there could also be other mechanisms aside from stigma that the data do not allow them to eliminate.[13] To our knowledge, no prior study has employed a quasi-experimental research strategy where the empirical approach by design is capable of identifying the impact of informal sanctions net of that of formal sanctions.

The rest of this paper is structured as follows. Section 2 describes the institutional background. Section 3 discusses the data and sample construction. We outline the empirical strategy in Section 4. The results are provided in Section 5 followed by conclusions in Section 6.

2 Background

In most countries, criminal background checks serve as a key tool utilized by various actors such as employers, landlords, and higher education institutions to address concerns surrounding public safety and the prevailing notion that individuals with criminal histories may be less inclined to fulfill their responsibilities. Additionally, these checks are often employed to mitigate

[11] The concept of labeling is also evident in the economic model of crime. Once an individual is labeled a criminal, the opportunity cost of refraining from crime may decrease, as potential employers may limit their access to job opportunities (Agan and Starr, 2018; Becker, 1968; Grogger, 1995; Holzer, Raphael and Stoll, 2006).

[12] Most empirical studies compare offenders either convicted or incarcerated with those not formally sanctioned e.g. Bernburg and Krohn (2003); Farrington, Osborn and West (1978); Murray et al. (2017).

[13] For instance, selective out-migration, re-sentence deterrence connected to the original offense and changes in monitoring by law enforcement in response to defendants' observable criminal histories could in some cases not fully be discarded.

potential liability, particularly in the realms of negligent hiring or negligent renting doctrines.[14] The criminal background information tend to be accessible through various channels, including direct provision by states and an increasing number of private companies, especially in the United States, offering background search services.

The databases used for the records checks compile data from diverse origins, often procured in bulk from public sources such as law enforcement agencies, state courts, or through web scraping technology from public websites (NLCS, 2019). The checks tend to reveal current criminal charges or past arrests, but may also encompass non-conviction data, including records of police interactions, mere allegations, withdrawn or stayed charges, as well as acquittals.[15] A background check typically contains limited details about listed offenses but generally includes the individual's name associated with the record, the jurisdiction of origin, the record's creation date, and a case number or law enforcement identifier (NLCS, 2019). While a brief description of the offense, such as "possession of marijuana", may be provided, details about the offense or related circumstances are usually omitted. Some public sources only display criminal convictions, whereas private firms may report arrests for misdemeanors that later result in dismissal (Bushway and Sweeten, 2007).

2.1 The Swedish setting

The regulations governing the request of criminal records in Sweden have traditionally been stringent, primarily for employment-related purposes (Backman, 2012). Employers are strictly limited in their ability to request a candidate's criminal record, typically permitted only at the concluding stages of the hiring process. Subsequently, the job applicant is required to formally request his criminal record from the police, who dispatches it in a sealed envelope. Upon receipt of the record, the applicant must then forward the envelope directly to the employer. This protocol often results in a significant time delay, with the entire process spanning several weeks. It is therefore not surprising that survey data from a large sample of employers collected in 2011 indicates a relatively low utilization of criminal background checks by employers in Sweden, with only approximately 14 percent of the 1,200 employers reported to have engaged in such practices (SOU, 2011). This conclusion is also supported by Figure A.1 showing that only about two percent of the job ads posted on the Public Employment Office's in 2011 mentioned criminal background checks. Consistent with this, the same year the ratio of the number of official background checks to the number of firms was 17 percent. Since then, there has been a slowly growing trend towards increased use of official background checks.

The information provided in the official criminal background checks is expunged after 3-10 years. The exact amount of time depends on factors such as the age of the offender when committing

[14] About 94 percent of employers in the United States conduct some form of criminal history check, and about 90 percent of landlords run background checks on prospective tenants (NLCS, 2019) In 2004, a survey from the United States showed that about 50 percent of the employers check the criminal background of job applicants (Holzer, Raphael and Stoll, 2006)

[15] Several US states have laws prohibiting the use of arrests that did not result in convictions in employment screenings, e.g. https://ccresourcecenter.org/state-restoration-profiles/50-state-comparisoncomparison-of-criminal-records-in-licensing-and-employment/

the crime and the sentence.

2.2 The database

In January 2014, the online platform Lexbase completely transformed the landscape of criminal background checks in Sweden. Instead of targeting solely employers, the platform offered unrestricted access to criminal records for the general public. The database allowed anyone to anonymously search for charged individuals by name, leveraging data sourced from court records mandated by law to be provided upon request. The check maps the name of the individual to current age and place of residence and reveals whether that specific person has been charged with a crime. This means that even for common names it was typically possible to uniquely identify an individual.

Unlike many other online criminal databases, which often specialize in specific crime types and cater exclusively to certain actors like employers or landlords, the platform presents a comprehensive repository of every criminal charge in Swedish courts, without any fees or user profile. For a nominal fee, at the time approximately USD 5, registered users could also obtain full court proceedings, gaining access to information about the number of charges, conviction and crime date, type of crime, circumstances surrounding the offense, the sentence, and the address and personal id number (cf. social security number) of the individual. However, acquiring the complete court proceedings seems to have been a rare occurrence. Data from the database's company indicate that 0.16 % of all background searches actually resulted in a purchased court proceeding in the first year (see Appendix Figure A.2). When it was introduced, there was no information on the website that indicated that the database also included acquittals, only that it contained the records of individuals convicted in court.[16] Appendix Figure A.3 show the homepage of the platform and Appendix Figure A.4 shows the different stages in the search process.

The platform immediately received large media attention and gained immense popularity. Data provided by the company behind the platform show that 32 million background checks were

[16]The following text is a translation of the official press release from the company sent on the day of the launch: "All criminal convictions from the last five years are now easily searchable - New unique site is launched today. Today, a beta version of a new site, Lexbase.se, is being launched, which makes all criminal convictions, five years or younger, easily searchable. Lexbase will be Sweden's largest database with legal information and a new ground-breaking service on the Swedish market." (https://www.lexbase.se/site/press/21?s=285b36b14745ae4aedb1d218ea9989c4). Reports in the media also highlighted that only criminal convictions were included. For instance, the following text is a translation of parts of a report in one of Sweden's largest newspapers called Expressen. "At Lexbase you could find out everything about the convictions of your friends, neighbors and acquaintances. Is the neighbor convicted of rape? Or maybe just for speeding? The only thing you needed to find the convictions was their name or social security number. The convictions were exposed as red dots on a map and could then be downloaded for a small fee. The new site was launched at a high-profile press conference on Monday. Already on the same day, lots of curious people started logging in, while harsh criticism was immediately directed at Lexbase. And immediately the first reports from desperate people came in to the Chancellor of Justice (JK). "I was so sick I couldn't go to work. I'm in that database. I've done everything to get back to a normal life and have been holding a job for the past two years. If someone searches for me, I'll lose my job immediately. You have to do something about this. My neighbors can see my house on a map with a red dot. I'm totally stressed out now by this! I feel so damn bad I might kill myself," one man writes to JK ." (https://www.expressen.se/nyheter/deras-liv-forstors-av-lexbase/)

conducted in just the first year. Appendix Figure A.2 shows that by 2019 when the first competing platform was introduced, the cumulative number of searches had soared to 103 million. Appendix Figure A.5 shows the sharp increase in the number of media articles written in printed pres mentioning the database. Appendix Figure A.6 shows that from the week after its inception and up until 2019, Lexbase consistently outnumbered other popular Google search terms in Sweden, surpassing common queries such as "new car" or "best restaurants".

Since the platform did not have any competitors until 2019, the massive use of the database cannot be explained by a shift in the use from other similar platforms. It is also seems unlikely that employers substituted away from official background checks by the police to instead using the platform. First, the introduction of the platform did not coincide with any meaningful shift in the long-run trend towards increased use of official criminal background checks (see Appendix Figure A.1). Secondly, the demand for regular background checks was relatively low among employers. As discussed earlier, a large survey of 1,200 firms just before the introduction of the platform revealed that 14 percent of employers reported that they check criminal records (SOU, 2011) and only two percent of the job ads posted on the Public Employment Office's mentioned criminal background checks. Third, the number of official background checks, 171,229 in 2011, is completely dwarfed by the about 20 million of searches conducted every year on the platform. Taken together, the large use of the platform is most likely explained by a high demand for the service in the general public.

The platform was closed down by the Internet service provider just two days after its launch, after the site came under heavy criticism from Swedish media and the general public for making criminal records too easily accessible. However, since April 2014 the platform has been fully operational. Several attempts by the Swedish government to shut town the platform have failed because the information is protected by the Swedish constitution.

While the information provided by the database in theory was accessible through public records even before 2014, the process was laborious, necessitating visits to multiple courts to potentially find a matching name. Just as in the case of most other online criminal databases, the introduction of this particular platform thus represented a significant reduction in the costs of accessing criminal background information, especially for the general public.

2.3 Mapping to empirical design

The company behind the platform was officially established as a business entity in 2012. During the two years preceding the public launch of the database, the company began acquiring criminal background data through direct engagement with individual courts, requesting access to criminal charges stored in the digital administrative system known as Vera. A government decree mandated courts to grant access to Vera to anyone seeking information on criminal charges for a given individual.[17] However, after five years, courts should expunge the informa-

[17] See decree 2001:639)

tion.[18] The courts had never before processed this type of large scale request for the universe of criminal cases. Together with some uncertainties surrounding the exact details regarding the expungement procedures, this created some variation among courts in the way they were able to respond to the request. For instance, while some courts were not able to strictly adhere to the exact expungement period, other courts not only delivered criminal cases but also civil law suits.[19] In the end, the request resulted in distinct cutoff dates for each court, marking the point at which criminal charges were integrated into the database. The specific cutoff date was determined by both the timing of the company's request for information and the length of time it took for each court to process the request and then deliver the records.

To identify the cutoff dates, we obtained data from the company detailing the number of cases added to the database for each court and month. Utilizing our comprehensive micro data covering all criminal charges in Sweden since 1986, we calculated the share of cases included in the database. Our analysis comprises the 46 courts that operated during the period preceding the database's launch (i.e., 2006-2013). For each court j and month t, we computed the change in the share of cases by calculating the difference between the average share of cases in the preceding 12 months and that of the subsequent 12 months. The cutoff date was identified as the month exhibiting the largest increase in the share of cases over the entire period. In the empirical analysis we focus on the 32 courts with observed cutoff dates before 2010 to ensure the representation of all courts in our sample within the database at its launch.[20] Additionally, we excluded four courts with no clear evidence of cutoff.[21]

While approximately 5-6 years passed between the initial charge and the release of the platform, the open access version of the database did not include the date of the charge. This omission meant that most employers, landlords, and ordinary citizens could not distinguish between charges based on their timing unless they actively registered on the platform and purchased the sentencing information. Consequently, from the perspective of most public agents, the launch of the platform provided new information about the individuals listed in the database.

Figure 1 shows the average share of cases included in the database around the cutoff for the courts in our sample. Consistent with not all courts strictly adhering to the stipulated timeline for expungement, the cutoff is not sharp. However, there is a clear discontinuity in the share of cases with the an estimated size of the cutoff of 0.64 (0.13).[22] [23] When interpreting this figure it should be noted that the estimates are affected by measurement error. First, our register based

[18]The courts were still obliged to preserve cases in printed form. In order to search for a specific case a person had to visit the court to manually search for the case number in a volume of books that compiles all the cases that have been processed by the court.

[19]For instance, in 2013 the Swedish Authority for Privacy Protection criticized DC Västmanland for not following the stipulated timeline (DNR 1319-2013). A similar statement was made for DC Attunda (DNR 1316-2013). Report 2012:1 by the Swedish Authority for Privacy Protection provide a detailed discussion about these issues.

[20]It is unclear whether all courts were actually included in the database at the time it was launched. Thus, it is possible that some courts were added at a later point in time.

[21]These courts, which represent 7.5 percent of the total number of cases, are DC Umeå, DC Gällivare, DC Haparanda, DC Värmland.

[22]The estimated size of the cutoff is 0.68 (0.12) when not weighting by the size of the courts as measured by the number of cases.

[23]Figure A.8 show the cutoff for each court in our sample.

micro data do not include the complete number of cases that exist in Lexbase since we miss information on individuals that are lacking a personal id number. Most of these individuals are asylum seekers. Second, in some cases, courts accidentally delivered not only criminal cases but also civil cases that were not made searchable on the platform.[24]

3 Data and sample construction

3.1 Administrative data

Our empirical analyses rely on micro data that originate from various administrative registers managed by Statistics Sweden. These registers contain information on the entire Swedish population aged 16 to 65, spanning the years from 1990 to 2016. The data holds demographic and socioeconomic information such as age, place at birth, gender, educational level, employment and income measures. To ensure comprehensive tracking of labor market outcomes for all individuals in our sample, we restrict the sample to individuals younger than 55 in the year of the initial charge.

These data have been linked to the Swedish Conviction Register maintained by the National Council for Crime Prevention (*Brottsförebyggande rådet* - BRÅ). Within these records, we have access to comprehensive details concerning criminal charges during this period. The data include information on the date of the conviction as well as the sentence. This information pertains to criminal charges in Swedish district courts, which is the court of first instance. A single criminal charge may encompass multiple crimes, and we observe all crimes within a given charge.[25] The data also include information on acquittals.

Appendix B describes how the variables in our analysis are constructed.

3.2 Classifying offenders

Utilizing the court-specific cutoff dates, we categorize defendants by exposure statuses employing a 12-month window. Defendants charged within 12 months after this date are classified as exposed, while those charged up to 12 months prior to this are categorized as non-exposed. We focus on first time defendants to ensure that each individual will only appear once in our estimation sample. In addition, this group is particularly important from a policy perspective as non-disclosure interventions are likely to have a greater chance to succeed for individuals before they have accumulated significant criminal experience. This is also the reason why non-disclosure programs typically target first-time defendants.

[24]Data provided by the company behind the database show that 20.81 percent of the total number of records delivered were civil cases.

[25]The data exclude minor offenses such as speeding tickets but include offenses such as driving without a license and DUI.

Table 1 columns (1) and (2) shows descriptive statistics for our sample separately for exposed and non-exposed defendants. These baseline characteristics are measured in one year before the initial trial. As expected, the background characteristics of the individuals are very similar in the two groups. We can also see that there are striking similarities in terms of the case characteristics and the sentence. This does not only show that the covariates in the two groups balance but also that our research design should be able to identify the role of stigma net of that from formal sanctions.

4 Research design

In our main analysis, we estimate the following conventional difference-in-differences model where we compare the outcomes of exposed and non-exposed individuals before and after the database was introduced

$$Outcome_{it} = \alpha + \beta \text{exposed}_i * \text{post}_t + \text{exposed}_i + \gamma_t + \lambda_c + \delta X_{it} + \epsilon_{it} \quad (1)$$

where $Outcome_{igt}$ is the relevant outcome of defendant i in calendar year t. $exposed_i$ is an indicator variable set to one for individuals who were charged after the cutoff date and zero otherwise; $post_t$ is an indicator variable set to one for years when the database was in place and zero otherwise. $exposed_i$ and court (λ_c) fixed effects control for persistent unobserved heterogeneity between exposed and non-exposed individuals as well as between courts. X_{it} is a vector of controls for baseline characteristics such as age, gender and immigrant. It also controls for time since the charge and seasonality in criminal charges by including calendar month fixed effects.[26] Since the reform affected everyone at the same time and remained in effect, the concerns typically associated with staggered difference-in-differences research designs (Roth et al., 2023) do not apply to the current setting.

This strategy relies on the standard assumption of parallel trends: the outcomes should not have evolved differently between exposed and non-exposed offenders in the absence of the database. We assess this by inspecting the estimated β coefficients for the years leading up to the launch of the platform.

The other main assumption is that of no anticipatory effects, which is violated if, for instance, courts or offenders somehow were able to foresee the introduction of the database and the exact cutoff date in each court. In this case, we would expect to see a change in the frequency of court cases around the cutoff date. For instance, courts might choose to have the trial after the cutoff date to ensure that an offender gets exposed in the database. We might also expect to see changes in the characteristics of offenders around the cutoff if some offenders, for instance, more highly educated ones, are better able to predict the future arrival of the database. In this case, there should be an increase in the share of highly educated offenders just before the cutoff.

[26] There are typically fewer cases processed by courts in the summer and around Christmas.

We already saw in Table 1 that the characteristics of exposed and non-exposed offenders are very similar. To examine this in more detail, we estimated the change in the frequency of court cases and in the characteristics of the offenders within a one-year window around the court-specific cutoff date. Figure 2 shows no evidence of such changes for the average number of cases in the courts and the predicted risk of recidivism.[27] These findings are, of course, as expected, as it is hard to think of a scenario where courts could predict the (timing of) arrival of a currently non-existing platform of this type, less which the cutoff dates would be.[28]

5 Results

This section presents the results from our empirical analysis. We start by providing the main results and robustness. We then show results for relevant subgroups of the population. We end by examining some of the underlying mechanisms.

5.1 Main results

Table 2 shows the difference-in-differences estimates of the effect of exposure on criminal behavior and labor supply. We can see that exposure has no statistically significant exposure effect on the probability that the individual is (re-)convicted. We can also see that there is no significant effect on employment. The estimates are relatively precise, ruling out larger increases in the probability of conviction than 0.9 percent points and larger decreases in the probability of being employed by 2.4 percentage points with 95 percent confidence. This corresponds to about one third of the causal effect of incarceration on the probability of being employed.[29] We do, however, find a that exposure leads to a statistically significant decrease in log earnings by about 9.3 percent.[30] While we cannot rule out adjustments in terms of hours worked, the absence of an effect on labor supply at the extensive margin provide suggestive evidence that open access to online criminal databases may primarily hurt the outcomes of defendants by reducing wages.

Consistent with the strong similarity in baseline characteristics shown in Table 1 the estimates

[27] We calculate our measure of recidivism risk from the predicted dependent variables for each individual from the following OLS regression

$$\text{Convicted}_i = \alpha + \beta X_i + \epsilon_i$$

where X_i is a vector of the following baseline defendant characteristics, age, immigrant, educational attainment, employment, log earnings. The risk score is then defined as

$$\hat{\alpha} + X_i\hat{\beta}$$

and captures a defendant's risk of recidivism over a five year period. Appendix Figure A.9 shows similar results when estimating the model separately for each individual characteristic included in the prediction.

[28] Consistent with this, Appendix Figure A.1 shows that there was no mentioning of Lexbase in Swedish media before Jan 2014.

[29] Dobbie et al. (2018) estimate the causal effect of incarceration on the probability of being employed in Sweden to be -7.6 percentage points.

[30] We replace zero earnings with 1 before taking logs. The results are similar when instead estimating the model using the inverse hyperbolic sine.

are almost completely invariant to adding baseline controls to the regressions. Figure A.10 plots dynamic difference-in-differences estimates that mimics those in Table 2 by year relative to 2013.[31] We find no evidence of differential pre-trends.

To further explore how exposure impacts labor market outcomes, Appendix Figure A.11 plots the difference-in-differences estimates based on model (1) and corresponding 90 percent confidence intervals of the effect of exposure on the probability of a defendant's earnings falling above various thresholds. The impact of on earnings is concentrated in the left tail of the earnings distribution, with little to no effect on the probability of earning above higher thresholds than the 80th percentile. These results suggest that exposure primarily affects earnings at the low to mid-end of the earnings distribution.

The exposure effects on earnings are smaller than those shown by Mueller-Smith and Schnepel (2021) who find that diverted felony defendants in Harris County, TX experiences increases in total earnings over the ten-year follow up period grow by 93 percent and improvements in quarterly employment rates by 49 percent. The effect is also smaller than those reported by Finlay (2008) who show that ex-prisoners in US states with open records to online criminal databases experience 18.7 percent reductions in earnings. On the other hand, Agan et al. (2024) find no significant benefits among individuals whose records were reduced proactively in San Joaquin County, CA.

To which extent can differences in the results be explained by the specific Swedish context? We examine this question by following Rose (2021) and estimate the following event study specification to examine the labor market impact of a criminal conviction

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \sum_{\text{s}\in [-5,5]} \delta_s \gamma D_{ist} + \epsilon_{it}, \tag{2}$$

where Y_{it} is employment or log earnings for individual i and year t, α_i is an individual fixed effect, γ_t is year fixed effects, X_{it} is a vector of time-varying annual age dummies, an D_{ist} is 1 when individual i is s quarters from his first conviction. We use dummies for $s \in [-5, 5]$ to estimate 5 years of dynamic effects focusing on individuals convicted for the first time between 1997 and 2010 so that the outcomes are observed for at least 5 years before and after conviction.[32] We take advantage of the fact that our data provides information on the full population of individuals charged with a crime. While the results should primarily be interpreted as a descriptive comparative exercise, the model controls for permanent unobserved heterogeneity

[31] We estinate the following dynamic difference-in-differences model

$$Outcome_{i,g,s} = \alpha + \sum_{\substack{k=2011 \\ k\neq 2013}}^{2016} \beta_k \cdot \mathbb{1}\{exposed\} \cdot \mathbb{1}\{s = k\} + \gamma \cdot \mathbb{1}\{exposed\} + \theta_s + \beta X_{i,g,s} + \epsilon_{i,l,s},$$

where the outcome $Outcome_{i,g,s}$ is observed for individual i, belonging to exposure group g in year s.

[32] We use +/- 7 years and bin the end points after +/- 5 years to be able to identify both the age and year fixed effects.

using individual fixed effects. The results for the two-way fixed effects model are reported in Appendix Figure A.12 and the results using the procedure by Sun and Abraham (2021) are reported in Appendix Figure A.13.

The results show a sharp reduction in both employment and earnings after the initial charge. The drop in earnings is about 38 percent one year after the charge. This effect size is similar to that reported by Rose (2021) who finds earnings losses by 44 percent for individuals convicted of a felony and about 26 percent for individuals convicted for a misdemeanor.[33]

5.2 Further robustness checks

Table A.1 provides results from further robustness checks. It is possible that unobserved seasonal breaks that potentially coincide with the timing of our cutoffs could confound with our results.[34] While we did not see any evidence of differences in the composition of crime types or sentences between exposed and non-exposed defendants in Table 1, Table A.1 provides additional support for the absence of seasonal confounders. The table presents the estimated effects of a placebo experiment where we adjust the cutoff dates to one year earlier or one year later than the actual court-specific cutoff dates. Panels A and B show that there are no significant discontinuities in these placebo exercises.

Panels C and D show the sensitivity of our results for decreasing the bandwidth for individuals included in the treatment and control group to +/-9 months and +/-6 months. We can see that, while the sample size decreases considerably, the estimates for log earnings are similar to baseline albeit slightly larger. The other estimates are also similar and not significantly different from baseline.

There are at least two reasons to believe that our exposure effects represent a lower bound of the actual effect of being listed in the database. First, it is possible that some of the defendants that we classify as non-exposed re-offended in the period between the initial charge and the introduction of the platform and therefore also were exposed. Second, society may place less emphasis on offenses that occurred in the distant past, as is the case in our setting. However, given that the recidivism rate within our sample—comprised of first-time offenders—is relatively low, and that the date of the charge is not revealed unless the full verdict is purchased, we anticipate that such bias is limited.[35] We empirically assess the role of this "bias" by taking advantage of the fact that the database excluded confession cases in the first two years after its launch. Confession cases generally involve less severe crimes and entail a defendant's admission of guilt, with the prosecutor, rather than a court, overseeing these cases. This institutional feature provides a way to examine whether the immediate exposure effects are different from

[33] As another comparison, Grenet, Grönqvist and Niknami (2024) show that avoiding a prison term by substituting from prison to electronic monitoring in Sweden increases log earnings by 22.1 percent.

[34] For instance, during holiday seasons courts may choose to process more urgent cases and postpone less serious ones.

[35] While the one-year recidivism rate is 8.8 percent for first-time defendants it is 26.1 percent among multiple-time defendants.

the intermediate exposure effects. Panel E shows estimates from this alternative difference-in-differences model where the "exposed" group includes defendants whose cases where processed in court, and the "non-exposed" group comprises defendants with a confession case. In this analysis we measure the outcomes one-year post-conviction and the repeated cross-sectional dataset runs from 2010 to 2015. All regressions control for the same set of covariates as in model (1). While the results in Panel E are similar to our baseline specification, these results should, be interpreted with some caution as confession cases represent special types of crimes.[36] [37]

5.3 Heterogeneity

Having established that our results are not sensitive to changes in the specification or sample we next examine heterogeneity in the effects of exposure based on two different criteria: (i) defendant baseline socioeconomic and demographic characteristics, (ii) case characteristics. The first criterion is motivated by the observation that the effect of criminal justice interactions often vary based on labor market attachment and demographics; e.g. Dobbie et al. (2018); Garin et al. (2023); Jordan, Kager and Neal (2023); Rose (2021); Doleac and Hansen (2020). The second criterion is motivated by providing additional insights about the role of stigma.

5.3.1 Defendant characteristics

Table 3 presents results from a variation of model (1) where we interact exposed*post with baseline defendant characteristics measured in the year before the initial trial. In this exercise we use indicators for male, above age 25 (the median age in our sample), immigrant and whether the defendant holds a high school (HS) degree. Each panel presents the results for a different dependent variable. Column (1) recapitulates our baseline estimates from Table 2. In column (2), we find that the average exposure effect on convictions masks significant differences by gender. While the effect of exposure is close to zero for males, exposure increases the risk of conviction for females by 1.2 percentage points. We also find significant stronger increases in the probability of conviction for defendants above age 25, and for those with at least a HS degree. Turning to labor market outcomes in Panels B and C we find significantly stronger adverse effects on both employment and earnings for males, defendants above age 25 and those holding a HS degree. We also find that the effect on earnings is significantly weaker for migrants compared to natives -3.6 and -11.4 percent, respectively. In summary, we find stronger adverse exposure effects in groups where the opportunity cost of engaging in crime is higher.

[36] This analysis is potentially also affected by general deterrence effects if the existence of the database reduced the likelihood of engaging in crime in the general population.

[37] As an alternative to model (1) we also estimated a regression discontinuity in time design (RDiT) where the running variable was time to the court-specific cutoff using the same dependent variables observed after the introduction of the platform. The estimates were, however, very imprecise to inform about the effect of interest. While a 95 percent confidence interval overlapped with our baseline estimates, the interval was not informative about the causal effect of interest. For instance, the exposure effects including 20 percent reductions in earnings could not be ruled out with 95 percent confidence.

5.3.2 Case characteristics

Table 4 presents estimates from regressions where we investigate potential heterogeneous exposure effects by case characteristics. Such effects are interesting in themselves as they may be informative about what type of information is relevant to include in criminal background checks. They potentially also contribute to shed light on the role of informal sanction costs. The labeling theory is one of the cornerstone models in criminology and posits that revealing one's criminal background can exacerbate criminal behavior. In this model, interactions with the criminal justice system labels the offender and cause him to adopt stigmatizing beliefs about himself and adopt a deviant identity (e.g. Becker, 1968; Lemert, 1967). Because of the severe difficulties involved in distinguishing informal sanction costs attached to the stigma of holding a criminal record from the imposed formal ones such as fines, community service or prison, we are not aware of any study where the empirical strategy by design is able to identify informal sanctions net of formal ones (Chalfin and McCrary, 2017).[38] Isolating the role of informal sanction costs is important as these mechanisms may operate in different direction and vary depending on the context. For instance, in places where the spatial concentration of criminals is higher, the role of stigma could be reduced.[39]

The fact that all individuals in our sample are charged with a crime but public exposure of criminal records is quasi-randomly determined presents a rare opportunity to identify the impact of informal sanctions independently of formal ones. While this setting is sufficient to identify the role of informal sanctions, an analysis of whether the exposure effect varies by case characteristics provides an additional way to investigate the importance of stigma.

The results in column (2) of Panel A show that the exposure effects on recidivism are significantly higher for defendants who were acquitted in court compared to defendants who were convicted. The estimates suggest that the probability of being convicted increases by almost 1 percentage point for those acquitted.

Panels B and C show that there is also a significantly larger labor market penalty for acquitted defendants compared to convicted defendants. We can also see that the labor market penalty of exposure is significantly higher for defendants who were convicted of a misdemeanor compared to those convicted with a non-misdemeanor. Many misdemeanor offenses do not involve violence or firearms; rather, they stem from the criminalization of everyday behaviors like possessing small amounts of banned substances for personal use, vandalism, disturbing the peace, minor theft, and trespassing.[40]

[38] Empirical studies in criminology tend to compare offenders either convicted or incarcerated with those not formally sanctioned e.g. Bernburg and Krohn (2003); Farrington, Osborn and West (1978); Murray et al. (2017).

[39] It can be useful to consider the (dis)utility of the offender, U, as a function of formal sanctions, F, and informal sanctions, I, so that $U = f(F, I)$. If informal sanctions lead to worse utility $\partial U/\partial I < 0$, while formal sanctions improve utility $\partial U/\partial F > 0$, for instance if prisons provide adequate rehabilitation programs, then the reduced form effect depends on the relative magnitudes of these partial derivatives and their interaction within the utility function.

[40] Misdemeanors are minor offenses, usually resulting in incarceration of one year or less. Felonies are more serious, such as robbery, theft over USD 1,000, and assault with a deadly weapon, carrying harsher penalties. While the grouping of offenses into misdemeanor and felonies is not used in Sweden, we classify misdemeanors as crimes for which no individual was sentenced to prison in our data.

Recall that the open access version of the database did not differentiate between severe crimes and misdemeanors. Additionally, it included acquitted defendants without indicating their not-guilty verdicts. Therefore, one possible interpretation for the stronger treatment effects is that the platform's information may have signaled that an individual was a more prolific criminal than they actually were.[41].

Several theoretical models predict that, all else equal, the degree of stigma in a given area is declining in the number of individuals with criminal records (e.g. Chalfin and McCrary, 2017; Weibull and Villa, 2005; Funk, 2004; Rose, 2021). To examine this we used our data to calculate the number of individuals convicted of a crime in the past five years living in each municipality. We then interacted *exposed* x *post* in model (1) with an indicator for the defendant living in an area with less than median criminals. The results show that the adverse labor market effects indeed are concentrated in the subsample of individuals who lives in neighborhoods with a low share of the population having a criminal conviction. For instance, while the decrease in earnings is 15.2 percent for those living in places with relatively few ex-convicts, it is 5.4 percent for defendants in neighborhoods with a higher concentration of ex-convicts.

5.4 Interpretation and additional results

Our main findings so far is that the massive use of the database is connected to significant adverse exposure effects on earnings but not on employment or criminal recidivism. However, there are significantly stronger detrimental effects on both labor market outcomes and recidivism in specific defendant subgroups such as those with at least a high school degree, acquitted individuals, and those living in areas with few criminals. These results are consistent with stigma attached to being marked as a criminal playing a key role in explaining the effects of interactions with the criminal justice system. In this section we provide results from additional analyses that aims to assess various explanations for our findings.

5.4.1 Dynamic scarring effects

Several studies show that criminal conviction is associated with a large drop in labor market outcomes (Rose, 2021). In Section 5.1 we showed that this is also the case in Sweden. If a conviction causes individuals to quickly accumulate labor market scars, such as lengthy employment gaps, employers may remain hesitant to hire these individuals, potentially making it more difficult to identify exposure effects (Agan et al., 2024). However, we do find a significant labor market penalty even net of this dynamic scaring effect. We also estimated our event study model (model (2)) focusing on the effect on labor market outcomes around the time of acquittal and misdemeanor charges. While imprecise, the point estimates are much closer to zero compared to that of convictions.[42] Since we find significantly stronger exposure effects for

[41] Appendix C provides a toy model that clarifies the conditions under which employers respond to limited information about criminal background

[42] The results are available on request.

acquittal and misdemeanor cases, it means that for those results, any dynamic scaring effects are likely to be small. The results in Panel E in Appendix Table A.1 where we examine the immediate effect of the introduction of the platform using confession cases as an alternative control group finding similar results as in baseline also seem to rule out that dynamic scarring effects matters, at least in the context of this paper.

5.4.2 Behavioral responses

To what extent can our results be explained by behavioral adjustments taken by the defendants listed in the database? We examine this by estimating exposure effects for several potentially important behavioral markers. Previous studies have shown that criminal justice interactions has meaningful effects on earnings through changes in family structure (Charles and Luoh, 2010; Chetty, Hendren and Porter, 2020; Svaver, 2011) and human capital investments (Finlay, Mueller-Smith and Street, 2023). Table A.2 shows the results for exposure effects on the probability of enrolling in college and the likelihood of starting a labor market training program. These outcomes could proxy for human capital investments that defendant may potentially take in order to avoid the potential harmful effects of having ones criminal records exposed. We also examine the probability of moving to a different neighborhood. Finally, we present estimates for family formation outcomes. With the exception for the probability of enrolling in college, we find no significant exposure effects. The results for college enrollment should also be interpreted with some caution since the effect size is small (0.4 percentage points) and only significant at the 10 percent level. One potential reason for why individuals not seem to respond to exposure may be that lack of knowledge or perceived low rewards. It also provide suggestive evidence that the reaction of employers to exposure is what explains the labor market penalty.

Industry sorting

Individuals could also respond to exposure by sorting into jobs that are more lenient towards hiring individuals with criminal records. To examine this we follow Mueller-Smith and Schnepel (2021) and classify each two digit SNI (NAICS) industrial sector as having low or high employment penetration among convicted individuals using the regression adjusted ratios of the sector-specific employment rate for those individuals with convictions versus those without in our analysis samples (see Table A.3). Sectors with ratios of 1.5 or less are high penetration industries relatively more accessible to those with criminal histories and the opposite is true for low penetration industries with ratios greater than 1.5. The results in Table A.4 show small and imprecise estimates but no strong evidence in favor of differential sorting across industries for any of the groups. For sectors with employment penetration ratios of 1.2 or higher we find some evidence that individuals are sorting away from high penetration industries but the estimates are small in magnitude and imprecise.

AKM sorting

Job ladder models propose that workers engage in on-the-job search throughout their careers, gradually moving to higher-paying employers. In these models, pay disparities among firms reflect differences in how the surplus is divided between workers and employers (Burdett and

Mortensen, 1998; Haltiwanger et al., 2018). Within this framework, exposed workers may experience earnings losses due to employer effects as they move down the job ladder to employers that offer lower pay. We examine to what extent the exposure effects can be explained by differential sorting of workers to firms that vary in their estimated earnings premia by following (Lachowska, Mas and Woodbury, 2020). We estimate the earnings premia in different firms using the methods suggested by Abowd, Kramarz and Margolis (1999) (AKM). In this exercise we define a firm as an establishment.

The AKM model can be written

$$logY_{itj} = \alpha_i + \beta_{j(i,t)} + \phi_t + X_{itj} + \epsilon_{itj} \tag{3}$$

where $logY_{itj}$ denotes earnings of worker i with employer j in year t; α_i is a worker-specific fixed effect; $\beta_{j(i,t)}$ is an employer-specific fixed effect; ϕ_t is a vector of calendar year indicators; X_{itj} consist of education-gender-specific age dummies (following Engbom, Moser and Sauerman (2023)), and ϵ_{itj} is the error component. The function $j(i,t)$ indexes the employer j effect for worker i in year t. Due to the problem of multicollinearity between age and year-fixed effects, we assume that the age effects are constant between ages 45 and 54.

This model controls for sorting of workers based on fixed differences in their underlying characteristics. The firm fixed effects capture firm characteristics that result in above- or below-average earnings for all workers at employer j and can be interpreted as representing the advantages derived from being employed by a given employer. We estimate the AKM model for firms in the connected set over the years 2005-2016, including workers aged 18-65. We exclude establishments with fewer than 5 movers when estimating the model but note that our results are robust to also including all establishments in the connected set.

We use the estimated $\hat{\beta_{j(i,t)}}$ as outcomes when we estimate exposure effects using model (1). The results are presented in Appendix Table A.5. In general, we find significantly stronger adverse exposure effects on the earnings premia in the subgroups of defendants where the labor market responses were concentrated. This result is consistent with the predictions in job ladder models that exposed workers will experience earnings losses due to employer effects, as they moves down the job ladder to a lower-paying employer.

In order to better understand whether a move down the job ladder is due to workers changing employer or workers staying in the same firm we estimated model (1) using as dependent variable an indicator set to one if the individual has changed firm from one year to another and zero if (s)he stayed in the firm. Individuals who are non-employed are missing in the data. Overall, the results shown in Appendix Table A.6 shows that workers at least age 25 are significantly more likely to change firm compared to younger workers, that acquitted workers are significantly more likely to change firm relative to convicted workers, and that workers charged with a misdemeanor are significantly more likely to move to a different firm relative to those charged with a more severe crime.

Taken together, it seems as if at least part of the reduction in the AKM earnings premium is due to workers transiting to lower-paying employers. While we cannot distinguish between voluntary and involuntary job losses, it is possible that exposed workers are fired or do not get their temporary contracts renewed. These workers may be forced to adjust their reservation wage downwards and in the process moving down the job ladder to a lower-paying employer.

5.4.3 Effects on children

In this section we investigate whether the introduction of the database affected the outcomes of the children of exposed defendants. The potential for significant yet previously overlooked spillover effects of parental criminal justice involvement is suggested by recent studies estimating the causal effect of parental incarceration (Arteaga, 2022; Bhuller et al., 2020; Dobbie et al., 2018; Norris, Pecenco and Weaver, 2021), but it is unclear how children are affected by increased public access to criminal records of their parents.[43] In this analysis, we focus on children who finishes grade 9 (typically at age 16) in the period 2010-2016 and link them to their parents. We also add information on court convictions for these children. We start by verifying in Appendix Table A.7 that the effects in the parental subsample are relatively similar to that in the full population, although the precision of the point estimates is lower for log earnings and higher for employment. Table 5 shows the difference-in-differences estimates from model (1) of the exposure effects on children. We find significantly stronger adverse exposure effects on compulsory school GPA for children belonging to the subgroups of defendants with the largest adverse exposure effects on the defendants themselves. It should be noted that the adverse exposure effect on children could be driven by both a direct effect of the children experiencing the stigma from having a parent in the database or by a reduction in the family's resources.

6 Concluding Remarks

We study the unexpected launch of Sweden's first online criminal database in 2014 that provided unrestricted, anonymous, and cost-free public access to the criminal records of every individual charged with a crime. To estimate the causal effect of public access to this database, we leverage rich register data and the fact that administrative rules only allowed the company behind the database to identify offenders sentenced before certain dates in the past.

We find significant adverse effects of exposure on earnings but not on employment or criminal

[43] While the findings are mixed, many of the estimates lack precision, resulting in overlapping confidence intervals for comparable outcomes. Using a random-judge design, Dobbie et al. (2018) report that parental incarceration leads to increased teen crime, lower school performance, and negative impacts on employment and earnings. Consistent with this, Grenet, Grönqvist and Niknami (2024) show quasi-experimental evidence that a reform were parents were able to substitute a prison sentence to instead serve the full sentence using electronic monitoring significantly improved child outcomes. Conversely, Norris, Pecenco and Weaver (2021) find that in the United States, parental incarceration reduces teen crime, does not affect teen parenthood, and increases the likelihood of children living in affluent neighborhoods as adults. Arteaga (2022) observes positive effects on children's educational attainment in Colombia. Bhuller et al. (2020) find no significant effects on school performance or children's propensity for criminal behavior.

recidivism. However, there are significantly stronger detrimental effects on both labor market outcomes and recidivism in defendant subgroups with a higher opportunity cost of engaging in criminal activity, acquitted individuals, those charged with a misdemeanor and those living in areas with fewer criminals. Our results highlight that the stigma associated with being marked as a criminal is a potentially important but previously unappreciated mechanism explaining behavioral responses to criminal justice interactions.

While we are unable to directly measure the long-run impact of open access criminal databases on defendants using our data, we place the magnitude of our estimates in perspective by considering a partial back-of-the-envelope calculation that accounts for the direct economic impacts on the defendant himself. To calculate the implied effect we make the conservative assumption that the treatment effect on log earnings is constant for a period of three years then completely fades away. We multiply the treatment effect with baseline average earnings (SEK 148,134) and estimate the three year earnings loss for the defendants in our sample to be SEK 44,436. Since 2014, about 47,000 first time defendants are added to the database every year. For instance, in 2018 48,183 first time defendants were convicted in court and consequently were listed in the database.[44] The combined earnings losses this year thus amounts to SEK 2.14 billion, or 23 % of total government spending on prison and rehabilitation that year.[45] It is, however, important to note that we are unable to determine the full welfare consequences of open access to online criminal records using our research design. It is possible that these databases have a general deterrence effect in the full population. Our analysis will, therefore, overstate the social costs of these databases if the threat of being included in the database decreases crime rates. The existing literature has struggled to provide definitive explanations for why the impact of stricter sentencing policies on offenders' outcomes appears to vary across different study contexts. Our findings potentially offer a partial explanation. In environments where stigma carries less weight, such as places with a high spatial concentration of criminals, informal sanctions may play a diminished role in explaining the reduced form effect of sanction policy relative to that of formal sanctions. On the other hand, in contexts where informal sanction costs matter more, the beneficial effects of more prison staff or rehabilitation programs could be offset. Our findings directly speak to the debate over what type of information on cases should be disclosed in criminal background checks. Our results suggest that reforms that increase leniency in the conviction dimension (e.g. higher evidentiary standards or not prosecuting marginal cases) potentially has social benefits. In many countries, the information documented in background checks has vastly expanded along with their public availability so that many record checks now include non-criminal details that presumptively mark individuals as risky (Zedner and Ashworth, 2019). However, hardly any research has investigated the consequences of shifts in the content and governance of these records.

[44] https://bra.se/statistik/kriminalstatistik/personer-lagforda-for-brott.html

[45] Total government spending on prison and rehabilitation was SEK 9.3 billion in 2018. https://www.kriminalvarden.se/globalassets/publikationer/ekonomi/kriminalvardens-arsredovisning-2018.pdf

References

Abowd, J., F. Kramarz, and N. Margolis. 1999. "High wage workers and high wage firms." *Econometrica*, 67: 251–333.

Abrams, David S. 2012. "Estimating the deterrent effect of incarceration using sentencing enhancements." *American Economic Journal: Applied Economics*, 4(4): 32–56.

Agan, A., J. Doleac, and A. Harvey. 2023. "Misdemeanor prosecution." *Quarterly Journal of Economics*, 138(3): 1453–1505.

Agan, Amanda, Andrew Garin, Dimitri Koustas, Alexandre Max, and Crystal Yang. 2024. "The Labor Market Impacts of Reducing Felony Convictions." *American Economic Review: Insights*.

Agan, Amanda, and Sonja Starr. 2018. "Ban the box, criminal records, and racial discrimination: A field experiment." *The Quarterly Journal of Economics*, 133(1): 191–235.

Aizer, Anna, and Joseph J Doyle Jr. 2015. "Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges." *The Quarterly Journal of Economics*, 130(2): 759–803.

Anker, Anne-Sofie, Jennifer Doleac, and Rasmus Landersø. 2021. "The Effects of DNA Databases on the Deterrence and Detection of Offenders." *American Economic Journal: Applied Economics*, 13(4): 194–225.

Arteaga, C. 2022. "Parental Incarceration and Children's Educational Attainment." *Review of Economics and Statistics*.

Backman, Cristel. 2012. "Criminal Records in Sweden: Regulation of Access to Criminal Records and the Use of Criminal Background Checks by Employers."

Barbarino, Alessandro, and Giovanni Mastrobuoni. 2014. "The incapacitation effect of incarceration: Evidence from several Italian collective pardons." *American Economic Journal: Economic Policy*, 6(1): 1–37.

Becker, Gary S. 1968. "Crime and punishment: An economic approach." *Journal of political economy*, 76(2): 169–217.

Bernburg, Jön Gunnar, and Marvin D Krohn. 2003. "Labeling, life chances, and adult crime: The direct and indirect effects of official intervention in adolescence on crime in early adulthood." *Criminology*, 41(4): 1287–1318.

Bhuller, Manudeep, Gordon B Dahl, Katrine V Løken, and Magne Mogstad. 2020. "Incarceration, recidivism, and employment." *Journal of Political Economy*, 128(4): 1269–1324.

Buonanno, Paolo, and Steven Raphael. 2013. "Incarceration and incapacitation: Evidence from the 2006 Italian collective pardon." *American Economic Review*, 103(6): 2437–2465.

Burdett, K., and D. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review*, 39(2): 257–273.

Bushway, Shawn D, and Gary Sweeten. 2007. "Abolish lifetime bans for ex-felons." *Criminology & Pub. Pol'y*, 6: 697.

Bushway, S., I. Cabreros, J. Welburn Paige, D. Schwam, and J. Wenger. 2022. "Barred from employment: More than half of unemployed men in their 30s had a criminal history of arrest." *Science Advances*, 8(7).

Chalfin, A., and J. McCrary. 2017. "Criminal Deterrence: A Review of the Literature." *Journal of Economic Literature*, 55(1): 5–48.

Charles, K., and M. Luoh. 2010. "Male incarceration, the marriage market, and female outcomes." *The Review of Economics and Statistics*, 92(3): 614–627.

Chetty, R., N. Hendren, and S. Porter. 2020. "Race and economic opportunity in the United States: An intergenerational perspective." *The Quarterly Journal of Economics*, 135(2): 711–783.

Cullen, Zoe, Will Dobbie, and Michael Hoffman. 2024. "Increasing the Demand for Workers with a Criminal Record." *Quarterly Journal of Economics*.

Dobbie, Will, Hans Grönqvist, Susan Niknami, Mårten Palme, and Mikael Priks. 2018. "The intergenerational effects of parental incarceration." National Bureau of Economic Research.

Dobbie, Will, Jacob Goldin, and Crystal S Yang. 2018. "The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges." *American Economic Review*, 108(2): 201–240.

Doleac, Jennifer L, and Benjamin Hansen. 2020. "The unintended consequences of "ban the box": Statistical discrimination and employment outcomes when criminal histories are hidden." *Journal of Labor Economics*, 38(2): 321–374.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova. 2009. "The deterrent effects of prison: Evidence from a natural experiment." *Journal of political Economy*, 117(2): 257–280.

Engbom, N., C. Moser, and J. Sauerman. 2023. "Firm pay dynamics." *Journal of Econometrics*, 233: 396–423.

Farrington, David P, Stan G Osborn, and Donald J West. 1978. "The persistence of labelling effects." *Brit. J. Criminology*, 18: 277.

Finlay, Keith. 2008. "Effect of employer access to criminal history data on the labor market outcomes of ex-offenders and non-offenders." National Bureau of Economic Research.

Finlay, Keith, Michael Mueller-Smith, and Brittany Street. 2023. "Children's Indirect Exposure to the U.S. Justice System: Evidence From Longitudinal Links between Survey and Administrative Data*." *The Quarterly Journal of Economics*, 138(4): 2181–2224.

Freeman, Richard B. 1992. "Crime and the Employment of Disadvantaged Youths." *Urban Labor Markets and Job Opportunity*. Washington DC:The Urban Institute.

Funk, Patricia. 2004. "On the Effective Use of Stigma as a Crime Deterrent." *European Economic Review*, 48(4): 715–725.

Garin, A., D. Koustas, C. McPherson, S. Norris, M. Pecenco, E. Rose, Y. Shem-Tov, and J. Weaver. 2023. "The Impact of Incarceration on Employment, Earnings, and Tax Filing." *University of Chicago, Becker Friedman Institute for Economics Working Paper*.

Grenet, Julien, Hans Grönqvist, and Susan Niknami. 2024. "The Effects of Electronic Monitoring on Offendes and Their Families." *Journal of Public Economics*.

Grogger, Jeffrey. 1995. "The effect of arrests on the employment and earnings of young men." *The Quarterly Journal of Economics*, 110(1): 51–71.

Grönqvist, Hans, Julien Grenet, and Susan Niknami. 2022. "The Effects of Electronic Monitoring on Offenders and Their Family Members." *Available at SSRN 4181510*.

Haltiwanger, J., H. Hyatt, L. Kahn, and E. McEntarfer. 2018. "Cyclical Job Ladders by Firm Size and Wages." *American Economic Journal: Macroeconomics*, 10(2): 52–85.

Helland, Eric, and Alexander Tabarrok. 2007. "Does three strikes deter? A nonparametric estimation." *Journal of human resources*, 42(2): 309–330.

Hjalmarsson, R., and M. Lindquist. 2022. "The Health Effects of Prison." *American Economic Journal: Applied Economics*, 14: 234–270.

Holzer, Harry J, Steven Raphael, and Michael A Stoll. 2006. "Perceived criminality, criminal background checks, and the racial hiring practices of employers." *The Journal of Law and Economics*, 49(2): 451–480.

Jackson, Osborne, and Bo Zhao. 2017. "The effect of changing employers' access to criminal histories on ex-offenders' labor market outcomes: evidence from the 2010–2012 Massachusetts CORI Reform."

Jordan, A., E. Kager, and D. Neal. 2023. "Heterogenous impacts of sentencing decisions." *NBER Working Paper*.

Kahan, Dan M, and Eric A Posner. 1999. "Shaming white-collar criminals: A proposal for reform of the federal sentencing guidelines." *The Journal of Law and Economics*, 42(S1): 365–392.

Kamat, V., S. Norris, and M. Pecenoc. 2024. "Conviction, Incarceration and Policy Effects in the Criminal Justice System." *SSRN Working Paper*.

Kessler, Daniel, and Steven D Levitt. 1999. "Using sentence enhancements to distinguish between deterrence and incapacitation." *The Journal of Law and Economics*, 42(S1): 343–364.

Kling, Jeffrey R. 2006. "Incarceration length, employment, and earnings." *American Economic Review*, 96(3): 863–876.

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

Lachowska, Marta, Alexandre Mas, and Stephen A. Woodbury. 2020. "Sources of Displaced Workers' Long-Term Earnings Losses." *American Economic Review*, 110(10): 3231–66.

Lee, David S, and Justin McCrary. 2017. "The deterrence effect of prison: Dynamic theory and evidence." In *Regression discontinuity designs*. Vol. 38, 73–146. Emerald Publishing Limited.

Lemert, Edwin McCarthy. 1967. *Human deviance, social problems, and social control.* Englewood Cliffs, NJ, Prentice-Hall.

Mueller-Smith, Michael. 2015. "The criminal and labor market impacts of incarceration." *Unpublished working paper*, 18.

Mueller-Smith, Michael, and Kevin Schnepel. 2021. "Diversion in the Criminal Justice System." *Review of Economic Studies*, 88(2): 883–936.

Mungan, Murat C. 2016. "A generalized model for reputational sanctions and the (ir) relevance of the interactions between legal and reputational sanctions." *International Review of Law and Economics*, 46: 86–92.

Murray, Joseph, Arjan Blokland, David P Farrington, and Delphine Theobald. 2017. "Long-term effects of conviction and incarceration on men in the Cambridge Study in Delinquent Development." In *Labeling theory*. 209–235. Routledge.

Nagin, Daniel S. 2013. "Deterrence: A review of the evidence by a criminologist for economists." *Annu. Rev. Econ.*, 5(1): 83–105.

Nagin, Daniel S, and Greg Pogarsky. 2003. "An experimental investigation of deterrence: Cheating, self-serving bias, and impulsivity." *Criminology*, 41(1): 167–194.

Norris, S., M. Pecenco, and J. Weaver. 2021. "The Effects of Parental and Sibling Incarceration: Evidence from Ohio." *American Economic Review*, 111(9): 2926–2963.

Ouss, Aurelie., John-Eric. Humphries, Kamelia. Stareva, Megan. Stevenson, and Winnie van Dijk. 2023. "Conviction, Incarceration and Recidivism: Understanding the Revolving Door."

Owens, Emily G. 2009. "More time, less crime? Estimating the incapacitative effect of sentence enhancements." *The Journal of Law and Economics*, 52(3): 551–579.

Pager, Devah. 2003. "The mark of a criminal record." *American journal of sociology*, 108(5): 937–975.

Polinsky, A Mitchell, and Steven Shavell. 2007. "The theory of public enforcement of law." *Handbook of law and economics*, 1: 403–454.

Prescott, J., and S. Starr. 2020. "Expungement of Criminal Convictions: An Empirical Study." *Harvard Law Review*, 133(8): 2460–2465.

Raphael, S. 2014. "The New Scarlet Letter? Negotiating the US Labor Market with a Criminal Record." W E Upjohn institute.

Rasmusen, Eric. 1996. "Stigma and self-fulfilling expectations of criminality." *The Journal of Law and Economics*, 39(2): 519–543.

Rose, E., and S-Y Yotam. 2021. "How Does Incarceration Affect Reoffending? Estimating the Dose-Response Function." *Journal of Political Economy*, 129(12): 3302–3356.

Rose, Evan K. 2021. "Does banning the box help ex-offenders get jobs? Evaluating the effects of a prominent example." *Journal of Labor Economics*, 39(1): 79–113.

Roth, J., P. Sant'Anna, A. Bilinski, and J. Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics*, 235(2): 2218–2244.

Selbin, J., J. McCrary, and J. Epstein. 2018. "Unmarked? Criminal record Clearing and Employment Outcomes." *Journal of Criminal Law and Criminology*, 108(1).

Shoagii, Daniel, and Stan Veugeriii. 2016. "No Woman No Crime Ban the Box, Employment, and Upskillingi."

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199. Themed Issue: Treatment Effect 1.

Svaver, M. 2011. "Crime and Partnerships." *Review of Economics of the Household*, 9: 307–325.

Uggen, C., M. Vuolo, S. Lageson, E. Ruhland, and K. Witham. 2014. "The edge of stigma: An experimental audit of the effects of low-level criminal records on employment." *Criminology*, 38(52): 627–654.

Weibull, Jörgen W, and Edgar Villa. 2005. "Crime, punishment and social norms." SSE/EFI Working Paper Series in Economics and Finance.

Western, B. 2002. "The Impact of Incarceration on Wage Mobility and Inequality." *American Sociological Review*, 67(4).

Zedner, L., and A. Ashworth. 2019. "The Rise and Restraint of the Preventive State." *Annual Review of Criminology*, 2: 429–450.

Zimring, Franklin E, Gordon Hawkins, and James Vorenberg. 1973. "Deterrence: The legal threat in crime control."

Figures and Tables

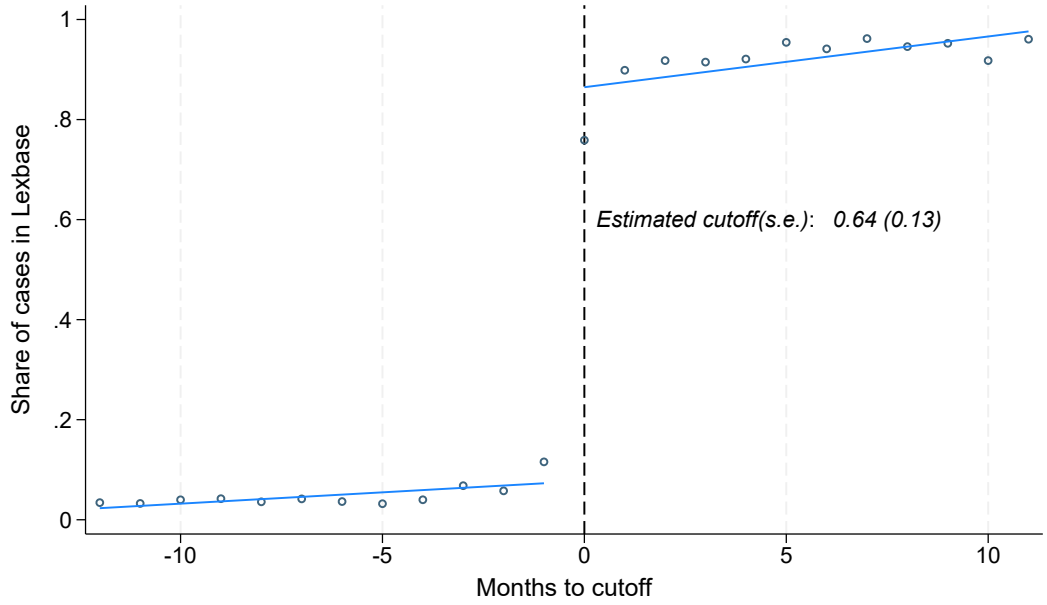


Figure 1: Change in the weighted share of cases around the cutoff

Notes: The figure plots the weighted average share of Lexbase cases across all district courts in our sample within a 12 month window around the court-specific cutoff date.

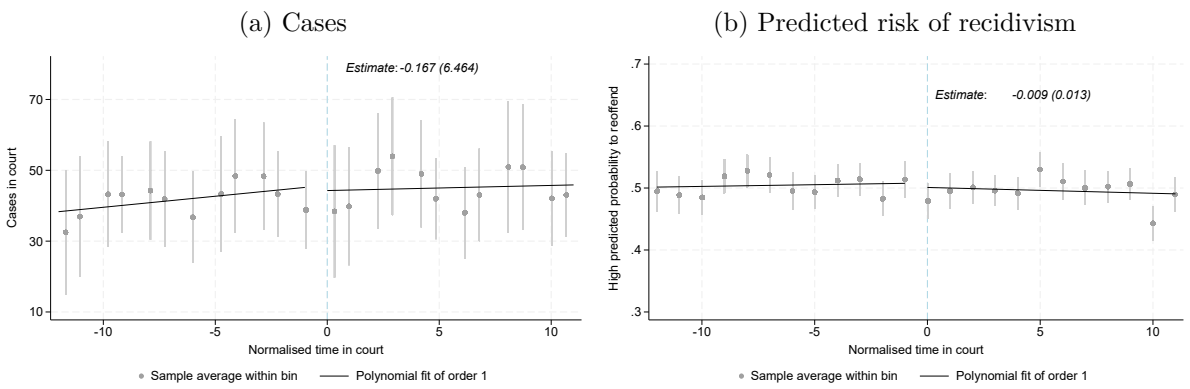


Figure 2: Bunching around the cutoff

Notes: Figure (a) shows the average number cases for the courts in our sample by month relative to the cutoff. Figure (b) shows the average predicted risk of recidivism of defendants by month relative to cutoff. The prediction is done using the baseline individual characteristics of the defendant listed in Table 1.

Table 1: Descriptive Statistics

Panel A: Baseline Characteristics	Exposed defendants	Non-exposed defendants
Male	0.779	0.777
	(0.415)	(0.416)
Age	29.627	29.133
	(11.966)	(12.031)
Immigrant	0.295	0.282
	(0.456)	(0.450)
High school degree	0.618	0.607
	(0.486)	(0.488)
Log earnings	5.068	4.968
	(3.375)	(3.363)
Employed	0.544	0.528
	(0.498)	(0.499)
Panel B: Case characteristics		
Acquitted	0.144	0.144
	(0.351)	(0.351)
Misdemeanor	0.171	0.157
	(0.377)	(0.363)
Felony	0.829	0.843
	(0.377)	(0.363)
DUI	0.086	0.081
	(0.280)	(0.273)
Drug offense	0.057	0.052
	(0.231)	(0.222)
Property offense	0.148	0.148
	(0.355)	(0.355)
Violent offense	0.173	0.182
	(0.378)	(0.385)
Traffic offense	0.157	0.147
	(0.363)	(0.354)
Other offense	0.381	0.390
	(0.486)	(0.488)
Panel C: Sentence		
Fines	0.499	0.488
	(0.500)	(0.500)
Probation	0.176	0.172
	(0.381)	(0.378)
Community service	0.214	0.217
	(0.410)	(0.412)
Prison	0.074	0.075
	(0.262)	(0.263)
Other	0.037	0.048
	(0.189)	(0.214)
Panel D: Outcomes		
Convicted	0.091	0.085
	(0.287)	(0.278)
Employed	0.659	0.673
	(0.474)	(0.469)
Log earnings	5.801	5.873
	(3.318)	(3.304)
Observations	97,374	89,466

Note: This table provides summary statistics for baseline characteristics of the estimation sample for exposed and non-exposed defendants. Exposed defendants were sentenced up to 12 months after the court-specific cutoff date, making them visible in the database, and non-exposed defendants were sentenced in a 12 month window before the cutoff. Baseline characteristics are measured one year before the initial trial.

Table 2: Main Results

	Convicted		Employed		Log earnings	
	(1)	(2)	(3)	(4)	(5)	(6)
Exposed*Post	0.003	0.003	-0.007	-0.008	-0.093**	-0.093**
	(0.003)	(0.003)	(0.008)	(0.008)	(0.042)	(0.042)
Sample mean	0.088	0.088	0.666	0.666	5.835	5.835
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Court FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	No	Yes	No	Yes	No	Yes
Observations	186,840	186,840	186,840	186,840	186,840	186,840

Note: This table reports difference-in-differences estimates based on model (1) for crime and labor market outcomes. Exposed is an indicator variable set to one for defendants sentenced up to 12 months after the court-specific cutoff date; zero otherwise. Post is an indicator variable set to one for the years the database was in place (i.e. 2014 and onwards); zero otherwise. The regressions control for age, gender, immigrant, time since charge, calendar month and the main effect of being exposed. Numbers in bracket show the mean of the dependent variable. Standard errors in parenthesis are clustered at the individual and court level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 3: Results by Demographic and Socioeconomic Characteristics

Panel A: Convicted	(1)	(2)	(3)	(4)	(5)
Exposed x Post	0.003	0.012***	-0.000	0.003	-0.002
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
x Male		-0.011***			
		(0.003)			
x Above age 25			0.007**		
			(0.003)		
x Immigrant				-0.001	
				(0.003)	
x HS degree					0.007**
					(0.003)
Sample mean	0.088	0.088	0.088	0.088	0.088
Panel B: Employed					
Exposed x Post	-0.008	0.005	0.016**	-0.010	0.006
	(0.008)	(0.010)	(0.007)	(0.008)	(0.007)
x Male		-0.017*			
		(0.009)			
x Above age 25			-0.047***		
			(0.006)		
x Immigrant				0.007	
				(0.007)	
x HS degree					-0.027***
					(0.008)
Sample mean	0.666	0.666	0.666	0.666	0.666
Panel C: Log Earnings					
Exposed x Post	-0.093**	-0.010	0.027	-0.114**	-0.010
	(0.042)	(0.064)	(0.046)	(0.047)	(0.042)
x Male		-0.107*			
		(0.060)			
x Above age 25			-0.237***		
			(0.048)		
x Immigrant				0.078*	
				(0.046)	
x HS degree					-0.167***
					(0.048)
Sample mean	5.835	5.835	5.835	5.835	5.835
Year FE	Yes	Yes	Yes	Yes	Yes
Court FE	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes
Observations	186,840	186,840	186,840	186,840	186,840

Note: This table reports difference-in-differences estimates based on model (1) of the effect of exposure on crime and labor market outcomes. The regressions control for the same variables as in Table 2. Standard errors in parenthesis are clustered at the individual and court level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 4: Case characteristics and labeling

Panel A: Convicted	(1)	(2)	(3)	(4)
Exposed x Post	0.003	0.002	0.001	0.002
	(0.003)	(0.003)	(0.004)	(0.003)
x Acquitted		0.007**		
		(0.004)		
x Misdemeanor			0.003	
			(0.003)	
x Neighborhood offenders				0.001
< median				(0.003)
Sample mean	0.088	0.088	0.088	0.088
Panel B: Employed				
Exposed x Post	-0.008	-0.004	0.001	-0.001
	(0.008)	(0.008)	(0.007)	(0.008)
x Acquitted		-0.029**		
		(0.011)		
x Misdemeanor			-0.036***	
			(0.009)	
x Neighborhood offenders				-0.017***
< median				(0.006)
Sample mean	0.666	0.666	0.666	0.666
Panel C: Log Earnings				
Exposed x Post	-0.093**	-0.065	-0.051	-0.054
	(0.042)	(0.046)	(0.043)	(0.044)
x Acquitted		-0.194**		
		(0.076)		
x Misdemeanor			-0.141**	
			(0.061)	
x Neighborhood offenders				-0.098**
< median				(0.039)
Sample mean	5.835	5.835	5.835	5.835
Year FE	Yes	Yes	Yes	Yes
Court FE	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes
Observations	186,840	186,840	186,840	186,840

Note: This table reports difference-in-differences estimates based on model (1) of the effect of exposure on crime and labor market outcomes. The regressions control for the same variables as in Table 2. Numbers in brackets show mean of the dependent variable. Standard errors in parenthesis are clustered at the individual and court level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 5: Effects on children's outcomes

Panel A: GPA (pct. rank)	(1)	(2)	(3)	(4)	(5)	(6)
Exposed x Post	-2.203	0.190	-1.043	-2.170	-0.799	-0.709
	(2.192)	(2.394)	(2.638)	(2.241)	(2.234)	(2.025)
x Defendant is the father		-3.368**				
		(1.671)				
x Parent HS degree			-1.828			
			(1.763)			
x Parent aqcuitted				-0.736		
				(1.493)		
x Parent misdemaneor					-3.827*	
					(1.972)	
x Neighborhood offenders < median						-3.706**
						(1.572)
Sample mean	42.572	42.572	42.572	42.572	42.572	42.572
Panel B: Court conviction						
Exposed x Post	0.009	-0.013	-0.000	0.009	-0.000	0.008
	(0.010)	(0.009)	(0.011)	(0.010)	(0.016)	(0.010)
x Defendant is the father		0.032***				
		(0.009)				
x Parent HS degree			0.012			
			(0.011)			
x Parent aqcuitted				0.0053		
				(0.013)		
x Parent misdemaneor					0.003	
					(0.011)	
x Neighborhood offenders < median						0.003
						(0.010)
Sample mean	0.023	0.023	0.023	0.023	0.023 2	0.023
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Court FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,119	5,119	5,119	5,119	5,119	5,119

Note: This table reports difference-in-differences estimates based on model (1) of the effect of exposure on compulsory school GPA (percentile rank) in grade 9. The sample consist of children to defendants in our main sample belonging to the cohorts who graduated from compulsory school 2010-2016. The regressions control for the same characteristics of the parents as in Table 2. Numbers in brackets show mean of the dependent variable. Standard errors in parenthesis are clustered at the individual and court level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix

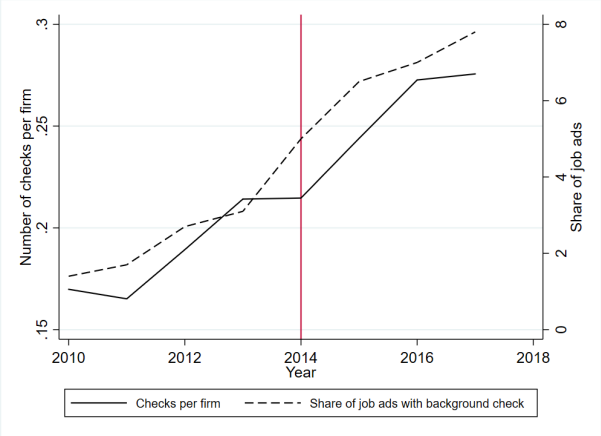


Figure A.1: Criminal background checks

Notes: This figure plots the number of official criminal background check inquires for personal use (likely to be submitted to a potential employer) per registered firm submitted to the Swedish police over the period 2010-2017 using data from the Swedish police and Statistics Sweden. The figure also plots the share of ads posted at the Public Employment Office that mentions criminal background checks using data from the Public Employment Office.

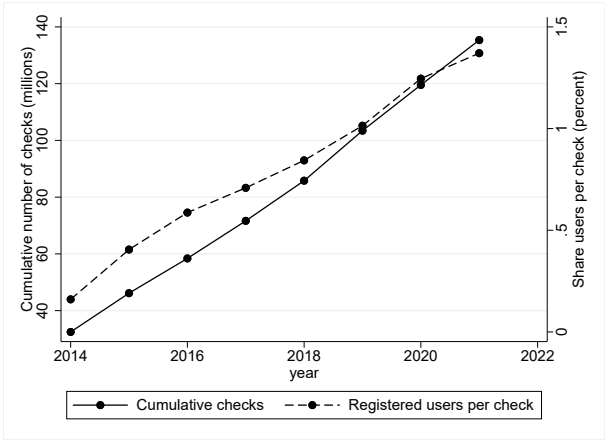


Figure A.2: Background checks using the platform

Notes: This figure plots the cumulative number of searches for criminal background using the platform. The figure also plots the number of registered users per background check conducted. The figure uses data provided by the company behind the platform.

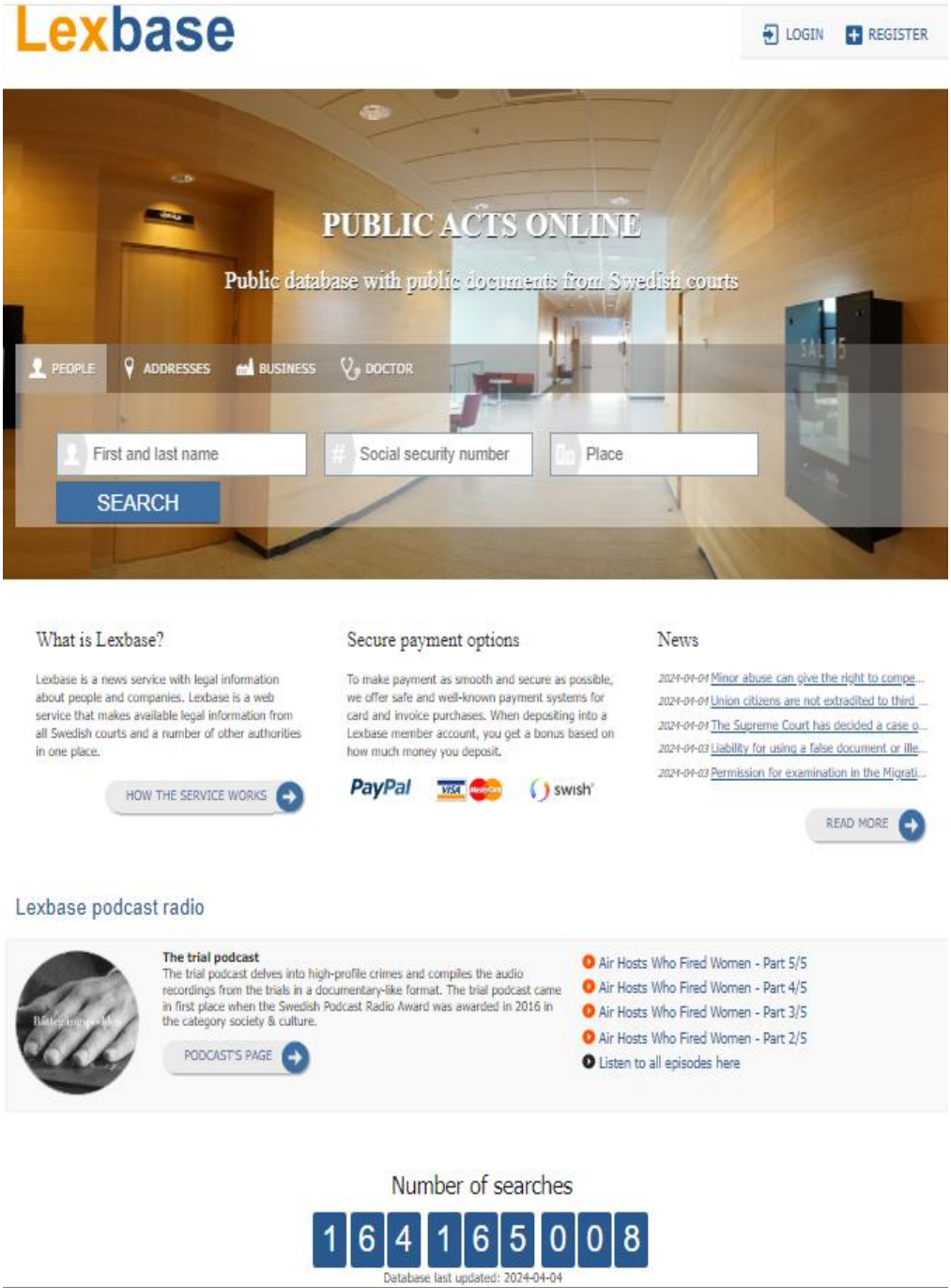


Figure A.3: Homepage Lexbase

Notes: This image shows the homepage of Lexbase and was retrieved on April 4, 2024.

How to use Lexbase

With a few single button presses, you can obtain useful decision-making information. You start by choosing a search method. The guide below explains how the service works.

1. Select search method

A) Search person
In this search function Search person you can specify your search using four fields; First name, last name, social security number and city. If you get a hit on a person, it means that he or she has been subject to a court hearing. If you do not get a hit, it is probably because the person does not appear in any legal process.

Example: Shows results for Anders Andersson in Stockholm

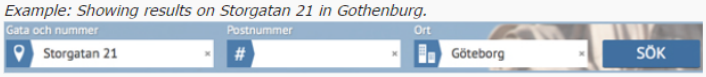


In the search fields, you can use asterisk (*) in your search text to replace part of a search term. You also don't need to search for an exact name of a person or company. Parts of the name in question are sufficient.

If you do not get a hit on a certain person, it is probably because they are not present in any legal process.

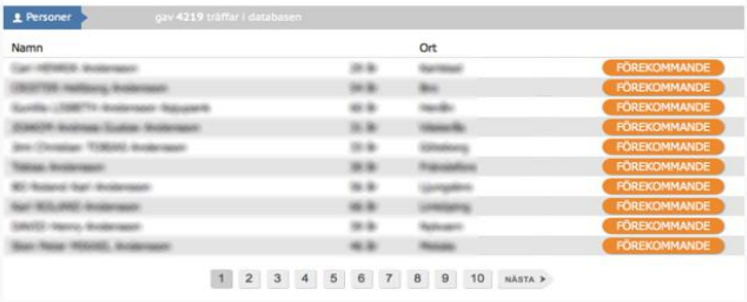
B) Search address
This search function helps you who want information about people living in a certain area or in your immediate area. The search tool sorts by Address, Postcode and City. The more fields you fill in, the narrower the hit picture you get.

Example: Showing results on Storgatan 21 in Gothenburg.



2. Search results are displayed

When you have made a search, the search result is presented in a list. To read a report, click on the button on the right which looks like this and is called FÖREKOMMANDE



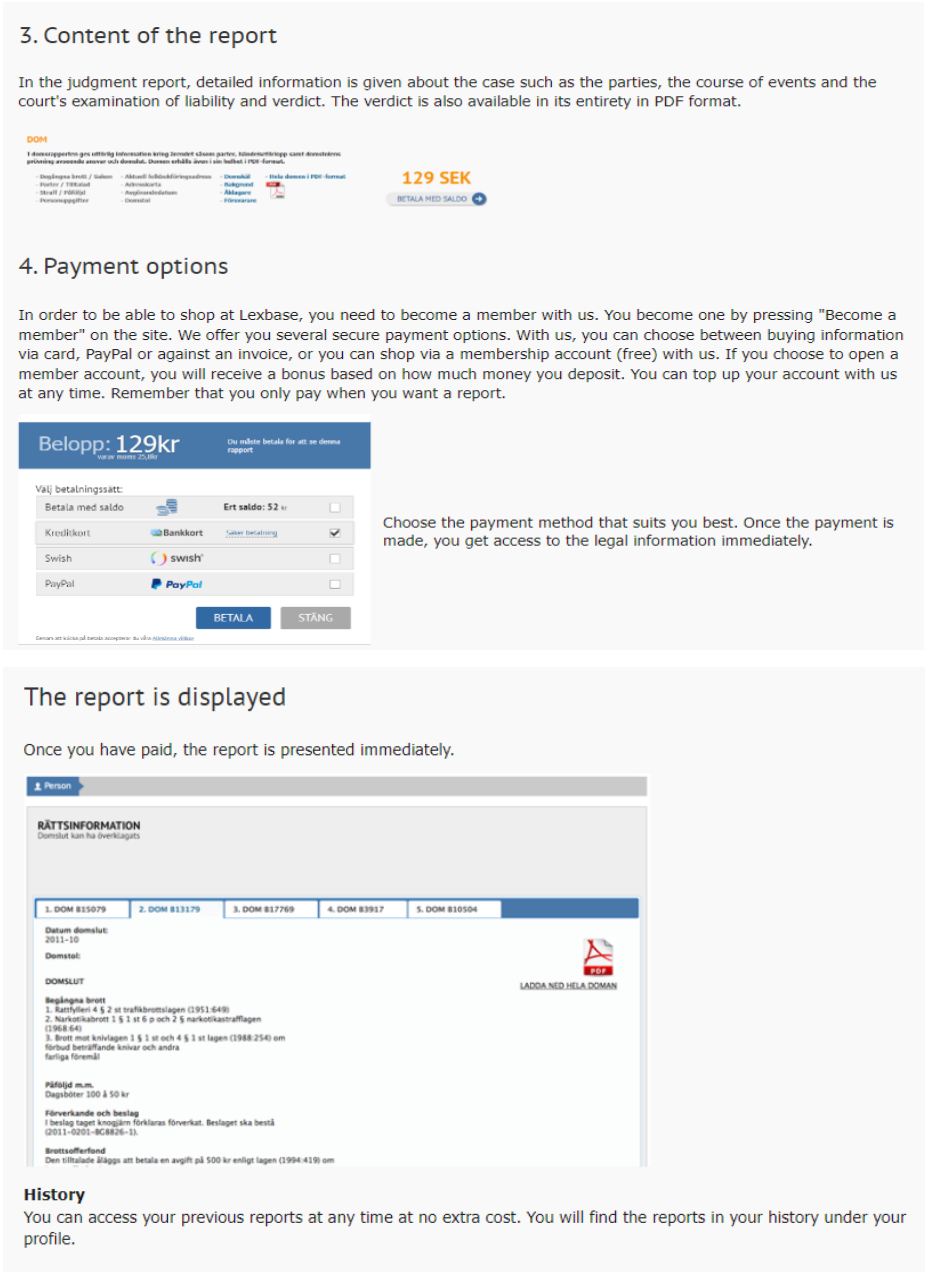


Figure A.4: How the Search-Engine at Lexbase Works

Notes: This image is from the homepage of Lexbase and provides the guide from the platform on how to use the search engine. The image was retrieved on April 4, 2024.

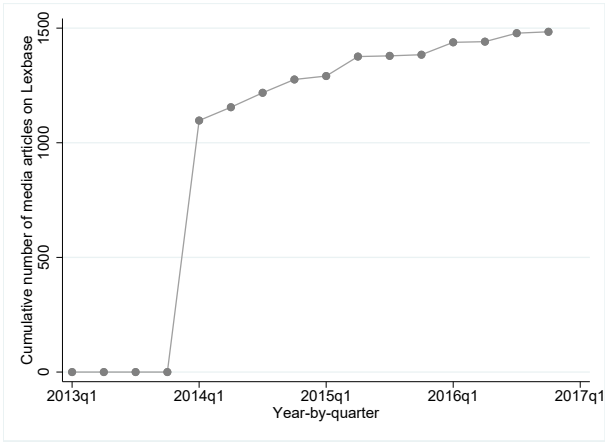


Figure A.5: Cumulative number of media articles covering Lexbase 2013-2016

Notes: This figure plots the cumulative number of media articles mentioning Lexbase in 2013-2016. Data on media coverage is from the Retriever/Mediearkivet database, which is a Nordic news archive with material from mainly major Swedish newspapers. The information was collected on Jan 20th, 2024.

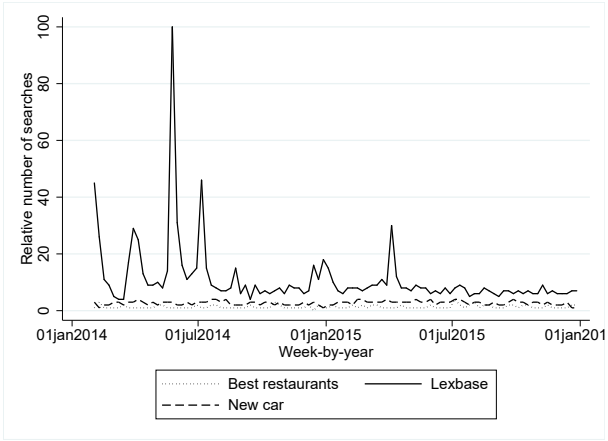
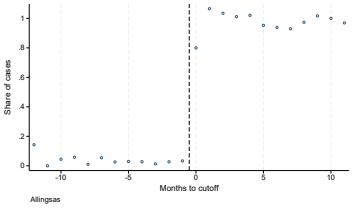


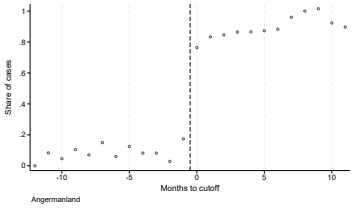
Figure A.6: Relative number of searches on Google

Notes: This figure plots the relative number of searches on Google for the terms "Lexbase", "Best restaurants" and "New car". The data from Google Trends was collected on Jan 21st, 2024. Google Trends normalizes search data to make comparisons between terms easier. Search results are normalized to the time and location of a query by dividing each data point by the total searches of the geography and time range it represents. The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics. To make the figure more easy to read we drop the days immediately following the launch of the database when the searches skyrocketed and focus on the period starting one week after the introduction and up until the last week of 2015.The information was collected on Jan 20th, 2024.

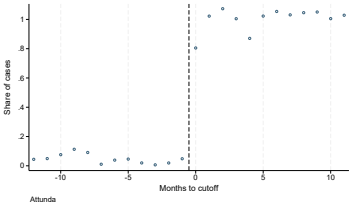
(a) DC Allingsås



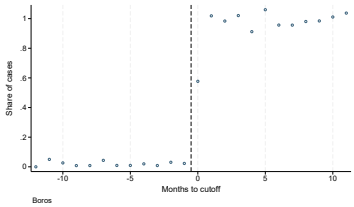
(b) DC Ångermanland



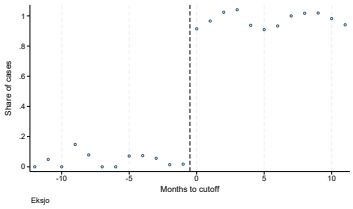
(c) DC Attunda



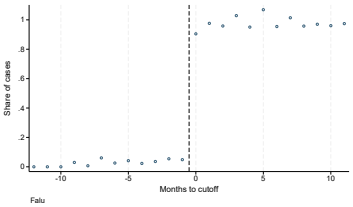
(d) DC Borås



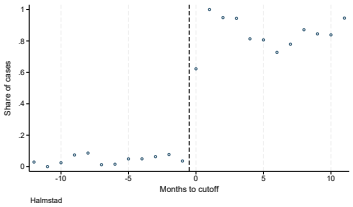
(e) DC Eksjö



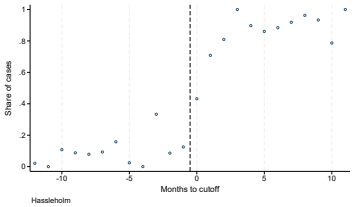
(f) DC Falun



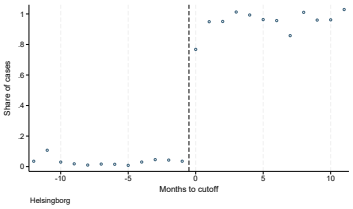
(g) DC Halmstad



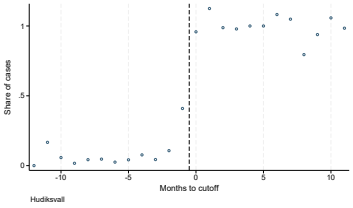
(h) DC Hässleholm



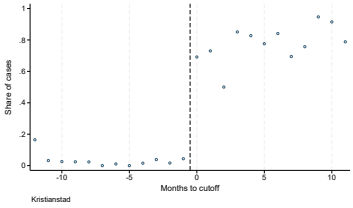
(i) DC Helsingborg



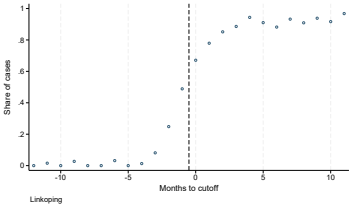
(j) DC Hudiksvall



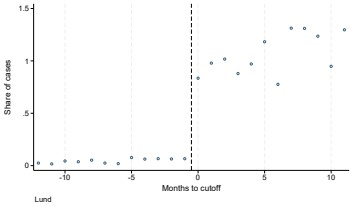
(k) DC Kristianstad



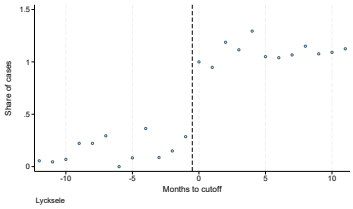
(l) DC Linköping



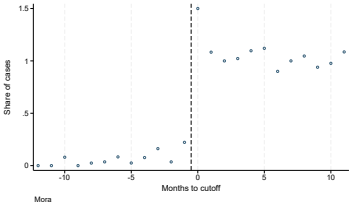
(m) DC Lund



(n) DC Lycksele



(o) DC Mora



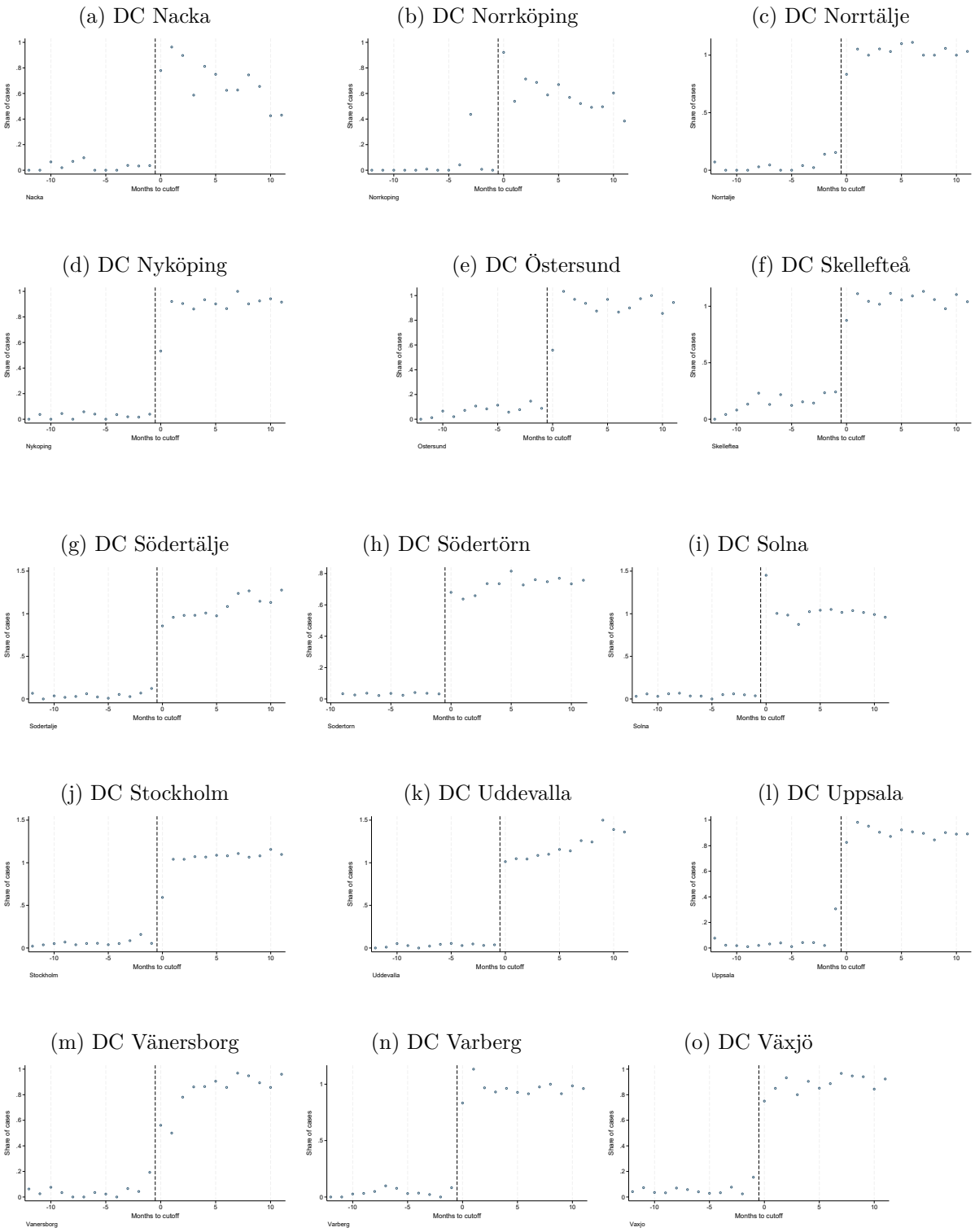
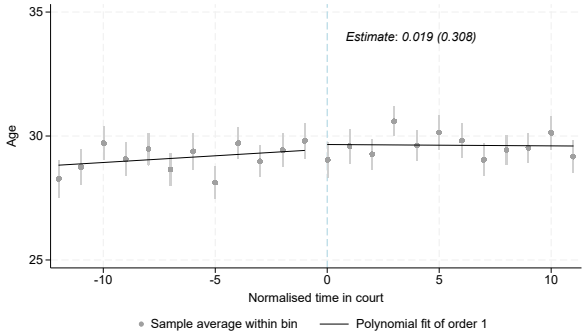


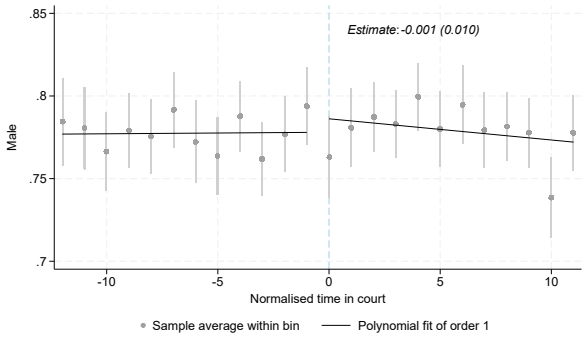
Figure A.8: Cases around the cutoff

Notes: The figure plots the share of Lexbase cases in our register data within a 12 month window around the cutoff separately for the district courts (DC) in our sample.

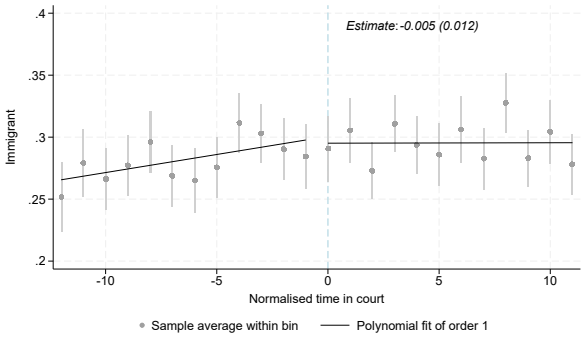
(a) Age



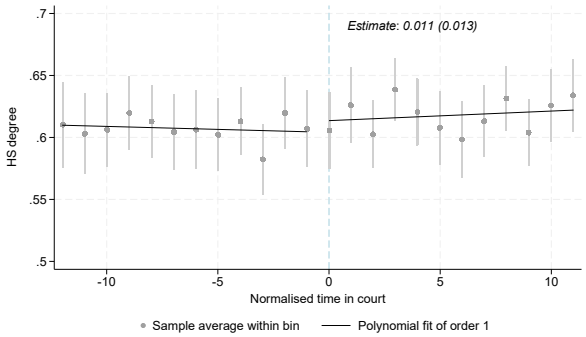
(b) Males



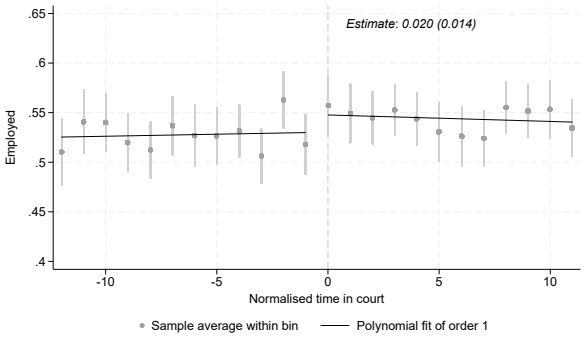
(c) Immigrants



(d) High school degree



(e) Employment



(f) Acquitted

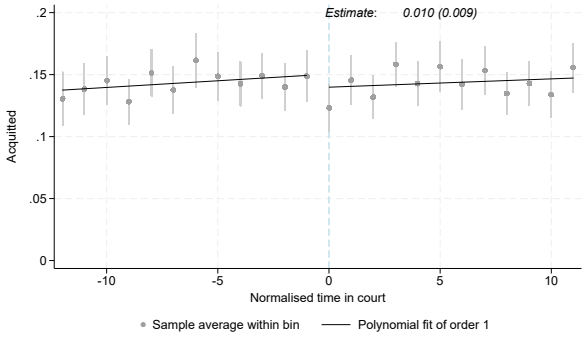


Figure A.9: Cases around the cutoff

Notes: The figure shows the average case and defendant characteristics by time relative to cutoff.

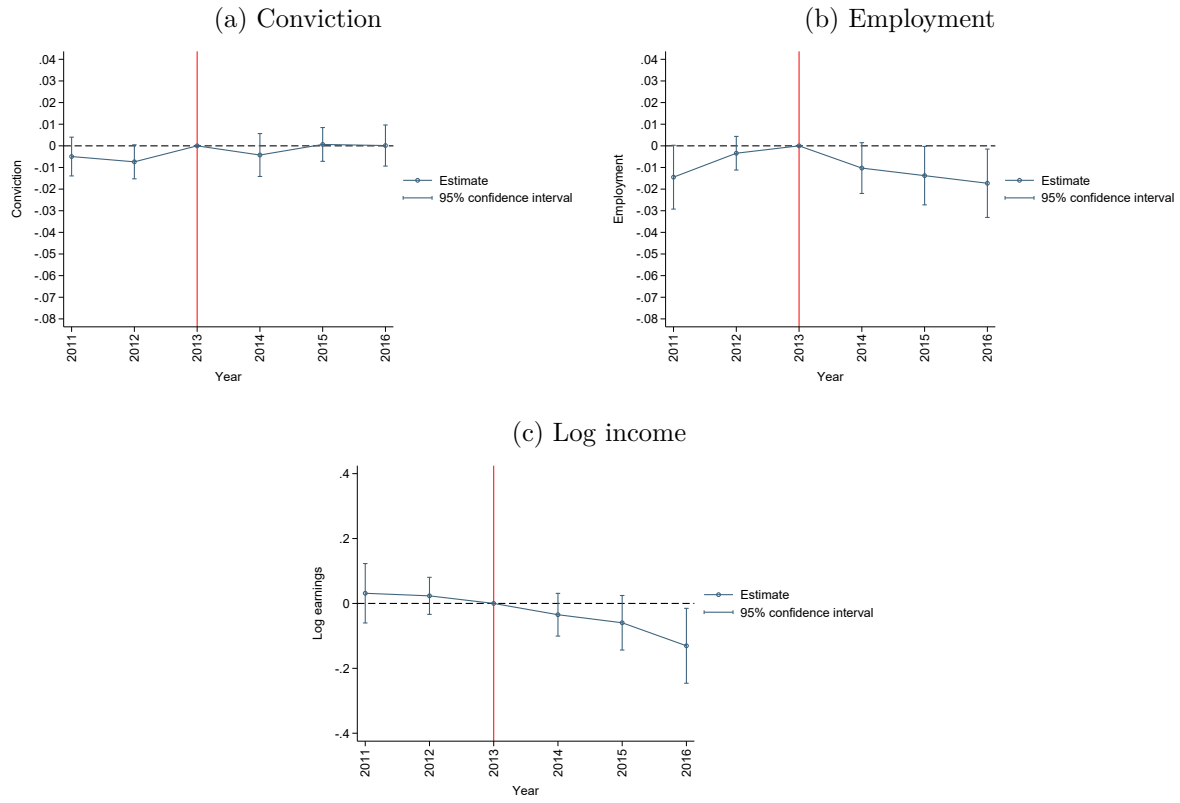


Figure A.10: Annual DD estimates of the exposure effects

Notes: This figure plots DD-estimates from model (1) by year relative to 2013. The regressions control for the same variables as in Table 2. The bars show the 95 percent confidence interval based on standard errors clustered at the individual and court level. Panel (a) shows conviction between 2011-2016, Panels (b)-(c) shows labor market outcomes in 2011-2016.

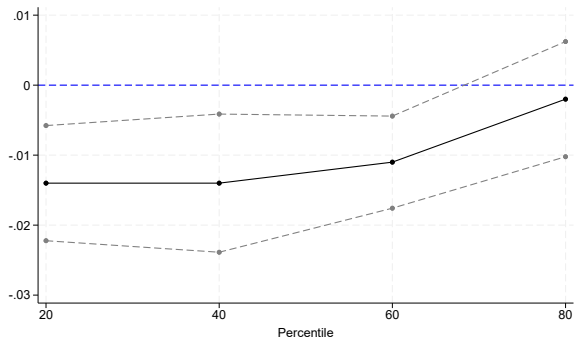


Figure A.11: Effect of exposure on position in the earnings distribution

Notes: This figure plots the estimated exposure effects on (solid line) along with the 90 percent confidence interval (dashed line) based on model (1) where the dependent variable is an indicator for the individual earning more than the i:th percentile in the earnings distribution. The regressions control for the same variables as in Table 2. Numbers in brackets show mean of the dependent variable. Standard errors in parenthesis are clustered at the individual and court level.

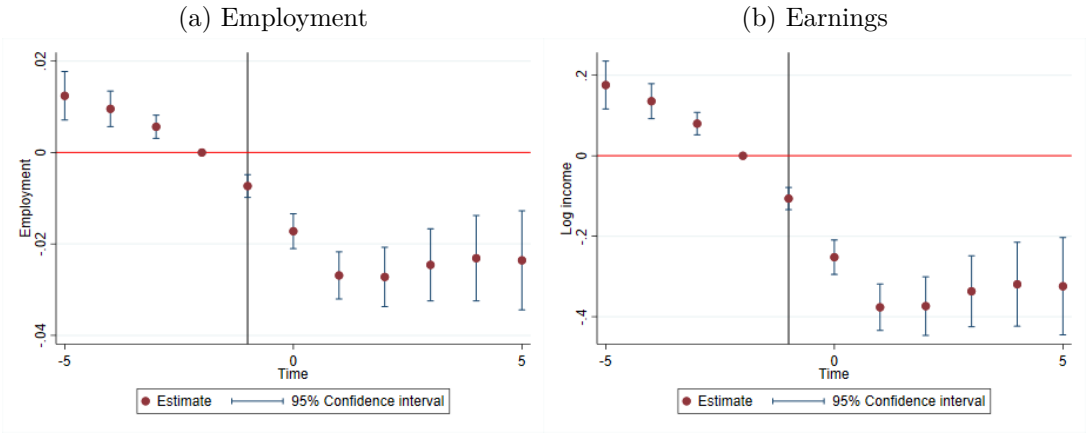


Figure A.12: Event study estimates of the effect of conviction.

Notes: This figure estimates event study coefficients +/- 7 years around the charge, where the end points after +/- 5 year are binned. Panel (a) gives the coefficients for employment, and Panel (b) gives the coefficients for log earnings. The sample is restricted to first-time offenders between 1997-2010 for those aged 25-57 at the time of the charge. All regressions control for age, year and individual fixed effects. The bars show the 95 percent confidence interval based on standard errors clustered at the individual level.

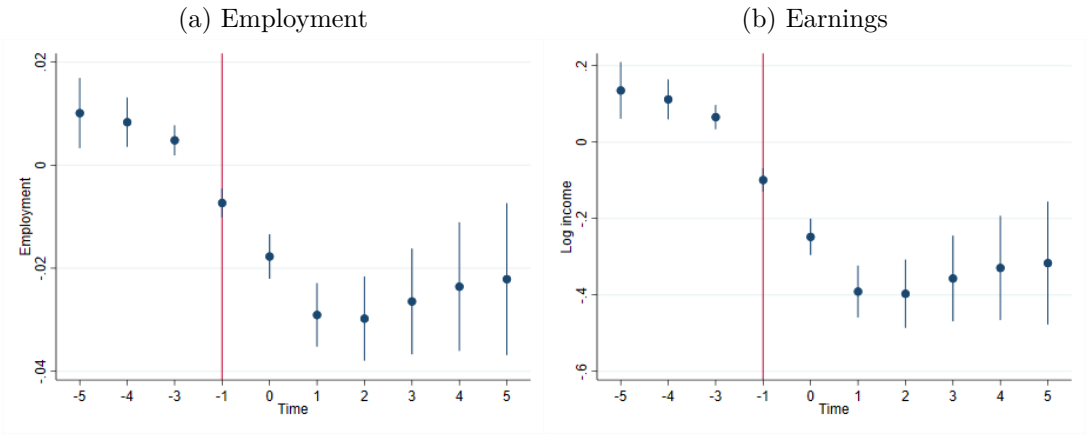


Figure A.13: Event study estimates of the effect of conviction following Sun and Abraham

Notes: This figure estimates event study coefficients +/- 7 years around the charge, where the end points after +/- 5 year are binned. The procedure from Sun and Abraham (2021) is employed, using the last treated unit as control, excluding time periods after the last unit received treatment. Panel (a) gives the coefficients for employment, and Panel (b) gives the coefficients for log earnings. The sample is restricted to first-time offenders between 1997-2010 for those aged 25-57 at the time of the charge. All regressions control for age, year and individual fixed effects. The bars show the 95 percent confidence interval based on standard errors clustered at the individual level.

Table A.1: Robustness

	Convicted	Employed	Log Earnings
Panel A: Cutoff -1 (N=172,548)	(1)	(2)	(3)
Exposed*Post	-0.000	0.003	0.023
	(0.003)	(0.006)	(0.045)
Outcome mean	0.084	0.674	5.875
Panel B: Cutoff +1 (N=206,364)			
Exposed*Post	-0.004	-0.003	-0.048
	(0.003)	(0.005)	(0.030)
Outcome mean	0.101	0.648	5.728
Panel C: Bandwidth 9 months (N=142,764)			
Exposed*Post	0.002	-0.011*	-0.120***
	(0.003)	(0.006)	(0.031)
Outcome mean	0.088	0.665	5.825
Panel D: Bandwidth 6 months (N=95,724)			
Exposed*Post	0.001	-0.010*	-0.135***
	(0.004)	(0.006)	(0.030)
Outcome mean	0.087	0.667	5.847
Panel E: Short-term effects (N=136,656)			
Exposed*Post	0.004	-0.007	-0.098***
	(0.004)	(0.006)	(0.035)
Outcome mean	0.101	0.532	5.118
Year FE	Yes	Yes	Yes
Court FE	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes

Note: This table reports difference-in-differences estimates based on model (1) of the effect of exposure on crime and labor market outcomes. Panels A and B show results from placebo cutoff dates one year before and one year after the actual cutoff date in a given court, respectively. Panels C and D show results using alternative bandwidths around the court-specific cutoff date. Panel E shows estimates from an alternative difference-in-differences model where the "exposed" group includes defendants who underwent a trial, while the "non-exposed" group comprises defendants who were directly convicted by a prosecutor for confession cases. These confession cases were not included in the database in 2014 and 2015. Outcomes are measured one year post-conviction and the repeated cross-sectional dataset runs from 2010 to 2015. The regressions control for the same variables as in Table 2. Numbers in brackets show mean of the dependent variable. Standard errors in parenthesis are clustered at the individual and court level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A.2: Additional outcomes

	All
	(1)
In college	0.004*
	(0.002)
	[0.029]
Labor market training	0.001
	(0.002)
	[0.011]
Moving	0.004
	(0.006)
	[0.227]
Single household	0.001
	(0.005)
	[0.703]
Get pregnant	-0.001
	(0.003)
	[0.052]
Year FE	Yes
Court FE	Yes
Baseline Controls	Yes
Observations	186,840

Note: This table reports difference-in-differences estimates based on model (1) for alternative outcomes. The regressions control for the same variables as in Table 2. Numbers in brackets show the mean of the dependent variables. Standard errors are clustered at the individual and court level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A.3: Employment penetration by industry

	Ratio	Penetration classification
	(1)	(2)
Banking	2.2	Low
Construction	0.4	High
Culture	1.2	High
Education	2.7	Low
Electricity	1.9	Low
Farming	1.5	High
Healthcare	1.6	Low
Hotel and Restaurants	0.6	High
Information	1.5	High
Law and Economics	1.3	High
Letting	0.6	High
Manufacturing	1.0	High
Mining	0.9	High
Other services	1.4	High
Public Administration	2.6	Low
Real estate	1.2	High
Retail	0.9	High
Transport	0.5	High
Water & Sanitation	0.6	High

Note: This table presents the ratio of employment in a particular industry for individuals without a court conviction relative to those with a court conviction in year 2013. No adjustments for baseline characteristics are done. Equal employment penetration among these two groups is expressed as a ratio equal to 1. We classify industries as low penetration if the ratio exceeds 1.5.

Table A.4: Industry effects

	All
	(1)
Empl. in high penetration industries 1.5	-0.010*
	(0.006)
	[0.584]
Empl. in low penetration industries 1.5	-0.003
	(0.003)
	[0.150]
Empl. in high penetration industries 1.2	-0.008*
	(0.004)
	[0.489]
Empl. in low penetration industries 1.2	-0.004
	(0.005)
	[0.245]
Empl. in high penetration industries 1	-0.005
	(0.005)
	[0.398]
Empl. in low penetration industries 1	-0.00
	(0.005)
	[0.336]
Year FE	Yes
Court FE	Yes
Baseline Controls	Yes
Observations	186,840

Note: This table reports difference-in-differences estimates based on model (1) of the effect of exposure on sorting across industries. The regressions control for the same variables as in Table 2. Numbers in brackets show mean of the dependent variable. Standard errors are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A.5: AKM sorting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exposed x Post	0.002	0.016	0.014*	-0.005	0.021*	0.006	0.011	0.016**
	(0.007)	(0.012)	(0.007)	(0.006)	(0.009)	(0.006)	(0.008)	(0.008)
x Male		-0.017*						
		(0.009)						
x At least age 25			-0.022**					
			(0.006)					
x Immigrant				0.033***				
				(0.009)				
x HS degree					-0.031***			
					(0.008)			
x Acquitted						-0.019**		
						(0.009)		
x Misdemaneor							-0.036***	
							(0.009)	
x Nbh offenders < med.								-0.032***
								(0.007)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Court FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	100,152	100,152	100,152	100,152	100,152	100,152	100,152	100,152

Note: This table reports difference-in-differences estimates based on model (1) of the effect of exposure on the sorting of workers to firms with different earnings premia. The dependent variable is the earnings premia estimated using the methods suggested by Abowd, Kramarz and Margolis (1999).. We define a firm as an establishment. To estimate the AKM model, we include firm fixed effects, individual fixed effects, and age and year fixed effects. The age dummies are education-gender-specific (following Engbom, Moser and Sauerman (2023)). Due to the problem of multicollinearity between age and year-fixed effects, we assume that the age effects are constant between ages 45 and 54. We estimate the AKM model for firms in the connected set over the years 2005-2016, including workers aged 18-65. We exclude establishments with fewer than 5 movers but note that the results are very similar also when including all establishments in the connected set. The regressions control for the same variables as in Table 2. Standard errors are clustered at the individual and court level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A.6: Firm switching

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exposed x Post		-0.002	-0.029***	-0.008	-0.023**	-0.010	-0.013*	-0.011
		(0.013)	(0.009)	(0.010)	(0.010)	(0.007)	(0.008)	(0.008)
x Male		-0.007						
		(0.011)						
x At least age 25			0.040***					
			(0.006)					
x Immigrant				0.003				
				(0.007)				
x HS degree					0.024***			
					(0.006)			
x Acquitted						0.017**		
						(0.007)		
x Misdemaneor							0.025***	
							(0.008)	
x Nbh offenders < med.								0.007
								(0.006)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Court FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	103,908	103,908	103,908	103,908	103,908	103,908	103,908	103,908

Note: This table reports difference-in-differences estimates based on model (1). The dependent variable is an indicator set to one if the individual has changed firm from one year to another and zero if (s)he stayed in the firm. Individuals who are non-employed are missing in the data. The regressions control for the same variables as in Table 2. Standard errors are clustered at the individual and court level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A.7: Results for parental subsample

	Convicted		Employed		Log earnings		Family log disp. inc.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exposed*Post	0.004	0.004	-0.023*	-0.023*	-0.104	-0.104	-0.036	-0.036
	(0.006)	(0.006)	(0.013)	(0.013)	(0.074)	(0.074)	(0.035)	(0.034)
Sample mean	0.049	0.049	0.735	0.735	6.221	6.221	8.084	8.084
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Court FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	23,388	23,388	23,388	23,388	23,388	23,388	23,388	23,388

Note: This table reports difference-in-differences estimates based on model (1) for crime and labor market outcomes restricted to the subsample of individuals in our main sample that were parents of children belonging to the grade 9 graduation cohorts 2010-2016. Exposed is an indicator variable set to one for defendants sentenced up to 12 months after the court-specific cutoff date; zero otherwise. Post is an indicator variable set to one for the years the database was in place (i.e. 2014 and onwards); zero otherwise. The regressions control for the same variables as in Table 2. Numbers in bracket show the mean of the dependent variable. Standard errors in parenthesis are clustered at the individual and court level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

B Data Appendix

Exposed: An indicator equal to one if the individual was sentenced within 12 months after the court-specific cutoff month.

Non-Exposed: An indicator equal to one if the individual was sentenced within 12 months before the court-specific cutoff month.

Male: An indicator equal to one if the individual is a male.

Immigrant: An indicator equal to one if the individual was born outside of Sweden.

Age at Initial Trial: We calculate age at initial trial as the year of the trial minus the calendar year when the individual is born.

High School Degree: An indicator for whether an individual has a high school degree or a higher education degree (e.g., college, university).

Convicted: An indicator set to one if an individual was charged and convicted in the initial trial.

Acquitted: An indicator set to one if an individual was charged and acquitted in the initial trial.

Misdemeanor: An indicator set to one if the the offender committed a crime that no defendant has ever been sentenced to prison for in our data.

Convictions per 10,000 inhabitants: The number of convictions per 10,000 inhabitants in the municipality of residence.

Criminal Conviction: An indicator for whether the charged individual is convicted of new crime in a given year.

Employment: Measures the employment status of the charged individual in a given year. A binary employment indicator set equal to one if the charged individual is registered as employed on Nov 1st.

Earnings ($1,000s)): Annual labor earnings. Earnings are deflated to 2015 and represented in U.S. dollars using the exchange rate SEK/$ = 9.25.

Mobility: An indicator set to one if the individual has changed place of residence in a given year.

In college: An indicator set to one if the individual was enrolled in college education in a given year.

Labor market training: An indicator set to one if the individual was enrolled in labor market training in a given year.

C Toy Model

Assume the employer has three different expected utility levels form hiring applicants of three given types: no record (N), a record of a misdemeanor (M), and a record of a felony (F),

$$E[U_N] > E[U_M] > E[U_F].$$

We further assume that the employer is interested in hiring all types. The population shares of the three types satisfy $p_N + p_M + p_F = 1$.

Assume a fixed cost for identifying the type of criminal record of $C_N = 0 < C_M < C_F$. We impose a zero-profit condition on firms (as a result of Bertrand competition in the labor market). This leads to a natural ranking of wages $W_N > W_M > W_F$ where

$$\begin{aligned} W_N &= E[U_N] \\ W_M &= E[U_M] - C_M \\ W_F &= E[U_F] - C_F. \end{aligned}$$

Now suppose we introduce the database that can, for free, separate N from $\{M, F\}$, but we cannot distinguish M from F.

Again, under a zero-profit condition, this would not affect wages for N. In contrast, M and F would now get the same wage and the employer saves the cost C_M and C_F. Let's call this new wage W_{db}. This should lead to higher wages for F and lower wages for M. Can we show this?

When using the database, a risk-neutral employer is willing to pay the wage W_{db} to any applicant who is in the database. This wage is equal to the employer's expected utility of

$$W_{db} = \frac{p_M}{p_M + p_F} E[U_M] + \frac{p_F}{p_M + p_F} E[U_F].$$

The wage for F will increase:

$$\begin{aligned} W_{db} &= \frac{p_M}{p_M + p_F} E[U_M] + \frac{p_F}{p_M + p_F} E[U_F] \\ &> E[U_F] > E[U_F] - C_F \\ &= W_F. \end{aligned}$$

The wage for M will decrease under the condition

$$\begin{aligned} W_{db} = \frac{p_M}{p_M + p_F} E[U_M] + \frac{p_F}{p_M + p_F} E[U_F] &< E[U_M] - C_M \\ \iff p_M E[U_M] + p_F E[U_F] &< (p_M + p_F) E[U_M] - (p_M + p_F) C_M \\ \iff C_M &< \frac{p_F}{p_M + p_F} (E[U_M] - E[U_F]), \end{aligned}$$

i.e., if the (weighted) utility increase from hiring an M rather than an F outweighs the cost of identifying an M outside of the database.

In summary, it is plausible that the introduction of the database will decrease the wages of M and increase the wages of F, where the magnitudes depend on assumptions.

We have neglected the question of whether the database will crowd out individual screening of candidates. This, again, depends on assumptions (but we haven't checked it properly). A firm

that uses the database saves on screening cost (C_M, C_F). The wages for N are unaffected, but M and F now get the same wage. This would mean an M gets a higher wage at a firm that does individual screening, whereas F are better off at a firm that uses the database.