Dissertations in Economics 2022:1

Stockholm University

Essays on Economic Disadvantage

Criminal Justice, Gender and Social Mobility

Ulrika Ahrsjö

Doctoral Thesis in Economics at Stockholm University, Sweden 2022

Essays on Economic Disadvantage

Criminal Justice, Gender and Social Mobility

Ulrika Ahrsjö

Academic dissertation for the Degree of Doctor of Philosophy in Economics at Stockholm University to be publicly defended on Tuesday 14 June 2022 at 10.00 in Nordenskiöldsalen, Geovetenskapens hus, Svante Arrhenius väg 12.

Abstract

Youth Crime, Community Service and Labor Market Outcomes
Can lifetime trajectories of youth offenders be improved through criminal justice policy? I evaluate the effects of a youth justice reform in Sweden that sharply increased the share of juveniles assigned to court-ordered community service --- i.e. unpaid, low-skilled work. On average, the reform did not affect post-conviction recidivism or labor market outcomes, but these average effects mask considerable heterogeneity depending on the most likely alternative sanction. In particular, post-reform recidivism and incarceration rates are lower for individuals for whom community service replaces fines. Applying a machine learning method for causal inference, I then evaluate the net financial effect of the policy conditional on observable characteristics and analyze how the program could be targeted for improved efficiency. The results suggest that community service can benefit youth offenders, but that it is not suitable as a universal program.

Intergenerational Mobility Trends and the Changing Role of Female Labor
We present new evidence on the existence and drivers of trends in intergenerational income mobility using administrative income data from Scandinavia along with survey data from the United States. Harmonizing the data from Sweden, Denmark and Norway, we first find that intergenerational rank associations in income have increased uniformly across Scandinavia for cohorts of children born between 1951 and 1979. Splitting the trends by gender, we find that father-son mobility has been stable in all three countries, while correlations involving females display substantial trends. Similar patterns are confirmed in the US data, albeit with slightly different timing. Utilizing information about individual occupation, education and income in the Scandinavian data, we find that intergenerational mobility in latent economic status has remained relatively constant for all gender combinations. This is found to be driven by increased female labor market participation at the intensive as well as the extensive margin. The observed decline in intergenerational mobility in Scandinavia is thus consistent with a socially desirable development where female skills are increasingly valued in the labor market.

Wage Inequality, Selection and the Evolution of the Gender Earnings Gap in Sweden
We estimate the change in the gender wage gap between 1968 and 2019 in Sweden accounting for (1) changes in the intensive margin of labor supply; (2) changes in the overall wage inequality; (3) changes in selection into the labor market using parametric and non-parametric selection corrections. Our results show that between 1968 and 1991, about half of the changes in the gender wage gap can be attributed to changes in the overall wage distribution. Conversely, changes in the wage distribution from 1991 to 2019 mask a larger closure of the gender wage gap. Our corrections for selection into the labor force suggest that uncorrected estimates miss about half of the around 20 percentage points decrease in the gender wage gap over the 1968-2019 period.

Identity in Court Decision-Making
We explore the role of identity along multiple dimensions in high-stakes decision-making. Our data set contains information about gender, ethnic background, age and socioeconomic indicators for randomly assigned jurors and defendants in a Swedish district court. Our results show that defendants are significantly less likely to get a prison sentence if they and the jurors belong to the same identity-forming group. For example, a defendant is 15 percent less likely to get a prison sentence if he or she has the same level of education as all three jurors compared to if none of them have the same educational attainments.

Keywords: *Labor economics, Economic inequality, Criminal justice, Youth crime, Intergenerational mobility, Gender wage gap, In-group bias.*

Stockholm 2022
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-204123

ISBN 978-91-7911-894-5
ISBN 978-91-7911-895-2
ISSN 1404-3491

Department of Economics

Stockholm University, 106 91 Stockholm

Stockholm University

ESSAYS ON ECONOMIC DISADVANTAGE

Ulrika Ahrsjö

Stockholm University

Essays on Economic Disadvantage

Criminal Justice, Gender and Social Mobility

Ulrika Ahrsjö

ISBN print 978-91-7911-894-5
ISBN PDF 978-91-7911-895-2
ISSN 1404-3491

Printed in Sweden by Universitetsservice US-AB, Stockholm 2022

Till Thomas,
En dag har du en egen.

Acknowledgments

Funny enough, I think the decision to apply for a bachelor program in business and economics might be the least rational choice I ever made. At age 20, I had no idea what economics was, and I was pretty sure I would not enjoy the business part. What does that say about me as an economist? While I might never fully understand this choice, I have since learned that maths isn't actually so bad, and that economics can — thankfully — be the science of other things than interest rates.

With that said, I would have never made it this far without the company of some key people. In particular, I'm incredibly grateful to my supervisors, who made my PhD studies a less lonely experience. I remember asking if I could join your project precisely because I realized I did not want to write an entire thesis on my own, and it turns out to have been a great decision. Mårten, thanks for being such a patient (and chatty) mentor! I think I have learned more from just talking to you than from any university course. And Susan, thanks for being the best possible role model. You make me believe in myself as a researcher!

Without colleagues, I also would have given up long ago, and so many of you deserve my warmest thanks! To Elin, Pernilla, Mikaela, Erik, Markus, Dave, Francesco, Agneta and Carolina for sharing the first semester of courses with me (and for calmly explaining difficult stuff despite me crying from pregnancy hormones). To Roza for being the wisest office mate I could have asked for. You always have the answers, the right amount of excitement for research and also a good sense of when it's time to crack a joke. To Louise, Nanna, Evelina, Xueping, Anna and Malin for sharing everyday life at the office, and for letting me whine all I want about my children; to Jens for being the in-house Wikipedia of data questions, and to Miika and Jakob for sharing your labor economics wisdom with me. To René and Joachim for being my best (and only) Zoom colleagues during the pandemic. I am grateful to Anna Tompsett, Mitch Downey and David & David for your generosity with time and encouragement. Special thanks to Anne for sharing your enthusiasm in general and about running in particular!

But the whole reason I even applied to the PhD program in the first place is of course because of you, Thomas. This thesis is dedicated to you, for your excitement about academia and research that somehow spilled over to me. Your calm belief in my ability to do whatever I set out to do is what has kept me going when I don't believe it myself. And a special thanks to Nora and Ellen, for being the best possible perspective on what matters in life.

Finally, a big thank you to my parents. You let me make my own decisions about education quite early on in life (I even recall choosing primary school myself!), but before that, I think that you gave me the two most important tools for getting this far. Mum, you patiently taught me to read and write very early, which I think cemented my self-confidence in school work. And my dad's highly annoying habit of answering every question with a question of his own is probably a key reason why I enjoy research. If you had had the patience for it, I'm certain you would have made an excellent researcher!

Ulrika Ahrsjö
Stockholm
May 2022

Contents

Introduction

Whether the economic fate of an individual is a matter for the state is a normative question of preferences. However, no matter one's political views, most modern societies — from the social democratic Scandinavian welfare states to the more individualistic United States — agree on the principle that all people should be able to make a good life for themselves if they make an effort. One prerequisite for a good life is the ability to sustain a decent income, and another is fair treatment as a citizen by representatives of the state. As is well known, there are many ways in which this principle is not upheld in practice.

This thesis touches upon three aspects of structural disadvantage in society, which transform into economic inequality. Two of these are well-known and extensively studied in economic literature: gender differences in labor earnings and the role of family heritage in labor market success. The third topic — the workings of the criminal justice system — is less obviously connected to economics. In recent years, however, a large literature has documented how income is affected by criminal prosecution and sanctioning (Aizer and Doyle, 2015, Bhuller et al., 2020, Rose, 2021, Mueller-Smith and Schnepel, 2020) and that justice system actors sometimes fail to uphold the standards of equality before the law (Anwar, Bayer and Hjalmarsson, 2022, Arnold, Dobbie and Yang, 2018).

The first two chapters examine the Swedish criminal justice system. In the first essay **Youth Crime, Community Service and Labor Market Outcomes**, I study to what extent the criminal sanction matters for young offenders. As opposed to previous literature, which has mainly focused on comparing outcomes for youth who were sent to prison for their crimes against youth given any other type of sanction, I study different forms of *community-based sanctions*. These are criminal sanctions used extensively for young offenders, with the dual goal of punishment and rehabilitation from deviant behavior. In fact, only around one to twelve percent of all juvenile offenders in Western countries receive a prison sentence, with the remaining share being sentenced to community-based sanctions. These come in many different forms, but some of the most commonly used are fines, probation, out-of-home placement and individual treatment programs.

Another common option is court-mandated community service, whereby the offender is required to perform unpaid low-skilled work. As such, community service constitutes a form of middle ground in between a punishment (unpaid work) and rehabilitation (a chance to "pay back" for the damage inflicted upon the local society). Depending on the quality of the work, doing community service might influence human capital, either by learning job-specific skills, or by demanding that participants demonstrate basic non-cognitive skills such showing up on time and taking orders. In this sense, community service resembles an employment training program, and could as such benefit young offenders by facilitating labor market entry. Apart from the unpaid work, youth community service is in the setting of this essay coupled with behavioral therapy, with the aim of reflecting upon future life opportunities and learning impulse control.

The essay centers around a reform to the juvenile criminal justice system in Sweden that was implemented on the first of January 2007, and which changed the extent to which different community-based sanctions were given to youth offenders. A new sanction was introduced: youth community service, and it immediately became the most common sanction for young offenders, mainly replacing fines and referral to child welfare services for treatment ("care"). In 2007, about half of all minors convicted in courts were given community service, while the remaining half were about equally likely to get either a monetary fine to pay, or be referred to the care sanction. I don't find that the reform affected either general deterrence among youth, or the composition of the youth offender population in terms of observable characteristics. The reform thus constitutes a sharp change in probabilities of different sanctions, which I use in a difference-in-discontinuity regression model, comparing youth convicted just before versus just after the reform, and differencing out seasonal variation in criminal cases with conviction from pre-reform years.

I evaluate the effects of this change in the sanctions composition — higher likelihood of community service, lower likelihood of fines and care — on relapse into crime, educational outcomes, and labor market outcomes in early adulthood. Surprisingly, given the large scale of the reform, I do not find that either of those outcomes are affected in the general population of youth offenders. In order to understand this average treatment effect of zero, I split the population into two disjoint subgroups: individuals whose most likely sanction in pre-reform years is a fine, and individuals whose most probable sanction in pre-reform years is care.[1] This heterogeneity analysis reveals that the average zero effect is a result of opposite effects in these two groups. Whereas the fines group react to the reform by lower rates of future recidivism and a higher probability of graduating high school, the care group become more likely to relapse into serious crime, increasing their risk of incarceration spells as young adults. Neither group is found to respond to the reform in terms of labor market earnings or employment probabilities, despite the changes in crime propensities.

When I summarize these different effects into a measure of "net financial effects" (costs of future crimes and sanctions and net state transfers from income taxes and welfare payments), I find that these opposite effects between the fines and care groups persist. I then evaluate if the targeting of the reform was correct: is community service directed at *all* youth offenders a good policy, or could criminal justice efficiency be improved by a more narrow targeting of the program? And in that case, who should get it? For this heterogeneity analysis, I use a novel machine learning approach, the *causal forest* (Athey and Imbens, 2016, Wager and Athey, 2018) to calculate

[1] This is done by predicting the likelihood of each sanction on pre-reform convictions (from years 2003-2006), and extrapolating these probabilities into the reform year-convictions. Individuals with a predicted probability above the 75th percentile are assigned to the respective groups: "fines" and "care".

individual-level causal effects. This is a data-driven approach to understanding treatment effect heterogeneity among a large set of possible covariates of interest.

Two results are worth mentioning. First, among the whole sample, the probability of fines (calculated from crime characteristics alone) is the single strongest predictor of positive treatment effects, meaning net cost savings. In other words, my results suggest that community service is a good alternative to monetary fines. Second, I find that individual characteristics that predict net cost savings are not the same within the fines group, as within the care group. In particular, for individuals in the fines group, having had any prior employment is positively correlated with net financial gains. This could mean that for this group, who are often convicted of relatively light crimes, the reform ameliorated the detrimental effect on future employment that a criminal conviction constitutes. On the other hand, among the care group, a small positive treatment effect is found among younger individuals from relatively more stable home and schooling conditions. One interpretation of this is that if community service is to replace individual treatment, it is crucial to consider the youth's family situation.

In sum, I show that the choice of criminal sanctions for youth offenders matters, even among different community-based alternatives. While the context here is in some regards unique — Sweden stands out internationally in its focus on restorative justice — the question of punishment or treatment programs for youth offenders is universal. That this by definition low-cost policy can reduce severe recidivism when targeted at the right group of offenders ought to motivate further policy experiments increasing the role of community service among community-based criminal sanctions for young offenders.

In the second essay, **Identity in Court Decision-Making**, focus is shifted from criminal sanctions to the process of decision-making in court trials. Sweden, like many other countries, employs a system of trial by jurors, whereby common people are involved in deciding court outcomes. This is meant to ensure representation and democratic values, as criminal offenders are judges by their peers. Crucially, the underlying idea is then that similarity between offenders and jurors ensures fairness in the criminal justice system.

A large literature in social psychology and behavioral economics documents so-called "in-group bias", i.e. that humans favor similar others in their decision-making (Tajfel and Turner, 1986, Huffman, Meier and Goette, 2006, Bernhard, Fischbacher and Fehr, 2006, Rand et al., 2009). Economics research on e.g. gender pay differences and occupational choice nowadays often take into account that identity and social norms shape a person's choice set in life (Akerlof and Kranton, 2000, Bertrand, 2011). In extension, identity-based decisions of one person might affect other people, a form of "identity externality". The juror system can be seen as institutionalized in-group bias, but to what extent is this evident in court outcomes?

This essay draws upon a comprehensive data set of identity traits of jurors in criminal court trials, which we have collected from transcripts of court hearings from the Stockholm District Court for the period 2000-2004 and then linked to Swedish administrative registers. Since jurors are randomly assigned to court cases, we can interpret effects of the juror composition on defendant outcomes as causal. We observe several attributes about the jurors and defendants, which we use to study if similarity between jurors and the defendant in a court hearing sways the outcome either in favor or or against the defendant. Of these, three pertain to demographic identity: gender, ethnic background, age; and three can be thought of as representing socioeconomic identity: educational background, family disposable income, and neighborhood of residence. We study identity effects from each of these characteristics separately, and summarize them into indices for demographic identity, socioeconomic identity, and finally an index of all six together.

Our results show that identity is an important factor in juror decision-making. Jurors who are randomly assigned to defendants that are more similar to themselves are more lenient in their decision-making. The previous literature on the subject has found evidence supporting in-group biases in both gender (Bagues, Sylos-Labini and Zinovyeva, 2017) and ethnicity (Glaeser et al., 2000), but our study is the first to extend the analysis to socioeconomic factors. Interestingly, our results would suggest that socioeconomic attributes are at least as — if not more — important for the formation of in-group biases, as demographic attributes. In particular, we find education to be a strong source of identity effects: defendants faced with a juror triplet where all members have the same educational attainments as themselves have a 15 percent lower risk of incarceration, and on average get prison sentences that are half as long.

To say something about how are these biases formed, we look at heterogeneous effects by subgroups. Gender identity creates a strong in-group bias, which interestingly is mainly a result of women being less prone to sentence other women to prison. Native jurors clearly favor native defendants, while there is no such significant effect on foreign born defendants. We also show that the identity effects are only present among defendants who personally attend their court hearing, and that the main results are driven by juror groups paired with a more lenient judge, where they presumably have more say in the sentencing decision.

The detailed individual information on competing identities in our data reflects many of the attributes along which people form group affiliations in everyday life. In extension, this suggests that in-groups biases would be found in other economically relevant rulings where the decision-maker and the subject interact, such as grades in the education sector, hiring decisions, and any type of administrative decisions by bureaucrats in the social sector.

With the third essay, this thesis instead turns toward the "grand convergence" of men and women at the labor market (Goldin, 2014). In **Intergenerational Mobility Trends and the Changing Role of Female Labor**, we study how the increased labor market participation of women in the last half-century, at both the extensive and intensive margin, has affected the transmission of economic disadvantage from parents to their children. We set out to estimate trends in intergenerational income mobility, meaning the correlation in labor income between pairs of parents and children, in Scandinavia for cohorts born between 1951 and 1979. A high correlation means that parental income is highly predictive of child outcomes, and in other words, that intergenerational mobility is low. Intuitively, if parental income is defined as maternal and paternal income combined, then the time trend in this measure will to some extent reflect women's increased labor market participation.

However, it is not *a priori* clear in what direction this would bias the estimates. One the one hand, when female labor supply increases, the relative position of a woman in the female earnings distribution reflects her underlying skills better. All else equal, this puts a downwards pressure on measures of intergenerational mobility, since the incomes of mothers and their children will appear more strongly correlated. One the other hand, the whole income distribution of women also shifts upwards and maternal earnings represent a larger share of joint parental earnings. If female earnings initially has a lower signal value than that of males, this puts an upwards pressure on measures of mobility. Due to constrains on the quality of linked survey data, it has proven difficult for researchers to estimate trends in correlations between males and females both separately and jointly (Chadwick and Solon, 2002, Björklund, Jäntti and Lindquist, 2009, Blanden et al., 2004), and the extent to which the secular trend in female labor supply have affected measures of intergenerational mobility is largely unexplored.

Like much of the recent literature, we estimate trends in intergenerational mobility in terms of *intergenerational rank associations* (Dahl and DeLeire, 2008, Chetty et al., 2014). This means that we estimate OLS regressions of child percentile rankings on parent percentile rankings[2] separately by birth year of the child, and look at how these correlations evolve across birth cohorts. The result is a clear trend toward lower mobility in all three countries, most prominent between the early 60's and late 70's. Over the entire range of birth cohorts, from 1951 to 1979, the total change for Norway is 7.8 rank points (50 %) and 4.6 rank points for Sweden (28 %). In Denmark, the rank association in income increased by 7.3 rank points (39 %) from 1962 to 1979. We also check that these trends persist across a range of different income specifications.

We then split the sample by gender of the children and parents, and find that while correlations between fathers and sons, who arguably are not subject to this great change in labor market attachment over time, remain unchanged, all combinations involving

[2]Parental income is defined as the average of father and mother labor earnings, and these are ranked separately by child birth cohort.

mothers or daughters trend upward over time. In order to see that this is not a uniquely Scandinavian phenomenon, we estimate the same trend in the PSID data set for the United States, with similar results. To the extent that father-son correlations, credibly measure equality of opportunity, it is thus hard to argue that an actual decline in opportunity has taken place over time in either Scandinavia or the US. Thinking of transmission of skills and values as something passive, this suggests that determinants of male income ranks, as well as the labor market valuation of skills that are passed on across generations, are unchanged over time. Instead, a close-at-hand explanation lies in that women's increasing integration into the labor force has changed the measured income correlations across generations.

We corroborate this in two ways. First, we build and calibrate a simple model of transmission of earnings potential between parents and children, which we use to understand to what extent the observed trend can be explained by factors such as the importance of skills in earnings, how strongly skills are transmitted across generations, and assortative mating among parents. The results of this exercise would suggest that while the two latter have not contributed significantly to the trend in intergenerational mobility, the former — the importance of skills for earnings of women — explains a large portion of the change over time. In other words, female wages have increasingly begun to reflect earnings potential over the course of our study period.

We reach this same conclusion empirically in our final exercise, where we combine the information about female earnings potential contained in her earned income, level of education and occupation, into a measure of *economic status*. Our argument, following previous work by e.g. Vosters and Nybom (2017), Vosters (2018) and Adermon, Lindahl and Palme (2021), is that prime-age earnings are good approximations of economic status for men, but less so for women. By pooling the information contained in the three different measures, and weighting them together in an optimal way according to a method developed by Lubotsky and Wittenberg (2006), we can attain a better measure of female economic status. In short, the resulting trend in rank correlations between earnings of sons and "economic status" of mothers is flat over time, or follows closely the development of rank-rank correlations for sons and fathers.

Our results clearly point to the importance of accounting for changes in female economic status when estimating trends in intergenerational mobility. The interpretation that higher rank associations in income or earnings between children and parents reflect a lower degree of social mobility or equality of opportunity is not easily applicable when labor market conditions change.

The fourth and final essay continues on the same broad theme, but considers instead the effect of women's increased labor market integration over time on the wage gap between men and women. In **Wage Inequality, Selection and the Evolution of the Gender Earnings Gap in Sweden**, we estimate the change between 1968 and 2019 in the gender wage gap when accounting for three aspects of structural change at the labor market. These are, in turn, the length of an average work week of women, the overall earnings dispersion, and the composition of the labor force. While the effect of each of these separately for gender pay differences have been studied before (see e.g. Olivetti and Petrongolo (2008), Blau et al. (2021), Mulligan and Rubinstein (2008), Edin and Richardson (2002)), no other study accounts for all three at once. As a consequence, previous conclusions about the development of male-female wage differences over time might reflect these structural changes, making it hard to asses changes in labor market conditions for women.

Since it allows us to study hourly wages and hours of work over a long period of time, we have chosen to use data from the Swedish Level of Living Survey (SLLS), which is a panel survey covering about 0.1 percent of the Swedish population, done in six waves from 1968 to 2010. In order to extend the analysis to more recent developments, we also use data from the Wage Structure Statistics (WSS) between 1995 and 2019. It contains measures of hourly wage rates and information on hours of work. However, it is not a random sample, but contains information on the entire public sector as well as the private sector employees in companies with more than 500 employees.

First, we address the question of weekly hours by comparing the gender gap in monthly wages to that in hourly wages. Measuring the gender gap using observed weekly earnings (excluding zeros) gives a change from 66 percent in 1968 to 24 percent in 2010. Until year 2019, we estimate that the gap has closed further, to a level of 20 percent. When using observed hourly wage instead, the narrowing of the gender gap is not as pronounced – a decrease from 27 percent in 1968 to 14 percent in 2010. In 2019, the estimated wage gap in hourly wages is at ten percent.

Second, we consider how the overall wage dispersion in Sweden has affected the gender wage gap. Correcting for changes in overall wage inequality has a strong impact. During an era of wage compression, between 1968 and 1991, our results show that the entire change in the gender wage gap can be attributed to changes in overall wage inequality. This analysis has previously been done by Edin and Richardson (2002), and we extend the analysis to 2010 using the SLLS data, and to 2019 using the WSS data. For this second era of increasing wage inequality, our results show that the change in the gender gap in observed wages from 19 to 11.5 percent is more than twice as large – from 27.5 to 14 percent- if one corrects for changes in the overall wage inequality.

In the final part, we ask how the gender wage gap would have evolved over time, if the composition of the labor force had stayed constant over time. We begin by showing that participation in market work has increased especially much among married, low-skilled women, indicating that the skill level of the average woman in the labor force has decreased over time. In other words, selection into market work would have become less positive. We study the effect of selection primarily from 1968 to 1991, since the great increase in female labor force participation in Sweden took place during that period. The 3.5 percentage points change in the gender gap in median distributional corrected wages between 1968 and 1991 is instead 16.5 p.p. when correcting for sample selection, under the assumption of positive selection into the labor market within age and education level groups.

While our results show that the inference on the evolution of the gender gap is highly dependent on the choice of wage concept, we remain agnostic about which we consider as the most relevant. In the end it all boils down to what type of gender gap measure one is interested in studying. Since the increase in female labor supply is a universal development in most industrial countries, we believe that these evidences have external validity to the development in other economies and not only shed light on an important historical development in Sweden.

So what can be learned from this thesis? I would argue that the first two essays are informative about criminal justice policy, in Sweden and elsewhere. The first chapter speaks to the importance of considering individual circumstances in the sentencing decision, if one is concerned with individual well-being of children who have broken the law as well as efficiency in state finances. It further suggests that for youth offenders, rehabilitative justice better prevents relapse into crime, compared to pure punishments such as fines. The second chapter shows that equality before the law in criminal trials is contingent upon equal representation of people with different demographics and backgrounds, among judicial decision makers.

From the two subsequent essays, we learn something about measuring the change in an entity over long periods of time, when that same time period has also seen great structural change. Specifically, chapters three and four are about measuring family transmission of economic disadvantage and gender pay differences, respectively, over the course of a time when labor supply of women has undergone historical changes. Both show how these statistics are not readily comparable across time unless the estimates are somehow adjusted to take female labor supply into account, and attempts to as best as possible make these adjustments, in order to arrive at a "true" change over time. In sum, the results suggest that while intergenerational income mobility has remained stable across the past half decade, female labor market conditions as measured by gender pay gaps have changed substantially more than what can be observed from raw statistics.

Bibliography

Adermon, Adrian, Mikael Lindahl, and Mårten Palme. 2021. "Dynastic Human Capital, Inequality, and Intergenerational Mobility." *American Economic Review*, 111(5): 1523–48.

Aizer, Anna, and Joseph J. Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *The Quarterly Journal of Economics*, 130(2): 759–803.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity*." *The Quarterly Journal of Economics*, 115(3): 715–753.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2022. "Unequal Jury Representation and Its Consequences (forthcoming)." *American Economic Review: Insights.*

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review*, 107(4): 1207–38.

Bernhard, Helen, Urs Fischbacher, and Ernst Fehr. 2006. "Parochial Altruism in Humans." *Nature*, 442: 912–5.

Bertrand, Marianne. 2011. "Chapter 17 - New Perspectives on Gender." In . Vol. 4 of *Handbook of Labor Economics*, , ed. David Card and Orley Ashenfelter, 1543–1590. Elsevier.

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

Björklund, Anders, Markus Jäntti, and Matthew J Lindquist. 2009. "Family background and income during the rise of the welfare state: brother correlations in income for Swedish men born 1932–1968." *Journal of Public Economics*, 93(5-6): 671–680.

Blanden, Jo, Alissa Goodman, Paul Gregg, and Stephen Machin. 2004. "Changes in intergenerational mobility in Britain." *Generational Income Mobility in North America and Europe*, , ed. MilesEditor Corak, 122–146. Cambridge University Press.

Blau, Francine D, Lawrence M Kahn, Nikolai Boboshko, and Matthew L Comey. 2021. "The Impact of Selection into the Labor Force on the Gender Wage Gap." National Bureau of Economic Research Working Paper 28855.

Chadwick, Laura, and Gary Solon. 2002. "Intergenerational income mobility among daughters." *American Economic Review*, 92(1): 335–344.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics*, 129(4): 1553–1623.

Dahl, Molly W, and Thomas DeLeire. 2008. "The association between children's earnings and fathers' lifetime earnings: estimates using administrative data." University of Wisconsin-Madison, Institute for Research on Poverty.

Edin, Per-Anders, and Katarina Richardson. 2002. "Swimming with the Tide: Solidary Wage Policy and the Gender Earnings Gap." *The Scandinavian Journal of Economics*, 104(1): 49–67.

Glaeser, Edward L., David I. Laibson, José A. Scheinkman, and Christine L. Soutter. 2000. "Measuring Trust." *The Quarterly Journal of Economics*, 115(3): 811–846.

Goldin, Claudia. 2014. "A grand gender convergence: Its last chapter." *American Economic Review*, 104(4): 1091–1119.

Huffman, David, Stephan Meier, and Lorenz Goette. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." *American Economic Review*, 96: 212–216.

Lubotsky, Darren, and Martin Wittenberg. 2006. "Interpretation of Regressions with Multiple Proxies." *The Review of Economics and Statistics*, 88(3): 549–562.

Mueller-Smith, Michael, and Kevin T. Schnepel. 2020. "Diversion in the Criminal Justice System." *The Review of Economic Studies*, 88(2): 883–936.

Mulligan, Casey B., and Yona Rubinstein. 2008. "Selection, Investment, and Women's Relative Wages over Time." *The Quarterly Journal of Economics*, 123(3): 1061–1110.

Olivetti, Claudia, and Barbara Petrongolo. 2008. "Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps." *Journal of Labor Economics*, 26(4): 621–654.

Rand, David G., Thomas Pfeiffer, Anna Dreber, Rachel W. Sheketoff, Nils C. Wernerfelt, and Yochai Benkler. 2009. "Dynamic remodeling of in-group bias during the 2008 presidential election." *Proceedings of the National Academy of Sciences*, 106(15): 6187–6191.

Rose, Evan K. 2021. "Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders." *The Quarterly Journal of Economics*, 136(2): 1199–1253.

Tajfel, Henri, and John C Turner. 1986. "The Social Identity Theory of Intergroup Behavior." In *Psychology of Intergroup Relation.* , ed. S. Worchel and W.G Austin, 7–24. Hall Publishers, Chicago.

Vosters, Kelly. 2018. "Is the Simple Law of Mobility Really a Law? Testing Clark's Hypothesis." *The Economic Journal*, 128(612): F404–F421.

Vosters, Kelly, and Martin Nybom. 2017. "Intergenerational Persistence in Latent Socioeconomic Status: Evidence from Sweden and the United States." *Journal of Labor Economics*, 35(3): 869–901.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association*, 113(523): 1228–1242.

Chapter 1

Youth Crime, Community Service and Labor Market Outcomes*

*A million thanks to my supervisors Susan Niknami and Mårten Palme, and also to Erik Lindqvist, David Seim, David Strömberg, Mitch Downey, Ines Helm and numerous seminar participants at Stockholm university, Tillburg university, Norwegian School of Economics, the Research Institute of Industrial Economics and Gothenburg university. Thanks also to the people at Framtid Stockholm, for patiently answering all my question and letting me share your work space!

1.1 Introduction

Most criminal offenders begin a pattern of deviance during their teenage years. Despite criminal prosecution and sanctions, the rate of re-offense for youth offenders is high: almost 40 percent relapse into crime within one year.[1] Apart from the large societal costs incurred by crime, evidence shows that criminal activity is often costly for the youth themselves, in terms of their economic outlooks in life (Aizer and Doyle, 2015, Hjalmarsson, 2008, Mueller-Smith and Schnepel, 2020, Rose, 2021). It is therefore surprising that so little is known about what sanctions work best in terms of improving the life opportunities of youth involved in crime.

In this paper, I evaluate how different criminal sanctions for youth offenders affect recidivism, education, and labor market attachment in early adulthood, and whether these outcomes can be improved by better matching of sanctions to offenders. When sentencing youth offenders, judges account for individual circumstances — either by implicit biases (Arnold, Dobbie and Yang, 2018) or through requirement by law (Minstry of Justice, 2015). Either way, empirical evidence on "what works for whom" could aid in their decision-making. However, for the same reason, a simple comparison of observably similar people with different sentencing outcomes will result in erroneous conclusions.

To overcome this identification challenge, I make use of a youth sentencing reform in Sweden from 2007, which sharply changed the probabilities of different sanctions. In particular, *community service* became the most common sanction for offenders aged 15-17, mainly replacing pecuniary fines and individualized treatment ("care").[2] This study thus compares outcomes from different forms of *community-based sanctions.* As the vast majority of youth offenders across the Western world are *not* given a prison sentence, this is a

[1] The one-year recidivism rate is from Shem-Tov, Raphael and Skog (2021) for the US; for the UK see Ministry of Justice (2020); for a set of European countries (Germany, the Netherlands, France, and Iceland) see Albrecht et al. (2014); and for Sweden see SBU (2020).

[2] This judicial outcome, usually labeled *rehabilitation* or *social care*, here refers to a broad specter of individualized treatment programs, including custodial care. See Section 1.2 for more details.

highly relevant margin of comparison.[3] Community service represents a “middle ground” between a punishment and a rehabilitation program, as its core lies in restoring individuals from delinquent behavior, while also containing a punitive element: unpaid work. The design of the program makes community service in this setting similar to an internship, as it provides participants with real work experience in private and public establishments. By undertaking community service, individuals can gain both practical and noncognitive skills that facilitate a change of lifestyle. Economic theory, as reflected in e.g. Becker (1968), would suggest that strengthening the labor market attachment of at-risk youth will also reduce recidivism, by increasing the opportunity cost of crime, implying gains for the individual as well as for society.

An advantage of the Swedish setting is rich individual-level data, linking a nationwide crime register to longitudinal data on individuals and their families. The comprehensive data not only makes it possible to study effects on recidivism, but also on schooling and labor market outcomes. I study to what extent the 2007 reform and its entailing increased focus on community service affected these outcomes. My empirical approach uses the difference in the probability of a community service sentence between 2006 and 2007 convictions, paired with information about the exact conviction date for each court case. This natural variation in sentencing outcomes is then used in a difference-in-discontinuity (“RD-DD”) design (Eggers et al., 2018, Persson and Rossin-Slater, 2021), where convictions from years prior to the reform are used as an additional control group to net out seasonality effects.

The results reveal that the sanction reform did not affect average behavioral responses in terms of recidivism, educational attainment, or employment. Effect sizes on all outcomes are close to zero, and 95 percent confidence intervals allow ruling out effects larger than five to ten percent compared to the variable means. However, splitting the sample according to individuals’ most likely counterfactual

[3]The annual share of youth cases with a prison sentence outcome varies from one percent in the Nordic countries (Lappi-Seppälä, 2011), to seven percent in the UK (Ministry of Justice, 2020), and eight percent in the US (Hockenberry and Puzzanchera, 2020). Among the rest, community-based sanctions make up about 70-90 percent.

sentencing outcome (fines or social service referral, “care”) uncovers important heterogeneous responses.[4] Replacing fines (a pure penalizing measure) with community service attains long-lasting lower rates of recidivism and early-adulthood incarceration, as well as a higher rate of high school completion. Community service instead of care, on the other hand, results in longer prison sentences in adulthood. I thus find individual benefits from implementing community service as an alternative to fines, but costs when it replaces pure rehabilitating measures.

Next, I ask whether behavioral outcomes for young offenders can be improved, by better matching of sanctions to offenders. Understanding heterogeneous responses along individual attributes can potentially inform courts of how to target community service among juveniles, in light of some societal goal function (see Kleinberg et al. (2017) for a deeper discussion on human decision-making aided by algorithms). For example, we could consider a criminal justice policy successful if it reduces the number of cases of severe recidivism, and preferably also makes ex-offenders more likely to find employment. This would amount to increasing the financial efficiency of the youth criminal justice system. As a first step, I calculate the *net costs* incurred by each individual in my sample through the sanction they are given and all subsequent criminal convictions, as well as their post-conviction earnings history.

I then apply a novel machine learning method, the *causal forest* (Athey and Imbens, 2016, Wager and Athey, 2018, Davis and Heller, 2020), which searches for combinations of covariates that predict differences in average treatment effects, among a large set of possible individual and neighborhood characteristics. In effect, this method estimates individualized treatment effects, through an iterative process similar to the standard random forest algorithm. The outcome variable used in this analysis is the individual-level net costs. This analysis unveils subgroups within the “fines” and “care” groups, for whom the community service introduction lead to net savings, meaning fewer crimes and/or improved labor market

[4] I predict the most likely counterfactual sentence among fines and care for individuals sentenced to community service, and estimate the difference-in-discontinuity model separately for the resulting two groups.

attachment. Among youth whose most likely alternative sanction is care, younger age and a less disadvantaged background predicts net savings. Among the "fines group", on the other hand, net savings are found among older juveniles with some previous labor market experience.

In sum, my findings characterize treatment effect heterogeneity along two dimensions. First, the severity of the crime and the most likely alternative criminal sanction matters. My results suggest that community service is primarily a good alternative to monetary fines. The second dimension regards personal characteristics. Here, the two most important predictors of heterogeneity are age and family socioeconomic status, and their interactions with crime severity. In fact, individual attributes predict treatment effects in opposite directions, depending on the alternative sanction. This indicates that *optimal targeting* of this policy is a two-step process: consider first the crime, and then individual circumstances.

My paper adds to our understanding of the effects of criminal justice policy on offender outcomes. I show that the type of sanction a person gets matters for their subsequent behavior, beyond the prison-or-not margin. Previous studies have established that incarceration of youth offenders has negative effects on employment and schooling (Aizer and Doyle, 2015, Hjalmarsson, 2008). Others have put forward evidence that formal prosecution rather than diversion practices leads to increased re-offending (Mueller-Smith and Schnepel, 2020, Shem-Tov, Raphael and Skog, 2021) and that young individuals on the margin of incarceration or community supervision fare better from the latter (Rose, 2021).[5] This paper is the first to show that different forms of community-based sanctions have different effects on offenders and in particular, I'm the first to study long-term effects of community service.[6]

[5] In Rose (2021), interest lies in the risk of having one's probation sentence revoked because of technical violations and being sent to prison. Mueller-Smith and Schnepel (2020) study sudden changes in the probability of diversion vs. felony conviction.

[6] A couple of smaller case studies, summarized in Dunkel et al. (2014), evaluate community service as an alternative to prison sentences, finding mixed evidence. Additionally, two psychology dissertations evaluate youth community service in the Swedish setting, finding that recidivism is widespread after completion of such a program, especially among juveniles with pronounced mental health

Second, I show that, with the right content, court-mandated community service can have effects similar to those of active labor market program for at-risk youth (at least more so than other community-based criminal sanctions). In the Swedish setting, community service consists of formal — albeit unpaid — work experience in private, public, or non-profit establishments. Evaluations of employment training programs find small or no effects, and programs that benefit participants are generally expensive and of long duration (Card, Kluve and Weber, 2018, Crépon and van den Berg, 2016, Aizer et al., 2020).[7] However, a growing literature (Heller, 2014, Heller et al., 2017, Modestino, 2019, Gelber, Isen and Kessler, 2016, Kessler et al., 2021) has shown that public summer youth employment programs (SYEPs), which are similar to community service in duration, cost, and target age, can have substantial positive effects on youth delinquency.[8] I show that short-term and essentially cost-free employment training for at-risk youth, who have not taken active measures to apply for such, can have crime-reducing effects.

The final contribution of my paper is to show the potential in accounting for individual circumstances in court decisions. While a controversial topic given the justice principle of equal treatment for equal crimes, this is a regular feature of youth justice. As shown by the literature using between-judge variation in strictness for empirical identification (see e.g. Arnold, Dobbie and Yang (2018), Bhuller et al. (2020), Dobbie et al. (2018), just to name a few), the current system leaves court sanctions up to the discretion of judges. My results show that these decisions can be improved by allowing for algorithmic guidance. This result generalizes to program evaluations broadly: given the growing awareness that employment and training programs can have important effects for *some* participants (Bitler, Gelbach and Hoynes, 2006), it is important from a policy perspective to understand what characterizes such individuals. The causal forest

issues and anti-social behavior (Ginner-Hau, 2010). In contrast, studying female adolescents with limited delinquency, Azad (2019) finds that recidivism is reduced after completing youth community service, but that educational achievements are not improved.

[7] One exception is the evidence in Fallesen et al. (2018), where *mandatory* program participation is found to reduce crime among Danish young men.

[8] See Davis and Heller (2020) for an excellent summary of the literature.

method facilitates this aim by searching for combinations of attributes predicting treatment effects while avoiding inference issues from multiple hypothesis testing.

This paper is structured in the following way. Section 1.2 provides an overview of the Swedish juvenile justice system and explains the details of the 2007 sanctions reform and youth community service. It ends with a conceptual framework for how community service could affect labor market outcomes differently, depending on individual characteristics. In Section 1.3, the empirical method is explained, followed by a description of the data sources. Results are presented in Section 1.4 and policy implications are discussed in Section 1.5. Section 1.6 concludes.

1.2 Juvenile Justice in Sweden

Across the Western world, juvenile criminal justice systems emphasize the restorative aspect of justice. This is motivated by the fact that a person's cognitive development is not yet complete in adolescence, meaning that minors should not be held fully responsible for their actions (Cohen and Casey, 2014). From another perspective, society is obliged to help juvenile offenders realign their lives towards socially productive occupations. For this reason, community-based sanctions such as probation, child welfare referral, and fines are the most frequent outcomes of court trials involving youth offenders. Diversion practices — whereby offenders are not formally prosecuted — are also common: in US juvenile courts, about 30 percent of cases are dismissed (Hockenberry and Puzzanchera, 2020). In international comparison, the incarceration rate in the Swedish juvenile justice system is low, at about one percent, which can be compared to eight percent in the US (Hockenberry and Puzzanchera, 2020, Lappi-Seppälä, 2011). The crime rate is however quite similar. During the period under study here, years 2006-07, the Swedish rate of juvenile convictions per 1,000 inhabitants aged 15-17 was around 23 (author's calculation), compared to US juvenile court statistics, where cases involving an offender aged 10-17 decreased from 50 per 1,000 in 2005 to 20 per 1,000 in 2008 (Hockenberry and Puzzanchera, 2020).

The age of criminal responsibility in Sweden is 15 and offenders are tried as adults from age 18. The majority of criminal cases are settled in one of the 48 district (first-level) courts (55 % during 2000-2016; author's calculation), where they are tried by one judge and three lay jurors ("nämndemän"). Defendants can either hire a lawyer privately or request a public defender. All defendants who are charged with a serious crime have the right to a public defender, and it is possible to request a specific lawyer for the job. Responsibility for sentence implementation for youth offenders lies with the local social services agency (as opposed to the Prison and Probation Service for adults), who are also tasked with formulating a suggested sanction to the court. There are no separate courts for minors — they are sentenced within the same system as adults, but are subject to different sentencing options.[9] Youth-specific sanctions include child welfare referral ("care"), youth community service, and in rare cases, juvenile incarceration. However, they can also be given fines and probation on similar premises as adults. The "care" sanction is an umbrella term for a great variety of treatment programs, the choice of which is referred from the courts to the local social services. The three most commonly applied are cognitive-behavioral programs to prevent further criminality (25 %), counseling sessions with a social worker (≈ 20 %) and custodial care (≈ 15 %) (Brå, 2011).

1.2.1 The Youth Justice Reform of 2007

In 2005, the Swedish government proposed several changes to the youth justice system, as a continuation of a series of reforms concerning young offenders beginning in the late 1990s. The reform comprised of changes to the penal code (*Brottsbalken* and *Law 1964:167*), as well as changes to other legal writings concerning the treatment of young offenders by justice system actors. This reform had two main motivations. First, to reduce the number of young offenders given monetary fines, since it has no restorative justice element and puts financial stress on young people, who might then enter adulthood indebted.[10] Second, to restrict the use of pure rehabilitating

[9] Moreover, young adult offenders (ages 18 to 21) are normally given a "sentencing rebate" that reduces sentences from their full length/impact.

[10] The average amount of fines given to minors in 2006-2007 was equivalent to about $190 (author's calculation). At the time, parents were not co-liable for

sanctions (child welfare referral, "care"), so that it is given only to those with special care needs, i.e. individuals with mental health problems, addiction problems, very unstable home situations, etc (Prop. 2005/06:165, p. 47). The proposition was voted through in Parliament on May 30, 2006.

The most prominent change from the 2007 reform was the introduction of community service for youth offenders. Youth community service had previously existed as a combination sentence, where the main outcome was child welfare referral. However, the government report forming the basis of the reform finds that community service was not systematically implemented within this combined sanction and that it should therefore be replaced by a stand-alone sanction (Prop. 2005/06:165, p. 48).

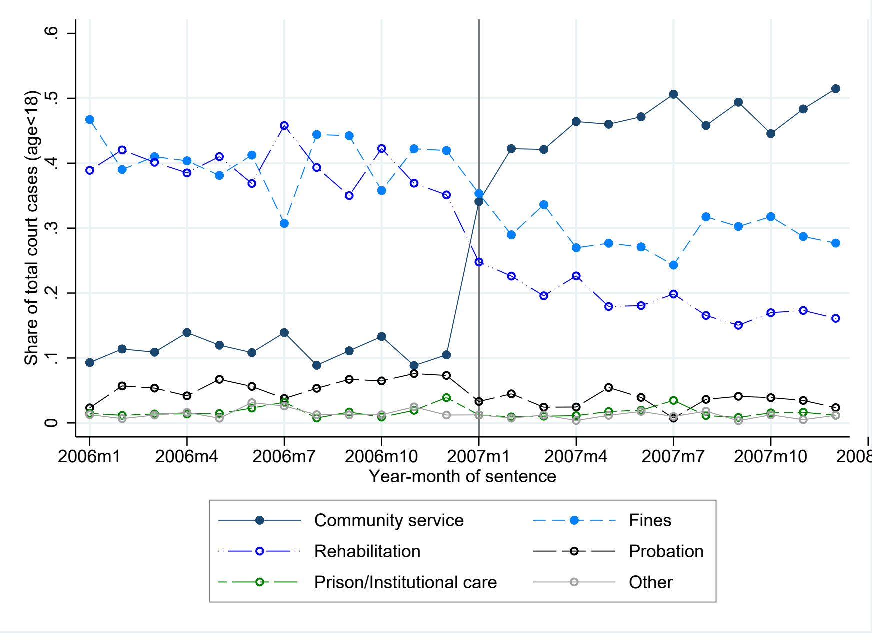
Figure 1.1 shows the change in the sanctions mix resulting from the reform. In 2006, about ten percent of minors were sentenced to the combined care-community service sentence. Immediately following implementation on January 1st, 2007, around 50 percent of juvenile offenders were given youth community service. The reform achieved its goals by decreasing the use of fines, from about 40 to 30 percent, and reducing the share of rehabilitation sentences from 40 to 20 percent. There were small reductions in the use of probation sentences and prison/institutional care, from already very low levels (three and one percent, respectively).

1.2.2 Youth Community Service

In the Swedish setting, youth community service as a criminal sanction is akin to a youth employment or training program: juvenile offenders given this sentence are required to perform unpaid work in public or private sector workplaces. They receive extra support at the workplace in form of a mentor, who can attest to their potential acquired skills after completed service. The second mandatory part of the sentence is a behavioral treatment program. This consists of a short series of talks with a social worker, usually following some written manuscript, and is mainly meant to make the youth reflect upon their actions and decisions leading to a criminal sentence, and

payment of the fines.

Figure 1.1: Criminal Sanctions for Juvenile Offenders in 2006 and 2007, by Month of Conviction.



Note: Juvenile court convictions by month and sanction type, as shares of all convictions among offenders aged 15-17. Unit of observation is individual-case. N=12,059.

their future goals.[11] At the start of the sentence implementation, the offender is also required to show up for treatment planning at the social services office, and a social worker monitors adherence to the plan. Individuals face immediate reactions if they fail to show up for work, and their cases can be sent back to prosecutors for re-evaluation.

[11]This set-up greatly resembles the setting for US summer youth employment programs, except for no wage payment. In these programs, youth are usually given a part-time minimum wage job for about six to eight weeks, with the additional support of a mentor. In some programs, participants are also given some form of behavioral/motivational treatment or additional job skills training (Heller et al., 2017).

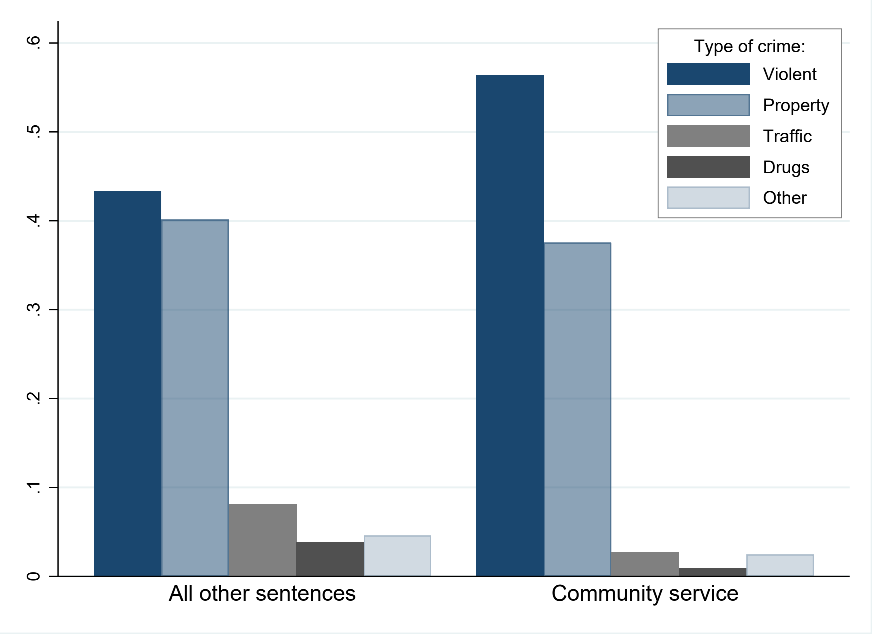
The law does not specify what crimes or what severity of crimes might merit community service. It states that community service should be used instead of a fine if it is “not too severe a sanction for the crime in question” and that it should be used instead of social services care whenever the youth is not considered in need of rehabilitation or out-of-home placement. It is thus an eligible option in essentially all court cases involving a minor. The extent of the punishment can range from 20 to 150 hours, where the lower limit applies to misdemeanors such as shoplifting and vandalism, and the upper limit can be applied to severe violent crimes. The work is to be undertaken weeknights and weekends to interfere minimally with school work, and the work hours are thus generally spread out over several weeks. Each social services office is responsible for contracting firms at which the unpaid work can be conducted and youths can request a certain type of workplace from the available options.

Figure 1.2 shows the relative frequency of juvenile crime types, separately for those receiving a youth community service sentence and those given any other penalty. The vast majority of crimes, around 80 percent, can be classified as either property crimes, including theft, property damage, burglary, and shoplifting, or violent crimes, including assault, robbery, coercion, manslaughter, and weapons and sex crimes. The violent crimes category consists mainly of assault cases (about 55-60 %). The remaining one-fifth are traffic violations (not including motor vehicle theft), drug crimes (use and distribution), and all other crimes (consisting mainly of different types of forgery, economic crimes, and alcohol-related misdemeanors). Cases given community service consist of a higher share of violent crimes, and a lower share of narcotics-related crimes and traffic violations.

1.2.3 A Conceptual Framework for Heterogeneous Effects of Community Service

This paper concerns the effect of a reform that increases the use of community service, at the expense of mainly fines and individualized treatment programs. Here follows a brief discussion of why community service would affect later life behavioral outcomes, in relation to the other two sanctions and depending on individual characteristics.

Figure 1.2: Youth Crime Categories.



Note: Type of crime in five categories: violent (incl. robbery, arms, assault and sex crimes), property (theft, property damage, shoplifting and fraud), traffic offences, narcotics (use and dealing) and other crimes (e.g. economic crimes, perjury). Right panel shows crime composition of cases receiving community service sentences (N=3,522), left panel shows crime composition of all other cases (N=8,537).

Criminal activity can be modeled as a choice arising when the opportunity cost of crime is perceived by the individual as sufficiently low (Becker, 1968). The Labor Economics insight, well established in the literature (see Hjalmarsson and Lindquist (2018) for an overview), is that education and employment training can help increase this opportunity cost, while simultaneously incapacitating further crime. Community service, as a form of employment training program, is unlikely to have an incapacitating effect, as the number of hours of unpaid work is typically low. On the other hand, youth sentenced to community service are monitored quite closely and risk harsher re-sentencing if they violate the adjudicated terms, which usually stipulate staying out of delinquent activities. The chances of continued employment at the assigned workplace are limited by the fact that each social services office re-uses the same firms for the placement of youth. Instead, the most likely gains from community service are either behavioral change or attested productivity from an employer.

Following standard theoretical framework, the cost of crime, C, can be thought of as depending on current levels of human capital H, the ability to acquire more human capital A, and the severity of punishment P.[12] ϵ_{it} represents a "noise" term, reflecting uncertainty over the true cost of crime. For example, adolescents might not be fully able to gauge the cost of crime, either because of an incomplete cognitive development process (Cohen and Casey, 2014) or because they have not yet learned about the economic returns to their skills. Entering these terms linearly, for the sake of simplicity, gives:

$$C_{it} = H_{it} + A_{it} + P_{it} + \epsilon_{it},$$

where i and t indexes individuals and time, respectively. Given program completion and meaningful content, community service can affect average C among the population of youth offenders by increasing H_{it} (through work experience) and A_{it} (by learning pro-social behavior such as showing up on time, following orders and controlling their impulses).[13] The short average duration of treatment (equivalent to one workweek) makes it more likely that individuals gain general noncognitive skills, such as showing up on time, taking orders, and being motivated to work, rather than occupation-specific practical skills. The effect of community service on P_{it} will likely vary with individual preferences. The noise component ϵ_{it} could also affect the cost of crime in either direction as the individual learns about her chances at the labor market.[14]

[12] One could, of course, imagine a wide range of other relevant factors, such as those related to peer effects and policing strategies. However, since these are less central to the crime-employment trade-off, I abstract from them here.

[13] The central role of noncognitive skills for criminal behavior is summarized in e.g. Cunha and Heckman (2008) and Almlund et al. (2011).

[14] One can imagine both a positive effect, such as if a person discovers she is good at a certain task, or a negative effect if, for example, a person is given a work task for which she has a strong distaste.

Compared to paying a fine, community service could be regarded as a harsher punishment for a young person with resources to pay the fine — either themselves if they are employed or with their parents' help — since it then requires participation. On the other hand, it can be regarded as a milder punishment for someone unable to provide the money for a fine. A third perspective is that a fine could also encourage someone to find a job in order to be able to pay it, whereby the potential employment-enhancing effect of community service is offset.

The many different forms of treatment available within the "care" sanction makes it hard to determine whether community service is a harsher or milder form of sanction. Compared to the dominant forms of care (therapy or mentorship programs), community service demands more active participation, but, on the other hand, the behavioral therapy component is less extensive. Any positive effect through A_{it} (non-cognitive skills) might thus be counteracted by unresolved behavioral issues.[15]

More generally, job training ought to have a larger effect on individuals who are initially the furthest removed from the labor market in terms of opportunities. Low perceived future pay-offs at the formal labor market lead youth from adverse socioeconomic backgrounds to under-invest in human capital relative to their abilities and become "under-educated" (Kearney and Levine, 2016). Previous research has also shown clearly that job opportunities are worse for people with a criminal record (Doleac and Hansen, 2020). Youth who complete the community service program are given a recommendation letter from their mentor and can use this person as a reference in future job search (see Heller and Kessler (2021) for evidence on the importance of reference letters in job search for marginalized youth). One last prediction is thus that people with an otherwise strong labor market attachment are helped by this policy, as it overcomes the stigma of having a criminal record.

[15] That cognitive behavioral therapy can effectively reduce re-offense among at-risk youth is found in Davis and Heller (2020) and Modestino (2019) in the setting on summer jobs paired with therapy programs, and also by Blattman, Jamison and Sheridan (2017).

1.3 Method and Data

1.3.1 Empirical Strategy

Youth community service is a sentencing form designated primarily for young offenders deemed capable of completing it, meaning that the justice system selects for this treatment those with relatively high ability. For example, youth with active substance abuse or with very unstable home environments are excluded. A simple comparison of youth given community service vs. other sanctions will thus reflect differences between the community service group and all other offenders in terms of unobserved characteristics.

To account for this potential omitted variable bias, I use the discontinuity in court outcomes arising from the juvenile justice reform on January 1st, 2007, which sharply increased the probability of being sentenced to community service, rather than any other sanction. The total time from arrest to hearing depends on the efficiency with which justice system actors handle cases, whereby hearing dates can differ between similar cases with the same arrest date. Given this arbitrary allocation of court dates, comparing youth sentenced before and after the new year 2006-2007 introduces random variation in sentences.[16]

Identifying the causal effect of the reform requires that youth convicted in the end of 2006 are comparable to youth convicted in the beginning of 2007. A test for balance across the threshold date on pre-determined covariates is shown in Figure 1.A1. Panel a) plots regression discontinuity coefficients for 2006 and 2007 convictions with the covariate as the dependent variable.[17] This shows that youth sentenced in early 2007 rather than in late 2006 are younger, more likely to have committed a violent crime, and their cases take longer for the justice system to process. One possible explanation for these differences is a seasonality effect in crimes.

I thus apply a difference-in-discontinuity design ("RD-DD"),

[16] While this strategy uses a discontinuity in time for identification of a causal effect, it relies on sample size in the number of court cases for estimation, rather than sample size in time series observations (Hausman and Rapson, 2018).

[17] Regression discontinuity model: $x_{itc} = \alpha_0 + \alpha_1 \text{Reform}_{it} + \alpha_2 f(\text{Day}_{it}) + \alpha_3 f(\text{Reform}_{it} \times \text{Day}_{it}) + \mu_c + \nu_{itc}$

where earlier cohorts of offenders are used as a control group for 2006/07 convictions (the regression discontinuity sample, "Reform").[18] Any potential discontinuity in outcome variables found in the reform sample is corrected for by the average discontinuity around the new year in pre-reform years. Three time periods are pooled to construct the control group: convictions from 2003/04, 2004/05 and 2005/06 ("Pre-reform"). The following model is estimated:

$$\begin{aligned}Y_{itpc} = \gamma_0 + \gamma_1 Treat_{itp} + \gamma_2 (Treat \times Reform)_{itp} + \gamma_3 f(Day_{itp}) + \\ \gamma_4 f(Day \times Treat)_{itp} + \mathbf{X}_{itpc}\beta + \mu_p + \mu_c + e_{itpc},\end{aligned} \tag{1.1}$$

where Y_{itc} represents recidivism, educational attainments or labor market outcomes for individual-case i convicted on date t in period p and court c. $Treat_{itpc}$ is an indicator variable taking the value one if individual-case i is convicted during months January to June of period p, zero if convicted in July to December. $Reform_{itp}$ is an indicator variable for being sentenced in the period 2006/07. The main coefficient of interest, γ_2, thus isolates the effect of an increased probability of community service, controlling for seasonal differences between fall and spring convictions.[19]

Polynomials of the running variable, conviction date re-centered around the new year (Day), are modeled separately before and after d. To account for between-court differences in the usage and execution of community service, the model is estimated with court fixed effects, μ_c. Period fixed effects, μ_p, are included, which also capture the main effect of the reform period. X_{itc} includes crime fixed effects in all specifications, and is also expanded with a set of individual and neighborhood characteristics.[20] Standard errors are clustered at court level, and e_{itpc} represents the error term.

[18]See Persson and Rossin-Slater (2021) for an earlier application of the method in a similar setting.

[19]Since the data does not contain any information on whether individuals actually started and completed their service, the estimate should be interpreted as an intention-to-treat (ITT) effect.

[20]Individuals: age at crime and conviction, gender, normalized family income, a family social welfare indicator, a single-parent household indicator, a dummy for missing family information, a dummy variable for living without parents or caretakers, indicators for any previous criminal record, for whether the individual

Under assumptions of excludability and monotonicity, the variation thus isolated could be used as an instrumental variable for assignment to community service. The most important change to the system was the new sanction youth community service. However, the 2007 juvenile justice reform contained other — albeit more minor — changes, such as new recommendations for social services and about pre-trial treatment of youth. The assumptions can thus be questioned, as it is not obvious that individuals who don't receive the community service sanction in 2007 are completely unaffected. This speaks in favor of studying the reduced-form effect rather than the full instrumental variable model. The reform was meant to change the way the social services treat juvenile offenders along several dimensions, with the unified aim of a stronger emphasis on rehabilitative criminal sanctions. As such, reduced form estimates from this specific setting contain useful information for policy discussions elsewhere.

The RD-DD strategy compares differences in outcomes for youth convicted before vs. after the new year 2006/07, to differences in outcomes for youth convicted before vs. after the three previous new years. Figure 1.A1 Panel b) runs the difference-in-discontinuity specification on the same pre-determined characteristics as above, finding no differences in discontinuities between the reform and pre-reform samples. Importantly, the variation in crime types and age between the control and treatment groups in the RD model no longer persist in this specification.

Two more identifying assumptions are required. First, as in any RD design, we require continuity in the running variable, which in this case means that individuals must be randomly assigned to court dates, with no endogenous sorting of cases on either side of the reform date. That is, defendants cannot choose their court date, in order to accommodate preferences for or against the new sentence regime, that might be correlated with the outcomes of interest. Figure 1.A2 presents case density by sentencing date, and a formal McCrary test of bunching at the threshold, for the reform sample (years 2006-2007) in Panel a), and the whole RD-DD analysis sample in Panel b).

has any previous work experience and for school enrollment, and parental education. Neighborhood: unemployment/nonemployment rate, share foreign-born, share social welfare recipients, crime rate, share with high school education and an indicator for being located in one of Sweden's three largest cities.

Case density varies smoothly around the threshold, suggesting that individuals are equally likely to be assigned a court date just before and just after January 1st in the reform year, as in pre-reform years.

Second, the *difference* element of identification - reform vs. non-reform years - requires the existence of *local parallel trends* (Eggers et al., 2018). Under this assumption, outcomes for individuals convicted during the spring of 2007 would without the reform have evolved similarly to those of individuals convicted in the spring of pre-reform years. Suggestive evidence of parallel pre-trends, based on fall convictions in all sample years, will be presented in Section 1.4. Taken together, these tests suggest that the difference-in-discontinuity strategy identifies the average treatment effect of being subject to the reformed sentencing scale.

In order to separate the control from the treatment group in each time period — i.e. those convicted before the new year and those convicted after the new year — the maximum bandwidth is six months on either side of the new year. An analysis of bandwidth sensitivity shows that the choice of bandwidth has only minor effects on the point estimates. In Figure 1.A3, the RD-DD model is estimated at bandwidths spanning from one month to six months in approximately two-week intervals, and with linear and quadratic polynomials of the running variable. The estimates remain stable across the different specifications, and are similar when using a quadratic instead of a linear polynomial of the running variable. When including only one month on either side of the cut-off, the point estimates deviate slightly for several outcome variables, something that I discuss further below. In my preferred specification, I use a bandwidth of six months and a linear polynomial of the running variable.

1.3.2 What is Community Service Replacing?

As previously discussed, in absence of the 2007 reform, the most likely sentence for juvenile offenders would have been a fine, followed by individualized care. The "treatment" induced by the reform can thus be thought of as a mix of two counterfactual scenarios: fines and care. Responses to treatment likely differ depending on the most probable alternative sanction.[21]

I perform a "counterfactual treatment analysis" as follows. Individuals are split into two groups: those for whom the most likely alternative sanction, had the reform not taken place, is monetary fines ("high probability fines") and those for whom the most likely counterfactual outcome is an individualized treatment program ("high probability care"). This is done by predicting the likelihood of fines and care, respectively, on youth convictions from years 2003-06, using observable characteristics about the crime they are charged with.[22] The predicted probability is then extrapolated onto post-reform convictions using the same observables. Observations with a predicted probability of one of the sanctions that exceeds the 75th percentile are assigned to the respective "high probability" group (and groups are mutually exclusive). Appendix Figure 1.B1 presents evidence that the prediction model performs well out-of-sample.

Summary statistics for the two groups are shown in Appendix Table 1.B1, along with the difference in means between them. The first section shows the share of each sample who receive the different sanctions. Among youth in the group with a high predicted probability of fines, 84 % actually receive fines, compared to 19 % in the care group. In the "care group", 60 % actually receive the care sanction, while the corresponding number in the "fines group" is 12 %. The probability of receiving community service differs between the two groups (19 vs. 49 percent post-reform, 0 vs. 10 percent pre-reform).[23] However, as displayed at the bottom of the table, the

[21] A second interpretation of this analysis lies within the classic penalizing vs. restorative justice theory, in that the reform reduces or increases the severity of punishment. For the group with a high predicted probability of fines, the treatment is interpretable as replacing a pure punishment with a restorative judicial measure. The reform could then be seen as lenient for these individuals. On the other hand, for people who would have received care in absence of the reform, the introduction of youth community service is arguably a reduction in leniency.

[22] I estimate: $pr(Fines_{it}) = \mathbf{X_{it}}\beta + e_{it}$, with probit regression, where $\mathbf{X_{it}}$ = criminal background, time between arrest and conviction, number of charges, main crime (31 categories) and court. An analogous model is estimated for the rehabilitation sanction. Results are similar when using alternative prediction methods including a linear probability model and Lasso (probit).

[23] The reader is reminded that community service in the Swedish setting is applied mainly to felony crimes.

change in probability of community service between 2006 and 2007 (the first stage) is substantial and statistically significant for both groups: 0.20 (se 0.02) for the fines group and 0.37 (se 0.04) for the care group.

The most prominent difference between the two groups is the share of violent crimes: 18 versus 43 percent. The single most common crime among the "fines group" is property damage, while it is assault in the "care group". These facts together strongly indicate that behavioral issues are more prominent among youth with a higher probability of receiving the care sanction. Youth in the group with a high predicted probability of fines are more likely to still be in school when committing the crime, to have some previous work experience and to be convicted for the first time. Their family backgrounds are however similar, indicating that any differences in the effects of community service between these two groups are not due to differences in socioeconomic background.

One might wonder whether this analysis by alternative sanctions picks up heterogeneous *treatment* or heterogeneous *responses to treatment*. Proportionality in criminal punishments requires that the number of hours of unpaid work increases in the severity of the crime. Thus, individuals in the fines group on average get fewer hours of community service, than youth in the care group. I estimate average sentence length in my sample to be 26 (fines) and 35 (care) hours respectively.[24] In Appendix Table 1.E1, I show that the effect of community service does not differ in an economically meaningful way within the fines and the care group, depending on the estimated length of the unpaid work. This indicates that heterogeneity across these two groups stems from differences in treatment effects.

1.3.3 Causal Forest Estimation

Community service for youth offenders was implemented as a "one size fits all" policy. However, as discussed in the conceptual framework above, the effects of community service are likely to vary depending

[24]Since I do not observe adjudicated hours of unpaid work in my main data set, these figures are estimated from the mean number of hours that individuals convicted of the same crime get in my smaller data set from Stockholm years 2010-2016.

on circumstances such as family support, noncognitive skills, and perceived labor market outlooks. Knowing which type of individuals are most likely to benefit from the program can help inform policymakers, and it might also shed light on underlying mechanisms for program success or failure.

To explore whether effects vary across offender subgroups, I use the novel machine learning technique *causal forest* developed by Athey and Imbens (2016) and Wager and Athey (2018). The advantage of this method over a more traditional split-sample heterogeneity analysis is that it allows the researcher to identify non-linear combinations of attributes that drive treatment effects and avoid inference issues from multiple hypothesis testing. The method uses the same logic as a decision tree algorithm, whereby a subset of the full sample - a training sample - is split sequentially into small subgroups ("leaves") sharing the same observable characteristics. The splits are chosen to maximize the variance of the treatment effect in the training sample while maintaining balance between control and treatment units. Once the tree has reached its final nodes, the *conditional average treatment effect* (CATE) is estimated at each final node.[25] Individual CATEs, defined as $\tau(x) = E(Y_1 - Y_0|X = x)$ in the potential outcome framework (Athey and Imbens, 2016), are calculated as the average effect among the full set of trees (the "forest"). For a thorough description of the algorithm, see Athey, Tibshirani and Wager (2019) and Davis and Heller (2017). In essence, the method identifies individual traits which predict positive (and negative) treatment effects and is thereby useful for informing policy discussions about which "types" of offenders to target community service at, given some societal goal function.

I estimate CATEs for a summary measure of all outcome variables ("Net costs"; see Section 1.5.1), using a large set of individual, neighborhood, and case characteristics. See Appendix 1.C for a complete description of my implementation. To harness the causal interpretation of the RD-DD model, the outcome variable is a residual after controlling for court fixed effects, the January-June conviction main

[25]The estimated conditional average treatment effects are best evaluated with *doubly robust scores* when the method is applied to observational data; see Wager and Athey (2018). I estimate doubly robust scores and the appropriate standard errors; results using these are available upon request.

effect ("treatment"), and period fixed effects (capturing the reform year main effect). The regression model is then run on the reform period sample, i.e. convictions from 2006-07. Appendix Table 1.C1 shows estimates using this residualized method on the reform period sample. Results are qualitatively the same as those from the RD-DD model, but the point estimates differ slightly.

1.3.4 Data

My main data source is the Swedish criminal convictions register administered by the National Council for Crime Prevention (Brå), which includes all criminal cases leading to convictions in first-instance courts. The convictions register includes the type of crime, the start and end date of crimes, date of conviction, type of sanction, and court location. In each criminal case, the type of crime is defined according to the most severe ("main") crime, since this is what determines the sentencing range. Youth sentences typically consist of a single penalty, with limited occurrences of combined penalties such as probation and fines. The sanction is defined as the first listed punishment. Court cases are connected on an individual level to population registers, holding annual information on e.g. highest attained level of education, employment, income from work at individual and family level, family situation, and area of residence.

The main sample consists of individuals aged 15-17 when committing a crime, with court convictions from July-December 2003-2006 and January-June 2004-2007. This gives 22,374 observations on individual-case level, of which roughly half are brought to court in January-June. Overall, criminal cases are evenly distributed between ages 15, 16, and 17. Table 1.1 shows summary statistics for this sample, comparing the full under-18 population in 2003-07 in Column 1 to all juvenile offenders sentenced in courts in Column 2, and to individuals sentenced to youth community service in Column 3. The table includes background characteristics at the individual- and family level, and criminal records.

Several things are worth noting from this table. Youth offenders come from families with substantially lower income than the general population, and with a markedly higher likelihood of relying on social benefits for support: 9 % in the general population vs. around 27

% among youth offenders. Their parents are more likely to be born outside of Sweden (24 vs. 40 %), and more likely to be separated — almost half of the individuals sentenced in courts live in a single-parent household. In the year of committing the crime, 86 % of the court convicted youth are registered as students on either lower- or upper secondary level, compared to 97 % in the general population. The vast majority are boys.

In these years, about 6 % of the Swedish underage population had ever been convicted of a crime. This number obviously represents a lower bound to criminal or delinquent behavior, and school surveys indicate that about 10 % of the population at age 15 self-report as ever having been arrested by the police for a crime they have committed (Brå, 2016). The average number of days between arrest and conviction is around 155-160, with substantial variation (se=99.75 for all court cases). A final relevant comparison is between juveniles sentenced to community service (15 % of the convicted youth population) and all individuals with court convictions. Note that individuals given community service generally come from slightly less disadvantaged backgrounds in terms of family social welfare dependence and the likelihood of still being enrolled in school.

In order to better understand the community service sanction, I have also gathered information from social services files for a smaller subset of youth convicted in 2012-2016.[26] These data contain more details about the sanction: mandated hours of unpaid work, workplace and whether or not the program was completed, all of which is summarized in Table 1.2. This shows that the mean number of hours is 40, that about 84 percent of youth given community service complete it, and that the three dominant types of workplaces are retail stores, grocery stores, and places in the service industry (e.g. McDonald's and coffee shops).

[26] Although the time frame for this sample does not overlap with the main analysis sample, the additional details about youth community service contributes to the institutional understanding. Data for the years under study in this paper was no longer available at the time of initiating the study.

Table 1.1: Summary Statistics

	(1) All 2003-2007	(2) Convicted	(3) Community service
Female	0.481	0.165	0.106
	(0.500)	(0.371)	(0.308)
School enrolment	0.966	0.858	0.894
	(0.182)	(0.349)	(0.308)
Family disp. income (00's)	4736.8	3939.6	3978.6
	(6082.6)	(2602.8)	(2115.0)
Family social ben. (0/1)	0.0891	0.277	0.269
	(0.285)	(0.446)	(0.442)
Either parent immigrant	0.240	0.371	0.385
	(0.427)	(0.483)	(0.487)
Single-parent HH	0.282	0.494	0.492
	(0.450)	(0.498)	(0.498)
Father working	0.804	0.577	0.581
	(0.397)	(0.494)	(0.494)
Mother high school	0.837	0.639	0.631
	(0.370)	(0.480)	(0.483)
Ever convicted	0.0561	1	1
	(0.230)	(0)	(0)
Community service	0.00653	0.148	1
	(0.0806)	(0.355)	(0)
Num. past crimes	0.0278	0.630	0.682
	(0.261)	(1.078)	(0.999)
Prosecution time		158.7	154.9
		(99.75)	(90.17)
Num. charges		1.967	2.375
		(1.834)	(2.170)
Age at crime		16.00	16.03
		(0.813)	(0.788)
Observations	730,304	22,374	3,302

Notes: Sample: Swedish residents, aged 15-17. In (1) population, in (2) all with a criminal conviction, in (3) those with a court sentence and in (4) those sentenced to youth community service. Family disposable income measured in 100's of 2017 SEK. Family social benefit is a dummy equal to one if the family receives any income from social welfare. Single-parent household defined as living with either mother or father only. "Community service" is equal to one if sentenced in court to community service. "School enrolment" equals one if individual is enlisted in education at any level in the fall semester. "Work experience" is a dummy for any pre-conviction employment. "Prosecution time" denotes time between arrest and sentencing.

Table 1.2: Stockholm Social Services Notes 2012-2016.

	(1) Mean	(2) SD
Hours sentenced	40.49	(26.55)
Hours completed	28.13	(22.64)
Completed service	0.836	(0.370)
Work = Grocery store	0.235	(0.424)
Work = Municipal	0.0340	(0.181)
Work = Non-profits	0.0525	(0.223)
Work = Other private	0.0154	(0.123)
Work = Second-hand store	0.370	(0.484)
Work = Service	0.167	(0.373)
Observations	324	

Notes: Summary statistics for cases assigned to community service in 2012-2016, copied from social services notes in Stockholm. Courtesy of *Framtid Stockholm*. Sample restricted to notes still in the archive (N=353) and with match in register data (N=324). Variable mean in Column 1, standard deviation in Column 2.

Outcome Variable Construction

Labor market outcomes are recorded in the income taxation data, listing each person's total annual earnings. From these, I construct indicator variables for employment. Employment one year after conviction is defined as one if a person has earnings exceeding 5,000 USD (2017) in that year; and likewise for years two to nine after conviction. I also estimate effects on annual labor earnings (in 100's 2017 SEK).[27] For the "reform sample" (2006-07 convictions), earnings one year after the reform is measured in 2008, and analogously for each of the "pre-reform samples" (2005 for 2003-04 convictions, 2006 for 2004-05 convictions and 2007 for 2005-06 convictions). Earnings in years two to nine after the reform are defined in the same way.[28] As a final evaluation of a person's labor market attachment, I also

[27] Because of the large number of zero data entries, I chose not to use log earnings.

[28] I have checked whether these measures are sensitive to the exact choice of

look at the probability of receiving income assistance (social welfare). This variable takes the value one if a person's household (with or without parents or guardians) receives any social welfare.

Recidivism at the one year-horizon is defined as a new arrest that leads to a conviction and that occurs within 365 days of the original court date. This is defined analogously for two to nine years after conviction and thus captures the cumulative risk of a new crime. I also construct the same measures, but for violent and property crimes (the two largest crime categories) separately. To study relapse into more serious crime, any new conviction within one to nine years that leads to a prison sentence is also constructed. Given that the youngest individuals in the sample will have turned 25 years old by 2016, this variable captures the cumulative risk of serious offences during "peak crime age" in early adulthood.

To study human capital formation, I look at two measures. *High school graduate* takes the value one if an individual is observed completing high school at any point within ten years.[29] The outcome variable post-secondary studies is an indicator equal to one if the individual is recorded as enrolled in studies at any level beyond secondary school (thus not including adult secondary education) at any point within ten years.

1.4 Results

In this section, I first present how the probability of court-mandated community service and other sanctions changed because of the reform. Then I evaluate individual-level effects of the reform in Section 1.4.2, starting off with recidivism, followed by educational outcomes and labor market outcomes. Each of these sections also covers effects

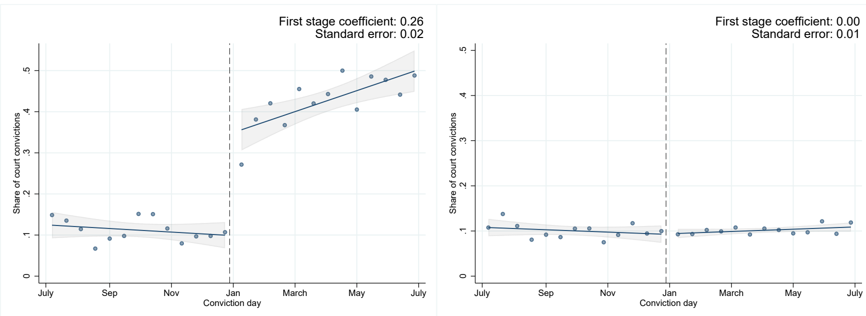
outcome specification, including the following: total annual earnings, labor force participation as recorded in the official registry (any employment in November) and employment defined as having annual earnings above other earnings thresholds. All specifications produce similar results. Results are available upon request.

[29]The result remains unchanged if I instead consider high school completion within a given time frame, such as five or ten years, or completion "on time" at age 19.

separately by the most likely alternative sanction.[30] Section 1.4.3 discusses sensitivity analyses.

1.4.1 Reform Effects on Youth Sanctions

Figure 1.3: First stage. Share of Community Service Sentences Among Juvenile Sentences.



(a) Reform year (b) Pre-reform years

Note: The figure shows a binned scatter plot (bin size approx. two weeks) of the share of youth community service sentences to juvenile offenders (ages 15-17) among all juvenile court convictions, along with a fitted linear trend on each side of the cut-off date (Jan 1st 2007, indicated by the dotted line) with 95 % CIs. Numbers in top left corners show the point estimate and standard error from OLS estimation of the discontinuity in the probability of court-ordered community service.

Panel a) of Figure 1.3 shows how, immediately following the reform implementation in 2007, community service as a share of all youth sentences increased from around 10 to 40%. During the first six months of 2007, it increased further to a level of around 50%. A formal test of difference in probability of community service across the threshold date is presented in the top left corner of the figure: 0.26

[30]Appendix Table 1.E3 also shows regression discontinuity estimates on the "reform sample": the effect of being sentenced in January-June of 2007 and thus subject to the juvenile sentence reform, rather than in July-December of 2006. Table 1.E2 shows OLS estimates of a community service sentence.

(0.02).[31] The change in treatment probability across the threshold is thus statistically significant at any conventional level. Panel b) shows the equivalent graph for pre-reform years, where no discontinuity across the new year is present.

Table 1.3 reports RD-DD regression estimates of the the reform-induced change in the probability of community service, as well as other youth sanctions. The coefficients displayed adhere to the indicator for being convicted after the new year in pre-reform years (*Jan-June)* and the indicator for being convicted in January-June in the reform year (*Jan-JuneXReform*), which is the main coefficient of interest. In Column 1, the dependent variable is an indicator for whether the case is diverted from a court trial — thus addressing the concern that the justice system might react to the reform by changing the proportion of juvenile offenders who are brought to court, and thereby subject to the new court sanction regime. Evidently, the reform does not change prosecutor behavior regarding diversion practices. Appendix Figure 1.E1 also plots the density in non-trial convictions in 2006-2007, showing that cases were not differentially kept out of court trials before and after the reform implementation. Together, this is evidence that the total number of processed court cases did not change as a consequence of the reform.

[31] The regression model is: $\mathrm{S}_{itc}^{k} = \alpha_0 + \alpha_1 \mathrm{Reform}_{it} + \alpha_2 f(\mathrm{Day}_{it}) + \alpha_3 f(\mathrm{Reform}_{it} \times \mathrm{Day}_{it}) + X_{itc}\gamma + \mu_c + \nu_{itc}$, where S^k denotes sanction type k, and $\mathrm{Reform}_{it} = 1(t > d)$ for $d =$ January 1st 2007.

Table 1.3: Reform Effects on Youth Criminal Sanctions.

	Diversion	Court sanctions					
	(1) Non-court	(2) Com. service	(3) Care	(4) Fines	(5) Probation	(6) Prison	(7) Other
Jan-June XReform	0.005	0.317***	-0.187***	-0.097***	-0.023***	-0.006*	-0.003
	(0.009)	(0.015)	(0.017)	(0.019)	(0.007)	(0.004)	(0.004)
Jan-June	-0.030***	-0.008	0.058***	-0.010	-0.024***	-0.009**	-0.007*
	(0.009)	(0.010)	(0.017)	(0.017)	(0.006)	(0.004)	(0.004)
Court FEs	No	Yes	Yes	Yes	Yes	Yes	Yes
Dep. mean	0.35	0.1	0.38	0.42	0.06	0.02	0.02
F-statistic	958.39	63.04	105.12	221.84	23.63	14.11	14.07
Observations	34,594	22,374	22,374	22,374	22,374	22,374	22,374

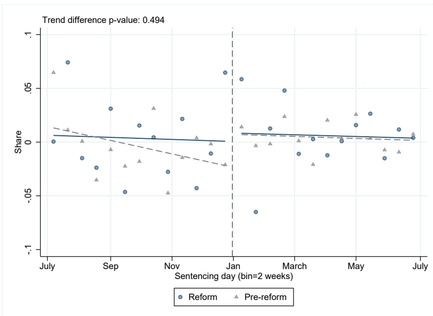
Notes: All Columns show difference-in-discontinuity estimates (eq. 1.1), and display the main effect of convictions from after the new year (January-June) and the interaction of January-June conviction in the reform year (Jan-JuneXReform). Dependent variables are indicators for different forms of youth sanctions, and "Dep. mean" shows their average values in July-December of pre-reform years. Column 1: all criminal convictions with offender age below 18. Column 2-7: all juvenile criminal cases settled in courts. Standard errors clustered at court level in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Columns 2-7 report estimated effects on, in turn, community service, care, fines, probation, prison (juvenile and adult) and any other type of sanction. The RD-DD estimate for the increase in community service, in Column 2, (0.317) differs slightly from the simple difference in means presented in Figure 1.3, as a result of the change in the underlying sample. The reduction in the use of care and fines make up for 90 % (-0.187 and -0.097, respectively) of the increase in community service. The remaining change stems from a smaller, but statistically significant reduction in the use of probation (-0.023). Also prison sentences display a statistically significant reduction, but the change is small enough (-0.006 compared to the pre-reform mean of two percent) to represent only a couple of individual cases.

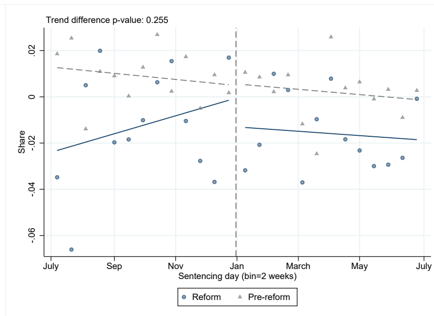
1.4.2 Reform Effects on Crime, Education and Labor Market Outcomes

Figure 1.4 provides a first glance at the difference-in-discontinuity results. It shows nonparametric visual evidence on the effects of being convicted in 2007, and thus substantially more likely to get court-mandated youth community service instead of fines or care, controlling for any potential seasonality effect. Four outcome variables are shown: recidivism, incarceration, high school completion and employment; all measured five years after conviction. Each outcome is a residual after controlling for the full set of variables mentioned in Section 1.3.1. They are plotted by conviction day binned into approximately two-week spans. Local linear trends of the conviction date are fitted on each side of the new year, for the reform sample (2006-07) and the pre-reform sample (2003/04, 2004/05 and 2005/06). The RD-DD estimate is in essence the difference between the discontinuity in the solid lines (reform year) minus the discontinuity in the dashed lines (pre-reform years). Evidently, the discontinuities at the threshold date — if any — are similar in the reform and pre-reform samples, indicating that the reform had limited effects on average.

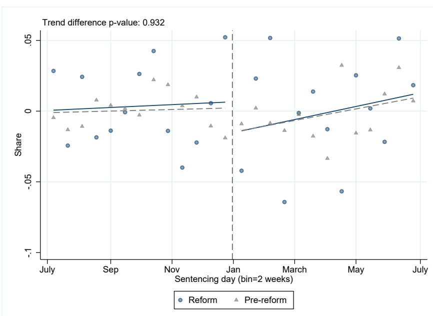
Figure 1.4: Regression Discontinuity Plots for Reform and Pre-reform Samples.



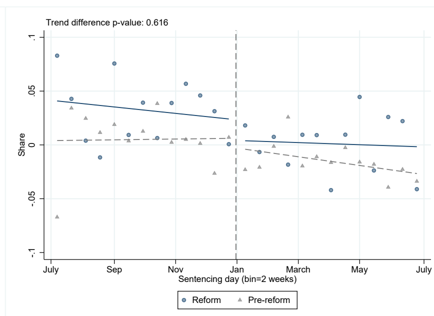
(a) Recidivism



(b) Prison sentence



(c) High school degree



(d) Employment

Note: Each panel shows a binned scatter plot (bin size approx. two weeks) of outcome variable means by conviction day, with linear fitted trends in conviction day on each side of the cut-off date (Jan 1st 2007 in the reform sample plots, Jan 1st in 2006, 2005 and 2004 in the pooled pre-reform sample). Bandwidth is six months on either side of threshold date. Panel (a): Any new conviction. (b): Any prison sentence. (c): High school degree. (d): Employed (employment = earnings>5,000 USD). All outcomes measured after 5 years.

Comparing the trend in each outcome variable between fall convictions of the reform and pre-reform samples, respectively, provides an informal test of the parallel trends assumption. If trends are observed to be similar among fall convictions, then pre-reform spring convictions presumably constitute a valid counterfactual scenario for spring convictions in 2007. The p-value of a test for differences in pre-trends is displayed in the top left corner of each graph. Evidently, fall convictions display a similar pattern in reform and pre-reform years in panels (a), (c) and (d). For prison sentences in panel (b), the slopes of the fitted trend lines diverge. However, a test for differences in slope coefficients does not reject the null hypothesis of no difference (p-value = 0.255).

In Appendix Figure 1.E2, the same visual evidence is presented for the fines and care groups separately. These groups are formed according to the predicted counterfactual sanction, as described in Section 1.3.2. The outcome variables are the same as in the previous figure. For Prison sentences (panel c) and Employment (panel g), trends diverge among the "fines group". Looking closer, the trends differ mainly because mean values for convictions from the last two weeks of December 2006 differ substantially from the other two-week averages. The same pattern is visible in all four chapter1/graphs, and implies that cases brought to court in the end of December and the beginning of January differ somewhat from other cases, and more so in the reform years than in the pre-reform period. In Section 1.4.3 I discuss implications of this for my results, and estimate "Donut-RD" versions of my results, which excludes these two weeks from the sample. From this graphical account, the reform appears to have affected certain key outcomes, and these effects differ depending on what type of sanction community service is replacing.

Recidivism

Table 1.4 shows difference-in-discontinuity estimates (eq. 1.1) of the reform effect on four measures of recidivism: any new conviction in Columns 1-2, any new conviction for a violent crime in Columns 3-4, any new conviction for a property crime in Columns 5-6, and any future conviction resulting in a prison sentence in Columns 7-8. These results are shown for the entire sample in Panel A, for individuals whose most likely alternative sanction is a fine in Panel B, and for individuals likely to have otherwise received individualized care in Panel C. All regression models control for crime and court fixed effects, and additional control variables are included in Columns 2, 4, 6 and 8. All outcome variables are measured from the conviction date and until five years later (cumulative risk), i.e. when the individuals are 20-25 years old.

Evidently, the average effects of the reform in the general youth offender population are small, and imprecisely estimated. With 95 % confidence intervals, I can rule out effects larger than five percent compared to the dependent variable means, for recidivism in general and for recidivism into violent crimes. For property crimes and prison sentences, the confidence spans are wider. Comparing Panel A to the subsets with a clearly identified counterfactual outcome in Panels B and C, these zero effects are a composite of two contrasting effects.

Among the "fines group", the number of subsequent convictions and the risk of a future incarceration spell decrease — by 10 and 59 percent respectively, when including control variables. This effect is driven by reductions in both violent and property crimes, although the effect on violent crimes is more precisely estimated. Among the "care group", on the other hand, point estimates are positive (though imprecise), indicating an increase in the recidivism rate as well as in the risk of violent crimes and property crimes. The estimated effects on the risk of incarceration is positive and significant; on average, the incarceration rate is three percent higher after the reform among this group, which is an increase of 23 percent over the mean.

In Figure 1.E3, RD-DD estimates at different time spans — one to nine years after conviction — are shown for two outcome variables: Any new conviction in Panel a) and any conviction resulting in a prison sentence in Panel b). Each figure plots these *dynamic effects* separately for the whole sample, and for the fines and care subsamples. These graphs show that the estimated effects for both the fines and care groups are similar (and similar to average treatment effect) one year after conviction. After that, the estimates start to diverge, with the fines group showing consistently negative effects while effects in the care group are positive or zero at all time horizons.

Table 1.4: Effects of the Sanction Reform on Recidivism

	Recidivism		Violent		Property		Prison	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: All								
Jan-June X Reform	0.000	-0.004	0.010	0.006	-0.021	-0.021	0.003	0.004
	(0.012)	(0.011)	(0.011)	(0.011)	(0.014)	(0.014)	(0.010)	(0.009)
Dep. mean	0.63	0.63	0.35	0.35	0.32	0.32	0.13	0.13
Observations	22,374	22,374	22,374	22,374	22,374	22,374	22,374	22,374
B: Fines								
Jan-June X Reform	-0.041	-0.050*	-0.037	-0.040*	-0.040	-0.043	-0.038**	-0.041***
	(0.028)	(0.026)	(0.024)	(0.023)	(0.028)	(0.026)	(0.015)	(0.015)
Dep. mean	0.52	0.52	0.25	0.25	0.26	0.26	0.07	0.07
Observations	5,594	5,594	5,594	5,594	5,594	5,594	5,594	5,594
C: Care								
Jan-June X Reform	0.024	0.018	0.004	-0.003	0.014	0.007	0.036*	0.033*
	(0.038)	(0.035)	(0.026)	(0.023)	(0.030)	(0.029)	(0.021)	(0.017)
Dep. mean	0.66	0.66	0.39	0.39	0.35	0.35	0.14	0.14
Observations	5,594	5,594	5,594	5,594	5,594	5,594	5,594	5,594
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: All Columns show difference-in-discontinuity estimates (eq. 1.1), and the independent variable is the interaction term Reform × January-June conviction. Panel A: All observations. Panel B: $\hat{Pr}(Fines) \geq$ 75th percentile. Panel C: $\hat{Pr}(Care) \geq$ 75th percentile. Court fixed effects and crime fixed effects included in all models. Bandwidth is six months before and after January 1st. "Dep. mean" shows control sample mean of dependent variable. Standard errors clustered at court level. * p<0.1, ** p<0.05, *** p<0.01.

Taken at face value, these results indicate that, overall, the reform failed at one of its main motivations, namely reducing the risk of future criminal convictions and continued risky behavior among juvenile offenders. Given the high rates of recidivism in general - about 35 % in the first year after conviction and a 13 % risk of ending up in prison within five years - this must be seen as negative news to the responsible policy makers. However, the reform did achieve this goal among an important subset of the youth offender population — those who would have been ordered to pay a fine instead of doing community service, in absence of the reform. One way to see this is that community service as a restorative justice program has certain advantages over a pure punishment such as fines. In the literature, restorative measures such as community service are often argued to primarily benefit youth convicted of milder forms of delinquency, and experiences from e.g. Finland suggest that restorative justice for offenders convicted of violent crimes are detrimental for future crime rates (Lappi-Seppälä, 2011). The evidence presented here points in the same direction: an increased juvenile justice focus on community service is found to reduce future severe criminality among youth convicted of milder forms of crimes, but to have the opposite effect for youth convicted of more severe crimes.

Educational Outcomes

Table 1.5 shows results from the same model specifications as those just presented, but for two educational outcomes: ever graduating from high school in Columns 1-2, and ever attending post-secondary education in Columns 3-4. They show that the reform did not increase the motivation of youth offenders to either complete high school, or go on to higher education or vocational schooling. The results in Panel A are quite precisely estimated zeroes ($\beta_1 = 0.003\,(0.017)$ in Column 2, $\beta_1 = 0.005\,(0.015)$ in Column 4).[32] Panels B and C show indications of heterogeneous responses in terms of the high school graduation rate, with an increased probability by eleven percent among the “fines” group, but a negative (but imprecise) effect on

[32] This corroborates previous work by (Azad, 2019), who finds that youth community service does not increase motivation to graduate from high school among a small sample of girls convicted of misdemeanour offences.

the "care" group. Neither group displays a statistically significant result on post-secondary education. Additional results, available upon request, shows that neither high school enrollment/completion not enrollment at any level of education are affected in the short run, for either group.

Table 1.5: Effects of the Sanction Reform on Education.

	High school completion		Any college	
	(1)	(2)	(3)	(4)
A: All				
Jan-June X Reform	0.004	0.003	0.003	0.005
	(0.018)	(0.017)	(0.014)	(0.015)
Dep. mean	0.56	0.56	0.22	0.22
Observations	22,374	22,374	22,374	22,374
B: Fines				
Jan-June X Reform	0.066	0.070**	-0.007	-0.005
	(0.041)	(0.034)	(0.032)	(0.030)
Dep. mean	0.61	0.61	0.26	0.26
Observations	5,594	5,594	5,594	5,594
C: Care				
Jan-June X Reform	-0.036	-0.032	-0.025	-0.022
	(0.026)	(0.027)	(0.026)	(0.025)
Dep. mean	0.53	0.53	0.22	0.22
Observations	5,594	5,594	5,594	5,594
Controls	No	Yes	No	Yes

Notes: All columns show difference-in-discontinuity estimates (eq. 1.1), and the independent variable is the interaction term Reform × January-June conviction. Panel A: All observations. Panel B: $Pr\hat{Fines} \geq$ 75th percentile. Panel C: $Pr\hat{Care} \geq$ 75th percentile. Court fixed effects and crime fixed effects included in all models. Bandwidth is six months before and after January 1st. "Dep. mean" shows control sample mean of dependent variable. Standard errors clustered at court level. * p<0.1, ** p<0.05, *** p<0.01.

Labor Market Outcomes

Next, effects of the reform on annual labor income (Columns 1-2), employment (indicator, Col. 3-4), and social welfare payments (indicator, Col. 7-8) are shown in Table 1.6. All are measured five years after conviction in order to capture labor market attachment in early adulthood. Average treatment effects, presented in Panel A, are small and not statistically significant. With 95 % confidence intervals, I can rule out effects larger than ten percent compared to the variable means. Youth in the "fines group" display a positive effect on earnings by about eight percent, but this is not statistically significant when adding control variables. On the other hand, their probability of receiving social welfare decreases by 16 % (-0.048, se 0.026). In contrast, the point estimate on labor income for the "care group" is negative, with zero effects on employment and welfare receipts.

Dynamic difference-in-discontinuity estimates for the same three outcome variables are shown in Figure 1.E3, Panels c)-e). Each point estimate represents an effect at between one and nine years after conviction, and this is plotted for the whole sample and for the two counterfactual outcomes. Again, the effects differ across the two groups: income and employment are unaffected or reduced among the care group, while positively affected in the long run among the fines group. The reduction in welfare payment reliance found above for the fines group seems to appear around three years after conviction, and persists throughout the studied time frame.

Table 1.6: Effects of the Sanction Reform on Employment and Earnings.

	Income		Employed		Welfare	
	(1)	(2)	(3)	(4)	(5)	(6)
A: All						
Jan-June X Reform	0.911	-23.930	-0.004	-0.013	-0.011	-0.003
	(26.347)	(25.189)	(0.014)	(0.014)	(0.013)	(0.012)
Dep. mean	988.43	988.43	0.52	0.52	0.35	0.35
Observations	22,374	22,374	22,374	22,374	22,374	22,374
B: Fines						
Jan-June X Reform	117.059*	82.228	0.041	0.029	-0.061**	-0.048*
	(66.697)	(65.141)	(0.033)	(0.032)	(0.028)	(0.026)
Dep. mean	1048.36	1048.36	0.56	0.56	0.30	0.30
Observations	5,594	5,594	5,594	5,594	5,594	5,594
C: Care						
Jan-June X Reform	-32.291	-33.042	-0.000	0.000	0.005	-0.000
	(57.717)	(51.633)	(0.025)	(0.025)	(0.035)	(0.030)
Dep. mean	902.51	902.51	0.48	0.48	0.39	0.39
Observations	5,594	5,594	5,594	5,594	5,594	5,594
Controls	No	Yes	No	Yes	No	Yes

Notes: All columns show difference-in-discontinuity estimates (eq. 1.1), and the independent variable is the interaction term Reform × January-June conviction. Panel A: All observations. Panel B: $Pr\hat{Fines} \geq$ 75th percentile. Panel C: $Pr\hat{Care} \geq$ 75th percentile. Court fixed effects and crime fixed effects included in all models. Bandwidth is six months before and after January 1st. "Dep. mean" shows control sample mean of dependent variable. Standard errors clustered at court level. * p<0.1, ** p<0.05, *** p<0.01.

1.4.3 Robustness Checks

The difference-in-discontinuity method only identifies the causal effect of the reform-induced change in juvenile sanctions if people convicted in the end of 2006 constitute a good control group for those convicted in the beginning of 2007. I will here perform a series of tests to validate this assumption.

Donut RD-DD

First, the holiday season affects the number of cases brought to trial in December and January. Possibly, cases of a more urgent type are then given priority, which would affect the case composition during these moths. The same pattern could arise if offenders were able to sort into and out of the treatment group, depending on individual preferences. If cases brought to court in December and January differ from other cases in ways that make them especially responsive to the reform, this could explain the treatment effect difference. In Appendix Table 1.E4, I show that running a "Donut-RD" specification, where court dates from the two last weeks of December and the two first weeks of January are dropped from the sample, does not alter the main results.

Selection bias

One might worry that the reform induced a compositional change in unobservable characteristics among the youth offender population. Next, I thus check for discontinuous variation in the amount of juvenile cases with observable characteristics that mark them as likely to be affected by the sentencing reform, around either the reform decision date or the reform implementation date. I estimate the probability of being given community service by case and individual characteristics in 2007, and predict the probability of community service based on the same characteristics for cases from 2006.[33] The

[33]The procedure is this: first, I regress an indicator for a community service sanction on a vector of individual and case characteristics. These are gender, parent immigrant status, single-head household, family (log) income, not living with the parents, age at sentencing, past number of crimes, number of charges, time between arrest and conviction, court and main and secondary (if any) crime

predicted probabilities around either threshold date are then evaluated. If individuals are able to sort according to preferences, this should be visible as a discontinuity in community service probability by conviction day around January 1st 2007. The results are displayed in Appendix Figure 1.E5 Panel a), and show no sudden changes around the threshold date. In Panel b), the probability is instead binned by *arrest date* around the time that the reform was voted through in the Swedish Parliament. If the composition of juvenile offenders changed as a result of at-risk youth anticipating the implementation of the reform after its adoption, we would expect a discontinuity in this predicted probability of community service for arrests made around May 30th 2006. No such evidence is found.

Placebo test

To verify that nothing else changed at the reform implementation date, which could confound the estimated effects on youth offenders, I perform a placebo exercise using offenders aged 18 to 20.[34] These were not affected by the sanctions reform, but constitute a group with arguably similar labor market prospects after conviction. Obviously, many things about the criminal justice system differ between minors and young adults. However, for young adults to constitute a valid placebo group in this setting, there can be no "spillover effects" of the reform, through for example a sudden increase in the probability of community service. Appendix Figure 1.E6, Panel a) shows the share of offenders aged 18-20 given community service (either youth service or the adult version, which differs from the youth version in its design), by day of conviction, with fitted lines on either side of January 1st 2007. While there is a slight increase in probability over time, there is no discontinuity at the cut-off date. Panel b) shows the trend in all sanctions given to offenders aged 18-20 in 2006-07. The four most common sanctions are fines, probation, prison and community

type. Next, I calculate fitted values (probabilities) for all individuals in the sample.

[34] Formally, the sentence form can be applied to offenders aged 18-20 under "extraordinary circumstances" (Prop. 2005/06:165, p. 48). This is rarely used, however - in 2007 only three percent of offenders aged 18 got youth community service, most commonly as a result of just barely having turned 18 when breaking the law.

service, none of which display a discontinuous change in probability around the time of the reform implementation. Table 1.E5 shows RD-DD results for this sample, on earnings, employment, high school completion, post-secondary education enrollment, recidivism and the incarceration risk — all measured after five years. All outcome variables show zero or imprecisely estimated effects, indicating that confounding effects for the treatment group are unlikely.

1.5 Policy Evaluation

From the set of results presented so far, the 2007 criminal justice policy reform primarily affected criminal behavior, but was less effective at addressing the weak labor market attachment of youth offenders. In this section, I do two things. First, I evaluate whether the policy constitutes a net fiscal cost or saving, by summarizing the individual-level responses documented above into a measure of "net costs". I then use this summary outcome to study how the effects of the policy differ across people from different backgrounds. For this purpose, I apply the *causal forest* method presented earlier.

1.5.1 Calculation of Net Costs

I evaluate the total accrued fiscal costs and revenues of the 2007 youth justice reform as follows. In a first step, I summarize for each individual the costs of all criminal convictions within five years after the focal conviction. This involves costs related to police work, prosecution, and court trials, plus the total cost of any incarceration spells. The second step is to calculate the *net transfers* accumulated by each person over the five years following conviction. In short, I estimate the total tax revenue from labor earnings and subtract from this the sum of social welfare payments (pure income assistance, not including earnings-related assistance such as unemployment benefits). On top of this, I add the cost (or revenue, in case of fines) of the first sanction. For details on these calculations, see Appendix 1.D.

The result is a measure of net costs accumulated by each individual in my analysis sample. This measure serves two purposes. First, it captures the efficiency gain or loss induced on state finances by the reform from differences in costs of subsequent offender behavior and from the difference in sanction costs. It is also, as mentioned above, a way to weigh together the effects on recidivism and labor market outcomes for each individual, to understand the net value for offenders.[35]

Figure 1.D1 illustrates the distribution of net costs, as well as the distribution of its two components: crime costs and net transfers. The crime costs in Panel a) exhibit a long right tail of extreme values, constituting repeat offenders given long prison sentences as adults. Panel b) shows that most youth offenders do not acquire taxable income in early adulthood, and a large fraction also relies on welfare payments, either individually or through their family. The net cost calculation in Panel c) shows that only about 25 % of the sample generate state income that exceeds the cost of their criminal behavior.

In Table 1.7, I estimate the effect of the reform on these aggregate outcome variables with the RD-DD model. Effects on crime costs without and with control variables are shown in Columns 1-2, net transfers in Columns 3-4, and the sum of these in Columns 5-6. Panel A holds results for the full sample, and these confirm the zero average treatment effect. Among the fines group (in Panel B), the reform is found to result in a positive net effect of 56,900 SEK (se=32.2), stemming from both reduced crime costs and increased net transfers. The reform did not alter the net transfers among the care group (in Panel C), but it increased crime costs, which produces a net negative effect of 89,000 SEK (se=52.4) per person on average. That the cost increase among the care group exceeds the cost savings among the fines group can be seen as bad news for the policy makers. Next, I explore how the sanction could be targeted based on observable characteristics for improved program efficiency.

[35] Obviously, this measure does not fully capture social costs and benefits of the reform. Given the uncertainty over e.g. victimization costs of crime, such a calculation is likely fraught with error, and left outside of the scope of this paper.

Table 1.7: Estimated Total Cost Impacts of the Reform (in 1,000 SEK 2017).

	Cost crimes (C)		Net transfers (G)		Net costs (G-C)	
	(1)	(2)	(3)	(4)	(5)	(6)
A: All						
Jan-June X Reform	9.990	13.779	3.091	0.429	-6.899	-13.350
	(24.008)	(23.735)	(2.395)	(2.078)	(24.089)	(23.730)
Dep. mean	339.83	339.83	3.68	3.68	-336.14	-336.14
Observations	22,374	22,374	22,374	22,374	22,374	22,374
B: Fines						
Jan-June X Reform	-46.342	-50.504	10.918**	6.380	57.259*	56.884*
	(32.055)	(31.668)	(5.386)	(4.878)	(32.970)	(32.210)
Dep. mean	173.87	173.87	13.21	13.21	-160.66	-160.66
Observations	5,594	5,594	5,594	5,594	5,594	5,594
C: Care						
Jan-June X Reform	91.433	86.358	-2.517	-2.391	-93.951*	-88.749*
	(55.260)	(52.022)	(6.150)	(6.085)	(55.723)	(52.375)
Dep. mean	378.57	378.57	-3.41	-3.41	-381.99	-381.99
Observations	5,594	5,594	5,594	5,594	5,594	5,594
Controls	No	Yes	No	Yes	No	Yes

Notes: All columns show difference-in-discontinuity estimates (eq. 1.1), and the independent variable is the interaction term Reform $\times$ January-June conviction. Panel A: All observations. Panel B: $Pr\hat{Fines} \geq$ 75th percentile. Panel C: $Pr\hat{Care} \geq$ 75th percentile. Court fixed effects and crime fixed effects included in all models. Bandwidth is six months before and after January 1st. "Dep. mean" shows control sample mean of dependent variable. Standard errors clustered at court level. * p<0.1, ** p<0.05, *** p<0.01.

1.5.2 Conditional Average Treatment Effects

To study how treatment effects vary with individual characteristics, I implement the causal forest analysis described in Section 1.3.3. Figure 1.5 provides a visual account of differences in means: its first Panel ("All") shows standardized differences in means of observable characteristics between the group with the 20 percent lowest CATEs (the first quintile, "q1") against the group with the 20 percent highest CATEs ("q5"), for the full sample. The covariates listed as those fed into the causal forest algorithm. In the remaining two panels, the same analysis is carried out for the two counterfactual outcome groups, comparing the lowest one-third CATEs to the highest one-third.[36] The average values and a t-test test for differences in means in non-standardized units are also presented in Appendix Table 1.C2.

First, does the algorithm uncover any meaningful heterogeneity in reform responses, i.e. do the included covariates actually predict any differences in treatment effects? From the first row of Figure 1.5, showing the predicted CATEs by percentile groups, it is evident that the differences in treatment effects are large — about two standard deviations — and statistically significant at the one percent level, for the whole sample as well as within the two subgroups. In the whole sample, the average effect for the group with the lowest (most negative) CATE is a net cost increase per person of about 130,000 SEK (around 13,000 USD). This is roughly equivalent in cost to one and a half months of incarceration. The average treatment effect in the fifth quintile amounts to a cost saving of about 33,000 SEK (3,300 USD) per person, which is comparable to the median monthly wage in Sweden.[37]

To understand what drives these differences, it is informative to look at two measures: the relative informativeness of each covariate for producing the causal forest splits (*variable importance*), and the differences in mean covariate values between quintiles for average treatment effects. Panel a of Appendix Figure 1.C1 shows variable

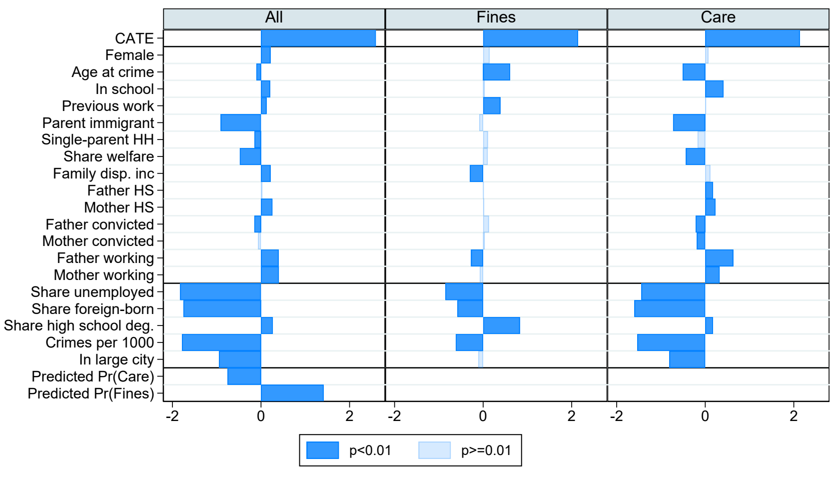
---

[36] The sample is divided into three percentile groups rather than five because of the smaller number of observations.

[37] These differences remain statistically significant when computed as *Augmented Inverse-Propensity Weighted (AIPW) Average Treatment Effect* (Wager and Athey, 2018).

importance measures for the full sample, clearly identifying the *probability of fines* as the most important covariate. Figure 1.5 also shows that in the whole sample, the difference between the first and fifth treatment effect quintiles in predicted $Pr(Fines)$ is one of the most prominent differences in means. Together, these results reinforce the findings in the previous section: cost savings from the reform are concentrated in the "fines group".

Figure 1.5: Causal Forest Covariate Heterogeneity.



Note: This figure shows differences in mean covariate values between the subsample for individuals with the 20 ("All") or 33 ("Fines" and "Care") percent lowest conditional average treatment effect (CATE) and those with the 20 (or 33) percent highest CATE, i.e. q1-q5 (or q1-q3). Variables for which the difference is statistically significant are depicted in darker blue, the insignificant differences in lighter blue. "Fines" is the group with a high predicted probability of fines; "Care" is the group with a high predicted probability referral to social services care. All variables are standardized (mean subtracted and divided by the standard deviation); effects are thus shown in standard deviations of the variables. Share unemployed, Share foreign-born, Share high school deg. and Crimes per 1000 refer to the neighborhood of residence.

Next, I compare results between the fines and the care groups, by looking at the corresponding panels of Figure 1.5. Variable important plots (Panels b and c of Appendix Figure 1.C1) show that within the two counterfactual groups, it is largely the same set of characteristics that predict treatment heterogeneity for the fines group, as for the care group. However, the relation between the covariates and the predicted CATEs are often of opposite signs. While older individual benefit more from community service in the fines group, younger age predicts positive treatment effects in the care group. Having an unemployed father predicts positive treatment effects for the fines group, but the opposite is true for the care group. Moreover, indicators of low socioeconomic status such as neighborhood crime rates and receiving social welfare are stronger predictors of treatment effect heterogeneity among the care group than among the fines group, indicating that program success for the former is more sensitive to the home environment.

Interestingly, the probability of having had any formal employment before conviction is positively related to CATEs for the fines group, but irrelevant for the care group. One interpretation of this is that the "value of a first job" is not per se the most important mechanism through which community service has an effect, but rather that the possibility of finding work again after a criminal conviction is improved. Another possible interpretation is that the crime-reducing effect for the fines group is a result of an increased level of punishment. Since these appear to be relatively well able to pay a fine, community service constitutes a more severe sanction, by actually requiring an effort. As such, it might deter from future crimes. Among the care group — for whom the reform entails community service rather than individualized treatment — the post-reform treatment consists of an arguably lower amount of care. That I find the largest cost increases within this group among individuals from unstable home conditions speaks to an unmet need for social interventions of their part. In other words, these are youth for whom the increased focus on punishment (unpaid work) rather than rehabilitation leads to worse outcomes.

Who Should Get Community Service?

Taken together, these results show that an increased focus on community service among juvenile sanctions can achieve the dual goal of increased employment and lower crime rates if targeted at specific subgroups of youth offenders. If we assume financial efficiency to be the long-term goal of criminal justice policy, the policy recommendation is clear: give community service to youth offenders whose attributes predict net costs above zero (net revenues). However, as recently discussed in (Athey and Wager, 2021), an "optimal policy" (defining a policy as a function that maps observable characteristics to treatment status of some program) might be one that considers just a few attributes. This makes it easy to apply by decision-makers, and transparent for public scrutiny.

The results suggest that a coarse way to practically improve net program revenues would be as follows. First, the judge would consider the type of crime in question — does it warrant a fine or a rehabilitative measure? If the former ("high probability of fines"), my results indicate that the two most important characteristics to consider are offender age and previous employment: community service is a suitable sanction for youth offenders with some work experience. In the other case ("high probability of care"), the recommendation is quite different: give community service to individuals from stable home and school conditions. If age is to be considered, preference should be given to younger individuals. Running the RD-DD model on this "target subsample" returns a precisely estimated per-person cost saving of 95,700 SEK over five years.[38] The majority of this net gain comes from prevented crime costs, which is unsurprising given the magnitude of cost savings from prevented incarceration relative to tax income on income for young people.

My results show that observable characteristics play a key role in how young offenders respond to sanctions. However, recommending that courts consider a young person's background at trial might raise ethical questions. One guiding principle of the criminal justice system is that punishments should be foreseeable, to reduce the risk of arbitrary court outcomes.[39] On the other hand, when it comes to youth offenders, this principle is partly set aside by the overarching focus on restorative justice. In the Swedish case, the law requires judges to consult social services about the juvenile's

[38]More precisely, I limit the sample to individuals with either a high predicted probability of fines *and* a previous employment, or a high predicted probability of care *and* residency in below-median crime neighborhoods *and* 16 years of age or below.

[39]See discussions in e.g. Arnold, Dobbie and Yang (2018) about the unequal treatment of black vs. non-black defendants in the US judicial system.

personality and home situation. Given that judges *do account for* individual circumstances (as is evident from the literature using variation in judge's decision-making, e.g. Aizer and Doyle (2015), Dobbie et al. (2018), Bhuller et al. (2020)), it is appropriate to inform these decisions with empirical facts.

1.6 Conclusions

How to lower the perceived pay-off from further crime for youth offenders? I evaluate effects of a juvenile justice reform which introduced court-ordered community service — in this context consisting of (unpaid) actual labor market experience for youth convicted of non-trivial crimes. People given this sentence are allocated to private, public, and non-profit workplaces to perform real work tasks. Paired with mentorship and behavioral therapy, the treatment has the potential to both be crime-preventive and encourage labor market participation among participants.

The results suggest that on average, the sentencing reform failed to improve offender behavioral outcomes. However, there are important heterogeneous responses. For offenders whose most likely sanction in absence of the reform is monetary fines, recidivism is reduced and the high school graduation rate increases. When the most likely counterfactual is instead individualized treatment, the risk of future incarceration spells increases. Moreover, the effects of the program differ substantially across different subgroups of offenders, above and beyond the type of crime they have committed. From an efficiency standpoint, net financial gains are found for youth convicted of milder crimes with an already-established labor market connection, but for younger offenders with stable home conditions among offenders convicted of more serious crimes.

In sum, my results suggest that providing at-risk youth with short spells of labor market experience does affect the perceived cost of crime for certain individuals, even in a setting where participation in the program is court-mandated rather than voluntary. While the context here is in some regards unique — Sweden stands out internationally in its focus on restorative justice — the question of punishment or treatment programs for youth offenders is universal.

That this by definition low-cost policy can reduce severe recidivism when targeted at the right group of offenders ought to motivate further policy experiments increasing the role of community service among community-based criminal sanctions for young offenders.

Bibliography

Aizer, Anna, and Joseph J. Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *The Quarterly Journal of Economics*, 130(2): 759–803.

Aizer, Anna, Shari Eli, Adriana Lleras-Muney, and Keyoung Lee. 2020. "Do Youth Employment Programs Work? Evidence from the New Deal." National Bureau of Economic Research Working Paper 27103.

Almlund, Mathilde, Angela Lee Duckworth, James J Heckman, and Tim D Kautz. 2011. "Personality Psychology and Economics." National Bureau of Economic Research Working Paper 16822.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Athey, Susan, and Stefan Wager. 2021. "Policy Learning With Observational Data." *Econometrica*, 89(1): 133–161.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *The Annals of Statistics*, 47(2): 1148 – 1178.

Azad, Azadé. 2019. "Characteristics of adolescent females with limited delinquency." PhD diss. Department of Psychology, Stockholm University.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *J. Pol. Econ.*, 76: 169.

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

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2006. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *American Economic Review*, 96(4): 988–1012.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review*, 107(4): 1165–1206.

Brå. 2011. "Ungdomsvård och ungdomstjänst. En utvärdering av 2007 års påföljdsreform för unga lagöverträdare." Brottsförebyggande Radet, Bra Report 10.

Brå. 2016. "School Survey on Crime 2015." Brottsförebyggande Radet, Bra Report 21.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Design." *Econometrica*, 82(6): 2295–2326.

Card, David, Jochen Kluve, and Andrea Weber. 2018. "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations." *Journal of the European Economic Association*, 16(3): 894–931.

Cohen, Alexandra O., and B.J. Casey. 2014. "Rewiring Juvenile Justice: The Intersection of Developmental Neuroscience and Legal Policy." *Trends in Cognitive Sciences*, 18(2): 63–65.

Crépon, Bruno, and Gerard van den Berg. 2016. "Active Labor Market Policies." *Annual Review of Economics*, 8(1): 521–546.

Cunha, Flavio, and James J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources*, 43(4): 738–782.

Davis, Jonathan M.V., and Sara B. Heller. 2017. "Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs." *American Economic Review*, 107(5): 546–50.

Davis, Jonathan M.V., and Sara B. Heller. 2020. "Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs." *The Review of Economics and Statistics*, 102(4): 664–677.

Dobbie, Will, Hans Grönqvist, Susan Niknami, Mårten Palme, and Mikael Priks. 2018. "The Intergenerational Effects of Parental Incarceration." National Bureau of Economic Research Working Paper 24186.

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.

Eggers, Andrew C., Ronny Freier, Veronica Grembi, and Tommaso Nannicini. 2018. "Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions." *American Journal of Political Science*, 62(1): 210–229.

Fallesen, Peter, Lars Pico Geerdsen, Susumu Imai, and Torben Tranaes. 2018. "The Effect of Active Labor Market Policies on Crime: Incapacitation and Program Effects." *Labour Economics*, 52: 263–286.

Gelber, Alexander, Adam Isen, and Judd B. Kessler. 2016. "The Effects of Youth Employment: Evidence from New York City Lotteries." *The Quarterly Journal of Economics*, 131(1): 423–460.

Ginner-Hau, Hanna. 2010. "Swedish young offenders in community-based rehabilitive programmes." PhD diss. Department of Psychology, Stockholm University.

Hausman, Catherine, and David S. Rapson. 2018. "Regression Discontinuity in Time: Considerations for Empirical Applications." *Annual Review of Resource Economics*, 10(1): 533–552.

Heller, Sara B. 2014. "Summer Jobs Reduce Violence Among Disadvantaged Youth." *Science*, 346(6214): 1219–1223.

Heller, Sara B., and Judd B. Kessler. 2021. “The Effects of Letters of Recommendation in the Youth Labor Market.” National Bureau of Economic Research Working Paper 29579.

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack. 2017. “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago.” *The Quarterly Journal of Economics*, 132(1): 1–54.

Hjalmarsson, Randi. 2008. “Criminal Justice Involvement and High School Completion.” *Journal of Urban Economics*, 63(2): 613–630.

Hjalmarsson, Randi, and Matthew J. Lindquist. 2018. “Labour Economics and Crime.” *Labour Economics*, 52: 147–148.

Hockenberry, Sarah, and Charles Puzzanchera. 2020. “Juvenile Court Statistics 2018.” National Center for Juvenile Justice. Pittsburgh, PA.

Kearney, Melissa S., and Phillip B. Levine. 2016. “Income Inequality, Social Mobility, and the Decision to Drop Out of High School.” *Brookings Papers on Economic Activity*, 333–380.

Kessler, Judd B, Sarah Tahamont, Alexander M Gelber, and Adam Isen. 2021. “The Effects of Youth Employment on Crime: Evidence from New York City Lotteries.” National Bureau of Economic Research Working Paper 28373.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. “Human Decisions and Machine Predictions.” *The Quarterly Journal of Economics*, 133(1): 237–293.

Lappi-Seppälä, Tapio. 2011. “Nordic Youth Justice.” *Crime and Justice*, 40: 199–264.

Ministry of Justice. 2020. “Youth Justice Statistics 2018/19.” Youth Justice Board of England and Wales/Ministry of Justice.

Minstry of Justice. 2015. "The Swedish Judicial System." Regeringskansliet.

Modestino, Alicia Sasser. 2019. "How Do Summer Youth Employment Programs Improve Criminal Justice Outcomes, and for Whom?" *Journal of Policy Analysis and Management*, 38(3): 600–628.

Mueller-Smith, Michael, and Kevin T. Schnepel. 2020. "Diversion in the Criminal Justice System." *The Review of Economic Studies*, 88(2): 883–936.

Persson, Petra, and Maya Rossin-Slater. 2021. "When Dad Can Stay Home: Fathers' Workplace Flexibility and Maternal Health." Working Paper.

Rose, Evan K. 2021. "Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders." *The Quarterly Journal of Economics*, 136(2): 1199–1253.

SBU. 2020. "Insatser i öppenvård för att förebygga ungdomars återfall i brott. En systematisk översikt och utvärdering av ekonomiska, sociala och etiska aspekter." Statens beredning för medicinsk och social utvärdering. SBU-rapport nr 308.

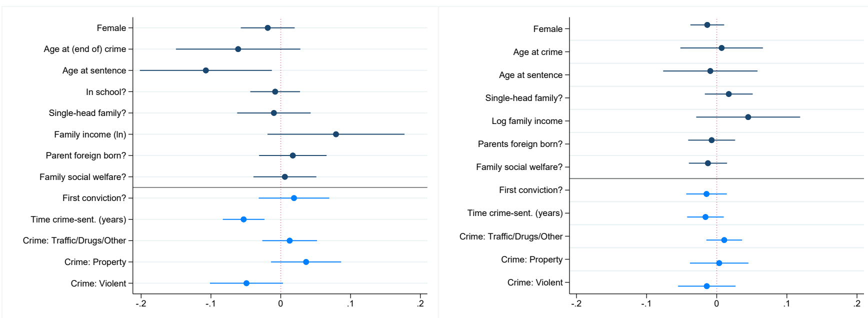
Shem-Tov, Yotam, Steven Raphael, and Alissa Skog. 2021. "Can Restorative Justice Conferencing Reduce Recidivism? Evidence From the Make-it-Right Program." National Bureau of Economic Research Working Paper 29150.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association*, 113(523): 1228–1242.

Appendices

1.A RD-DD Model Evaluation

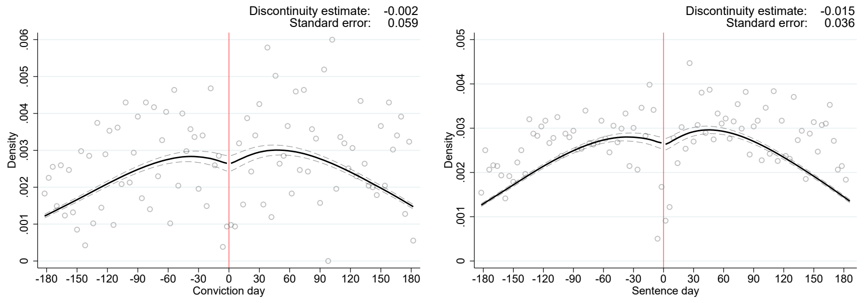
Figure 1.A1: Covariate Balance Between Control and Treatment samples.



(a) Reform sample. (b) RD-DD sample.

Note: Panel a): Each row represents a regression discontinuity estimate ($x_{itc} = \alpha_0 + \alpha_1 \text{Reform}_{it} + \alpha_2 f(\text{Day}_{it}) + \alpha_3 f(\text{Reform}_{it} \times \text{Day}_{it}) + \mu_c + \nu_{itc}$) with the pre-determined covariate as the dependent variable. Panel b): Each row represents estimating the RD-DD model (eq. 1.1) with the pre-determined covariate as the dependent variable. Linear polynomials of the running variable (conviction day) fitted separately before and after the cut-off date, and both models include court fixed effects. Standard errors clustered at court level.

Figure 1.A2: McCrary Test for Discontinuity in Running Variable (conviction day) at Threshold (1st of January).

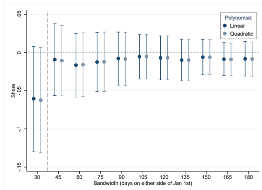


(a) Reform sample.

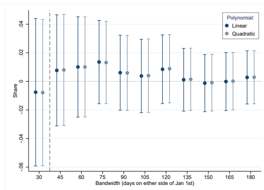
(b) RD-DD sample.

Note: The figure shows a binned scatter plot of density in the running variable, conviction date. The black line fits a polynomial to the running variable density, with 95 % confidence intervals in dashed lines. Numbers in top right corner are from a formal test of a discontinuity in the running variable at the cut-off. Panel a): Conviction dates between July 1st 2006 and June 30th 2007. The red line indicates the cut-off date: January 1st 2007, i.e. day of reform implementation. $N = 5,876$. Panel b): Conviction dates between July 1st 2003 and June 30th 2007. The red line indicates January 1st of 2004-07. $N = 22,374$.

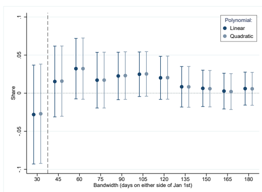
Figure 1.A3: RD-DD Estimates at Different Bandwidths.



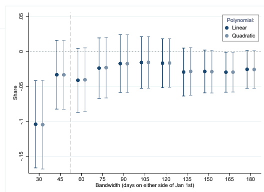
(a) Any new convictions



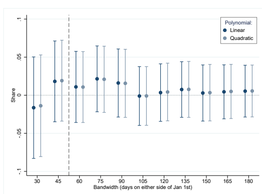
(b) Any prison sentence



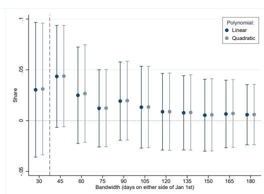
(c) Any violent crime



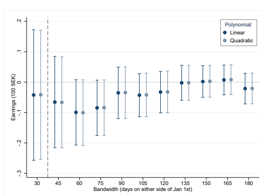
(d) Any property crime



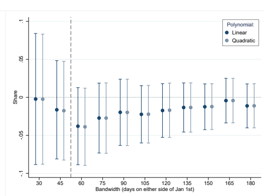
(e) High school completion



(f) Any post-secondary education



(g) Labor earnings

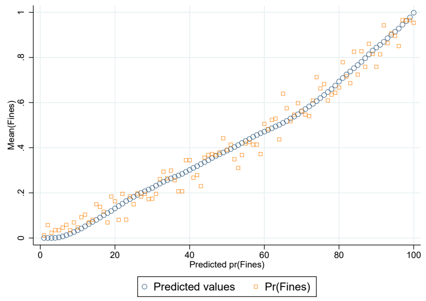


(h) Employment

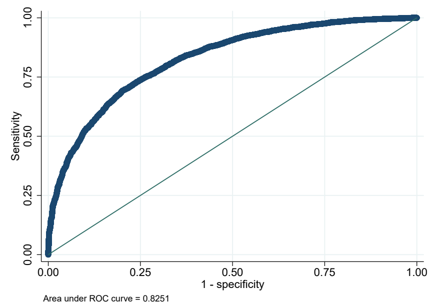
Note: Each panel shows regression estimates of eq. 1.1 at different bandwidths (time before and after Jan 1st 2007), with a linear (darker color) and quadratic (lighter color) function of the running variable respectively. X-axis shows number of days on either side of the cut-off included in sample. CCT optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014) marked with a dashed line. All models include court and crime fixed effects, and additional controls. Standard errors clustered at court level displayed in parentheses.

1.B Counterfactual outcome prediction

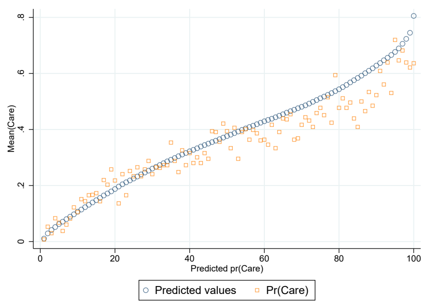
Figure 1.B1: Predicted Probability of Fines and Rehabilitation — Calibration Tests.



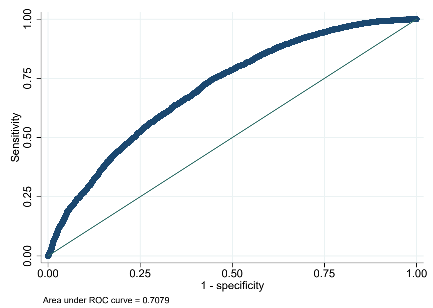
(a) Predicted vs. actual Pr(Fines).



(b) lroc Pr(Fines).



(c) Predicted vs. actual Pr(Care).



(d) lroc Pr(Care).

Note: Panels (a) and (c) show the correlation between the predicted and actual probabilities. Panels (b) and (d) show the area under ROC curve "lroc", i.e. a test for share of correct classifications. Both calibration tests evaluated on hold-out test sample of 50 % of observations. Probabilities of each sanction are predicted on individuals sentenced in 2003-2006 (pre-reform) on observable crime characteristics using a 50 % training sample. Predicted probabilities are then estimated for the reform-sample observations.

Table 1.B1: Summary Statistics for Low vs. High Predicted Probability of Fines.

	(1) High Pr(Fines)	(2) High Pr(Care)	(3) Diff (1)-(2)
Sanctions:			
Community service	0.053	0.155	0.10***
Rehabilitation	0.138	0.585	0.45***
Fines	0.788	0.194	-0.59***
Probation	0.017	0.045	0.03***
Prison	0.000	0.010	0.01***
Crime:			
Crime: Drugs	0.073	0.018	-0.05***
Crime: Other	0.093	0.018	-0.07***
Crime: Property	0.474	0.545	0.07***
Crime: Traffic	0.186	0.000	-0.19***
Crime: Violent	0.174	0.418	0.24***
Hours unpaid work (pred.)	26.381	35.003	8.62***
Defendant:			
Female	0.218	0.162	-0.06***
Immigrant parent	0.327	0.368	0.04***
Family disp. inc. (100's SEK)	4085.196	3866.377	-218.57***
In school	0.901	0.851	-0.05***
Previous work	0.519	0.467	-0.05***
Family welfare (0/1)	0.236	0.285	0.05***
Age crime	16.432	16.409	-0.02
Age sentence	16.930	16.741	-0.19***
First conviction	0.755	0.646	-0.11***
First stage:			
Jan-June X Reform	0.16***	0.31***	
	(0.03)	(0.05)	
Observations	5,594	5,594	11,184

Notes: Summary statistics for the subsamples of individuals with a high (> 75 percentile) predicted probability of being sentenced to fines and care, respectively. Probabilities predicted out-of-sample on individuals sentenced in 2003-2006, and imputed in-sample on observable pre-determined crime characteristics and criminal history: number of previous convictions, time between arrest and conviction, number of charges, main crime (31 categories), second listed charge and a set of court indicators. * p<0.1, ** p<0.05, *** p<0.01.

1.C Causal Forest

I apply the causal forest algorithm as implemented in the R-package *Generalised Random Forest* (Athey, Tibshirani and Wager, 2019). Causal forest estimation specifics: Minimum leaf size = 10; training sample share = 0.5; maximum imbalance of split = 0.05; number of trees = 5,000 when using whole sample, 2,000 when estimating effects for the counterfactual subsamples separately.

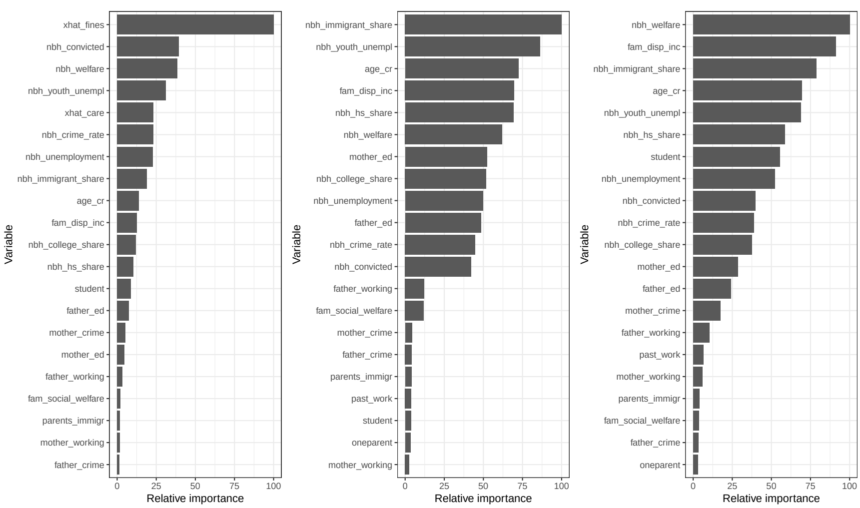
Covariates used are the following. Individual: gender, age at sentencing, age at crime, parent immigrant status, single-parent household, family social welfare, family disposable income (standardized), school enrolment at crime, any past work experience, mother/father with a criminal conviction, mother/father employed and mother/father highest attained level of education. Neighbourhood: indicator for location in the three largest cities, unemployment rate, youth unemployment rate, share foreign-born, share with a high school degree, share with a college degree, share welfare recipients, crime rate, share of population with a criminal record. All variables pertain to the year when the crime is committed. Crime: main crime (30+ categories), second crime (second listed charge, in 30+ categories incl. missing category), number of past convictions and time between arrest and conviction.

Table 1.C1: RD-DD Estimates with Residualized Outcome Variables.

	Sum costs		Residuals	
	(1)	(2)	(3)	(4)
A: All				
Jan-June X Reform	-6.465	-13.350	-4.642	-14.477
	(23.990)	(23.730)	(22.662)	(21.427)
Observations	22,374	22,374	5,876	5,876
B: Fines				
Jan-June X Reform	50.048	56.884*	47.859	50.547
	(32.820)	(32.210)	(31.599)	(32.331)
Observations	5,594	5,594	1,488	1,488
C: Care				
Jan-June X Reform	-89.709	-88.749*	-48.911	-62.775
	(56.338)	(52.375)	(51.460)	(48.822)
Controls	No	Yes	No	Yes
Observations	5,594	5,594	1,407	1,407

Notes: Each column shows estimates for the 2006-07 ("reform") sample, of the following model: $(Y-\hat{Y})_{it} = \beta_0 + \beta_1 Jan-June + \mathbf{X}_{it}\beta_2 + \epsilon_{it}$, where $(Y-\hat{Y})_{it}$ denotes residuals after controlling for a January-June main effect, period fixed effects (i.e. 2004-05, 2005-06 and 2006-07 sample indicators, with 2003-04 as the omitted category) and court fixed effects. Standard errors clustered at court level in parentheses.

Figure 1.C1: Causal Forest - Variable Importance.



(a) All (b) Fines (c) Care

Note: Variable importance is the rate at which a variable is used by the random forest algorithm for partitioning the sample, across all trees in the forest (N= 5,000 for the whole sample, 2,000 for the counterfactual subsamples). These are shown in relative numbers, compared to the variable with the highest importance score. “Fines” denotes the group with a high predicted probability of fines; “Care” denotes the group with a high predicted probability of referral to social services care.

Table 1.C2: Causal Forest Conditional Average Treatment Effects, 1st and 5th/3rd Quintile Comparison.

	All			Fines			Care		
	q1	q5	Diff	q1	q3	Diff	q1	q3	Diff
CATE (tSEK)	-129.535	33.374	162.91***	-0.954	60.556	61.51***	-143.729	10.882	154.61***
Individual:									
Female	0.105	0.185	0.08***	0.169	0.222	0.05*	0.140	0.167	0.03
Age at crime	16.512	16.422	-0.09**	16.232	16.751	0.52***	16.549	16.116	-0.43***
In school	0.825	0.894	0.07***	0.900	0.912	0.01	0.811	0.951	0.14***
Any work experience	0.539	0.602	0.06**	0.495	0.688	0.19***	0.511	0.521	0.01
Family:									
Parent immigrant	0.603	0.167	-0.44***	0.360	0.318	-0.04	0.543	0.194	-0.35***
Single-parent household	0.449	0.306	-0.14**	0.350	0.453	0.10	0.445	0.278	-0.17*
Share welfare	0.365	0.157	-0.21***	0.204	0.249	0.05	0.351	0.162	-0.19***
Family disp. inc (std)	-240.066	193.576	433.64***	465.547	-135.296	-600.84***	-297.140	-67.069	230.07
Father high school	0.214	0.227	0.01	0.234	0.239	0.00	0.189	0.263	0.07**
Mother high school	0.241	0.359	0.12***	0.326	0.335	0.01	0.240	0.346	0.11***
Father convicted	0.629	0.554	-0.08***	0.540	0.604	0.06*	0.645	0.538	-0.11***
Mother convicted	0.304	0.273	-0.03	0.246	0.263	0.02	0.338	0.252	-0.09**
Father working	0.500	0.695	0.20***	0.715	0.582	-0.13***	0.440	0.750	0.31***
Mother working	0.505	0.701	0.20***	0.680	0.643	-0.04	0.532	0.690	0.16***
Neighbourhood:									
Share unemployed	0.255	0.124	-0.13***	0.203	0.142	-0.06***	0.226	0.123	-0.10***
Share foreign-born	0.347	0.096	-0.25***	0.229	0.146	-0.08***	0.315	0.086	-0.23***
Share high school deg.	0.699	0.717	0.02***	0.699	0.753	0.05***	0.710	0.722	0.01**
Crimes per 1000	26.638	12.661	-13.98***	20.561	15.711	-4.85***	25.269	13.248	-12.02***
In large city	0.514	0.106	-0.41***	0.305	0.259	-0.05	0.466	0.118	-0.35***
Crime:									
Predicted Pr(Care)	0.429	0.289	-0.14***						
Predicted Pr(Fines)	0.192	0.587	0.40***						
Observations	1,176	1,175	2,351	491	490	981	470	468	938

Notes: CATE = conditional average treatment effect. * p<0.1, ** p<0.05, *** p<0.01.

1.D Calculating Net Costs

I evaluate the total costs and benefits incurred by the 2007 youth sentencing reform as the difference between accumulated earnings net of transfers, and costs associated with crimes. I calculate this on the individual level, in order to use the Causal forest algorithm to understand heterogeneous responses to the reform."Net costs" is thus both reflecting the social costs, and a way to aggregate the sum of negative and positive individual effects on different outcomes.[40] For each person, this is calculated in the following steps:

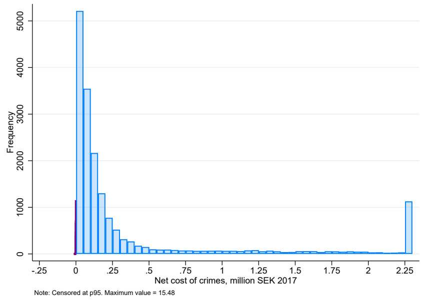
1. **Net transfers (G) =**
   - Taxes on earnings: $\sum_{t=1}^{5} Earnings_t \times 0.3$[41]
   - $-$ Welfare payments: $\sum_{t=1}^{5} FamilyWelfare_t$
2. **Cost of future crimes for** $t \in (0, 5)$ **(**C_1**) =**
   - Cost of police work: \$4,000 per violent crime, \$2,000 per property crime, \$1,300 per crime for other types.
   - Prosecutor and court costs: \$2,150 $\times$ number of crimes.
   - Cost of incarceration (adjudicated, not necessarily served): \$360 $\times \sum_{t=1}^{5} PrisonDays_t$.
3. **Costs of sanction at** $t = 0$ **(**C_2**) =**
   - $\approx$ \$360 if community service
   - $\approx$ \$2,700 if care
   - $\approx$ \$630$\times$ number of days if juvenile prison
   - Average fine: \$160 (subtracted from the cost).
4. **Net costs =** Net transfers - Cost of future crimes - Cost of sanction at $t = 0$
   $= G - C_1 - C_2$

[40] A number of other factors would need to be added for a full calculation of social costs and benefits, such as victimization costs, deadweight loss of taxation, employer surpluses from free labor, opportunity costs of unpaid work, etc. In its current version, this calculation is more reflective of net public transfers.

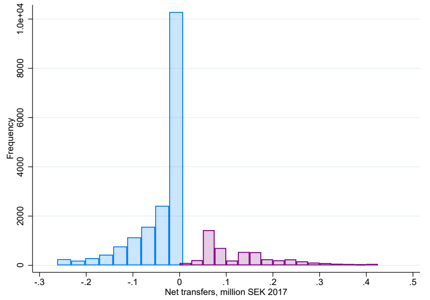
[41] Annual earnings below the income taxation threshold (\$2,000 in 2017) are not included.

The resulting estimates are summarized in Figure 1.D1 below.

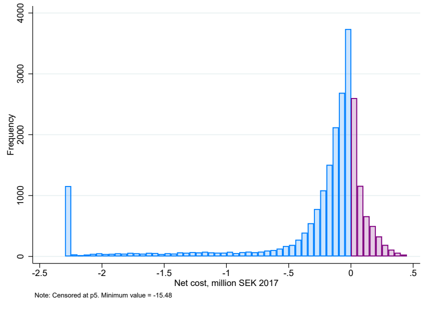
Figure 1.D1: Histograms - Costs of Crime and Net Transfers



(a) Cost of future crimes



(b) Net transfers

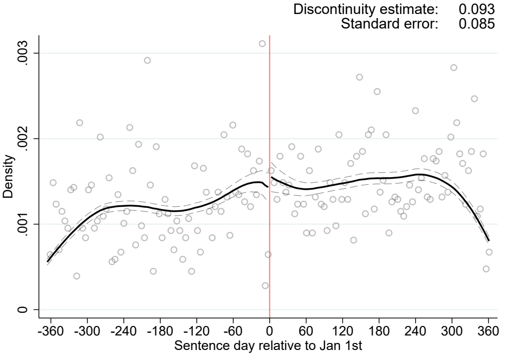


(c) Total: Net transfers - Costs

Note: All recidivism variables — number of future crimes in total and property, violent and other crimes separately, as well as total number of prison days — are top-coded at the 99th percentile. Purple = incomes exceed costs; Blue = costs exceed incomes.

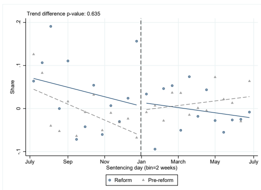
1.E Additional Figures and Tables

Figure 1.E1: Density in Non-trial Convictions.

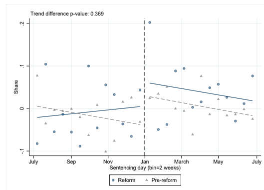


Note: Figure shows the density of non-trial convictions (diversion and direct fines) by day of conviction, between January 1st 2006 and December 31st 2007.

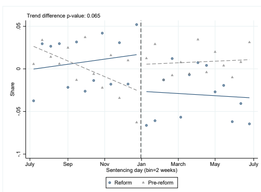
Figure 1.E2: Regression Discontinuity Plots by Predicted Counterfactual Sanction.



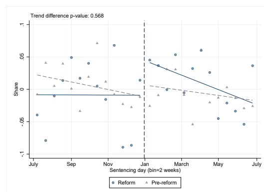
(a) Fines - Recidivism



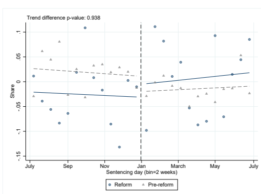
(b) Care - Recidivism



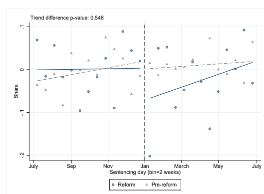
(c) Fines - Prison sentence



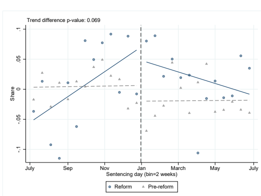
(d) Care - Prison sentence



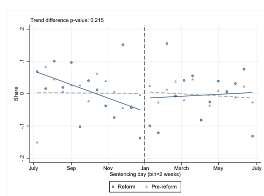
(e) Fines - High school



(f) Care - High school



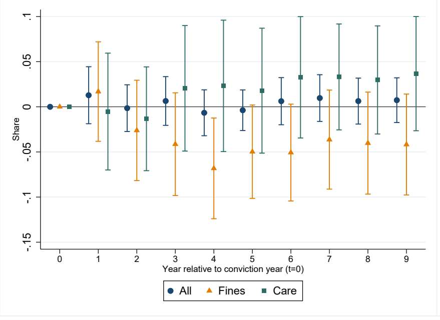
(g) Fines - Employment



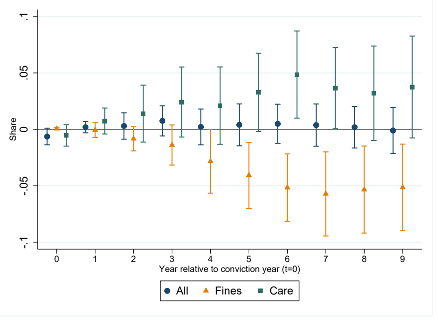
(h) Care - Employment

Note: Each panel shows a binned scatter plot (bin size approx. two weeks) of outcome variable means by conviction day, with fitted linear trends on each side of the cut-off date (Jan 1st 2007 in the reform sample plots, Jan 1st in 2006, 2005 and 2004 in the pooled pre-reform sample). All outcomes measured after 5 years.

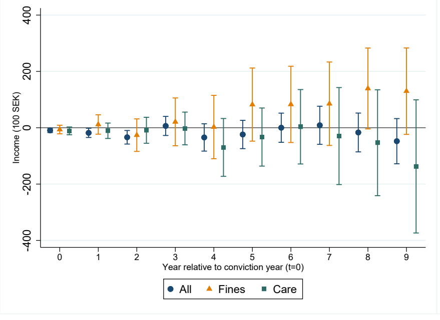
Figure 1.E3: RD-DD Estimates at Different Time Horizons



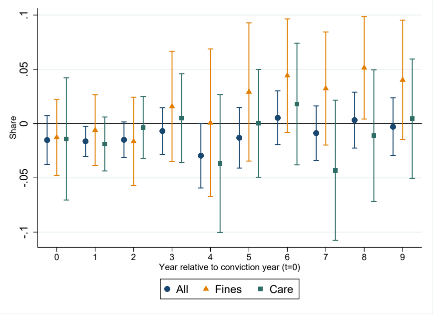
(a) Any new conviction



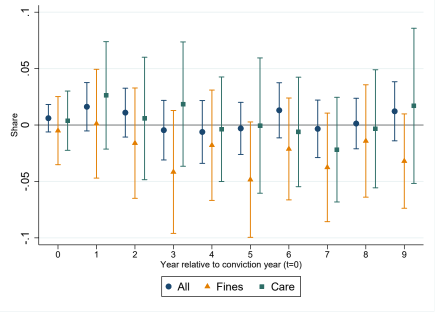
(b) Any prison sentence



(c) Labor income



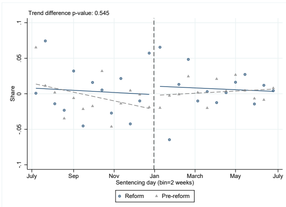
(d) Employment



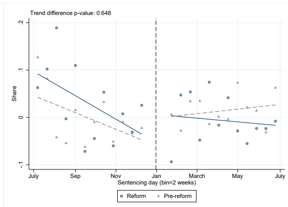
(e) Welfare payments

Note: Each row represents a separate regression estimate (with 95 % CI) from running eq. 1.1 with dependent variables measured as *cumulative risk* at different time intervals from conviction. Panel (a): Any new conviction, Panel (b): Any prison sentence, Panel (c) Average monthly income in 100's of 2017 SEK; in Panel (d) Employment (earnings>5,000 USD); and in Panel (e) Indicator for any individual-level social welfare payments. Linear polynomials of the running variable (conviction day) fitted separately below and above the cut-off date. Court and crime fixed effects and additional control variables are included. Standard errors clustered at court level.

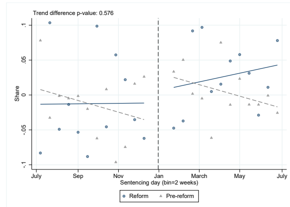
Figure 1.E4: Regression Discontinuity Plots, Donut-RD.



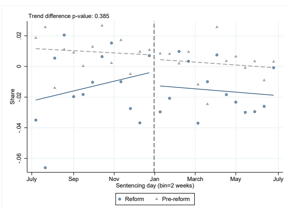
(a) Recidivism



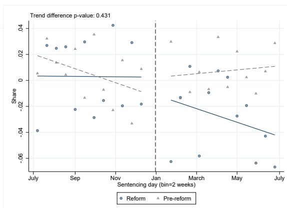
(b) Fines - Recidivism



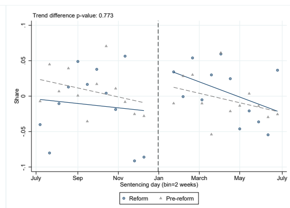
(c) Care - Recidivism



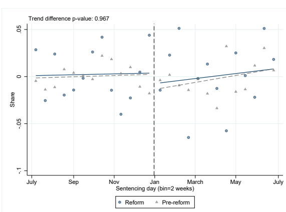
(d) Prison sentence



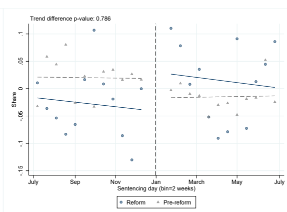
(e) Fines - Prison sentence



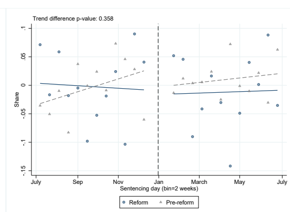
(f) Care - Prison sentence



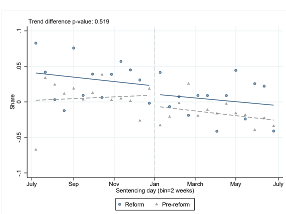
(g) High school graduate



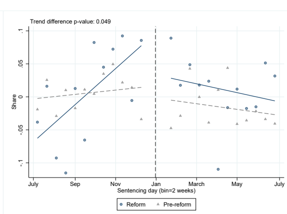
(h) Fines - High school



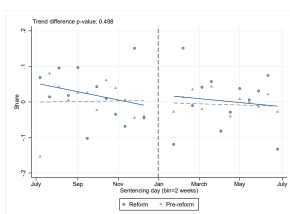
(i) Care - High school



(j) Employment



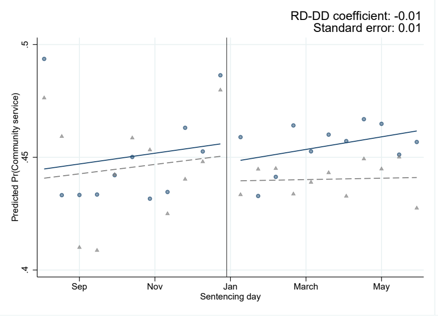
(k) Fines - Employment



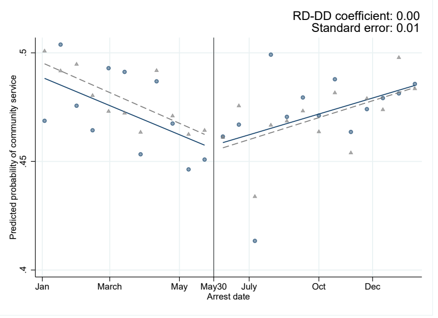
(l) Care - Employment

Note: Each panel shows a binned scatter plot (bin size approx. two weeks) of outcome variable means by conviction day, with linear fitted trends in conviction day on each side of the cut-off date (Jan 1st 2007 in the reform sample plots, Jan 1st in 2006, 2005 and 2004 in the pooled pre-reform sample). Donut: excl. ±2 weeks. Panel (a): Any new conviction. (b): Any prison sentence. (c): High school degree. (d): Employed (employment = earnings>5,000 USD). All outcomes measured after 5 years. In each graph, the first and last week around the new year are excluded.

Figure 1.E5: Regression Discontinuity Plots for Predicted Probability of Community Service.



(a) Date of conviction



(b) Date of arrest

Note: Figure shows the average predicted probability of community service by date of conviction and arrest, respectively. Panel (a): Running variable is date of conviction in two week-bins around the new year. Panel (b): Running variable is date of crime, in two week-bins around reform decision date, May 30th. Predicted probabilities are calculated on observations in 2007 as described in section 1.4.3.

Table 1.E1: Heterogeneity by Predicted Hours of Community Service.

	Income		High school		Recidivism		Prison	
	(1) Low	(2) High	(3) Low	(4) High	(5) Low	(6) High	(7) Low	(8) High
A: All								
Jan-June X Reform	-28.109	-10.830	0.013	-0.007	-0.016	0.016	-0.011	0.025
	(30.665)	(51.137)	(0.021)	(0.020)	(0.017)	(0.018)	(0.012)	(0.016)
Dep. mean	867.49	849.96	0.56	0.55	0.62	0.65	0.13	0.16
Observations	13,462	8,912	13,462	8,912	13,462	8,912	13,462	8,912
B: Fines								
Jan-June X Reform	25.228	425.615*	0.065*	0.130	-0.053*	-0.061	-0.038**	-0.059
	(62.521)	(226.390)	(0.038)	(0.101)	(0.029)	(0.086)	(0.016)	(0.049)
Dep. mean	955.25	1153.12	0.62	0.69	0.52	0.48	0.07	0.06
Observations	4,966	628	4,966	628	4,966	628	4,966	628
C: Care								
Jan-June X Reform	5.415	-91.034	-0.030	-0.031	-0.013	0.063	0.030	0.031
	(92.279)	(68.289)	(0.035)	(0.046)	(0.042)	(0.045)	(0.027)	(0.028)
Dep. mean	814.26	816.12	0.52	0.54	0.68	0.65	0.17	0.13
Observations	3,476	2,118	3,476	2,118	3,476	2,118	3,476	2,118

Notes: Each column shows difference-in-discontinuity estimates (eq. 1.1) for the interaction term between January-June and reform years (2006-07). Number of hours of community service imputed from data for years 2010-2016, from case characteristics. Low: $<$ 35 hours, High: $\geq$ 35 hours. Standard errors clustered at court level in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1.E2: OLS Estimates of Effects of Community Service, Sample July 2006-June 2007.

	Earnings	Employment	HS grad.	Post-sec.	Recidivism	Prison
A: All						
Community service	68.274**	0.034***	0.031**	-0.012	0.000	-0.032***
	(32.815)	(0.012)	(0.014)	(0.012)	(0.016)	(0.007)
Dep. mean	862.94	0.48	0.57	0.24	0.62	0.11
Observations	6,317	6,317	6,317	6,317	6,317	6,317
B: Fines						
Community service	86.078	0.013	0.013	-0.070***	0.037	-0.005
	(67.350)	(0.027)	(0.025)	(0.025)	(0.033)	(0.012)
Dep. mean	1037.56	0.55	0.66	0.27	0.47	0.05
Observations	1,580	1,580	1,580	1,580	1,580	1,580
C: Care						
Community service	138.571*	0.053**	0.046	0.012	-0.033	-0.058***
	(71.420)	(0.024)	(0.030)	(0.023)	(0.026)	(0.020)
Dep. mean	780.59	0.45	0.54	0.20	0.67	0.14
Observations	1,581	1,581	1,581	1,581	1,581	1,581

Notes: OLS Estimates of the effect of begin sentenced to community service (indicator variable). All models include court fixed effects, crime fixed effects, and a full set of additional control variables (described in Section 1.3.1). "Dep. mean" shows control sample mean of dependent variable. Standard errors clustered at court level in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1.E3: Regression Discontinuity Estimates, Sample July 2006-June 2007.

	(1) Earnings	(2) Employment	(3) HS grad.	(4) Post-sec.	(5) Recidivism	(6) Prison
A: All						
Jan-June	-45.108	-0.026	-0.014	0.031	0.002	-0.002
	(53.392)	(0.026)	(0.026)	(0.025)	(0.024)	(0.019)
Dep. mean	941.37	0.50	0.56	0.23	0.63	0.12
Observations	5,876	5,876	5,876	5,876	5,876	5,876
B: Fines						
Jan-June	-15.957	-0.056	0.058	-0.024	-0.003	-0.029
	(122.248)	(0.045)	(0.053)	(0.044)	(0.050)	(0.024)
Dep. mean	1062.97	0.56	0.63	0.26	0.51	0.06
Observations	1,469	1,469	1,469	1,469	1,469	1,469
C: Care						
Jan-June	76.306	0.058	-0.041	0.073*	-0.001	0.054
	(90.661)	(0.050)	(0.045)	(0.044)	(0.054)	(0.036)
Dep. mean	868.86	0.48	0.53	0.21	0.68	0.15
Observations	1,469	1,469	1,469	1,469	1,469	1,469

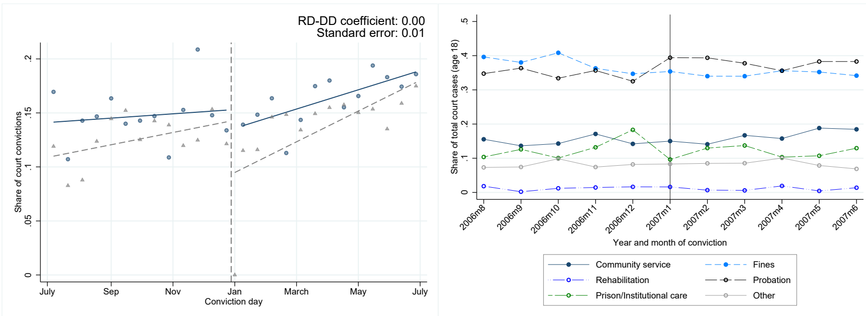
Notes: Each column shows regression discontinuity estimates; coefficients on the effect of being convicted in January-June 2007 (treatment) rather than July-December 2006 (control). All models include court fixed effects, crime fixed effects and a full set of additional control variables. Standard errors clustered at court level in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1.E4: Difference-in-Discontinuity “Donut” Estimates.

	(1) Earnings	(2) Employment	(3) HS grad.	(4) Post-sec.	(5) Recidivism	(6) Prison
A: All						
Jan-June X Reform	-23.873	-0.017	0.010	0.003	-0.000	0.008
	(26.242)	(0.014)	(0.018)	(0.015)	(0.012)	(0.009)
Dep. mean	861.61	0.48	0.55	0.22	0.63	0.14
Observations	21161	21161	21161	21161	21161	21161
B: Fines						
Jan-June X Reform	75.345	0.022	0.073**	-0.013	-0.045	-0.033**
	(69.984)	(0.033)	(0.036)	(0.030)	(0.027)	(0.015)
Dep. mean	981.10	0.53	0.63	0.27	0.51	0.07
Observations	5,294	5,294	5,294	5,294	5,294	5,294
C: Care						
Jan-June X Reform	-32.306	-0.008	-0.024	-0.016	0.018	0.036**
	(57.043)	(0.024)	(0.026)	(0.024)	(0.032)	(0.016)
Dep. mean	818.08	0.46	0.53	0.20	0.67	0.16
Observations	5,291	5,291	5,291	5,291	5,291	5,291

Notes: Each column shows difference-in-discontinuity estimates (eq. 1.1); coefficients on the conviction in January-June main effect and the interaction term between January-June and reform years (2006-07). Convictions from the two last weeks of December and the first two weeks of January are excluded. Standard errors clustered at court level in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Figure 1.E6: Criminal Sanctions for Offenders Aged 18-20.



(a) Probability of community service. (b) Sanctions composition.

Note: Panel a: Share of court convictions given court-ordered community service or *youth* community service, for offenders aged 18-20 when committing a crime, and convicted between July 2003 and June 2007 ($N = 18,557$) by month of conviction, with a fitted line and 95 % CIs on either side of cut-off. Panel b: Share of all court convictions of offenders aged 18-20 given different sanctions, by month of conviction.

Table 1.E5: Placebo Test, Reform Effect on Offenders Aged 18-20.

	(1) Earnings	(2) Empl.	(3) HS grad.	(4) Post-sec.	(5) Recidivism	(6) Prison
Jan-June X Reform	-68.574 (48.571)	-0.023 (0.016)	-0.006 (0.016)	-0.005 (0.014)	0.022 (0.014)	0.025 (0.015)
Jan-June	103.070*** (36.041)	0.017 (0.014)	-0.009 (0.012)	- 0.024** (0.012)	-0.008 (0.013)	-0.010 (0.014)
Dep. mean	1116.02	0.54	0.54	0.20	0.65	0.28
Observations	18,557	18,557	18,557	18,557	18,557	18,557

Notes: Each column shows difference-in-discontinuity estimates (eq. 1.1); coefficients on the conviction in January-June main effect and the interaction term between January-June and reform sample (2006-07), for a sample of offenders aged 18-20 when committing a crime and thus not subject to the reform. * p<0.1, ** p<0.05, *** p<0.01.

Chapter 2

Identity in Court Decision-Making*

*This chapter is written together with Susan Niknami (Swedish Institute for Social Research) and Mårten Palme (Department of Economics, Stockholm University). We thank Pegah Hakuei, Cecilia Palme, Lizette Romero Niknami, Leila Sultan, Felicia Westring and Clara Östling Palme for excellent research assistance, and Johan Steinholtz at Stockholm district court for answering questions about the institutional context.

2.1 Introduction

In recent years there has been a growing awareness in the economic literature of the role of identity and social interaction for economic decisions. Research on e.g. gender pay differences and occupational choice nowadays often take into account that identity and social norms shape a person's choice set in life (Akerlof and Kranton, 2000, Bertrand, 2011). In extension, identity-based decisions of one person might affect other people, a form of "identity externality". A large literature in social psychology and behavioral economics documents so-called "in-group bias", i.e. that humans favor similar others in their decision-making (Tajfel and Turner, 1986, Huffman, Meier and Goette, 2006, Bernhard, Fischbacher and Fehr, 2006, Rand et al., 2009). Examples of real-world group favoritism include gender (Bagues, Sylos-Labini and Zinovyeva, 2017) and ethnicity (Glaeser et al., 2000). However, each of these traits represents only a single dimension along which identity may be shaped, and as such paint an incomplete picture of the role of identity in high-stakes decision-making.

In this paper, we collect what we believe to be the most comprehensive data set to date on identity traits, and investigate the importance of shared identity in the context of criminal court decisions. Our main source of data comes from transcripts of court hearings in the Stockholm District Court for the period 2000-2004. From these documents we extract information about the case, defendant, judge and jurors. We link these data to longitudinal administrative registers for the whole Swedish population dating back to 1985 to get detailed background information on both jurors and defendants. These data allow us to explore similarities along both demographic dimensions, including age, gender and ethnicity, and socioeconomic factors such as education, income and residential area. While many scholars argue that socioeconomic status is one of the most fundamental factors behind shaping identity (Easterbrook, Kuppens and Manstead, 2020), lack of data has prevented previous studies to explore its role in decision-making.[1]

[1] An important exception is Anwar, Bayer and Hjalmarsson (2022), where race and income jointly are used to study bias formation. They are, however, unable to disentangle the two.

Trial by juror is a cornerstone in most justice systems around the world. The idea behind the inclusion of jurors in the decision-making process is to ensure representation, and the juror system thus institutionalizes in-group biases. Ideally, the variation in characteristics between jury members making a joint decision should be sufficiently large to balance out their potential biases. In practice, however, jury groups are either typically quite small, or more homogeneous than society at large, meaning that the presence of in-group bias is likely to systematically affect court decisions.

The detailed individual information on competing identities in our data reflects many of the attributes along which people form group affiliations in everyday life. Our results may thus be valid for decision-making in several other settings, such as the education sector, hiring decisions, and any type of administrative decisions by low-level bureaucrats in the social sector. However, several features make the juror setting especially suitable for investigating how shared social identity affects high-stakes decision-making. First, in contrast to trained profession judges, who are expected to apply the law blindly, jurors may be more likely to let identity shape their decisions. Second, since jurors are ordinary citizens, there is more variation in their background characteristics compared to judges.

The result that jury composition affects court decision-making has been obtained in several previous studies (Anwar, Bayer and Hjalmarsson, 2012, 2019*a*). However, studying one characteristic alone does not capture the full representativity of juries, meaning that the total effect of in-group biases can not be estimated. Previous studies that have taken into account several identity-shaping factors have been limited to gender and ethnicity (Schanzenbach, 2005, Lim, Silveira and Snyder, 2016, Bar and Zussman, 2019), or looked at quite specific contexts such as sports judges (Sandberg, 2018). In this paper, we are able to study the effect of identity along several dimensions, on an outcome of high economic relevance to the adjudged party.

Criminal cases in Sweden are decided by a judge in collaboration with three jurors. At the heart of our research design is the ability to exploit the random assignment of cases to jurors, who are randomly drawn from a pool of eligible jurors in the district. We start by documenting evidence consistent with random assignment of cases to jurors, finding that defendant and juror characteristics are uncorrelated. This random assignment ensures that unobservable characteristics of defendants are the same across jurors. Any systematic differences in sentencing outcomes can therefore be attributed to juror-defendant similarities rather than criminal case selection.

Our results show that identity is an important factor in juror decision-making. Jurors that are randomly assigned to defendants that are more similar to themselves are more lenient in their decision-making. We further find that socioeconomic identity is more important than demographic identity. Holding fixed the variation between individual judges, a defendant is 11 percent less likely to be sentenced to prison if assigned to a juror group with the same socioeconomic background. In particular, sharing educational background is found to reduce the risk of prison, and defendants faced with a juror triplet where all members have the same educational attainments as themselves have a 15 percent lower risk of incarceration, and on average get prison sentences that are half as long.

Sharing the same broadly defined residential area as the jurors does not affect sentences. While gender identity is found to be important, combining the three demographic characteristics gender, ethnicity and age into a demography index produces only a small and imprecisely estimated effect.

We then ask whether identity effects are more prevailing for certain groups within each attribute. The effect of gender identity is mainly a result of women being less prone to sentence other women to prison. Native jurors clearly favor native defendants, while there is no such significant effect on foreign born defendants. Moreover, we provide evidence that the identity effects are only present among defendants who personally attend their court hearing, and that the main results are driven by juror groups paired with a more lenient judge, where they presumably have more say in the sentencing decision. Finally, the type of crime the defendant is accused of matters for the strength of the in-group bias, with the strongest effect found for violent crimes.

Our paper is closely related to the interdisciplinary literature on social identity.[2] These studies are almost exclusively based on

[2]Economics: see e.g. Glaeser et al. (2000), Huffman, Meier and Goette (2006), Charness, Rigotti and Rustichini (2007), Chen and Li (2009), Leider et al. (2009),

laboratory experiments where group identity is either natural (e.g. ethnicity, gender) or artificially created. The experiments demonstrate that people are more lenient towards members of their own group (in-group favoritism), even if the group is artificially created. Previous studies have also shown how in-group biases are stronger if group identity is more salient. Moreover, in-group favoritism is affected by norms and group status; Bernhard, Fischbacher and Fehr (2006) find much higher willingness to punish norm violations if the victim of the violation belongs to the punisher's group, and Tanaka and Camerer (2009) find strong in-group favoritism among poor minorities in Vietnam, while rich minorities and majority ethnic group do not show in-group favoritism when matched to the poor minority.

More recent studies use non-experimental data. Shayo and Zussman (2011) use data on judges and plaintiffs from Israeli claims courts where the assignment of a case to an Arab or Israeli judge is random. They find evidence that a claim was more likely to be accepted if assigned to a judge who was of the same ethnicity as the plaintiff. Bagues, Sylos-Labini and Zinovyeva (2017) analyze how a larger presence of female evaluators affects committee decision-making in academia if the applicant is female. They find no evidence for in-group favoritism among female evaluators. They do however show that male evaluators become less favorable toward female candidates when a female evaluator join the committee. Depew, Eren and Mocan (2017) investigate racial in-group favoritism in juvenile courts and find evidence for negative racial in-group bias, where black (white) juveniles who are randomly assigned to black (white) judges are more likely to get longer sentences. Sandberg (2018) uses data from the Olympic sport of dressage and finds that judges favor athletes of the same ethnicity but not of the same gender. Bar and Zussman (2019) analyze Israeli driving tests, and find Arabic/Jewish in-group favoritism but gender out-group favoritism. While students are more likely to pass the test when the tester is from the same ethnic group, students are less likely to pass if the tester is of the same gender.

Beyond the literature on social identity, this paper is also closely

Rand et al. (2009), Fong and Luttmer (2009). Social Psychology: see survey articles by Tajfel and Turner (1986), McDermott (2009).

related to the literature on discrimination in courts. The first strand examines the impact of defendant ethnicity for trial outcomes (Mustard, 2001, Abrams, Bertrand and Mullainathan, 2012, Arnold, Dobbie and Yang, 2018, McConnell and Rasul, 2021) and the second strand examine whether judge and juror characteristics affect court decisions (Peresie, 2005, Boyd, Epstein and Martin, 2010, Anwar, Bayer and Hjalmarsson, 2012, 2014, 2019*a*,*b*, Glynn and Sen, 2015, Cohen and Yang, 2019). Two recent studies further analyze whether juror decisions are biased by media coverage and find significant effects (Lim, Snyder Jr and Strömberg, 2015, Philippe and Ouss, 2018).

We make several contributions to the literature. First, this paper is unique in that we use high quality register data to study how identity shapes economic outcomes. Second, these data enable us to provide the first causal evidence on the role of identity along multiple dimensions in high-stakes decision-making. While past studies have focused on investigating a single trait, in real life, people identify with multiple groups. Group affiliations may also be overlapping, i.e. ethnicity and socioeconomic status, making it difficult to determine the underlying mechanisms producing effects in a single setting. Third, while many scholars have argued that socioeconomic background is an important attribute for decision-making (see Dal Bó et al. (2017) for an example), it is important to highlight that ours is the first paper to look at socioeconomic status in both the identity and the court context. Fourth, with our identification strategy we determine whether discriminatory behavior among jurors are produced by preferences or statistical discrimination.

The rest of the paper is structured as follows. Section 2.2 provides a brief conceptual framework, followed by institutional details about the Swedish court and juror system in 2.3. Section 2.4 describes our data and Section 2.5.2 our empirical strategy. The results are provided in Section 2.6, and Section 2.7 concludes.

2.2 Conceptual framework

The social context in our setting is a criminal court hearing, where jurors are tasked with deciding the socially optimal punishment for a given crime. A person derives utility ("a sense of justice") in her work as a juror from aligning sanction S with her preferences for sanction strictness P, which are formed from norms regarding her social group. Moreover, the utility from a sentence depends on the identity of the central person to the decision: the defendant. In particular, juror utility is a function of the distance between the juror's social group and the defendant's social group, or otherwise put, the distance in *identity*. This relationship could arise for several reasons, such as the juror wanting to shield members of their own in-group from punishment, or, on the contrary, jurors trying to distance themselves from norm-breaking members of their own group by way of harsher punishment. In addition, different defendant characteristics might be collectively associated by jurors with different norms. For example, women might generally be met with greater compassion in criminal sentencing.

The utility of juror j from deciding sanction S for defendant d can be described by the following expression:

$$U(S)_{jd} = P_j + f(|Identity_j - Identity_d|) + \varepsilon_{jd} \tag{2.1}$$

Since there are three jurors, a decision is formed from the sum of the preferences and distances in identity-shaping attributes among the triplet members. Jurors' basic preferences for sentence strictness remain constant across defendants, and the defendant characteristics are not systematically correlated with jurors. Thus, only the identity distance function varies with the defendant. Moreover, our empirical approach holds the identity and preferences of the judge constant. Importantly, juror triplet homogeneity in itself is controlled for in our setting, but juror preferences may vary depending on who the other members in the group are. Preferences might also differ depending on case characteristics, that could make the defendant's identity more or less salient.

2.3 Institutional setting

In this section, we describe the aspects of the criminal justice system in Sweden that are most relevant for our study. We also describe the lay juror system at the Stockholm district court.

2.3.1 The Swedish criminal justice system

The first step in the justice process after suspicion of a crime, is an investigation undertaken by the police or a prosecutor. After investigation, the prosecutor decides whether the criminal case should advance to a court trial. The criminal court system in Sweden consists of three levels: the district court, the court of appeal, and the supreme court. The vast majority of criminal cases are settled at the district court level, and each district court is generally responsible for all cases originating in its jurisdiction. When a case is taken to court, a computer program at the court randomly assigns the case to a section (rotel).[3] Each section consists of one judge, one clerk, and a number of administrative personnel.

Each district court also maintains a large pool of appointed lay jurors (nämndemän) that serve a similar function as juries in the American or British systems. Lay jurors are appointed to the courts for a four-year term by the municipality councils after being nominated by a political party. The distribution of seats is proportional to the political party representation in the last local election. Lay jurors are not required to be politically active and every Swedish citizen over the age of 18 is eligible for nomination as a juror (with exception for employees within the justice system). Jurors are randomly assigned to criminal cases and each juror works approximately 10 to 15 days per year. The randomization process varies across courts. In the next section, we will describe this process at Stockholm District court.

In most district court trials, both the verdict and sentence are decided jointly by the judge and the three lay jurors. Following the hearing, the judge summarizes the facts of the case and any relevant laws for the three lay jurors. The judge and the three lay jurors then discuss the possible decisions, including the verdict and sentence. If the judge and the lay jurors disagree on the verdict, a vote is held to determine the outcome of the case. The votes of the judge and lay jurors have equal weight, but the judge holds the tiebreaker if there is no clear majority. If a defendant is found guilty, there is a second

[3] The computer program allows for some exceptions; including cases involving youth defendants, the least serious crimes (e.g., traffic offenses), and the most serious crimes (e.g., murder, rape). As a result, the random assignment of cases to judges occurred within age and crime type cells in most district courts.

vote to determine the sentence, with the least severe option chosen if there is an even split between different sentencing options. If the severity ranking of the different options is unclear, then the judge holds the tiebreaker.

2.3.2 Stockholm District Court

Stockholm district court is the largest court in Sweden in terms of criminal caseload. The number of reported crimes in Stockholm in year 2000 was more than 20,000 per 100,000 inhabitants and thus more than four times higher than the national average on 4,670 per 100,000.[4] In 2000-2004, Stockholm district court was divided into eleven divisions. Criminal cases were distributed across four divisions (divisions 11, 12, 13 and 14), while civil cases were reserved for divisions 4, 6, 7, 9 (although these divisions could sometimes also get criminal cases). The court had around 600 lay jurors.

Lay jurors are appointed to the courts for a four-year term after election. Our data spans two election periods; the 1998 election and the 2002 election. After election, the central coordinators at Stockholm court receives lists with names, personal identification numbers and political party of the lay jurors. Coordinators use the list to form triplets with some attempt to balance gender, age, and political party. The jurors within a triplet then work together the upcoming four years. Triplets randomly receives a group number and are thereafter evenly distributed across divisions. Each division received about 50 groups.

Central coordinators then assign groups to different dates by going down the list of triplets in order. The head of each division assigns sections different days of the week for hearings. These two schedules are then merged and given to lay juror coordinators at each division. The schedules are updated each semester. The schedule rotates in the sense that the first triplet group scheduled in Fall, is the one with group number next to the last triplet group scheduled in spring. If a juror is unable to attend on a certain date, for instance due to sickness, the lay juror coordinator at the division calls in the

[4] Source: The Swedish National Council for Crime Prevention, 2022. `https://statistik.bra.se/solwebb/action/index` (2022-03-18).

next lay juror on the list. While we do not have explicit data on absences, one should note that these cannot be correlated with case characteristics, since lay jurors do not get to know the case before arriving to the court.

Figure 1 shows the rotating schedule for division 11 in the fall of 2002. Triplet groups 1-50 were assigned to this division. Triplet group 1, for instance, was scheduled to section 3 every Wednesday in September and to section 9 every Tuesday in December. This implied that group 1 was working with the same judge in a given month, but with different judges across months. Each Section had hearings two days of the week. Section 3, for instance, had hearings on Mondays and Wednesdays.

At the heart of our research strategy is the ability to exploit the random assignment of triplets to cases, since triplets were assigned dates before they knew about the cases to be tried on those dates. This random assignment ensures that unobservable characteristics of cases and defendants are the same across triplets. We will test this in Section 2.5.1.

2.4 Data

To characterize trial outcomes, we assemble what we believe to be the most detailed and comprehensive data set on the topic to date. In this section we briefly summarize data sources, key variables, sample construction and descriptive statistics.

2.4.1 Data sources and sample construction

Our empirical analysis is based on individual-level data from various sources. Our main source consists of transcripts of criminal court decisions in Stockholm District Court for 2000-2004. From these documents we extract information about the defendant(s) and the case. Defendant information includes personal identity number, country of citizenship and residential address. Case information, among other things, includes date of decision, judge name, juror names, number of defendants, the charges on which the defendant(s) was acquitted or convicted, damages requested and awarded, and the sentence. We use lists kept by Stockholm county to obtain juror political affiliations and juror personal identity number.

We link these data to several administrative registers for the whole population. Administrative data on criminal behavior is provided by Swedish National Council for Crime Prevention. The crime data include information on all court cases between 1985 and 2017, including cases that did not end in a conviction. We observe the date of the crime, the date of conviction, the type of crime committed, the sentence imposed by the court, whether there are any co-offenders and unique identifiers for district courts and defendants. Crime outcomes are available from age 15.

Administrative data from Statistics Sweden contain detailed information on family linkages contained in the multi-generation register and background characteristics from LISA register. The multi-generation register contains the personal identification numbers for all individuals born in Sweden starting in 1932, along with the personal identification numbers of each individual's parents and children. These data allow us to identify the children and parents of defendants, judges and jury members.

The LISA data contains rich longitudinal data that includes outcomes for every Swedish resident at least 16 years old from 1990 to 2016. For each year, the data contain demographic and socioeconomic information (e.g., age, country of origin, county of birth, gender, marital status, area of residence, education level, occupation and income measures). We use data on family disposable income to calculate income percentile rank by gender and age cohorts (using a 1 to 100 scale). All values are weighted by the number of family members and deflation adjusted.

We make two key restrictions to our estimation sample. First, we restrict the sample to cases including unique identifiers for judges and jurors. Second, we restrict the main sample to cases with non-missing information on country of origin (less than one percent of observations).

2.4.2 Descriptive statistics

Table 2.1 reports summary statistics for our estimation sample. Panel A presents characteristics of the lay juror triples. In our triplets, 46 % are females and the average age is just under 60 years old. More than 89 % are Swedish born, almost six percent are born in other Nordic or western countries, close to four percent are born in the Middle East and just under one percent are born in Africa. Education levels are relatively high among lay juror triplets: 55 % have some form of post-secondary education and 34 % are high school graduates. Income levels are also high among lay jurors: almost 60 % have incomes above the 75th percentile in the income distribution, and less than ten percent have incomes below the 25th percentile. Almost seven percent of the jurors have been convicted themselves in the past. Finally, almost 43 % in our triplets belong to the left-wing of the political spectrum and over 50 % are appointed by right-wing parties.

Panel B present the characteristics of defendants. Over 85 % of the defendants are male and the average age is just over 35 years. In terms of both gender and age, defendants are thus very different from jurors. While almost 36 % of defendants are foreign born, only four percent are not Swedish citizens. Around twelve percent are born in other Nordic or western countries, twelve percent in the Middle East, six percent in Africa, five percent in Latin America and two percent in Asia. In contrast to jurors, education levels are low among defendants. In our sample, 46 % of defendants have less than high school education, 42 % a high school education and twelve percent some post-secondary education. Income levels are equally low, with 50 % of incomes below the 25th percentile in the income distribution and only 14 above the 75th percentile. Most defendants, 75 %, are previously convicted, and 32 % of the defendants have a prior prison sentence.

Panel C presents summary statistics of the criminal cases. The most common offences are violent crimes (26 %), property crimes (26 %), traffic violations (14 %), narcotic crimes (12 %) and drunk driving (8 %). In 8 % of the criminal cases, defendant admits guilt, and in 17 % of the cases, crimes are committed with at least one co-offender. Eight percent of defendants have a pre-trial detention. Table 2.1 further shows that 94 % of the charges result in conviction, and 23 % in a prison sentence. Average prison sentence length is about two months in the overall sample and about seven months among incarcerated defendants.

Table 2.1: Descriptive Statistics.

	Count	Mean	SD
Lay juror Triplet:			
Average age	9,999	59.586	6.775
Male jurors	9,999	0.430	0.162
Born in Sweden	9,999	0.891	0.183
Born in Other Nordic	9,999	0.034	0.107
Born in Other Western	9,999	0.025	0.091
Born in Middle East	9,999	0.036	0.108
Born in Africa	9,999	0.009	0.055
Born in Asia	9,999	0.001	0.019
Born in Latin America	9,999	0.003	0.033
Less than high school	9,999	0.112	0.182
High school degree	9,999	0.342	0.275
Post-secondary education	9,999	0.545	0.293
Low income rank	9,999	0.081	0.156
Middle income rank	9,999	0.331	0.276
High income rank	9,999	0.588	0.286
Ever convicted	9,999	0.071	0.143
Left-wing party	9,999	0.427	0.256
Right-wing party	9,999	0.506	0.256
Defendant:			
Age at trial	9,999	35.352	12.687
Defendant male	9,999	0.853	0.354
Birth country: Africa	9,999	0.055	0.227
Birth country: Asia	9,999	0.015	0.120
Birth country: Latin America	9,999	0.048	0.215
Birth country: Middle East	9,999	0.119	0.324
Birth country: Other Nordic	9,999	0.052	0.222
Birth country: Sweden	9,999	0.635	0.481
Birth country: Other Western	9,999	0.077	0.266
Education: Less than HS	9,074	0.458	0.498
Education: High school	9,074	0.418	0.493
Education: Post-secondary	9,074	0.124	0.330
Disp. income rank < 25	9,620	0.499	0.500
Disp. income rank mid-50	9,620	0.365	0.482
Disp. income rank ≥ 75	9,620	0.136	0.343
Ever convicted	9,999	0.741	0.438
Ever prison sentence	9,999	0.320	0.466
Foreign citizen	9,929	0.036	0.185
Court case:			

Table 2.1: Descriptive Statistics (continued).

	Count	Mean	SD
Violent crime	9,999	0.262	0.440
Property crime	9,999	0.260	0.439
Drunk driving	9,999	0.081	0.273
Traffic offence	9,999	0.140	0.347
Narcotics crime	9,999	0.122	0.327
Other crime	9,999	0.106	0.308
Crime not classified	9,999	0.030	0.169
Admits guilt	9,832	0.069	0.254
Pre-trial detention	9,930	0.080	0.271
Co-offender	9,999	0.173	0.378
Guilty verdict	9,999	0.940	0.237
Prison sentence	9,999	0.232	0.422
Sentence length (months)	9,999	1.713	6.870

Notes: Juror triplet characteristics refer to share of triplet with the respective characteristic. SD = standard deviation.

2.5 Research Design

2.5.1 Random assignment

Our empirical strategy rests on the fact that jurors are randomly assigned to criminal cases. Evidence that criminal cases are randomly assigned to lay jurors is provided in Table 2.2, where observable characteristics of the jurors are regressed on a large vector of defendant and case characteristics. We look at the following five juror triplet characteristics: share of male jurors, average age, share born in Sweden, average years of education and average income rank. We run a stacked regression model, where the five outcome variables are fully interacted with the set of defendant and case characteristics, controlling for a full set of year by court department and judge fixed effects. We then perform a F-test of joint significance on all coefficients, which does not reject the null hypothesis of no systematic selection of jurors to cases on defendant characteristics.

Table 2.2: Randomization of Lay Judges to Criminal Cases.

	Coefficient	SE
Male	-0.015	(0.014)
Age at crime	0.001**	(0.000)
Age 18 or above	-0.008	(0.029)
Age 21 or above	-0.039*	(0.021)
Foreign born	-0.012	(0.009)
Foreign citizen	0.031	(0.022)
Less than high school	-0.008	(0.011)
Post-secondary education	-0.023	(0.016)
Disposable income rank	0.000	(0.000)
Employment	-0.004	(0.013)
Welfare payments	0.003	(0.009)
Non-single	0.007	(0.017)
Number children in house	-0.003	(0.014)
Missing in t-1	-0.008	(0.028)
Convicted last 3 years	-0.001	(0.010)
Prison sentence last 3 years	-0.003	(0.014)
Crime confessed	-0.029	(0.018)
Violent crime	-0.007	(0.018)
Property crime	0.006	(0.017)
Traffic offence	-0.026	(0.019)
Narcotics crime	-0.014	(0.022)
Economic crime	-0.013	(0.022)
Observations	49,995	
Joint F-test p-value	0.298	

Notes: The stacked dependent variable contains five characteristics of the juror triplets: share male, average age, share born in Sweden, average years of education and average family disposable income rank. The model includes year-by-division and judge fixed effects. Standard errors clustered at judge level in parentheses in column 2.

2.5.2 Empirical model

Consider a simplified court setting, where a single juror decides the outcome, and where identity is formed along a single characteristic k that can take on several different values: $k_i \in k_1, k_2, ..., k_n$. Our interest in this paper is to understand the effect on decisions, from similarity in k between juror j and defendant i, controlling for the main effects of that same characteristic when observed in the defendant and the juror, respectively:

$$\text{Decision}_{ij} = \alpha_1 k_i + \alpha_2 k_j + \alpha_3 I[k_i = k_j] + \epsilon_{ij}. \tag{2.2}$$

Here, α_3 would capture the effect of sharing the same observable characteristic, or *shared identity*, on the court outcome. In practice, however, there are three jurors and multiple observable attributes along which social groups are formed. We construct a variable indicating how many of the jurors that share a given characteristic k (for example, gender) with the defendant:

$$\text{Identity}^k = \frac{\sum_{j=1}^{3} 1[k_i = k_j]}{3} \tag{2.3}$$

The simplest example is the variable "gender identity", which measures the share of males (females) in the juror triplet, if the defendant is male (female). It takes the value zero if the defendant is a male (female) and all three jurors are females (males) and it is one if all jurors and the defendant are males (females).

We look at similarity in six attributes: gender, country of birth, age, education, income and neighborhood of residence. Gender and ethnicity are known to divide people into social groups and are thus obvious choices (Akerlof and Kranton, 2000). In recent years, the concept of "ageism" has emerged in the collective awareness, denoting discrimination of older people in e.g. work places (Ahmed, Andersson and Hammarstedt, 2012, Neumark, Burn and Button, 2016). Social psychologists have also identified age as an identity-shaping factor, although less prominent than e.g. gender and race (McNamara et al., 2016). This motivates us to explore whether jurors identify with their own age group in sentencing decisions. We use education and income as two indicators of socioeconomic status — a well-known factor for social divides. Our motivation for including residential areas is that it too reflects socioeconomic status. In Stockholm, like in many urban areas, districts are segregated with respect to income of the residents. Moreover, the choice of one area over another of similar status and price level might contain additional information about the social group that a person identifies herself with. Table 2.3 lists the groups within each characteristic used to define "similarity" in the baseline specification. In Section 2.6 below, we discuss alternative ways to define these groups.

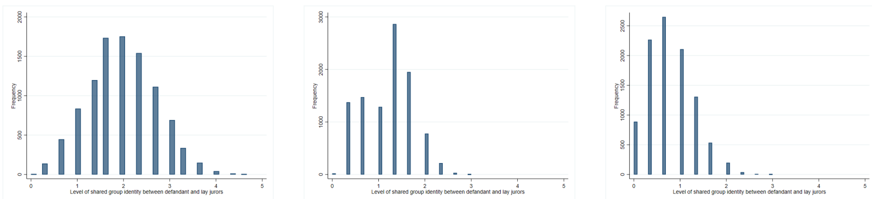
Table 2.3: Identity-Shaping Observable Characteristics

Gender	Man, Woman
Country of birth	Sweden, Other Nordic, Other European, Other Western, Middle East, North Africa, Sub-Saharan Africa, Latin America, Asia.
Age	Indicator for age difference between a juror and defendant less than ± 10 years.
Education	Less than high school, High school degree, Post-secondary education.
Income*	Less than 25th percentile, Middle 50 percentiles, Above 75th percentile.
Neighborhood**	Below region median income, above region median income.

Notes: *Income is family disposable annual income in year before trial date. **Neighborhoods defined as districts within Stockholm municipality (e.g. Kungsholmen, Södermalm, Farsta) or municipalities in the Stockholm region (e.g. Solna, Sollentuna, Botkyrka); in total 42 categories. "Region median income" refers to Stockholm region median income in years 2000-2004.

As a second step, we combine the identity variables into an index measure for demographic similarities, an index measure for socioeconomic similarities, and an index measure for similarities in all traits. We do this by simply summing over the identity variables. The first two indices take a value between zero (no shared characteristics between the defendant and the lay jurors) and three (the defendant belongs to the same group as all three lay jurors, across all three characteristics) in steps of one third. The overall index instead takes a value between zero and six. Figure 2.1 shows there is substantial variation in the similarity indices. The distribution of the overall identity index is approximately bell-shaped with a slightly longer right tail. It takes on a maximum value below five, indicating that no set of juror triplets exactly match a defendant in group affiliations. Summary statistics of all identity variables are shown in Table 2.A1.

Figure 2.1: Distribution of Identity Indices



(a) All characteristics (b) Demographics (c) Socioeconomic

Notes: Index in Panel A constructed from gender, ethnicity, age, education, income and neighborhood. Panel B: gender, ethnicity and age. Panel C: education, income and neighborhood.

We estimate the following regression model:

$$Y_{ijgtd} = \alpha_{td} + \gamma_g + \delta \text{Identity}_{ij}^{k} + \mathbf{X}_{it}\beta_1 + \mathbf{Z}_{jt}\beta_2 + \varepsilon_{ijgtd}, \quad (2.4)$$

where Y_{ijgtd} represents the outcome for defendant i sentenced by juror triplet j and judge g in year t and court division d. Our outcomes of interest are whether or not the defendant is sentenced to prison and the sentence length. The terms α_{td} and γ_g denote year-by-division and judge fixed effects, respectively. $\mathbf{X}_{it}$ is a vector of defendant and case controls, including gender, age fixed effects (in 5-year groups), native born, citizenship status, education, income rank, neighborhood of residence, and crime fixed effects. $\mathbf{Z}_{jt}$ comprises a vector of juror triplet controls, including gender, age, native born, education, income and neighborhood of residence. $Identity_{ij}^{k}$ is either of the identity variables (measuring the extent to which observable characteristic k is shared between the defendant and the jurors) and coefficient δ thus captures the effect of shared identity on trial outcomes.

2.6 Results

2.6.1 Identity effects in sentencing decisions

Table 2.4 shows results from estimating equation 2.4, with a binary outcome of whether or not the defendant is sentenced to prison in Column 1, and the prison sentence length in months as the dependent variable in Column 2. Each row represents a separate regression, and all models include year-by-division fixed effects, judge fixed effects and defendant and juror triplet controls. To account for the exploratory nature of the analysis, we show Romano-Wolf stepdown adjusted *p*-values in brackets for each index component.[5] Standard errors are clustered at the judge level. Panel A presents results for demographic characteristics and Panel B shows the results for socioeconomic characteristics. In Panel C, all characteristics are summarized into an overall index.

Beginning with demographic characteristics, gender is an obvious candidate. Our results show that more lay jurors of the same sex as the defendant significantly reduces the risk of incarceration. Specifically, a given defendant faced with a juror triplet who are all of the same sex as (s)he is 6.4 percentage points less likely to end up in prison, compared to if all jurors were of opposite sex. This amounts to a reduction by 27 %, indicating the presence of a strong in-group bias in gender identity. A closer look at this result reveals that this effect is mainly a result of women being less prone to sentence other women to prison. In Table 2.5, Panel A, we split the gender identity-effect by gender of the defendant, to show that the point estimate for women is almost four times as strong as that of men. In fact, an all-female triplet renders the risk of incarceration close to zero for women. The corresponding effect for men is weaker and not statistically significant.

The effect of similarity in terms of ethnic origin on judicial decision-making is less clear. Our results show a weak and imprecisely estimated negative effect of ethnic identity on prison sentences. Splitting the sample by native and non-native origin of the defendant, we do find a strong and significant effect among native-born Swedes;

[5]This multiple testing correction is done within each group of characteristics.

Swedish-born jurors are more lenient toward Swedish defendants, while jurors with a non-native origin do not display the same bias.

We do not find any evidence that similarity in age sways sentencing outcomes in either a positive or negative direction. The coefficient on age is an imprecisely estimated zero. Splitting the sample into younger and older defendants (below age 30 vs. age 30 and above), reveals that younger jurors tend to be more lenient towards younger defendants, but that the same is not true for the older age group. When summarizing gender, ethnicity and age into an index, we find that demographic characteristics together have a negative but imprecisely estimated effect on the incarceration decision ($\delta = -0.022$, se=0.014).

Panels B displays the results of three characteristics which reflect socioeconomic status: level of education, family disposable income and residential area. Our results show that education creates a bias, while we can not decisively say the same for the two other characteristics. Defendants faced with a juror triplet where all members have the same educational attainments as themselves have a 15 % lower risk of incarceration, compared to an all-different triplet. Panel D in Table 2.5 shows that this effect does not originate from any one particular educational group. Point estimates are not statistically different from one another between defendants with less than high school, a high school degree or post-secondary education. Jurors are not more prone to give preferential treatment to defendants of similar income level, or to defendants residing in similar parts of town as them. As shown in Panels E-F of Table 2.5, the income identity effect does not differ across the income distribution of defendants, but a stronger negative neighborhood effect is found among defendants from areas in Stockholm with below-median income.

Table 2.4: Similar Observable Identity and Prison Sentences.

	(1) Prison (0/1)	(2) Months prison
Panel A: Demographic characteristics		
Gender	-0.068**	-0.470
	(0.031)	(0.385)
	[0.008]	[0.100]
Ethnicity	-0.013	-0.331
	(0.031)	(0.573)
	[0.610]	[0.462]
Age	-0.006	-0.173
	(0.022)	(0.394)
	[0.721]	[0.562]
Demographic identity index	-0.022	-0.278
	(0.014)	(0.267)
Panel B: Socioeconomic characteristics		
Education	-0.035*	-1.141***
	(0.018)	(0.432)
	[0.008]	[0.004]
Income	-0.014	0.163
	(0.016)	(0.280)
	[0.195]	[0.434]
Residential area	-0.033	0.253
	(0.028)	(0.438)
	[0.080]	[0.355]
Socioeconomic identity index	-0.026***	-0.318
	(0.009)	(0.227)
Panel C: All		
Identity index	-0.025***	-0.308
	(0.008)	(0.187)
Observations	9,999	9,999

Notes: Dependent variables are: indicator for prison sentence in Column 1, prison sentence length in months in Column 2. All specifications contain year times court division fixed effects, judge fixed effects, and control variables. Defendant controls: male, age at trial fixed effects (5-year groups), born outside Sweden, foreign citizenship, level of education, income rank (low/high) and a set of 20 dummy variables for type of crime. Juror controls: triplet mean age, share male, share native-born, share with less than high school and high school education, share with low and middle disposable income. Standard errors clustered at judge level in parentheses. Romano-Wolf stepdown adjusted p-values in brackets.

The index constructed from these three variables together is however found to substantially reduce the risk of prison ($\delta = 0.026$, se=0.009). Taken at face value, our results thus show that socioeconomic identity affects judicial decisions more than demographic identity. When we summarize all six measures of identity into a single index, the estimated effect of identity on prison sentences is a precisely estimated reduction by 9 % compared to mean incarceration rate ($\delta = 0.025$, se=0.008). In Appendix Table 2.B1, we show that these results are robust to changing the treatment of missing demographic information (excluding missing values on education and income as opposed to assigning them to the "low" category). Table 2.B2 shows that the results remain virtually unchanged when entering all indices, or all index components, into the same regression model, which suggests that each identity variable holds important information about social groups. Tables 2.B3-2.B6 show how results change when changing the definitions of sub-categories within the characteristics education, income, age and residential area. In general, these tables convey that identity effects arise in our data when social groups are defined not too broadly, but also not too fine-grained.

Column 2 of Table 2.4 shows the same model specifications, but with prison sentence length as the outcome variable.[6] Here, none of the demographic characteristics are found to affect the sentencing outcome; the estimated effects are imprecisely estimated and vary in sign. However, socioeconomic indicators — and in particular education — does create a bias. Defendants faced with a juror triplet of all the same level of education as themselves get shorter prison sentences by about 1.1 months, which is a reduction by 65 % compared to the mean prison sentence length. Appendix Table 2.A2 further shows estimated identity effects on the conviction margin, which are all small and imprecisely estimated for the three indices.

[6] In a sense, this outcome variable captures both sanction type and sentence length, since it takes the value zero for individuals who are not given a prison sentence.

Table 2.5: Heterogeneous Effects of Identity-Shaping Characteristics.

A. Gender:	Men	Women	
Gender	-0.025	-0.138**	
	(0.029)	(0.058)	
Observations	8,521	1,420	
B. Country of birth:	Foreign-born	Native-born	
Ethnicity	-0.042	-0.085***	
	(0.084)	(0.030)	
Observations	3,617	6,332	
C. Age:	Young	Older	
Age	-0.067	0.001	
	(0.061)	(0.022)	
Observations	4,047	5,887	
D. Education:	Low	Middle	High
Level of Education	0.046	-0.022	0.150
	(0.039)	(0.027)	(0.192)
Observations	5,055	3,748	1,069
E. Income rank:	Low	Middle	High
Disposable income rank	0.008	-0.011	-0.002
	(0.036)	(0.020)	(0.037)
Observations	5,145	3,472	1,256
F. Neighborhood:	Low incoke	High income	
Neighborhood	-0.029	-0.004	
	(0.025)	(0.033)	
Observations	5,613	4,322	

Notes: Dependent variable is an indicator for prison sentence. All specifications contain year times court division fixed effects, and judge fixed effects. Standard errors clustered at judge level in parentheses. Control variables as described in table 2.4.

2.6.2 Heterogeneity by defendant and case characteristics

In Table 2.6, we show how identity effects vary across subgroups with similar court trial and case attributes. First, we ask if it matters whether the defendant has confessed to the crime or not. Panel A, Column 1 shows confessed crimes, and Column 2 shows non-confessed crimes, which make up the vast majority. Evidently, the main results presented above are driven by non-confessed crimes. The overall identity index, as well as the two indices for demographic and socioeconomic identity, all affect the risk of prison sentences negatively among this group. Here, the effects of similarity in demographic traits and in socioeconomic status are found to be of similar magnitude.

Next, we look separately at cases where identity is arguably more salient, namely trials where the defendant is present in court. As would be expected, the point estimates in Column 3 (defendant present) are similar to the main results, while the estimated effects of identity in Column 4 (not present) are close to zero. This indicates that similarities between jurors and defendants only influence jurors' sentencing preferences when they can actually observe the defendant in person.

Columns 5-6 attempt to discern what role the influence of jurors play in relation to the judge. For example, is might be that juror composition effectively counterbalances a judge with a preference for stricter sentences, and thus upholds some balance in court outcomes. To study this, we create a canonical measure of judge strictness (Kling, 2006, Aizer and Doyle, 2015, Bhuller et al., 2020, Dobbie et al., 2018), and divide the judges in our sample into two groups: Judges with a below-median propensity of incarceration ("mild") and judges with an above-median incarceration rate ("strict"). Our results show that a favourable juror composition from the perspective of the defendant renders sentencing outcomes *even more* mild, beyond the effect of being allocated to a less strict judge. Interestingly, however, socioeconomic identity influences the court decision even when in the presence of strict judges; point estimates for socioeconomic identity are almost identical between mild and strict judges.

In Panel B, we look at identity effects by type of crime. Column 1 displays the main result for all crime types, and Column 2-6 looks at, in turn, violent crimes, property crimes, drug crimes (use and distribution), drunk driving and all other types together (including economic crimes and traffic offences). Evidently, the main effect of juror identity is driven by violent crime cases. For defendants accused of a violent crime, each unit increase in the identity index results in a reduced risk of prison by 20 %. For property and drug crimes, similar effect are found, but these are imprecisely estimated. Interestingly, the “other crimes” category displays a zero effect, indicating that jurors are not swayed by in-group bias in these milder forms of crimes.

Appendix Table 2.A3 shows heterogeneous effects of the three identity indices, for gender, ethnic background and educational subgroups. Column 1 displays the main result for the three identity indices (demography, socioeconomic and overall). In Columns 2-3, we split the sample by gender of the defendant. This shows that, while the gender identity effect above was found to be concentrated among women, the negative treatment effect of composite identity on prison sentences is found solely for men. Women are thus less subject to identity effects than men, or even subject to a negative bias along some dimensions.

Table 2.6: Effects of Identity on Prison Sentences, by Case Characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Court case	Confessed	Not confessed	Present	Not present	Mild judge	Strict judge
Identity	-0.001	-0.029***	-0.022**	0.002	-0.033**	-0.015
	(0.007)	(0.008)	(0.009)	(0.019)	(0.013)	(0.011)
Demographic identity index	-0.003	-0.032**	-0.019	0.004	-0.051***	0.007
	(0.004)	(0.015)	(0.019)	(0.030)	(0.019)	(0.021)
Socioeconomic identity index	-0.000	-0.027***	-0.024**	0.000	-0.023	-0.025**
	(0.010)	(0.010)	(0.012)	(0.027)	(0.015)	(0.012)
Observations	627	9,136	6,862	1,219	5,032	4,967
Panel B: Crime type	All	Violent	Property	Drugs	Drunk driving	Other
Identity	-0.025***	-0.046**	-0.019	-0.032	0.006	-0.001
	(0.008)	(0.019)	(0.020)	(0.033)	(0.038)	(0.011)
Demographic identity index	-0.022	-0.053	-0.000	-0.086	-0.008	-0.005
	(0.014)	(0.032)	(0.030)	(0.061)	(0.069)	(0.022)
Socioeconomic identity index	-0.026***	-0.041	-0.028	-0.008	0.012	0.000
	(0.009)	(0.026)	(0.024)	(0.045)	(0.047)	(0.013)
Observations	9,999	2,580	2,565	1,156	772	2,713

Notes: Dependent variable is an indicator for prison sentence. All specifications contain year times court division fixed effects, and judge fixed effects. Standard errors clustered at judge level in parentheses. Control variables as described in table 2.4.

Dividing the sample into native-born and foreign-born defendants, in Columns 4-5, we find that the bulk of the identity effects is found among defendants born outside of Sweden. However, point estimates remain negative also for natives. Finally, Columns 6-8 split the sample by level of education, and show that identity effects are similar across these three groups.

2.7 Conclusions

Designing judicial systems that ensures unbiased court decisions is a key objective in all open societies around the world. In this study we have provided strong evidence that identity of the jury members in relation to the defendant affects court outcomes. The magnitudes of these biases are far from negligible. For example, defendants are 15 percent less likely to get a prison sentence if they have the same education level as all three lay jurors assigned to their case, compared to if none of them have the same educational attainments.

We can also conclude that previous studies that restrict the analysis to only one identity - such as ethnicity or gender - may underestimate the importance of the multidimensional nature of how people form identities in decision-making. Our results suggest that socioeconomic identities are more important than demographic factors, including ethnicity, which have been more extensively studied in the previous literature within this field of research. In extension, this suggests that similar effects would be found in other economically relevant decisions where the decision-maker and the subject interact.

Bibliography

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" *The Journal of Legal Studies*, 41(2): 347–383.

Ahmed, Ali M., Lina Andersson, and Mats Hammarstedt. 2012. "Does age matter for Employability? A field experiment on ageism in the Swedish labour market." *Applied Economics Letters*, 19(4): 403–406.

Aizer, Anna, and Joseph J. Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *The Quarterly Journal of Economics*, 130(2): 759–803.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of Economics*, 115(3): 715–753.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *The Quarterly Journal of Economics*, 127(2): 1017–1055.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2014. "The Role of Age in Jury Selection and Trial Outcomes." *The Journal of Law & Economics*, 57(4): 1001–1030.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2019*a*. "A Jury of Her Peers: The Impact of the First Female Jurors on Criminal Convictions." *The Economic Journal*, 129.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2019*b*. "Politics in the courtroom: Political ideology and jury decision making." *Journal of the European Economic Association*, 17(3): 834–875.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2022. "Unequal Jury Representation and Its Consequences (forthcoming)." *American Economic Review: Insights*.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review*, 107(4): 1207–38.

Bar, Revital, and Asaf Zussman. 2019. "Identity and Bias: Insights from Driving Tests." *The Economic Journal*, 130(625): 1–23.

Bernhard, Helen, Urs Fischbacher, and Ernst Fehr. 2006. "Parochial Altruism in Humans." *Nature*, 442: 912–5.

Bertrand, Marianne. 2011. "Chapter 17 - New Perspectives on Gender." In . Vol. 4 of *Handbook of Labor Economics*, , ed. David Card and Orley Ashenfelter, 1543–1590. Elsevier.

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

Boyd, Christina L., Lee Epstein, and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science*, 54(2): 389–411.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review*, 97(4): 1340–1352.

Chen, Yan, and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review*, 99(1): 431–57.

Cohen, Alma, and Crystal S. Yang. 2019. "Judicial Politics and Sentencing Decisions." *American Economic Journal: Economic Policy*, 11(1): 160–91.

Dal Bó, Ernesto, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne. 2017. "Who becomes a politician?" *The Quarterly Journal of Economics*, 132(4): 1877–1914.

Depew, Briggs, Ozkan Eren, and Naci Mocan. 2017. "Judges, Juveniles, and In-Group Bias." *The Journal of Law and Economics*, 60(2): 209–239.

Dobbie, Will, Hans Grönqvist, Susan Niknami, Mårten Palme, and Mikael Priks. 2018. "The Intergenerational Effects of Parental Incarceration." National Bureau of Economic Research Working Paper 24186.

Easterbrook, Matthew J., Toon Kuppens, and Antony S. R. Manstead. 2020. "Socioeconomic status and the structure of the self-concept." *British Journal of Social Psychology*, 59(1): 66–86.

Fong, Christina M., and Erzo F.P. Luttmer. 2009. "What determines giving to hurricane Katrina victims? Experimental evidence on racial group loyalty." *American Economic Journal: Applied Economics*, 1(2): 64–87.

Glaeser, Edward L., David I. Laibson, José A. Scheinkman, and Christine L. Soutter. 2000. "Measuring Trust." *The Quarterly Journal of Economics*, 115(3): 811–846.

Glynn, Adam N., and Maya Sen. 2015. "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?" *American Journal of Political Science*, 59(1): 37–54.

Huffman, David, Stephan Meier, and Lorenz Goette. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." *American Economic Review*, 96: 212–216.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review*, 96(3): 863–876.

Leider, Stephen, Markus M. Möbius, Tanya Rosenblat, and Quoc-Anh Do. 2009. "Directed Altruism and Enforced Reciprocity in Social Networks." *The Quarterly Journal of Economics*, 124(4): 1815–1851.

Lim, Claire S.H., Bernardo S. Silveira, and James M. Snyder. 2016. "Do Judges' Characteristics Matter? Ethnicity, Gender, and Partisanship in Texas State Trial Courts." *American Law and Economics Review*, 18(2): 302–357.

Lim, Claire SH, James M Snyder Jr, and David Strömberg. 2015. "The judge, the politician, and the press: newspaper coverage and criminal sentencing across electoral systems." *American Economic Journal: Applied Economics*, 7(4): 103–35.

McConnell, Brendon, and Imran Rasul. 2021. "Contagious Animosity in the Field: Evidence from the Federal Criminal Justice System." *Journal of Labor Economics*, 39(3): 739–785.

McDermott, Rose. 2009. "Psychological approaches to identity: Experimentation and application." *Measuring identity: A guide for social scientists*, 345–367.

McNamara, Tay K., Marcie Pitt-Catsouphes, Natasha Sarkisian, Elyssa Besen, and Miwako Kidahashi. 2016. "Age Bias in the Workplace: Cultural Stereotypes and In-Group Favoritism." *The International Journal of Aging and Human Development*, 83(2): 156–183. PMID: 27199491.

Mustard, David B. 2001. "Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts." *The Journal of Law & Economics*, 44(1): 285–314.

Neumark, David, Ian Burn, and Patrick Button. 2016. "Experimental Age Discrimination Evidence and the Heckman Critique." *American Economic Review*, 106(5): 303–08.

Peresie, Jennifer L. 2005. "Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts." *The Yale Law Journal*, 114(7): 1759–1790.

Philippe, Arnaud, and Aurélie Ouss. 2018. ""No hatred or malice, fear or affection": media and sentencing." *Journal of Political Economy*, 126(5): 2134–2178.

Rand, David G., Thomas Pfeiffer, Anna Dreber, Rachel W. Sheketoff, Nils C. Wernerfelt, and Yochai Benkler. 2009.

"Dynamic remodeling of in-group bias during the 2008 presidential election." *Proceedings of the National Academy of Sciences*, 106(15): 6187–6191.

Sandberg, Anna. 2018. "Competing Identities: A Field Study of In-group Bias Among Professional Evaluators." *Economic Journal*, 128(613): 2131–2159.

Schanzenbach, Max. 2005. "Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics." *The Journal of Legal Studies*, 34(1): 57–92.

Shayo, Moses, and Asaf Zussman. 2011. "Judicial Ingroup Bias in the Shadow of Terrorism." *The Quarterly Journal of Economics*, 126(3): 1447–1484.

Tajfel, Henri, and John C Turner. 1986. "The Social Identity Theory of Intergroup Behavior." In *Psychology of Intergroup Relation.* , ed. S. Worchel and W.G Austin, 7–24. Hall Publishers, Chicago.

Tanaka, Tomomi, and Colin F. Camerer. 2009. "Status and Ethnicity in Vietnam: Evidence from Experimental Games." 1–2. Boston, MA:Springer US.

Appendix

2.A Additional tables

Table 2.A1: Summary Statistics of Identity Variables.

	Mean	SD	Min	Max
Gender	0.451	0.169	0.000	1.000
Ethnicity	0.574	0.446	0.000	1.000
Age	0.170	0.249	0.000	1.000
Level of Education	0.237	0.279	0.000	1.000
Disposable income rank	0.231	0.282	0.000	1.000
Residential area	0.299	0.258	0.000	1.000
Demographic identity index	1.195	0.534	0.000	3.000
Socioeconomic identity index	0.767	0.490	0.000	3.000
Identity index	1.962	0.739	0.000	4.667
Observations	9,999			

Notes: Categories within each characteristic are listed in Table 2.3. "Identity index" is the sum of all six characteristics; "Demographic identity index is the sum of the gender, ethnicity and age variables; "Socioeconomic identity index" is the sum of the eduation, income and neighborhood variables. SD = standard deviation.

Table 2.A2: Effects of Identity on Guilty Verdicts (Convictions).

Panel A: Demographic characteristics	
Gender	-0.021
	(0.016)
Ethnicity	0.024*
	(0.014)
Age	0.006
	(0.011)
Demographic identity index	0.004
	(0.007)
Panel B: Socioeconomic characteristics	
Education	-0.002
	(0.009)
Income	0.024***
	(0.007)
Residential area	-0.005
	(0.009)
Socioeconomic identity index	0.007
	(0.006)
Panel C: All	
Identity index	0.006
	(0.005)
Observations	9,999

Notes: Each row contains a separate regression model, with the row title indicating the identity variable. Dependent variable is an indicator for guilty verdict. All specifications contain year times court division fixed effects, judge fixed effects and control variables as described in Table 2.4. Standard errors clustered at judge level in parentheses.

Table 2.A3: Heterogeneous Effects of Identity on Prison Sentences.

	(1) Baseline	(2) Men	(3) Women	(4) Non-native	(5) Native	(6) Edu low	(7) Edu HS	(8) Edu high
Identity	-0.025***	-0.029***	0.030	-0.043***	-0.018	-0.014	-0.018	-0.028
	(0.008)	(0.009)	(0.021)	(0.012)	(0.011)	(0.015)	(0.015)	(0.023)
Demographic identity index	-0.022	-0.028	0.125**	-0.044	-0.006	-0.016	-0.016	-0.042
	(0.014)	(0.019)	(0.051)	(0.027)	(0.022)	(0.021)	(0.024)	(0.035)
Socioeconomic identity index	-0.026***	-0.029***	-0.002	-0.042***	-0.021*	-0.013	-0.019	-0.017
	(0.009)	(0.011)	(0.025)	(0.016)	(0.012)	(0.022)	(0.019)	(0.030)
Observations	9,999	8,521	1,420	3,617	6,332	5,055	3,748	1,069

Notes: Dependent variable is an indicator for prison sentence. All specifications contain year times court division and judge fixed effects. Standard errors clustered at judge level in parentheses. Control variables as described in table 2.4.

2.B Specification check results

Table 2.B1: Identity Effects on Prison Sentences When Excluding Missing Education and Income Observations.

	(1) Demography	(2) Socioeconomic	(3) All identities
Identity	-0.023	-0.025**	-0.025***
	(0.014)	(0.010)	(0.009)
Observations	9,073	9,073	9,073

Notes: Dependent variable: prison sentence. Each column shows a separate regression model, and contain year times court division and judge fixed effects. Control variables as described in table 2.4.

Table 2.B2: Identity Effects on Prison Sentences, Multiple Regression Results.

	(1) Prison (0/1)	(2) Months prison
Panel A: Demographic characteristics		
Demographic identity index	-0.022	-0.282
	(0.014)	(0.267)
Socioeconomic identity index	-0.026***	-0.320
	(0.009)	(0.227)
Panel B: Index components		
Gender	-0.068**	-0.484
	(0.031)	(0.386)
Ethnicity	-0.011	-0.295
	(0.031)	(0.570)
Age	-0.006	-0.203
	(0.022)	(0.399)
Level of Education	-0.034*	-1.149***
	(0.018)	(0.430)
Disposable income rank	-0.014	0.163
	(0.015)	(0.281)
Residential area	-0.032	0.277
	(0.028)	(0.435)
Observations	9,999	9,999

Notes: Dependent variable is an indicator for prison sentence in Column 1, prison sentence length in months in Column 2. All specifications contain year times court division fixed effects, and judge fixed effects. Standard errors clustered at judge level in parentheses. Control variables as described in table 2.4.

Table 2.B3: Education Identity Effects on Prison Sentences, Sensitivity Checks.

	(1) 3 groups	(2) Low/High	(3) Low/high v2	(4) 7 groups
Education	-0.035*	-0.002	0.005	-0.013
	(0.018)	(0.026)	(0.019)	(0.023)
Mean	0.244	0.484	0.456	0.125
Socioeconomic	-0.026***	-0.011	-0.010	-0.014
identity index	(0.009)	(0.012)	(0.010)	(0.010)
Mean	0.762	0.942	0.914	0.583
Observations	9,999	9,999	9,999	9,999

Notes: 3 groups: primary, secondary, any college (main definition). Low/high: low = primary, high = secondary and any college. Low/high v2: low = primary and secondary, high = any college. 7 groups: 1-digit ISCED codes (primary <9 years, compulsory, lower secondary, upper secondary, post-secondary, bachelor, doctorate). Dependent variable: prison sentence. Each column shows a separate regression model, and contain year times court division and judge fixed effects. Control variables as described in table 2.4.

Table 2.B4: Income Identity Effects on Prison Sentences, Sensitivity Checks.

	(1) 3 groups	(2) Low/High	(3) 4 groups
Income	-0.014	0.011	0.038
	(0.016)	(0.017)	(0.075)
Mean	0.224	0.365	0.023
Socioeconomic identity index	-0.026***	-0.013	-0.023**
	(0.009)	(0.009)	(0.011)
Mean	0.762	0.842	0.500
Observations	9,999	9,999	9,999

Notes: 3 groups: low= rank $\leq$ 25, middle= rank >25 and $\leq$ 75, high = rank >75. Low/High: below or above median. 4 groups: quartiles. Dependent variable: prison sentence. Each column shows a separate regression model, and contain year times court division and judge fixed effects. Control variables as described in table 2.4.

Table 2.B5: Alternative Definition of Age Identity; Age Groups Old, Middle and Young.

	(1) Age	(2) Demography	(3) Socioeconomic	(4) All identities
Identity	-0.027	-0.033**	-0.028***	-0.030***
	(0.024)	(0.014)	(0.009)	(0.008)
Observations	9,999	9,999	9,999	9,999

Notes: Dependent variable is prison sentences. Each column holds results of a separate regression, with the column headline indicating the identity measure used as independent variable. Age groups: young (below 30), middle (between 30 and 50) and old (above 50). Control variables as described in table 2.4.

Table 2.B6: Different Definitions of Neighborhood Identity.

	(1) District	(2) Crime rate	(3) Income	(4) Income 3gr	(5) Crime 3gr	(6) Zones	(7) Areas
Neighborhood	0.001	-0.019	-0.033	-0.005	-0.014	-0.025*	-0.024
	(0.030)	(0.019)	(0.028)	(0.015)	(0.015)	(0.015)	(0.018)
Mean	0.047	0.234	0.294	0.355	0.281	0.262	0.142
Socioeconomic identity index	-0.021**	-0.023**	-0.026***	-0.016*	-0.019*	-0.024***	-0.024***
	(0.010)	(0.009)	(0.009)	(0.009)	(0.010)	(0.008)	(0.009)
Mean	0.515	0.702	0.762	0.823	0.749	0.730	0.610
Observations	9,999	9,999	9,999	9,999	9,999	9,999	9,999

Notes: District = same residential area. Crime rate = districts with similar crime rate (above/below median). Income = districts with similar income level (above/below median). Income 3gr = districts with similar income level (below 25th, between 25th and 75th, above 75th percentiles). Crime 3gr = districts with similar crime rate (below 25th, between 25th and 75th, above 75th percentiles). Zones = inner city, near suburbs, far-out suburbs, other municipalities in region, other. Areas = geographically close and socioeconomically similar groups of neighborhoods, e.g. "Bromma-Ekerö", "Söderort", "Nacka-Värmdö". Dependent variable is prison sentences. Control variables as described in table 2.4.

2.C Other

Figure 2.C1: Lay Judge Rotation Scheme

TJÄNSTGÖRINGSI AN för n ındemän
Augusti – december 2002
Avdelning 11

Grupp 1 - 50

Rotel	Ordinarie Sessionsdag	Juli	Augusti Fr o m 19/8	September	Oktober	November	December
3	Måndag 9.30		26	44	12	30	46
4	Måndag 9.30		27	45	13	31	47
10	Måndag 9.30		28	46	14		
1	Tisdag 9.30		29	47	15	32	48
5	Tisdag 9.30		30	48	16	33	49
9	Tisdag 9.00		31	49	17	34	50
2	Tisdag 9.00		32	50	18	35	1
3	Onsdag 9.30		33	1	19	36	2
4	Onsdag 9.30		34	2	20	37	3
8	Onsdag 9.30		35	3	21	38	4
6	Onsdag 9.30		37 obs	4	22	39	5
1	Torsdag 9.30		36 obs	5	23	40	6
5	Torsdag 9.30		38	6	24	41	7
2	Torsdag 9.00		39	7	25	42	8
9	Torsdag 9.00		40	8	26	43	9
8	Fredag 9.30		41	9	27	44	10
6	Fredag 9.30		42	10	28	45	11
10	Fredag 9.30		43	11	29		
Reserv	Kallas till avd 2 och 7 vid behov		12,13,14	15,16,17	30,31,32	18,19,20	21,22,23

Kontaktperson på avdelning 11 är Christoffer Dahlgren tfn 657 5529
Om du inte får tag på kan du alltid ringa växeln 657 50 00 och be att få bli kopplad till avdelningens kansli.

Chapter 3

Intergenerational Mobility Trends and the Changing Role of Female Labor*

*This chapter is written together with René Karadakic (Department of Economics, Norwegian School of Economics) and Joachim Kahr Rasmussen (CEBI, Department of Economics, University of Copenhagen). We are grateful for advice and comments from Aline Bütikofer, Claus Thustrup Kreiner, Søren Leth-Petersen, Mårten Palme, Kjell G. Salvanes, David Seim, Jósef Sigurdsson, Alexander Willén and participants at seminars in Bergen, Copenhagen and Stockholm. Activities at NHH are funded by the Research Council of Norway through project No. 275800 and through its Centres of Excellence Scheme, FAIR project No. 262675 (Karadakic). The activities at CEBI are financed by the Danish National Research Foundation, Grant DNRF134 (Rasmussen). All errors are our own.

3.1 Introduction

A central question in the social sciences is how the childhood family environment shapes economic fortune in adulthood. If the family environment plays an important role in determining outcomes in adulthood, a common interpretation is that children are not born with equal opportunities in life. Early work by for instance Becker and Tomes (1979) and Solon (1999) highlight that when measuring the influence of family environment by way of estimating the empirical relationship between the earnings of parents and their children, it is essential to account for the role of idiosyncratic labor market conditions. Accordingly, variation in labor market conditions may be an important determinant of variation in estimates of intergenerational mobility across space and potentially also time (Corak, 2013). While spatial variation in intergenerational mobility is well documented (see e.g. Solon (2002), Chetty et al. (2014*a*) and Bratberg et al. (2017) for an overview), far less is known about the intertemporal aspect (see Lee and Solon (2009*a*), Olivetti and Paserman (2015), Chetty et al. (2014*b*) and Song et al. (2020) for notable exceptions).

In this paper, we ask what implications the "grand convergence" (Goldin, 2014) between men and women in labor market conditions has had for intergenerational income mobility. Over the past 50 years, women in all Western economies have become more likely to participate in market work (Olivetti and Petrongolo, 2016) and occupational segregation of men and women has decreased (Blau, Brummund and Liu, 2013). While it is widely acknowledged that economy-wide changes in female labor supply (on the intensive as well as the extensive margin) may change the precision with which female earnings indicate economic status (Chadwick and Solon, 2002), the implications of this change for intergenerational mobility as measured by household income is *a priori* unclear due to two opposing forces. One the one hand, when female labor supply increases, the relative position of a woman in the female earnings distribution arguably reflects her underlying skills better. All else equal, this puts a downwards pressure on measures of intergenerational mobility. One the other hand, the whole income distribution of women also shifts upwards and maternal earnings represent a larger share of joint parental earnings. If female earnings initially has a lower signal value than that of males, this puts an upwards pressure on measures of mobility. Due to constrains on the quality of linked survey data, it has proven difficult for researchers to estimate trends in correlations between males and females both separately and jointly (Chadwick and Solon, 2002, Björklund, Jäntti and Lindquist, 2009, Blanden et al., 2004), and the extent to which the secular trend in female labor supply have affected measures of intergenerational mobility is largely unexplored.

We address this issue by turning to the three Scandinavian countries. The high quality of Scandinavian administrative data allows us to follow how the changing patterns in female labor supply may have affected earnings at the individual level for both men and women. Scandinavia provides an ideal setting for understanding how the changing role of women at the labor market can affect intergenerational mobility, as the development toward gender equality precedes that in other countries (Kleven, Landais and Søgaard, 2019). First, we document trends in intergenerational earnings mobility in Sweden, Denmark, and Norway for cohorts of children born between 1951 and 1979 leveraging administrative earnings data from 1968 up until 2017. By applying a unified approach to long panels of full-population administrative data for three different countries, we can investigate the extent to which intergenerational mobility follows similar trends across countries that have been subject to different political and demographic developments, and we can ensure that any differences in findings are not related to the choice of data period or income definition.

Our results reveal a substantial decline in intergenerational mobility across Scandinavia that remains robust across a large set of common empirical specifications. In particular, we show that the results are largely unchanged when studying intergenerational correlations in log earnings rather than within-cohort earnings ranks, and when considering intergenerational correlations in gross or net-of-tax income rather than earnings. This suggests that the observed mobility trends were not driven by simultaneous rank-distorting changes in taxes or transfers across Scandinavia.

Second, after having documented that mobility has followed similar declining patterns across Scandinavia, we then turn our attention towards understanding how changes in female labor market conditions and access to education have affected estimates of intergenerational mobility over time. When breaking mobility trends down by the gender of parents and children, it is evident that earnings of children have become increasingly correlated with maternal earnings over time, while the correlation with paternal earnings has remained close to constant. In the earliest cohorts in our analysis, child earnings — in particular earnings of sons — were virtually uncorrelated with earnings of mothers while exhibiting a clear and economically significant correlation with earnings of fathers. Over time, these parent-specific mobility estimates between children and their mothers and fathers, respectively, have all converged to similar levels. Conducting a similar analysis on Panel Study of Income Dynamics (PSID) survey data from the US, we find similar patterns, albeit with a slightly lagged timing. This suggests that the observed patterns are not solely a Scandinavian issue.

Third, we build a simple model of gender-specific mobility and latent productivity that rationalizes the empirical patterns that we observe in the data. Inspired by some of the building blocks in the model by Becker and Tomes (1979), we assume that income is determined by an inheritable component, say skills or productivity, and a non-inheritable, idiosyncratic determinant. Doing so, we decompose the observational trend in intergenerational mobility into determinants associated with assortative mating (correlations in parental skills), gender-neutral skills transmission, gender-specific skills transmission and gender-specific return on skill. Calibrating our model to country-specific aggregate data, we show that the observational downwards trend in intergenerational mobility is largely compatible with a trend of increasing return on inheritable skills among women relative to men and that this phenomenon explains an increase in the intergenerational rank association of five to six rank points in all three countries for cohorts of children born from 1962 to 1979. Most of this trend is driven by mothers rather than daughters. To build intuition for this rise in gender-specific return on skills and the associated implications for mobility, we can think of an early period where a woman with a significant cognitive endowment is more likely to become a secretary than an equally skilled man with similar preferences who sorts into university and obtains a job that requires an academic degree. In this case, the female skills are arguably less well reflected by her earnings, which effectively attenuates the association between her earnings and that of her children. If this segregation becomes smaller over time, the observational relationship between maternal earnings and child earnings will increase. Bridging the model with this example, the decomposition suggests that the observed trends in income mobility could simply be an artifact of changes in how women participate in the labor market.

In the final part of the paper, we corroborate this decomposition exercise empirically, by showing that gender-specific intergenerational correlations in *economic status* — measured by combining own income, years of education, and occupation using the proxy variable method developed by Lubotsky and Wittenberg (2006) — remained constant over time, or are only weakly increasing. Mobility also remains at a constant level when correlating sons with their maternal uncles, as another way to proxy for maternal skills. Hence, our evidence suggests that the observed trends in intergenerational income mobility can be interpreted as a result of income rank correlations between children and parents — and in particular mothers — becoming gradually less attenuated by frictions caused by gender-specific segregation in the labor market. In other words, our results suggest that intergenerational mobility in income did in fact decline consistently in Scandinavia across cohorts born between 1951 and 1979, but they also suggest that this was almost solely driven by female earnings becoming more reflective of their actual skills. In other words, the return on latent productivity of women has converged towards that of men. Hence, female skills have increasingly become valued in the labor market in the same way as those of males and thus, the observed development in intergenerational earnings correlations can potentially be thought of as a natural implication of a socially desirable development rather than a sign of actually declining equality of opportunity.

With this paper, we make three contributions to the understanding of time variation in intergenerational earnings mobility. The first contribution is related to a series of recent empirical studies from Western economies which indicate that intergenerational mobility may have been declining in the past few decades, in turn suggesting that income inequality to a higher degree persists between generations. The results, however, are not conclusive, and the estimated trends show quantitatively large variation across the existing literature. In particular, Connolly, Haeck and Laliberté (2020), Harding and Munk (2020) and Markussen and Røed (2020) all find that intergenerational mobility has declined rapidly for cohorts of children born between 1960 and 1980 in the US, Canada, Denmark and Norway, respectively. Another set of recent studies, Pekkarinen, Salvanes and Sarvimäki (2017), Song et al. (2020) and Brandén and Nybom (2019) are only capable of detecting weakly declining — or even stable — trends in a similar set of countries. Davis and Mazumder (2020) find declining mobility in the US for children born between 1950 and 1960, while Chetty et al. (2014*c*) find no change in rank associations between children born in 1971 and later cohorts. In this paper, we provide clear evidence of a uniform decline in intergenerational mobility across Scandinavia for cohorts born between 1951 and 1979. In addition, we show that this trend persists across a range of common empirical specifications in the literature, and that the trends that have been observed in the existing literature are not simply a result of certain empirical specifications or country-specific policies. We also provide suggestive evidence of a similar pattern in the US from panels of linked survey data. To our knowledge, we are the first to estimate and compare trends in relative mobility across multiple countries, thereby providing suggestive evidence of a general phenomenon in Western economies.

The second contribution lies in explicitly documenting substantial gender-variation in mobility trends and showing that gender-specific mobility trends are surprisingly similar across a range of Western economies. A noteworthy strand in the mobility literature has previously highlighted that cross-sectional estimates of intergenerational mobility may differ substantially by gender due to different opportunities for men and women in the labor market (Corak, 2013, Lee and Solon, 2009*a*). With this paper, we show that mobility has seemingly been stable for father-son relations during the last few decades, while it has been declining considerably whenever female earnings are taken into account — a pattern that, to our knowledge, has only been documented in a Swedish setting by Engzell and Mood (2021) and Brandén and Nybom (2019). These findings suggest that not only do mobility *levels* vary by gender, but secular changes in gender-specific earnings determinants have also caused *trends* to differ substantially, in turn causing levels to converge. These patterns are present across all countries in our analysis, suggesting that one explanation why the recent literature has been reaching different conclusions in regards to the existence of mobility trends is choices in regards to dealing with female earnings.

The third and final contribution of this paper is that we provide an explanation for the observed pattern of declining mobility that is compatible with the gender-specific trends that we observe in Denmark, Sweden, and Norway. In recent studies, various explanations for downward trends in mobility have been proposed, none of which are consolidated across countries and specifications. One dominant explanation put forward by Davis and Mazumder (2020) is that the return on education has increased. Given that education and human capital are significant channels for the transmission of income across generations, this has led to a decline in mobility. A similar explanation put forward by Connolly, Haeck and Laliberté (2020) is that the degree to which women obtain secondary education has increased. Observing that conditional on parental income, income in the child generation is 'boosted' by a higher level of education among parents, the authors conclude that this upward trend in mothers' level of education must have led to a decline in social mobility. However, the underlying mechanism of this relationship remains unclear. Finally, Harding and Munk (2020) suggest other explanations, such as changes in family structure including marital status, assortative mating, and childbearing among women. While the importance of changes in educational attainment has thus already been discussed in the context of mobility trends by Davis and Mazumder (2020) and Connolly, Haeck and Laliberté (2020), our paper is the first to explicitly show a connection to meritocracy and valuation of female skills in the labor market.[1] In other words, our paper is the first to show that changes in female labor market conditions have caused

[1] This hypothesis is also put forward, but not further investigated by Engzell and Mood (2021).

parental earnings to be substantially better reflected in child earnings — hence causing a *real* downwards shift in intergenerational mobility — in spite of the between-generation correlation in latent skills likely being fairly constant over the period that is considered in our study.

The remainder of the paper is structured as follows. Section 3.2 provides a brief overview of the key features of the Scandinavian welfare states, and section 3.3 lists our data sources. In section 3.4 we describe the common methodology used to estimate intergenerational income mobility and present our main results. Next, section 3.5 builds and estimates a model for the connection between intergenerational rank associations and increasing female labor force attachment. In section 3.6, we finally show how our estimated trends change when we use a measure for maternal economic status that better captures female earnings potential, before we conclude with Section 3.7.

3.2 Institutional Context

Denmark, Norway, and Sweden share similar traits in terms of economic development, political culture, and institutions. The welfare state in all three countries is of universal character which means access to social security benefits, health care, subsidized childcare, and tuition-free higher education (Baldacchinoel and Wivel, 2020) for the whole population. In order to finance the provision of these public goods, marginal tax rates at the top of the income distribution, as well as the average tax burden, are substantially higher in Scandinavia than in other developed countries (Kleven, 2014). Employees are to a large degree organized in unions and wages are often collectively bargained (Pareliussen et al., 2018). Historically, all three countries have also been characterized by low levels of inequality and high levels of income mobility, in comparison to other Western countries (Søgaard, 2018, Bratberg et al., 2017).

During the second half of the 20th century the role of women in society, and in the labor market, in particular, experienced a "grand convergence" towards the position of men (Goldin, 2014). Contributing to this development were the individualization of the tax system (Selin, 2014), the introduction and expansion of paid paternity leave (Ruhm, 1998), and the expansion of

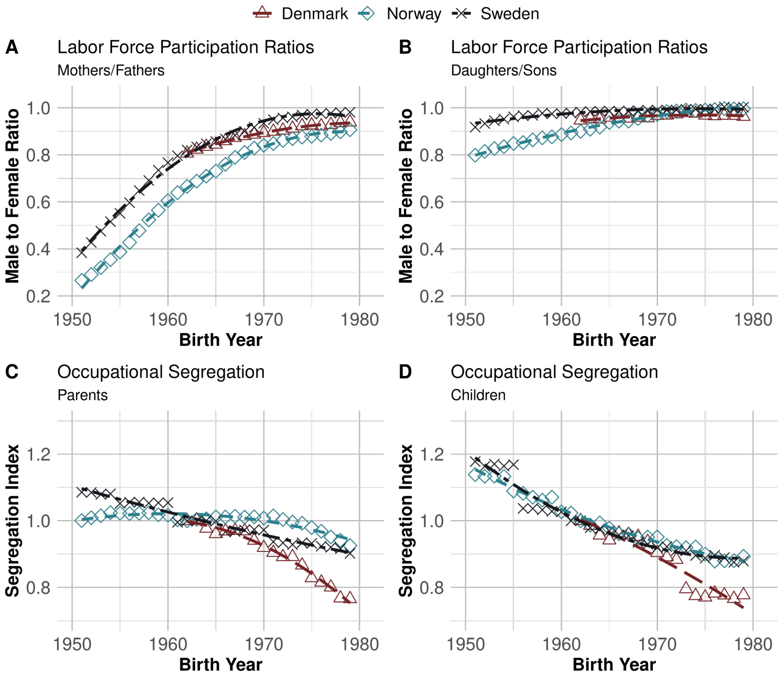


Figure 3.1: Labor Market Developments.

Note: Panel A and B depict female-to-male ratios of labor force participation in our main samples. Panel A shows participation ratios for parents by birth year of the child. Panel B shows participation ratios for children by birth year. Panel C and D depict an index for labor market segregation separately for parents and children respectively. The index is normalized to the base year 1962. In some years, Danish occupational codes have been imputed from other variables — therefore, the Danish trend in occupational segregation should be interpreted with caution.

compulsory and higher education (Meghir and Palme, 2005, Black, Devereux and Salvanes, 2005). As a result, female labor force participation increased from the early 1950s and is currently higher in Scandinavia than in most other Western economies.[2] Over the same period occupational segregation strongly decreased, indicating that women increasingly entered occupations that were previously

[2]See Appendix Figure 3.B1 for a comparison of labor force participation rates across Scandinavia and the United States or Figure 3.B2 for the development of labor force participation as defined in our samples.

male-dominated. In Figure 3.1 we provide some descriptive evidence on the development of female labor in the countries under study.

In Panels A and B we show how labor force participation rates of women converged to the male level.[3] Participation rates of mothers with children born in the 1950's were less than half the rate of fathers, but this gap had closed almost entirely for mothers of children born in the 1970s. It is even less pronounced when we compare sons and daughters of a given birth year. Even though the extensive margin labor supply gap narrowed considerably, women still work substantially more in part-time positions than men (Blau and Kahn, 2017). Panels C and D of Figure 3.1 show the development of occupational segregation across birth cohorts, capturing the extent to which men and women work in the same occupations. The segregation index is calculated as the difference in the share of all women and men in the labor force who work in a given occupation, summed over all observed occupations. To make comparisons of trends easier, we normalize the index to the base year 1962, allowing for an interpretation of occupational segregation relative to the 1962 level.[4] Evidently, occupational segregation has seen a substantial and persistent decline over time, similar to development documented by Blau, Brummund and Liu (2013) and Blau and Kahn (2017). In contrast to the development of female labor force participation, the decline in occupational segregation is to a larger extent present in the child generation, rather than the parent generation.

3.3 Data

For our main analysis, we rely on register data from Denmark, Norway, and Sweden that cover the whole population of each country from 1968 for Norway and Sweden and from 1980 for Denmark, and up

[3]The labor force participation rate for men and women is based on the income definitions we use in our later analysis and always relates to the birth year of the child. A person is considered "in the labor force" in a given year if they have annual earnings exceeding the equivalent of 10,000 USD (2017).

[4]The occupational segregation index is defined by three-digit occupation codes for Norway and Sweden and one-digit codes for Denmark due to data limitations. Therefore, the cross-country difference in trends should not be interpreted as hard evidence of deviating patterns of occupational segregation.

until 2017. The data consist of linked administrative records that provide a variety of information, including birth year, educational attainment, earnings and other income measures, family status, and various demographic variables. Individuals can be linked to their parents. This allows us to create three unique data sets containing all child-parent pairs in a given time frame, with relevant individual income measures. For more details about the registers used, see Appendix 3.A.

Our Scandinavian estimation sample consists of all children born between 1951 (1962 for Denmark) and 1979, who (i) have a valid personal identifier, and (ii) have at least one parent with a valid identifier. As this means that we remove a significant share of immigrants from our samples — in particular in early years — we remove all individuals who are immigrants or are children of immigrants. Sample sizes per birth year are approximately 70,000 child-parent pairs in Denmark, 60,000 pairs in Norway, and 100,000 pairs in Sweden, with variation over time. The results involving US data are based on the Panel Study of Income Dynamics (PSID). The PSID is a nationally representative survey that covers information on employment, income, occupation, education, and family links, starting from 1968. The PSID follows families and individuals across time and has a relatively low attrition rate. With this data, we create a sample of child-parent pairs and measure rank-rank correlations for the US in a comparable, yet more limited, fashion than our analysis on the main Scandinavian samples. In total, the US PSID sample contains about 5,000 child-parent pairs. See e.g. Lee and Solon (2009*a*), Vosters (2018) for previous applications of the PSID to intergenerational mobility estimation.

The main income specifications are chosen for easy comparisons with much of the recent literature.[5] Child income is defined as three-year averages of annual labor income.[6] See Appendix Table 3.B2 for an overview of the earnings components, and how these compare across countries. This is measured at ages 35-37, which balances the needs for a measure of permanent income rank with the needs for

[5]See e.g. Chetty et al. (2014*a*) and Lee and Solon (2009*b*).

[6]Averages are calculated including zeroes. Individuals with one or more missing observations in the years averaged over are dropped from the sample.

measuring child incomes relatively early in order to maximize the number of cohorts that can be included in the analysis (Nybom and Stuhler, 2016, Bhuller, Mogstad and Salvanes, 2017).

Parental income is defined as the average of maternal and paternal individual labor earnings, measured as three-year averages of annual income around age 18 of the child. In general, this means measuring the parents' income at age 40 or later, which is considered a meaningful proxy for lifetime income in the literature (Nybom and Stuhler, 2016). In our Appendix, we provide robustness checks to different income definitions for child and parent income variables, such as estimating trends in total factor (gross) income or net-of-tax income and evaluating the sensitivity to the exact age at which we measure child income. Finally, due to the fact that we measure parent income at age 18 of the child, parent age may vary substantially in our main specification. In particular, parents who get children at a younger age mechanically have their income measured at a younger age as well. Ranking parent income within both child birth year *and* parental birth year jointly, we are able to verify that the observed mobility trends are not driven by this measurement issue.

3.4 Trends in Intergenerational Mobility

In this section, we first describe the empirical method we apply for measuring child-parent rank associations, and present the trend for Scandinavia. We then analyze rank associations when we split the sample into sons, daughter, mothers and fathers, and compare our Scandinavian results to suggestive US estimates. Finally, we decompose the observed mobility trends into gender-specific contributions by calibrating a simple model of latent skill transmission to the data in the three countries.

3.4.1 Empirical Method

In order to measure the intergenerational income persistence, we transform observed income into cohort-specific ranks, as in Dahl and DeLeire (2008) and Chetty et al. (2014*a*). Using ranks, rather than levels or logs, offers certain advantages in this context. First, estimated rank correlations have proven to be less prone to life-cycle bias than other measures (Nybom and Stuhler, 2017), and in addition, the use of ranks enables the inclusion of zero incomes. However, in order to ensure that our results are not driven by the rank transformation, we also present mobility trends in intergenerational income elasticities (IGE) in the Appendix.

Trends in intergenerational income mobility are estimated with the following regression, separately by birth cohort and country:

$$\text{Rank}_{it}^{\text{C}} = \alpha_t + \beta_t \text{Rank}_{it}^{\text{P}} + \varepsilon_{it} \text{ with } t = 1951, ..., 1979, \tag{3.1}$$

where $\text{Rank}_{it}^{\text{C}}$ is the percentile rank of child i's average income at age 35-37 within the distribution of all children born in year t. When we analyze sons and daughters separately, we calculate their income rank separately by gender. $\text{Rank}_{it}^{\text{P}}$ is the percentile rank of the same child's parents' income within the distribution of all parents with children in birth cohort t, averaged over ages 17-19 of the child. The coefficient β_t captures the average cohort-specific parent-child correlation in income ranks, sometimes referred to as the intergenerational rank association (IRA). Lower values of β_t are interpreted as lower rank-associations in income, and thus higher levels of intergenerational mobility.

Intuitively, one can think of the IRA as the correlation in inheritable skills and values that are transmitted across generations. These are attenuated by earnings determinants that cannot be passed on to children, which reduce the signal value of parental income. Such "noise" may stem from individual-specific idiosyncratic shocks to the parental earnings process or time-specific characteristics of the labor market. In particular, changes in the IRA over time are not necessarily driven by the way that skills are transmitted, but rather by the importance of earnings determinants that cannot be passed on to children. In the context of analyzing how changing female labor market participation may have affected the intergenerational association in income, this is a relevant consideration.

3.4.2 Estimated Trends

In Figure 3.2, we present estimates for country-specific trends in intergenerational rank associations in individual labor income. Each point in the graph represents a slope parameter for a cohort-specific regression of equation (3.1) with linear trends estimated separately for 1951-1961 and 1962-1979. We provide fitted lines separately to facilitate comparisons between Denmark, Norway, and Sweden for the cohorts where all countries have available data.[7]

From Figure 3.2, it appears that intergenerational mobility, measured using the IRA, has declined in all three countries, with the fastest rate of decline in Denmark. There, the rank association in income increased by 7.3 rank points (39 %) from 1962 to 1979 — equivalent to an average annual increase of 0.5 rank points. While smaller than in Denmark, the trends in Norway and Sweden are by no means negligible. From 1962 to 1979, the rank association in income increased by 6.4 and 3.4 rank points (38 vs. 19 %) in Norway and Sweden, respectively, yielding annual increases of 0.4 and 0.2. Over the entire range of birth cohorts, from 1951 to 1979, the total change in IRA for Norway is 7.8 rank points (50 %) and 4.6 rank points for Sweden (28 %).

One may wonder what it actually means, in economic terms, that the rank association in income increased by up to 0.5 rank points per year in Scandinavia. Abstracting from nonlinearities in the relationship between parent and child income ranks, a straightforward interpretation is the following: for two children born by parents in the bottom versus the top percentile, the difference in the conditional expectation of their income ranks as adults increased by 0.5 each year — amounting to as much as five rank points over a decade. Taking the Norwegian results as an example, another interpretation of the observed trends is that in the earliest observed birth cohort, a ten rank points difference in parental income corresponded to an average difference in income ranks of 1.6 between their children. In contrast, the same difference was 2.3 rank points for children born in the latest cohort. While still indicating relatively high levels of mobility by international standards, such changes over relatively short periods of time are by all means economically substantial.

In order to ensure that the trends are robust and reflect structural

[7] In addition to providing graphical illustrations of the trends in the IRA, Appendix Table 3.B6 provides an overview of IRA coefficients for different specifications and tests whether trends are statistically different across countries.

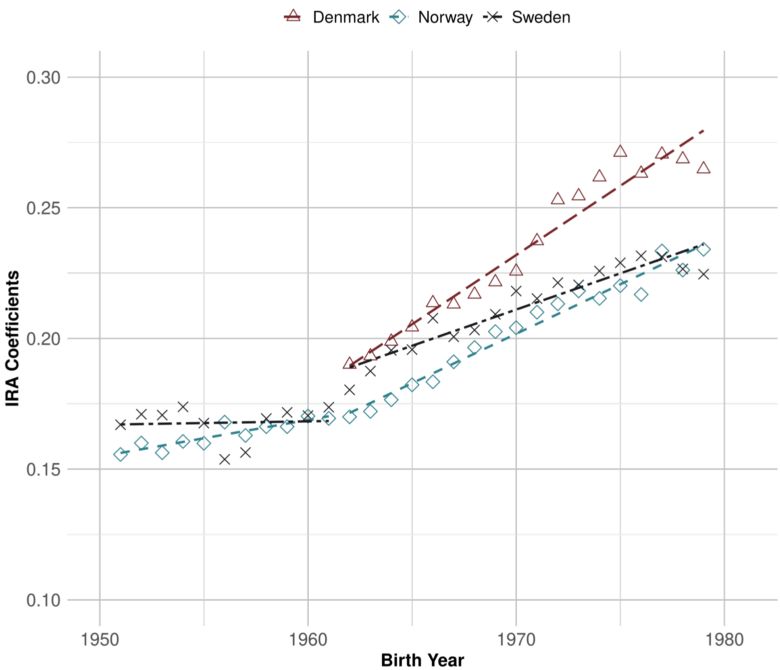


Figure 3.2: Trends in Intergenerational Mobility in Individual Labor Income.

Note: The figure plots the coefficients for the intergenerational rank association in individual labor income for Sweden, Denmark and Norway over the period from 1951 (1962) to 1979. Each panel shows fitted trend lines separately for the period 1951 to 1962 and 1962 to 1979.

changes in the economy (as opposed to being something that purely exists within a narrow set of specifications), we document similar trends for a large set of different specifications in Appendix 3.B. Most importantly, we show that the trends remain largely similar when measured in net-of-tax- and gross income (Figure 3.B4), and when measuring child income at various ages (Figure 3.B5).[8]

[8] We also tested a specification where we rank parental income within both child cohort *and* their own cohort in order to account for potential changes in life-cycle behavior. The trends remain stable, but the results are not presented in the current version of the paper.

In Figure 3.B6, we restrict the sample to parent-child pairs with labor earnings surpassing 10,000 USD (2017). In other words, we calculate rank associations for the subset of the population that is fully active in the labor market. In general, the mobility trends persist and are similar in magnitude in this specification. However, some cross-country differences are also revealed. Rank associations in Denmark and Norway are markedly lower when excluding non-participating workers from our samples, indicating that *intergenerational correlations in labor market participation* contribute greatly to intergenerational persistence in income — or at least that children of non-participating parents do disproportionately bad in the labor market themselves. In Sweden, on the other hand, the level of mobility largely remains the same after excluding non-participating parents from the estimation sample (panel B), and even increases slightly when excluding both non-participating parents and children (panel C).

3.4.3 Trends by Gender of the Child and Parent

Figure 3.3 presents estimates of country-specific IRA coefficients for pairs consisting of, in turn, sons and fathers (panel A), sons and mothers (panel B), daughters and fathers (panel C), and daughters and mothers (panel D). Each point is again obtained by separately estimating equation (3.1) for the respective combination of child and parent. In Appendix Table 3.B6, we provide results for several hypothesis tests regarding the trends and also report slope coefficients for different specifications.

The four sets of graphs make clear that — at least from 1962 and onward — the trends in IRA for all combinations of child and parent are similar in Sweden, Denmark, and Norway. Estimates for birth cohorts 1951-1979 are strikingly similar in Norway and Sweden: the trends are statistically indistinguishable for all combinations and years except for the trend in the mother-daughter IRAs after 1961. Across all panels, however, there are also several distinct differences. Most importantly, we see that the rank association between fathers and sons is generally *decreasing* (Sweden, Norway) or displays a much flatter trend over time, compared to all other graphs that display a clear upwards trend after 1962 (Denmark). The strongest trends

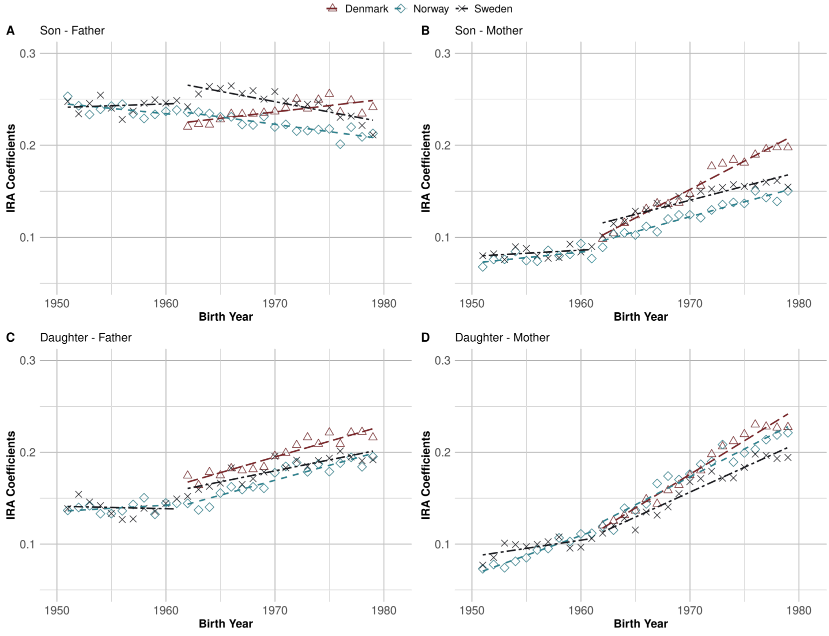


Figure 3.3: Trends in Intergenerational Mobility by Gender of Parent and Child.

Note: The four panels plot the coefficients for the intergenerational rank association in individual income for Denmark, Sweden and Norway over the period from 1951 (1962) to 1979. Each panel provides estimates separately by gender of the parent and child. Each marker indicates the coefficient of a separate regression and each line indicates fitted trend lines separately for the period 1951 to 1962 and 1962 to 1979.

in IRAs are found among mother and daughter correlations, closely followed by mother-son correlations. Father-daughter correlations depict slightly weaker trends.

In order to rule out that the mobility patterns that we observe in Scandinavia are just local phenomena, we compute comparable mobility estimates for the US for cohorts born between 1947 and 1983. Results from this exercise are presented in Table 3.1.[9]

[9]In Appendix Table 3.B3, we provide similar estimates with alternative sample specifications and weighting procedures. In Table 3.B5, we document the cohort-specific number of parent-child pairs used to compute these trends.

Table 3.1: IRA Coefficients and Trends (United States)

	Parents	**Father**		**Mother**	
	Child	Son	Daughter	Son	Daughter
Pooled IRA	0.317***	0.336***	0.195***	0.097***	0.137***
	(0.017)	(.022)	(0.031)	(0.025)	(0.029)
Trend × 100	0.603***	-0.240	0.980***	0.136	1.047***
	(0.149)	(0.205)	(0.277)	(0.253)	(0.292)
Observations	5,392	2,272	1,637	2,477	2,205

Notes: The table presents estimates of the IRA and linear trends in the IRA separately for different child-parent combinations. Due to the small sample sizes, trends have been estimated directly on the underlying micro data by regressing cohort-specific child ranks on cohort-specific parent ranks interacted with a linear time trend. The trend coefficients and corresponding standard errors have been multiplied by 100 in order to avoid too many digits after the separator. Estimates are based on the full sample of individuals in the PSID born between 1947 and 1983 using PSID sample weights. Standard errors are in parentheses. P-values indicated by * $<$ 0.1, ** $<$ 0.05, *** $<$ 0.01.

Similar to Scandinavia, US mobility trends are steepest for pairs involving mothers and — in particular — daughters. One interesting difference between gender-specific trends in the US and Scandinavia lies in the fact that the upwards trend in mother-son correlations in earnings ranks is not statistically significant in the US. Father-son rank associations appear to be relatively constant in the US, suggesting a comparable development as that observed in Scandinavia.[10]

To the extent that father-son correlations, which are stable over time, credibly measure equality of opportunity, it is hard to argue that an actual decline in opportunity has taken place over time in either Scandinavia or the US. Thinking of transmission of skills and values as something passive, this suggests that determinants of male income ranks, as well as the labor market valuation of skills that are passed on across generations, are unchanged over time. Instead, since all combinations of parent-child income that do yield upwards trends in IRAs (panels B-D) involve women,[11] a close-at-hand explanation lies in that women's increasing integration into the labor force has changed the way that incomes are correlated across generations.[12]

The fact that son-mother and daughter-mother coefficients are below 0.15 among children born before the 1970s suggests that maternal income ranks did not well reflect earnings potential. Higher participation and earnings over time among women may also be the key driver of the trend in the daughter-father rank association. The difference in maternal trends between the US and Scandinavia would also be in line with such an explanation, as developments concerning decreases in occupational segregation and increases in female labor

[10]Recent evidence by Song et al. (2020) supports relatively stable father-son trends for the relevant cohorts in our samples. Moreover, the IRA estimates provided in Song et al. (2020) are similar in magnitude for cohorts between 1950 and 1980.

[11]Notably, father-son correlations in Denmark display a weakly increasing pattern in 1962-1975. The source of this deviant pattern compared to Sweden and Denmark is a question we leave open for future research.

[12]The weak link between maternal income and skills for the earliest birth cohorts is also suggested by patterns of assortative mating on individual income. Appendix figure 3.B9 shows an increasing correlation in parental income over time, suggesting that mothers' income becomes more predictive of their true social status. An alternative explanation for the pattern in Figure 3.B9 would involve rapid and strong changes in underlying mating patterns, which appear to be unlikely given recent research by e.g. Bratsberg et al. (2018).

force participation started later in the United States and therefore likely impacted mothers only for later-born cohorts, while having a potentially larger impact through changing labor market equality for daughters.[13]

One last feature of Figure 3.3 and Table 3.1 is that incomes are more strongly related for parent-child pairs within gender (i.e., son-father and daughter-mother) than across gender (i.e., son-mother and daughter-father). In fact, while the association in income ranks is generally higher among sons and fathers than among any other combination of genders, the daughter-mother correlation reaches almost the same level towards the end of the considered period in Scandinavia. For the US, we only provide a pooled IRA coefficient due to the small sample. Nevertheless, the pattern that within-gender correlations are stronger than cross-gender correlations and that father-child correlations exceed mother-child correlations is also found in the US sample. This finding could have several reasons, such as intergenerational occupational mobility being lower within- than across gender, and the general tendency of men and women to sort into different occupations (see e.g., Blau and Kahn (2017) for a review on this latter point). Altonji and Dunn (2000) also find within-gender correlations in work hour preferences between parents and children and a recent working paper by Galassi, Koll and Mayr (2021) highlights how employment correlates between mothers and their children, especially so for daughters.

3.5 Decomposition by Earnings Determinants

In the previous section, we documented that the intergenerational rank association in earnings has increased rapidly in Scandinavia, but that this phenomenon is found almost exclusively for parent-child

[13]The validity of this explanation is confirmed in Table 3.B4. Here, we estimate child incomes around age 30 rather than 36, allowing us to compute gender-specific rank-correlations for cohorts of children born in 1953 to 1989 rather than 1947 to 1983. Looking at this set of children born slightly later, we find that rank-correlations that include mothers exhibit a clear and significant upwards trend.

pairs involving mothers or daughters. The exact mechanism driving this upward trend in intergenerational income correlations is, however, unknown. We cannot *a priori* distinguish a trend in the extent to which skills are transmitted across generations (for example if having a mother working *per se* generates higher child-mother correlations) from a trend in the extent to which female incomes reflect their inherent earnings potential ("skills"). In addition, our analysis might be influeced by any potential changes in assortative mating among parents.[14] However, we can use the gender-specific variation in mobility trends, along with correlations in parental earnings, to quantify the importance of these two potential mechanisms for our observed trends. In this section, we build a simple model that exactly allows us to quantify the importance of these channels through a decomposition exercise.

3.5.1 Model Setup and Calibration

In our framework, individual gender-specific earnings at time t, y_{it}^k, are determined by two factors; inheritable skills, x_{it}^k, and a non-inheritable determinant ε_{it}^k. This generalizes to all fathers, mothers, sons, and daughters, i.e. all $k \in \{F, M, S, D\}$. Interpreting the setup in the context of a highly simplified version of the frameworks formulated by Becker and Tomes (1979) and Solon (2004), we can think of x_{it}^k as representing an aggregate measure of earnings determinants that can be transmitted across generations such as skills, values, and connections, while ε_{it}^k represents the value of all other income determinants that are uncorrelated to skills that can be transmitted across generations (it may be instructive — yet slightly naïve — to think of this as luck).

We assume that inheritable skills in the parental generation follow a bivariate Gaussian distribution on the following form:

$$\begin{pmatrix} x_{it}^F \\ x_{it}^M \end{pmatrix} = \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_t\right), \ \mathbf{\Sigma}_t = \begin{pmatrix} 1 & \\ \frac{\psi_t}{\sqrt{\psi_t^2+(1-\psi_t)^2}} & 1 \end{pmatrix}$$

[14]However, the influence of changes in assortative mating on intergenerational income associations is found to be small in Holmlund (2022), studying the case of Swedes born in 1945-1965.

where $\mathbf{\Sigma}_t$ denotes the cohort-specific covariance matrix that summarizes the joint mean-zero distribution of parental skills. Standardizing the variance of skills to one, ψ_t reflects cohort-specific correlations in parental skills, thus measuring assortative mating in the model.

We assume that skills are transmitted passively from the parental generation to the child generation on the following form:

$$x_{it}^k = \begin{cases} \left(\kappa_t \left[\alpha_t x_{it}^F + (1-\alpha_t)\, x_{it}^M\right] + (1-\kappa_t)\, u_{it}\right)/\Gamma_t, & \text{for } k = S \\ \left(\kappa_t \left[\alpha_t x_{it}^M + (1-\alpha_t)\, x_{it}^F\right] + (1-\kappa_t)\, u_{it}\right)/\Gamma_t, & \text{for } k = D \end{cases}$$

Here, κ_t is a measure of correlation in inheritable skills — or the rate at which skills are transmitted — across generations within a given cohort of children, and α_t is a coefficient that allows the transmission of skills within gender to be stronger than skills across gender. Finally, Γ_t is a trivial scaling coefficient that ensures that the distribution of skills is standard normal.

Individual income is a monotone transformation of a linear index composed of inheritable and non-inheritable determinants:

$$y_{it}^k = \hat{F}_t^k \left(\tilde{\phi}_t^k x_{it}^k + \left(1 - \tilde{\phi}_t^k\right) \varepsilon_{it}^k\right), \quad \text{for } k \in \{F, M, S, D\}$$

Here $\tilde{\phi}_t^k = \phi_t^k / \max\left(\phi_t^F, \phi_t^M\right)$ for $k \in \{M, F\}$ in the parental generation and $\tilde{\phi}_t^k = \phi_t^k / \max\left(\phi_t^S, \phi_t^D\right)$ for $k \in \{S, D\}$ in the child generation, respectively. Hence, ϕ_t^F, ϕ_t^M, ϕ_t^S and ϕ_t^D reflect the relative importance of inheritable skills in the income process for fathers, mothers, sons and daughters, respectively. Making the simple assumption that the distribution of non-inheritable determinants can be summarized by a standard normal distribution, $\varepsilon_{it}^m \sim \mathcal{N}(0,1)$, the individual earnings index is standard normal[15].

When measuring gender-specific intergenerational mobility in income ranks, the functional form of the monotone transformation function, $\hat{F}_t^k(\cdot)$, is essentially unimportant; as long as it is monotone in the earnings index, any rank transformation of the earnings index will yield the same result as a rank transformation of earnings. However,

[15] Through simulations, it can be verified that composing the individual income index of two sets of Gaussian components, one inheritable and one non-inheritable, replicates the aggregate functional relationship between parental and child income ranks remarkably well.

in order to find both a pooled measure of child income ranks and a measure of joint parental earnings, such functional form can no longer be disregarded without also disregarding potentially non-negligible differences in gender-specific earnings distributions. Fortunately, we can obtain the functional forms directly from the data. Exploiting the assumed monotone relationship between the earnings index and earnings, we match index ranks to the earnings distribution observed in the data. This allows us to compute pooled earnings ranks in the child generation as well as a measure of joint parental earnings, y_{it}^{P}, that takes the true earnings distribution into account:

$$y_{it}^{P} = \hat{F}_t^{F}\left(\tilde{\phi}_t^{F} x_{it}^{F} + \left(1 - \tilde{\phi}_t^{F}\right)\varepsilon_{it}^{F}\right) + \hat{F}_t^{M}\left(\tilde{\phi}_t^{M} x_{it}^{M} + \left(1 - \tilde{\phi}_t^{M}\right)\varepsilon_{it}^{M}\right)$$

Here, $\hat{F}_t^{F}(\cdot)$ and $\hat{F}_t^{M}(\cdot)$ are year-specific estimates of the functions that map the earnings index to the earnings distribution observed in the data.

For each country and cohort, we are currently calibrating a vector of seven decomposition parameters, $\begin{bmatrix} \psi_t & \kappa_t & \alpha_t & \phi_t^{F} & \phi_t^{M} & \phi_t^{S} & \phi_t^{D} \end{bmatrix}'$, from only five equations. In order to avoid underidentification, we make two adjustments. First, we set $\phi_t^{F} = \phi_t^{S}$ such that the skill importance in earnings for mothers and daughters, ϕ_t^{M} and ϕ_t^{D}, must be interpreted relative to that of fathers and sons respectively — i.e. a generation-specific gender bias in the importance of skills for determination of earnings. Secondly, we set both ϕ_t^{F} and ϕ_t^{S} equal to one, thereby effectively pinning down the level around which κ_t trends over time[16]. Finally, the vector of decomposition parameters that are now left for us to calibrate across countries and years is given by: $\begin{bmatrix} \psi_t & \kappa_t & \alpha_t & 1 & \phi_t^{M} & 1 & \phi_t^{D} \end{bmatrix}'$. The calibration procedure is explained in Appendix section 3.C, where we also document the quality of the calibration exercise for each set of country-year combinations of parameters.

[16] The more skills are reflected in earnings, the less skills need to be transmitted across generations in order to obtain a given correlation in earnings over time. Fixing the importance of skills for earnings among males therefore effectively pins down the skill transmission rate across time for a given intergenerational correlation in earnings.

3.5.2 Decomposition

By calibrating the model, we are eventually interested in understanding how country-specific changes in intergenerational mobility can be decomposed into changes in the rate at which inheritable skills manifest themselves as labor earnings among mothers and daughters relative to fathers and sons, and changes in assortative mating on skills among parents. Before doing so, we first investigate how the parameters associated with these channels have changed over time in our calibration exercise. Parameters for selected years are displayed in Table 3.2.[17]

Table 3.2: Decomposition Parameters

	1951			1962			1979		
	SE	DK	NO	SE	DK	NO	SE	DK	NO
ψ_t	0.131	-	0.147	0.289	0.189	0.171	0.249	0.186	0.174
κ_t	0.301	-	0.300	0.257	0.267	0.274	0.261	0.290	0.286
α_t	0.580	-	0.603	0.632	0.582	0.626	0.561	0.560	0.564
ϕ_t^M	0.286	-	0.260	0.368	0.398	0.371	0.594	0.701	0.622
ϕ_t^D	0.511	-	0.501	0.591	0.721	0.619	0.935	1.011	0.951

Note: The table presents calibrated decomposition parameters for Sweden, Denmark, and Norway in three selected years. The coefficients have been obtained by matching a simulated version of the aforementioned model to empirical gender-specific IRA-coefficients as well as the relation between father and mother income.

Several noteworthy features of our calibration exercise stand out. First, the decomposition parameters generally evolve very similar across countries. This observation adds credibility to the decomposition approach. In particular, the parameters associated with skill-importance in earnings for mothers and daughters, ϕ_t^M and ϕ_t^D, have increased at a somewhat similar pace across all three countries. This, in turn, suggests that female earnings may have become more reflective of inheritable skills in both the parent and child generations. Second, the parameter associated with assortative mating is relatively constant over time in all three countries (at least from the early 1960s and onward) in spite of strongly increasing associations in maternal and paternal earnings over time. This may

[17] The full set of parameters is available upon request.

be an implication of the fact that maternal earnings have become more reflective of maternal inheritable skills, thereby mechanically increasing the observational correlation in father and mother earnings for a given correlation in skills. Third, within-gender correlations in skill do in fact seem to be stronger than cross-gender correlations in skills — α_t is approximately 0.6 across all countries but slowly declining from the early 1960s and onward. Finally, the coefficient associated with non-gendered skill-transmission is slowly downwards trending in both Sweden and Norway, while exhibiting a weak but robust upwards trend in Denmark.

While the trends in decomposition parameters are generally similar across countries, the direction and extent to which their changes may affect the intergenerational rank association in earnings between parents and children is unknown. In order to decompose changes in this main parameter into effects associated with changes in the modeling parameters, we compute "counterfactual" income associations holding one parameter fixed over time, while allowing the aggregate gender-specific income distributions that we obtained from the data to vary over time.

We do this by first defining $\tilde{\beta}_t$ as the rank association between joint parental and child earnings obtained from the calibrated set of parameters in the model stated above subject to a simulated set of data such that $\tilde{\beta}_t \equiv \beta\left(\psi_t, \kappa_t, \alpha_t \phi_t^M, \phi_t^D\right)$. Then we define $\tilde{\beta}^b_{t,\underline{t}}$ in a similar fashion, but we fix parameter $b_t \in \left(\psi_t, \kappa_t, \alpha_t \phi_t^M, \phi_t^D\right)$ to the calibrated value in period $\underline{t}$ such that for instance $\tilde{\beta}_t^{\psi_{\underline{t}}} \equiv \beta\left(\psi_{\underline{t}}, \kappa_t, \alpha_t \phi_t^M, \phi_t^D\right)$. Finally, the part of the trend in $\tilde{\beta}_t$ that can be attributed to parameter b is simply the difference in trend between $\tilde{\beta}_t$ and $\tilde{\beta}_t^{b_{\underline{t}}}$, while the part of the actual trend in β_t that can jointly be attributed other factors than decomposition parameters and changes in the aggregate gender-specific income distributions is the difference in trend between β_t and $\tilde{\beta}_t$. The results from this exercise are documented in Table 3.3.

Table 3.3: Decomposition Results

	1952-1961			1962-1979		
	SE	DK	NO	SE	DK	NO
Trend in β_t	0.013	-	0.140	0.277	0.530	0.379
Trend in $\tilde{\beta}_t$	0.068	-	0.138	0.240	0.527	0.327
Due to ψ_t	0.189	-	0.000	-0.056	-0.001	0.001
Due to κ_t	-0.343	-	-0.164	-0.041	0.242	-0.035
Due to α_t	0.009	-	0.007	0.002	0.000	0.001
Due to ϕ_t^M	0.054	-	0.117	0.138	0.220	0.161
Due to ϕ_t^D	0.020	-	0.130	0.158	0.069	0.119

Note: The table presents trends in observational IRA coefficients, β_t, in the three countries as well as trends in IRA coefficients obtained from the calibrated models in the three countries, $\tilde{\beta}_t$. The contribution from each parameter is computed as the difference in $\tilde{\beta}_t$ that is obtained from holding one calibrated parameter fixed at a time. The sum of contributions from each parameter need not sum to the trend in $\tilde{\beta}_t$ as part of the trend will be driven by changes in the scale of gender-specific income distributions which is not modeled.

As the rank associations in earnings did not exhibit any clear upwards trend for cohorts born between 1952 and 1961 in Sweden and Norway, there is not much to be explained by the decomposition parameters. However, there *are* certain noteworthy patterns in this period. In particular, the parameter associated with non-gendered skills transmission, κ_t, contributes negatively to the IRA over time, while the opposite is the case for the parameters associated with the extent to which female earnings are reflective of parental skills, ϕ_t^M, and ϕ_t^D. This could possibly suggest that skills transmission may in fact have declined over time, thereby pushing mobility up, but that this effect was mitigated by the increasing extent to which women's individual income reflects their earnings potential, i.e. inheritable skills. However, one should be careful with drawing too strong conclusions based on this evidence.

For birth cohorts 1962 to 1979, the simple decomposition model captures the fact that IRAs are increasing uniformly across Scandinavia well. While both ψ_t and α_t generally contribute little to mobility trends in this period, our results suggest a bigger role for κ_t — at least in Denmark, where this component explains almost half of the observed trend in mobility. In both Sweden and Norway, however, the contribution of κ_t is negative and the importance is somewhat negligible. Finally, changes in the extent to which female earnings, and particularly maternal earnings, are reflective of inheritable skills are found to be important drivers of downwards trends in mobility across Denmark, Norway, and Sweden. These effects jointly contribute to a yearly increase in the earnings IRA of between 0.28 and 0.30 rank points in all three countries, amounting to a total increase in the IRA of between 5 and 6 rank points over the period. In the next section, we show that it is indeed plausible to interpret this phenomenon as female earnings becoming more reflective of inheritable skills over time.

3.6 Intergenerational Correlation in Latent Economic Status

In this section, we present estimates of trends in intergenerational mobility, when we use information on parents' education and occupation to supplement the information about earnings potential contained in labor income. The intuition here is that income, education and occupation all constitute imperfect measures about a person's underlying, or "latent", socioeconomic status, but that a less attenuated measure of parental economic status can be constructed from a weighted average of the three. While we speak here of *socioeconomic status* rather than, as before, inheritable skill-based earnings potential, we argue that for our application to female individual labor earnings, socioeconomic status is more or less equivalent to potential earnings.

3.6.1 Measuring Latent Economic Status

Income correlations between mothers and their children are complicated by the fact that female labor earnings are a poor measure of their earnings potential during most of our studied time frame. Estimates of the model in equation 3.1 for maternal income will not capture the intergenerational relationship between maternal and child labor market skills, which is the main interest in this paper. To fix ideas, denote the underlying relationship of interest as:

$$x_{it}^{*C} = \alpha_t + x_{it}^{*P} + \varepsilon_{it},$$

where x_{it}^{*} is a person's true economic status, unobserved by the researcher. In our setting, it is reasonable to assume that lifetime earnings alone are a good measure of economic status among sons and fathers, but less so for mothers and daughters. We follow recent work by Vosters and Nybom (2017), Vosters (2018) and Adermon, Lindahl and Palme (2021) and apply the Lubotsky-Wittenberg (Lubotsky and Wittenberg (2006), from now on "LW") method in the intergenerational mobility context. In essence, this method amounts to using a set of proxy variables that together represent a single latent variable, economic status, and weighting these together optimally, given some outcome variable (in our case, child income percentile ranks). These optimal weights have been shown to result in an estimator which minimizes attenuation bias among its class of estimators (Lubotsky and Wittenberg (2006), p.552). The procedure requires the theoretical assumption that each proxy measure affects the left-hand side variable — child economic status — only through latent economic status, but it does not assume independence between the proxy variables.

The proxy variables for parental economic status that we use are income ranks, years of education and occupation: these are denoted $x_j, \quad j \in 1,..,k$. The LW estimator is constructed as follows:

$$\beta_{LW} = \sum_{j=1}^{k} \rho_j b_j, \tag{3.2}$$

where $\rho_j = \frac{cov(Rank_{it}^{\mathrm{C}}, x_{jit})}{cov(Rank_{it}^{\mathrm{C}}, Rank_{it}^{\mathrm{P}})}$, and the $\mathbf{b_j}$'s are OLS coefficients from a multiple regression of child income rank on the set of proxy variables.

This method has previously been used to estimate mother-child intergenerational income elasticities for Swedish birth cohorts 1951-1961 in Vosters and Nybom (2017). Our application uses the same set of proxy variables and the same methodology, with two exceptions. First, we calculate year-specific LW estimates, in order to study the time trend in latent economic status mobility. We also extend the analysis to later-born cohorts, which necessitates measuring parental income somewhat earlier in life than in Vosters and Nybom (2017).

Second, we make use of the explicit index construction mentioned in Lubotsky and Wittenberg (2006), p.554:

$$x_{it}^{\rho,P} = \frac{1}{\beta_{LW}} \sum_{j=1}^{k} x_j b_j. \tag{3.3}$$

We calculate LW index values for each mother-son and father-son pair using the logarithm of child and parental labor income and then transform these into percentile ranks.[18] Finally, we regress the child income ranks on these parental index ranks, for mothers and fathers separately. In order to keep the interpretation as close as possible to that in our main analysis, we assign individuals with zero labor income a token low level of log earnings.[19]

The method described so far addresses the problem of unrepresentative maternal earnings. If trends in intergenerational rank correlations in latent economic status between mothers and sons resemble those found between fathers and sons, it stands to reason that the upwards trend in mother-son earnings correlations are attributable to increased economic opportunities of women, and subsequently less attenuation bias in rank correlations. In order to understand whether daughter-father correlations are subject to the same issue (and bias in estimation), we repeat the above procedure for daughters and approximate their economic status with income, education and occupations. Since the LW method deals with measurement error in the right-hand-side (independent) variable, this requires "flipping" the intergenerational model (eq. 3.1), and estimating rank associations between fathers and their daughters. This has only minor impacts on the year-specific IRA estimates and does not alter the trend. Apart from this first step, the analysis proceeds in an identical manner as for son-parent estimates.

[18]This procedure serves the purpose of staying as close as possible to the previous literature on methodology with the use of log income rather than income ranks, while still achieving coefficients that are readily comparable to our main specification of rank correlations.

[19]Sensitivity checks show that the exact level of earnings assigned does not alter the conclusions from this analysis. Results are available on request.

3.6.2 Results

Here, we first present trends in the IRA and Lubotsky-Wittenberg (LW) coefficients for sons and their fathers and mothers, allowing us to isolate to what extent increased labor force attachment among mothers drives the observed trends. Second, we provide estimates of the trend in daughter-father IRA and LW coefficients, in order to investigate the extent to which changes in female market labor among the child generation influence trends in intergenerational mobility.

Table 3.4 provides estimates of the trend in IRA and LW estimates for the years 1962 to 1979, separately by country. We also report the difference between the trend estimates, which tests whether trends in intergenerational mobility are statistically distinguishable between the IRA and LW approaches. For a visual representation of the trends and corresponding estimates see Appendix Figure 3.B10. The son-father trends obtained from the LW method correspond well to the son-father IRA trends as suggested by Panel A in Table 3.4. Even though there are small differences between the estimated trends across all countries these differences are not statistically distinguishable from zero, indicating that son-father trends for the IRA and LW coefficients are similar. For Norway and Sweden, IRA and LW trends are negative, indicating a development towards *increased* mobility, while Denmark's decline in mobility is supported by both the IRA and LW methods. Our interpretation of this similarity in estimated trends is that father and son income ranks provide a reasonably stable measure of socioeconomic status.

Table 3.4: Comparison of Trends 1962 - 1979

	Denmark	Norway	Sweden
Panel A: Son - Father			
IRA	0.1385	-0.1598	-0.2243
	(0.0349)	(0.0222)	(0.0605)
LW	0.1504	-0.2062	-0.1898
	(0.0239)	(0.0350)	(0.0613)
Difference	-0.0118	0.0464	-0.0346
	(0.0423)	(0.0414)	(0.0861)
Panel B: Son - Mother			
IRA	0.6186	0.3244	0.3069
	(0.0256)	(0.0262)	(0.0408)
LW	0.2994	-0.1200	0.0175
	(0.0353)	(0.0273)	(0.0495)
Difference	0.3192	0.4444	0.2894
	(0.0436)	(0.0379)	(0.0642)
Panel C: Daughter - Father			
IRA	0.3416	0.3247	0.2388
	(0.0309)	(0.0318)	(0.0339)
LW	0.0658	-0.0385	-0.0897
	(0.0324)	(0.0314)	(0.0201)
Difference	0.2758	0.3632	0.3285
	(0.0447)	(0.0447)	(0.0394)

Note: IRA indicates linear trends estimated through all coefficients of the intergenerational rank association. LW specifies linear trends estimated through all coefficients obtained from applying the Lubotsky-Wittenberg method. The trend coefficients and corresponding standard errors have been multiplied by 100 in order to avoid too many digits after the separator. Difference indicates differences between LW and IRA trends and tests the null-hypothesis of equality in trends between the IRA and LW coefficients. Heteroskedasticity robust standard errors are in parentheses.

Panel B presents son-mother estimates. Compared to the IRA trend, our LW trend estimates are noticeably smaller, and in the case of Norway and Sweden even negative. This suggests a development similar to that of father-son estimates, with a development towards increased mobility in Norway and Sweden and a less pronounced decline (compared to rank associations in income alone) in mobility for Denmark. The difference between the trends of the IRA and LW coefficients is statistically meaningful and different from zero, and are similar in magnitudes across all three countries, which indicates that the LW method mitigates attenuation bias in a similar manner in the different settings. Evidently, when using mothers' years of education and occupations - rather than just labor earnings - to proxy for their latent economic status, the extent to which male children achieve similar economic success as their parents has remained relatively constant over time.

In Panel C of Table 3.4, we additionally present the comparison between trends in the LW and IRA coefficients for daughters and fathers. Similar to Panel B the trends in the IRA are significantly steeper than what the LW trends suggest. The differences between IRA and LW trends by country are almost identical across countries, suggesting that the use of additional proxy variables in the LW approach captures latent economic status in a similar fashion across all three countries. For Denmark, the adjusted trend still indicates that over time mobility in economic status decreases, however at a significantly lower rate, in Norway the relationship is stable, while in Sweden daughters experience a small increase in mobility over time.

In summary, Table 3.4 provides three important takeaways. First, in all three countries trends between sons and fathers are similar for the IRA and the LW approach, indicating that the IRA reasonably captures actual developments of intergenerational mobility in latent economic status. Second, trends in the son-mother and daughter-father IRA appear to overestimate declines in mobility and, third, differences in trends between the IRA and LW method are comparable across countries. In addition to the comparison of trends, the levels of the son-father, son-mother, and daughter-father LW coefficients are more similar to the IRA coefficients of son-father pairs which is what would be expected when accounting for attenuation in the coefficients and is also supported by findings in e.g. Vosters and Nybom (2017). Estimating rank associations in latent economic status by birth cohort shows that over time, father-daughter correlations have remained roughly constant at a level just below 0.3. The transmission of economic potential between parents and their female children, as well as their male children, has thus seen little change across birth cohorts from 1962 to 1979. That girls are not over time increasingly "invested in" by their parents might reflect the particular setting, with schooling relatively equally distributed among boys and girls already among individuals born in the 1950s. On the other hand, the fact that father-daughter correlations are as high as the father-son ones suggests that whatever skills relevant to economic success are transmitted between parents and their children, these are gender-neutral.

By estimating correlations in “latent economic status” rather than observed income, our goal is a measure that better approximates the transmission of income-generating skills between parents and their children. One could argue, however, that occupational and educational choices are so strongly correlated with realized income, that the approach adds little by way of intuition. This would also invalidate the primary assumption behind the LW approach — that of independence between the proxy variables. To corroborate the LW results, we also estimate the intergenerational rank association in labor income between sons and their maternal uncles. Given a constant level of brother-sister correlation in earnings potential, this estimated trend captures changes in the importance of parental earnings potential for child outcomes.[20] Using observed skills of maternal uncles to proxy for unobserved female values is a strategy previously used by e.g. Grönqvist, Öckert and Vlachos (2017). Due to high data demands needed for parental generation sibling links the sample size used to estimate the IRAs is relatively low, particularly for the earliest birth cohorts. Appendix Figure 3.B8, Panel A, presents the results, which reveals a constant level of rank associations over time. Panel B shows the original mother-son associations for comparison, and in Panels C-D, the same results are shown for daughters and maternal uncles. Daughter-uncle trends are substantially flatter than daughter-mother trends, indicating that a certain part of the mother-daughter trends is driven by the mothers. However, the remaining IRA trend shows that increased labor force attachment by daughters over time also contributes to the observed mobility trend.

To further clarify the intuition behind our central theme, Appendix Figure 3.B9 provides evidence that maternal “skills” and income are virtually unrelated in the early period of our sample.

[20] Using Swedish data, Björklund, Jäntti and Lindquist (2009) show that brother correlations in income remain similar for cohorts born between 1953 and 1968.

The figure plots maternal income ranks by ventiles (5 percentile rank bins) of the paternal income distribution, for 1951, 1962, and 1979 samples, respectively. Mothers of the 1951 cohort evidently earned the same (low) level of income irrespective of their husband's earnings: the average rank hovers around 50 across the whole of the fathers' income distribution. In 1962 — and even more so in 1979 — however, maternal income rises almost monotonically in paternal income. Assuming a time-invariant pattern of assortative mating, this is evidence favoring our hypothesis of female incomes better reflecting underlying skills over time.[21].

3.7 Conclusion

In this paper, we have documented trends in intergenerational income mobility in Denmark, Norway, and Sweden, for children born in 1951 (1962) to 1979. Harmonizing data and definitions, we have shown that the intergenerational rank association between parents and children in individual income has increased significantly in all three countries. These trends are robust to using different types of income measures, as well as to restricting the analysis to labor market active individuals. Splitting trends by gender of parents and children, son-father correlations exhibit the weakest trend in all three countries, whereas all correlations involving mothers and daughters increase over time. The strongest trend is found between mothers and daughters. To extrapolate our findings to countries outside of Scandinavia, we show that similar patterns can be found for US parent-child pairs from the PSID. In line with the Scandinavian results that are based on more detailed data of higher quality, we find a similar, but delayed, development in changes of the IRA in the US. Our results suggest that rising female labor supply and participation results in higher child-parent rank associations through better manifestation of maternal skills in income, such that the intergenerational correlation

[21] Whether assortative mating in income and education has declined or inclined over time is a topic of recent research by e.g. Eika, Mogstad and Zafar (2019) and Bratsberg et al. (2018), with the latter suggesting that trends in assortative mating by social class have stayed considerably more constant than assortative mating by education.

in "potential income", or *latent economic status* is revealed. In other words, the fact that maternal economic status was poorly reflected in maternal income among early cohorts of our sample caused rank associations between child income and joint parental income to be an attenuated measure of mobility of economic status or opportunity. Over time, as female labor supply and participation has increased, this attenuation has declined accordingly.

Our results clearly point to the importance of accounting for changes in female economic status when estimating trends in intergenerational mobility. The interpretation that higher rank associations in income or earnings between children and parents reflect a lower degree of social mobility or equality of opportunity is not always easily applicable when labor market conditions change substantially. In particular, our findings suggest that women's income over time is to a larger extent determined by their earnings potential, meaning that the traits and norms that women inherit from their parents are also better reflected in their income. While such a development must be seen as a necessary side-effect of increased gender equality in the labor market, it is *a priori* unclear whether such development should be seen as a reduction or advancement in equality of opportunity.

Bibliography

Adermon, Adrian, Mikael Lindahl, and Mårten Palme. 2021. "Dynastic Human Capital, Inequality, and Intergenerational Mobility." *American Economic Review*, 111(5): 1523–48.

Altonji, Joseph, and Thomas A. Dunn. 2000. "An Intergenerational Model of Wages, Hours, and Earnings." *Journal of Human Resources*, 35(2): 221–258.

Baldacchinoel, Godfrey, and Anders Wivel. 2020. *Handbook on the Politics of Small States.* Cheltenham, Gloucestershire, UNITED KINGDOM:Edward Elgar Publishing Limited.

Becker, Gary S., and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy*, 87(6): 1153–1189.

Bhuller, Manudeep, Magne Mogstad, and Kjell G Salvanes. 2017. "Life-cycle earnings, education premiums, and internal rates of return, mortality." *Journal of Labor Economics*, 35(4): 993–1030.

Björklund, Anders, Markus Jäntti, and Matthew J Lindquist. 2009. "Family background and income during the rise of the welfare state: brother correlations in income for Swedish men born 1932–1968." *Journal of Public Economics*, 93(5-6): 671–680.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital." *American economic review*, 95(1): 437–449.

Blanden, Jo, Alissa Goodman, Paul Gregg, and Stephen Machin. 2004. "Changes in intergenerational mobility in Britain." *Generational Income Mobility in North America and Europe*, , ed. Miles Corak, 122–146. Cambridge University Press.

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." *Journal of Economic Literature*, 55(3): 789–865.

Blau, Francine D., Peter Brummund, and Albert Yung-Hsu Liu. 2013. "Trends in Occupational Segregation by Gender 1970-2009: Adjusting for the Impact of Changes in the Occupational Coding System." *Demography*, 50(2): 471–494.

Brandén, Gunnar, and Martin Nybom. 2019. "Utvecklingen av intergenerationell rörlighet i Sverige."

Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel D. Schnitzlein, and Kjell Vaage. 2017. "A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US." *The Scandinavian Journal of Economics*, 119(1): 72–101.

Bratsberg, Bernt, Simen Markussen, Oddbjorn Raaum, Knut Roed, and Ole Jorgen Røgeberg. 2018. "Trends in assortative mating and offspring outcomes."

Chadwick, Laura, and Gary Solon. 2002. "Intergenerational income mobility among daughters." *American Economic Review*, 92(1): 335–344.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014*a*. "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics*, 129(4): 1553–1623.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014*b*. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *The American Economic Review*, 104(5): 141–147.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014*c*. "Is the United States still a land of opportunity? Recent trends in intergenerational mobility." *American Economic Review*, 104(5): 141–47.

Connolly, Marie, Catherine Haeck, and Jean-William P Laliberté. 2020. "Parental Education Mitigates the Rising Transmission of Income between Generations." In *Measuring and Under-*

standing the Distribution and Intra/Inter-Generational Mobility of Income and Wealth. University of Chicago Press.

Corak, Miles. 2013. "Income inequality, equality of opportunity, and intergenerational mobility." *Journal of Economic Perspectives*, 27(3): 79–102.

Dahl, Molly W, and Thomas DeLeire. 2008. "The association between children's earnings and fathers' lifetime earnings: estimates using administrative data." University of Wisconsin-Madison, Institute for Research on Poverty.

Davis, Jonathan, and Bhashkar Mazumder. 2020. "The Decline in Intergenerational Mobility After 1980."

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2019. "Educational assortative mating and household income inequality." *Journal of Political Economy*, 127(6): 2795–2835.

Engzell, Per, and Carina Mood. 2021. "How Robust are Estimates of Intergenerational Income Mobility?"

Galassi, Gabriela, David Koll, and Lukas Mayr. 2021. "The Intergenerational Correlation of Employment." University of Bonn and University of Mannheim, Germany CRC TR 224 Discussion Paper Series.

Goldin, Claudia. 2014. "A grand gender convergence: Its last chapter." *American Economic Review*, 104(4): 1091–1119.

Grönqvist, Erik, Björn Öckert, and Jonas Vlachos. 2017. "The intergenerational transmission of cognitive and noncognitive abilities." *Journal of Human Resources*, 52(4): 887–918.

Harding, David J, and Martin D Munk. 2020. "The Decline of Intergenerational Income Mobility in Denmark: Returns to Education, Demographic Change, and Labor Market Experience." *Social Forces*, 98(4): 1436–1464.

Holmlund, Helena. 2022. "How much does marital sorting contribute to intergenerational socioeconomic persistence?" *Journal of Human Resources*, 57(2): 372–399.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and gender inequality: Evidence from Denmark." *American Economic Journal: Applied Economics*, 11(4): 181–209.

Kleven, Henrik Jacobsen. 2014. "How can Scandinavians tax so much?" *Journal of Economic Perspectives*, 28(4): 77–98.

Lee, Chul-In, and Gary Solon. 2009*a*. "Trends in Intergenerational Income Mobility." *The Review of Economics and Statistics*, 91(4): 766–772.

Lee, Chul-In, and Gary Solon. 2009*b*. "Trends in intergenerational income mobility." *The Review of Economics and Statistics*, 91(4): 766–772.

Lubotsky, Darren, and Martin Wittenberg. 2006. "Interpretation of Regressions with Multiple Proxies." *The Review of Economics and Statistics*, 88(3): 549–562.

Markussen, Simen, and Knut Røed. 2020. "Economic mobility under pressure." *Journal of the European Economic Association*, 18(4): 1844–1885.

Meghir, Costas, and Mårten Palme. 2005. "Educational reform, ability, and family background." *American Economic Review*, 95(1): 414–424.

Nybom, Martin, and Jan Stuhler. 2016. "Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation." *Journal of Human Resources*, 51(1): 239–268.

Nybom, Martin, and Jan Stuhler. 2017. "Biases in Standard Measures of Intergenerational Income Dependence." *The Journal of Human Resources*, 52(3): 800 – 825.

OECD. 2021. "LFS by Sex and Age." *https: // stats. oecd. org/ Index. aspx? DataSetCode= lfs_ sexage_ i_ r*, Accessed: 2021-02-25.

Olivetti, Claudia, and Barbara Petrongolo. 2016. “The Evolution of Gender Gaps in Industrialized Countries.” *Annual Review of Economics*, 8(1): 405–434.

Olivetti, Claudia, and M. Daniele Paserman. 2015. “In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940.” *American Economic Review*, 105(8): 2695–2724.

Pareliussen, Jon Kristian, Mikkel Hermansen, Christophe André, and Orsetta Causa. 2018. *Nordic Economic Policy Review*, 17–57.

Pekkarinen, Tuomas, Kjell G Salvanes, and Matti Sarvimäki. 2017. “The evolution of social mobility: Norway during the twentieth century.” *The Scandinavian Journal of Economics*, 119(1): 5–33.

Ruhm, Christopher J. 1998. “The economic consequences of parental leave mandates: Lessons from Europe.” *The Quarterly Journal of Economics*, 113(1): 285–317.

Selin, Håkan. 2014. “The rise in female employment and the role of tax incentives. An empirical analysis of the Swedish individual tax reform of 1971.” *International Tax and Public Finance*, 21(5): 894–922.

Søgaard, Jakob Egholt. 2018. “Top incomes in Scandinavia–recent developments and the role of capital income.” *Nordic Economic Policy Review 2018: Increasing Income Inequality in the Nordics*, 66–94.

Solon, Gary. 1999. “Chapter 29 - Intergenerational Mobility in the Labor Market.” In . Vol. 3 of *Handbook of Labor Economics*, , ed. Orley C. Ashenfelter and David Card, 1761–1800. Elsevier.

Solon, Gary. 2002. “Cross-Country Differences in Intergenerational Earnings Mobility.” *Journal of Economic Perspectives*, 16(3): 59–66.

Solon, Gary. 2004. "A model of intergenerational mobility variation over time and place." *Generational Income Mobility in North America and Europe*, , ed. Miles Corak, 38–47. Cambridge University Press.

Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie. 2020. "Long-term decline in intergenerational mobility in the United States since the 1850s." *Proceedings of the National Academy of Sciences*, 117(1): 251–258.

Vosters, Kelly. 2018. "Is the Simple Law of Mobility Really a Law? Testing Clark's Hypothesis." *The Economic Journal*, 128(612): F404–F421.

Vosters, Kelly, and Martin Nybom. 2017. "Intergenerational Persistence in Latent Socioeconomic Status: Evidence from Sweden and the United States." *Journal of Labor Economics*, 35(3): 869–901.

Appendix

3.A Data Registers and Variable Definitions

Denmark

The Danish income registries start in 1980 and contain detailed information on the individual income composition of Danish adults. The registries are based on information from the Danish tax authorities and supplemented with information from other Danish authorities, including unemployment insurance funds and the municipalities.

The measure of labor income that is being used in this paper consists of wage payments (incl. perks, non-taxable wage payments, stock options, and more) and any net surplus from own, private company. Gross income is equal to labor income, transfers, property income, and any other non-classifiable income that the individual may have received throughout the year. Net-of-tax income is finally equivalent to gross income net of all taxes that have been paid to either the government, municipalities, or other public authorities. Individuals with no parents in the sample (generally people who moved to Denmark, whose parents have moved abroad, or whose parents do not live anymore) are naturally dropped from the sample.

When constructing household income measures, individuals are being linked to their spouses. In the Danish sample, a spouse is generally defined by marriage, registered partnership or simply from the fact that they are registered as a cohabiting couple. Matching individuals to spouses as well as parents is based on the population registries of Denmark.

Norway

For the Norwegian part of the analysis, we are able to include birth cohorts from 1951 onward. We combine information from the central population registry with information about income and earnings from the tax registry. Income data in Norway is available from 1967 to 2018. Labor income, which includes payments related to employment, including overtime pay, taxable sickness, parental leave, short-term disability, and rehabilitation benefits, is top-coded for a few years in the 1970s at the maximum amount for contributions to the national social security scheme (folketrygden). Gross income is the sum of labor income and taxable and non-taxable transfers and income from capital. Disposable income is defined as gross income minus taxes and is also sometimes referred to as net-of-tax income. The definitions mentioned change to some degree over time due to reforms of the benefit, insurance, and tax system. For the net-of-tax and gross income variable, the data series ends in 2014, which is why these income measures are then constructed from more detailed income data only available from 1993. Spouses are linked through their personal identifiers and include married couples as well as couples in civil unions.

The occupation data used for implementing the method proposed by Lubotsky and Wittenberg (2006) is pooled from matched employer-employee data (Registerbasert sysselsettingsstatistikk) available annually starting with the year 2000. In addition occupation data from the censuses 1960, 1970, and 1980 are added. To achieve a comparable classification of occupations we use the STYRK-08 one-digit code to group individuals into broad occupational groups (see Table 3.B1). Individuals are assigned the occupation they have at age 36. In cases where this is not possible we use the closest applicable occupation we observe in the data. Due to the long break in occupational data between the 1980 census and the start of the employer-employee data, there might be some differences in the age at which we observe occupations for individuals that are also connected to the relevant birth year.

The educational data for the LW method is also pooled from different registries. Most individuals we observe are included in the national education database available from 1970. These data include variables for the highest achieved education of all individuals which we can link via personal identifiers. For individuals who are not included in the national education database, we try to obtain information about their educational attainment via census data from 1960, 1970, and 1980.

Sweden

The Swedish Income and Taxation registry starts in 1968 and holds official records of income for all individuals with any recorded income. In general, it contains all earned income from employment or businesses, capital income, taxable (mostly social insurances), and non-taxable transfers (social welfare, educational grants, child benefits, etc.). Identifiers for biological or adoptive parents are linked to the child identifier through the multi-generational register. Households are constructed by linking individuals (children, mothers, and fathers) to their spouses. This is available only for married couples (and those in registered partnerships) and thus excludes households formed by cohabiting partners.

Data on occupations are taken from two sources. First, the population censuses (Folk- och bostadsräkningarna) contain occupational codes corresponding to the ISCO-58 classification system. This information is available from 1960, and then every five years between 1970 and 1990 for the whole adult population. Individuals without an occupational code can be either classified as "undefined" or have a missing value. In our applications, both these are coded as missing. The census data are used to infer occupations for all parents in our Swedish sample, and we assign each parent an occupational code from the census closest in time to when the child is 18 years old (for example, a mother with a child born in 1951 will primarily be assigned an occupational code from the 1970 census, and occupations for fathers with children born in 1975 will be taken from the 1990 census). If no occupations is observed in this year, we search iteratively through the second and third closest waves, and so on. Parents who are missing an occupational code after this procedure, and who are at least 18 years old in 1960, are assigned occupations from that year's census. This mainly serves to capture occupations of women who are out of the labor force continuously after the birth of their first child; about 6.5 percent of the mother sample (3 percent of the fathers).

Occupational codes for the child generation are taken from the 1990 census for individuals born in the years 1951-1955, and from population register data for those born between 1956 and 1979. The population occupations register uses an adapted version of the ISCO-08 classifications, called SSYK 2012, and is available in our data for the years 2012-2017. As a result, the age at which occupations are observed among the child sample varies between 35 and 56, which might induce noise in between-birth cohort comparisons. On the other hand, this age span corresponds to prime working age, and occupational choice is relatively constant, especially given the broad classes we use in our analysis.

The highest attained level of education is observed in the 1970 census, and in the annual population registers that start in 1990. Each person is assigned the level of education that he or she displays in the year closest in time to when income is observed (age 36 for children; age 18 of the child for the parents). Years of education is then inferred from these categorical data (e.g. completing a three-year secondary education program is coded as twelve years of education, or eleven years if the person completed primary school when it was still only seven years in duration).

3.B Additional Figures and Tables

Code	Definition
Norway	
0	Armed forces and unspecified
1	Managers
2	Professionals
3	Technicians and associate professionals
4	Clerical support workers
5	Service and sales workers
6	Skilled agricultural, forestry and fishery workers
7	Craft and related trades workers
8	Plant and machine operators and assemblers
9	Elementary occupations
Sweden	
1	Professional work (arts and sciences)
2	Managerial work
3	Clerical Work
4	Wholesale, retail and commerce
5	Agriculture, forestry, hunting, and fishing
6	Mining and quarrying
7	Transportation and communication
8	Manufacturing
9	Services
10	Military/Armed Forces
Denmark	
0	Military work
1	Management work
2	Work that requires knowledge at the highest level in the area in question
3	Work that requires knowledge at intermediate level
4	Ordinary office and customer service work
5	Service and service work
6	Work in agriculture, forestry and fisheries
7	Craft and related trades workers
8	Operator and assembly work, transport work
9	Elementary occupations

Table 3.B1: Occupation Classification by Country

Note: Occupational categories for Norway are assigned using the STYRK-08 classificaiton provided by SSB. For Sweden the classification follows SSYK-2012 similar to Vosters and Nybom (2017). For Denmark, we use the first integer from the Danish ISCO classication (link). In the Danish case, note that this variable is not available for all years in the data. For this reason, we generate it from a set of other available occupation related variables. Code is available upon request.

Table 3.B2: Overview of Income Definitions by Country

		Denmark	Norway	Sweden
1	Salary	taxable salary incl. fringe benefits, tax-free salary, anniversary and severance pay and value of stock options	all payments related to employment including overtime pay	all payments from employment
2	Net Profit	net profit from self-employment incl. profit of foreign company and net income as employed spouse	net income from self-employment and income from other businesses	net profit from self-employment, income from other businesses
3	Transfers	cash benefits, unemployment insurance benefits, sickness benefits, unemployment benefits, pensions, child allowance, and more	taxable sickness benefits, parental leave benefits, unemployment benefits, short-term disability payments, rehabilitation benefits	sickness benefit from employer (sjuklön), value of e.g. car, travel expenses (förmånsvärden)
		Earnings/Labor Income = Combination of 1+2+3		

4	Transfers	cash benefits, unemployment insurance benefits, sickness benefits, unemployment benefits, pensions, child allowance, and more	taxable transfers: benefits from the national insurance scheme (disability insurance, pensions, etc.) non-taxable: child benefits, housing allowance, scholarships, parental leave benefits, social assistance payments	taxable: social insurances (unemployment, parental leave etc), private pension income, stipends etc. non-taxable: pensions and annuities housing support child support social welfare alimony conscript support grants and loans for students
5	Property Income	capital and wealth income excl. calculated rental value of real estate	gross interest income, dividend income, return on life insurance, net realised capital gains (e.g. shares, house, land), other capital income (taxable rental income)	capital income (gross pre-2004, net post-2004) and after-tax rental income
		Gross Income = Combination of Earnings +4+5		
6	Other Income	other non-classifiable income		
7	Taxes	taxes on earnings, wealth taxes, property value tax, tax on share dividends/gains and more	taxes, maintenance paid, mandatory insurance premia	all taxes
		Disposable Income = Combination of Gross Income +6-7		

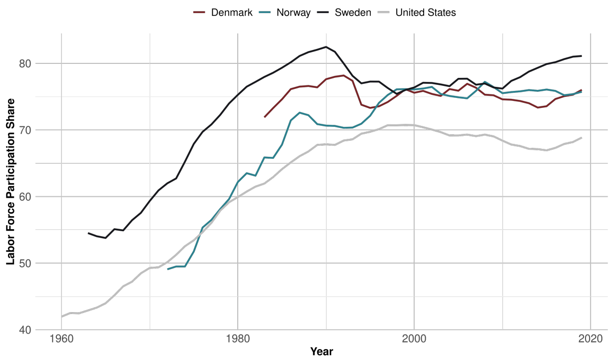


Figure 3.B1: Labor Force Participation Rate.
Note: The figure depicts the labor force participation rates of women aged 15 to 64 for Denmark, Norway, Sweden and the United States. The data was obtained from the OECD (2021) and covers all years available for the respective countries.

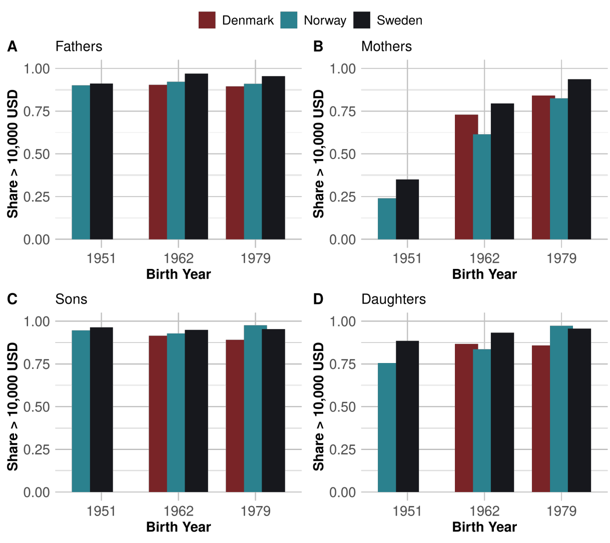


Figure 3.B2: Labor Force Participation.
Note: Each panel depicts shares of individuals with labor income exceeding 10,000 USD (2017) in Sweden, Denmark and Norway for the years 1951, 1962 and 1979. Panel A provides information for fathers, panel B mothers, panel C sons and panel D daughters.

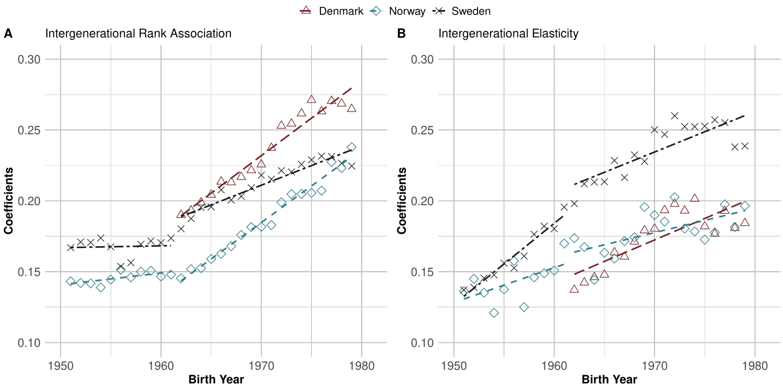


Figure 3.B3: Estimates of IRA and IGE in Labor Income.

Note: Panel A depicts intergenerational rank associations between parents and children for Sweden, Norway and Denmark, estimated as in eq. (1). Panel B shows intergenerational income elasticities, i.e. correlations in log income between parent and child pairs (with zero incomes excluded from analysis). Parental income averaged over child ages 17-19, and child income averaged over ages 35-37 in all estimates.

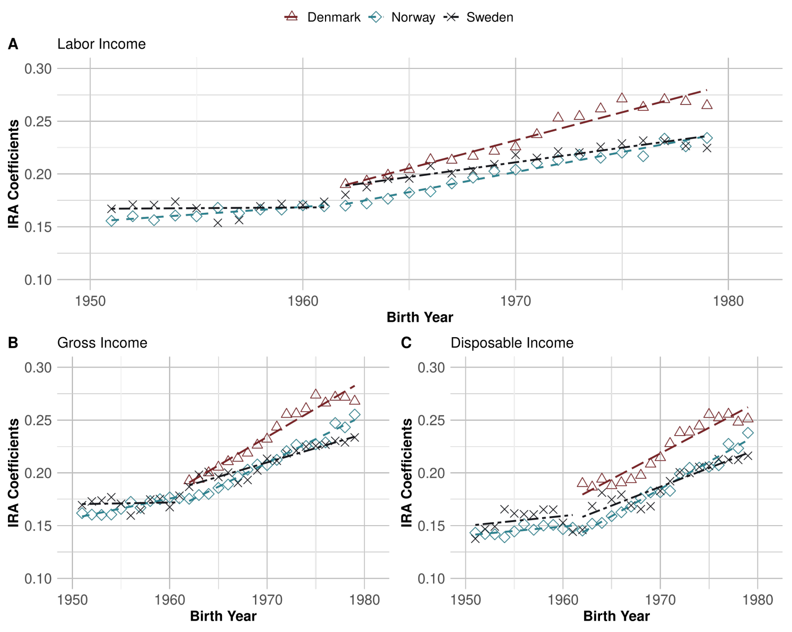


Figure 3.B4: Estimates of IRA in Net-of-tax, Gross and Labor income.

Note: Each panel depicts intergenerational rank associations between parents and children, estimated as in equation (1), for each country. Panel A shows estimates of the main specification: net-of-tax income. In panel B, total factor (gross) income is used, and panel C depicts labor earnings. Parental income averaged over child ages 17-19, and child income averaged over ages 35-37 in all estimates.

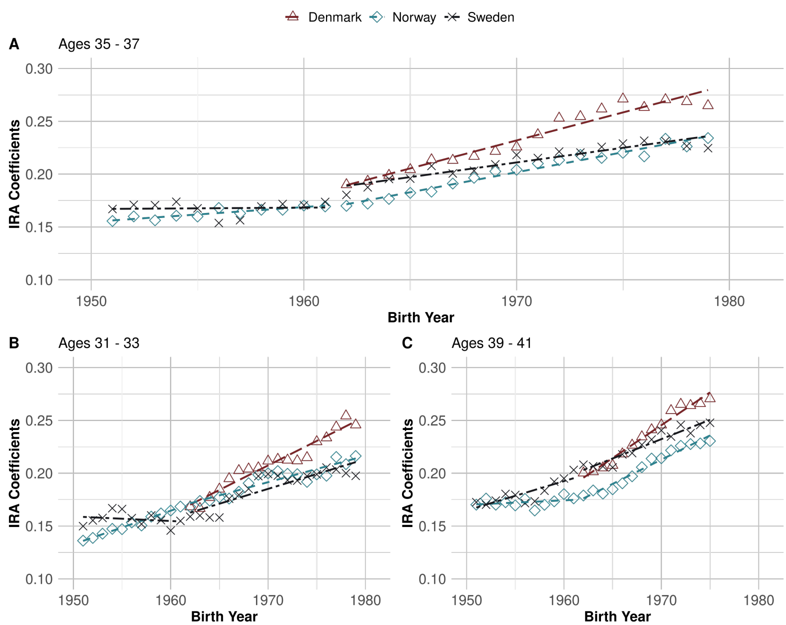


Figure 3.B5: Estimates of IRA at Different Ages of the Child (Labor Income).

Note: Each panel depicts intergenerational rank associations between parents and children, estimated as in equation (1), for each country. Panel A shows estimates of the main specification: average income at child ages 35-37. In panel B, child income is measured at ages 31-33, and in panel C, it is measured at ages 39-41. Parental income averaged over child ages 17-19 in all estimations.

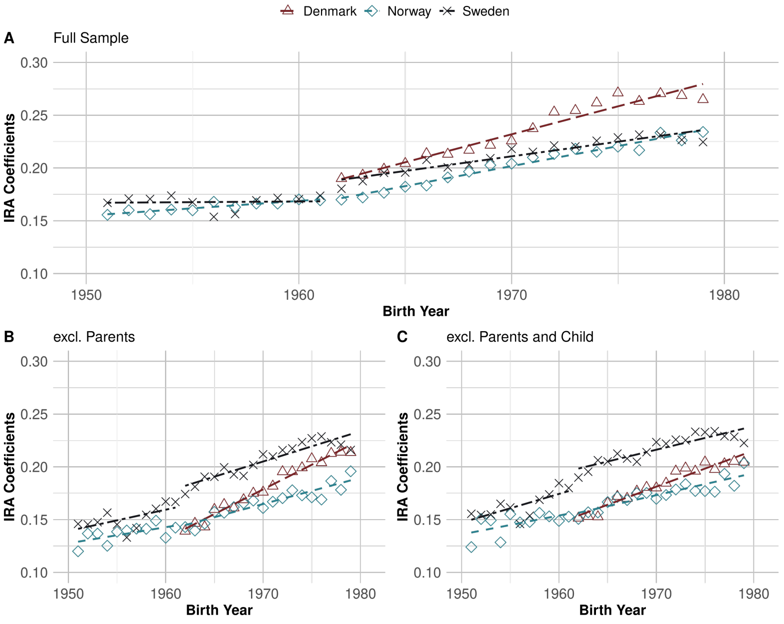


Figure 3.B6: Estimates of IRA, Labor Force Participants Only (Labor Income).

Note: Panel A depicts intergenerational rank associations between parents and children, estimated as in eq. (1), for each country. Panel B shows equivalent estimates of IRA, when excluding child-parent pairs where either parent earns less than 10,000 USD (2017) in a given year. In panel C, we additionally exclude child-parent pairs where both the child and the parents have incomes below the 10,000 USD threshold. Parental income averaged over child ages 17-19, and child income averaged over ages 35-37 in all estimates.

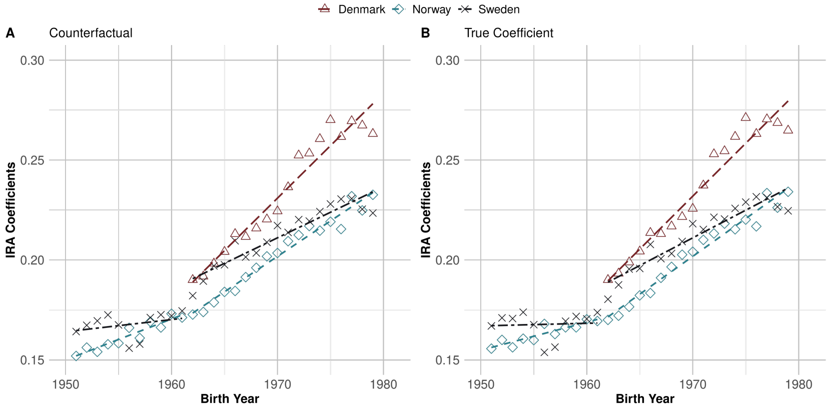


Figure 3.B7: IRA Estimates Accounting for Participation Differences.
Note: The two panels depict IRA coefficients by year for the the counterfactual and the true relationship between child and parental income for each country. Panel A presents the plot for the counterfactual where maternal incomes are changed to the corresponding percentile income in 1979. Panel B shows the true coefficients estimated from the data.

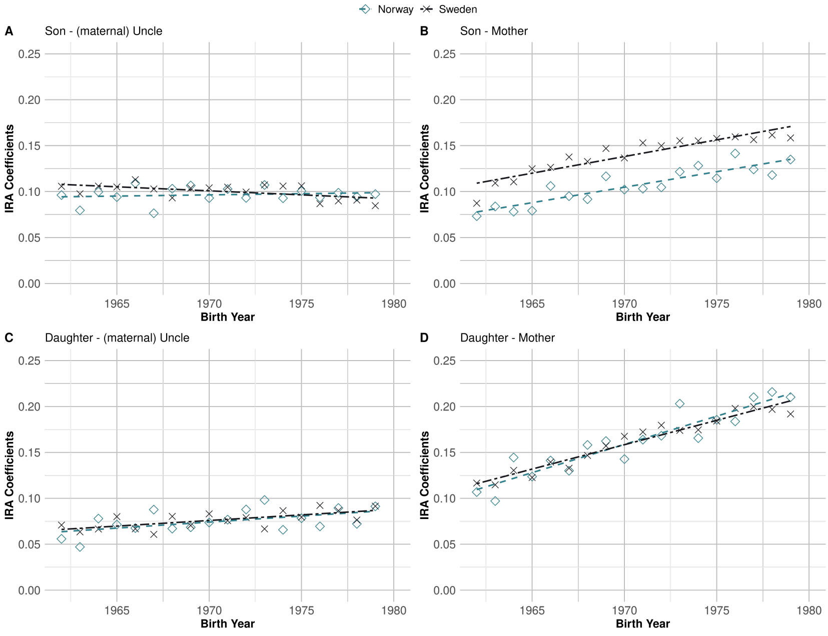


Figure 3.B8: IRA Estimates Between Children and Their Maternal Uncles.

Note: The four panels depict IRA coefficients by year for the income association between sons (Panel A) and daughters (Panel C) and their mothers' brothers. i.e. maternal uncles. Panels B and D show the estimated IRA between sons and daughters and their mothers for the sample where maternal brothers are applicable. Estimates are birth-year specific. Each panel depicts these measures separately by country for the years 1951, 1962 and 1979.

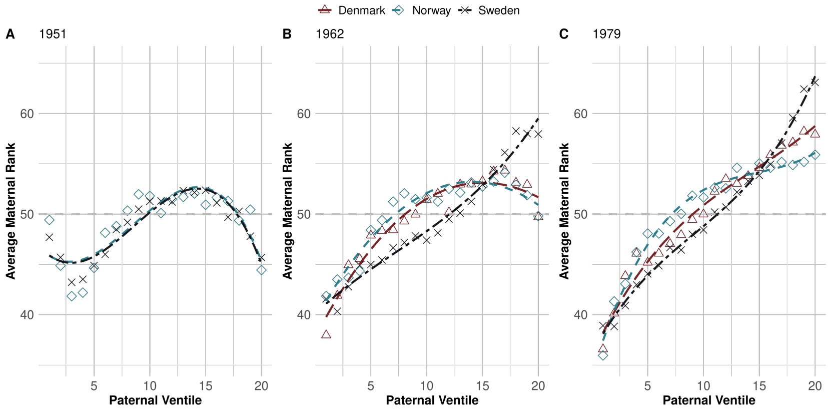


Figure 3.B9: Average Maternal Ventile Rank by Paternal Ventile.
Note: The three panels show the average maternal income rank of mothers with children in the same birth cohort, by paternal (within parental pairs) income ventile. Each panel depicts these measures separately by country for the years 1951, 1962 and 1979. The fitted lines in panel A to B are estimated with local polynomial (third order) regressions.

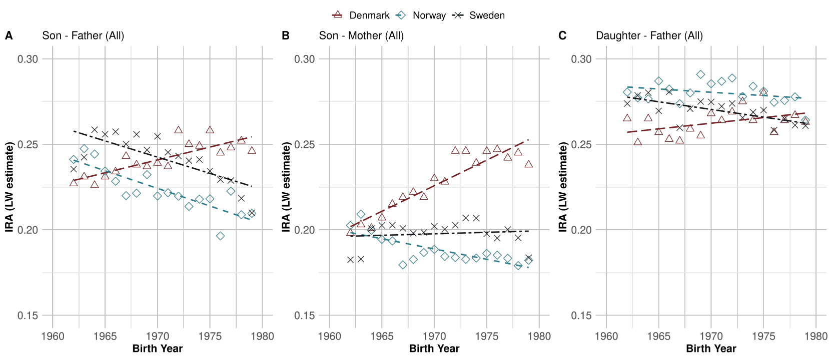


Figure 3.B10: Trends in Intergenerational Mobility in Latent Economic Status.
Note: The three panels plot coefficients for intergenerational rank associations in latent economic status for Denmark, Sweden and Norway over the period from 1951 (1962) to 1979. Panel A shows son-father correlations, panel B son-mother correlations and panel C daughter-father correlations. Each marker indicates the coefficient of a separate regression and each line indicates fitted trend lines for the period 1962 to 1979.

Table 3.B3: IRA Coefficients and Trends (United States)

	Parents	Father		Mother	
		Son	Daughter	Son	Daughter
Panel A:					
Pooled IRA	0.317***	0.336***	0.195***	0.097***	0.137***
	(0.017)	(.022)	(0.031)	(0.025)	(0.029)
Trend × 100	0.603***	-0.240	0.980***	0.136	1.047***
	(0.149)	(0.205)	(0.277)	(0.253)	(0.292)
N	5,392	2,272	1,637	2,477	2,205
Panel B:					
Pooled IRA	0.335***	0.360***	0.237***	0.107***	.152***
	(0.013)	(0.020)	(0.025)	(0.021)	(0.022)
Trend × 100	0.449***	-0.263*	0.728***	0.268	0.917***
	(0.118)	(0.178)	(0.229)	(0.202)	(0.213)
N	5,392	2,272	1,637	2,477	2,205
Panel C:					
Pooled IRA	0.294***	0.327***	0.192***	0.098***	0.126***
	(0.018)	(0.023)	(0.0353)	(0.026)	(0.032)
Trend × 100	0.433**	-0.393	1.156***	0.180	0.727
	(0.162)	(0.218)	(0.305)	(0.266)	(0.327)
N	2,927	1,583	904	1,497	1,001

Note: The table presents estimates of the IRA and linear trends in the IRA separately for different child-parent combinations. Due to the small sample sizes, trends have been estimated directly on the underlying micro data by regressing cohort-specific child ranks on cohort-specific parent ranks interacted with a linear time trend. The trend coefficients and standard errors have been multiplied by 100 in order to avoid too many digits after the separator. Panel A contains estimates for the full PSID sample using provided sample weights, Panel B uses the full sample without weights and Panel C includes estimates on the nationally representative SRC sample. Standard errors are in parentheses. P-values indicated by * < 0.1, ** < 0.05, *** < 0.01.

Table 3.B4: IRA Coefficients and Trends - Age 30 (United States)

	Parents	Father		Mother	
	Child	Son	Daughter	Son	Daughter
Panel A:					
Pooled IRA	0.327***	0.318***	0.222***	0.120***	0.151***
	(0.015)	(0.022)	(0.026)	(0.025)	(0.027)
Trend × 100	0.643***	0.133	0.661***	0.610**	0.571**
	(0.129)	(0.193)	(0.220)	(0.239)	(0.262)
N	6,652	2,664	2,109	2,685	2,611
Panel B:					
Pooled IRA	0.345***	0.341***	0.263***	0.148***	0.168***
	(0.012)	(0.018)	(0.021)	(0.019)	(0.020)
Trend × 100	0.457***	0.102	0.510***	0.429**	0.567***
	(0.101)	(0.59)	(0.183)	(0.176)	(0.181)
N	6,652	2,663	2,109	2,686	2,611
Panel C:					
Pooled IRA	0.303***	0.310***	0.225***	0.097***	0.133***
	(0.016)	(0.023)	(0.028)	(0.027)	(0.030)
Trend × 100	0.528***	0.020	0.586**	0.661**	0.352
	(0.146)	(0.210)	(0.245)	(0.261)	(0.307)
N	3,451	1,757	1,161	1,460	1,142

Note: The table presents estimates of the IRA and linear trends in the IRA separately for different child-parent combinations Children's income is measure at age 30. Due to the small sample sizes, trends have been estimated directly on the underlying micro data by regressing cohort-specific child ranks on cohort-specific parent ranks interacted with a linear time trend. The trend coefficients and standard errors have been multiplied by 100 in order to avoid too many digits after the separator. Panel A contains estimates for the full PSID sample using provided sample weights, Panel B uses the full sample without weights and Panel C includes estimates on the nationally representative SRC sample. Standard errors are in parentheses. P-values indicated by * < 0.1, ** < 0.05, *** < 0.01.

Table 3.B5: PSID Parent-child Links by Cohort, Main Spec.

	Parents	**Father**		**Mother**	
Birth year		Son	Daughter	Son	Daughter
1947	76	27	26	35	39
1948	107	38	39	48	56
1949	143	52	44	73	65
1950	171	57	72	68	99
1951	218	76	81	101	108
1952	193	70	78	88	102
1953	239	87	98	116	117
1954	235	78	96	101	128
1955	267	103	98	122	136
1956	263	86	106	107	147
1957	247	95	85	126	115
1958	220	75	95	95	119
1959	159	79	37	107	46
1960	177	96	34	121	54
1961	105	54	27	62	38
1962	117	50	37	63	53
1963	125	49	41	64	56
1964	100	47	30	56	43
1965	91	42	20	49	40
1966	88	39	17	52	34
1967	95	49	21	60	33
1968	66	32	17	39	26
1969	99	49	32	55	42
1970	87	40	20	50	35
1971	92	40	21	60	32
1972	111	49	22	63	43
1973	107	50	20	65	36
1974	117	51	25	64	48
1975	128	55	36	70	48
1976	132	58	37	63	53
1977	130	66	23	41	21
1978	138	66	34	27	30
1979	179	93	43	29	41
1980	142	65	36	33	26
1981	148	77	30	40	24
1982	120	58	27	26	30
1983	160	74	32	38	42
Total	5,392	2,272	1,637	2,477	2,205

Note: The table presents the number of cohort-specific parent-child links that were used to produce the main results from the PSID survey data.

Table 3.B6: IRA Coefficients, Trends and Differences Across Countries and Time

	1951		1962			1979			Trend 1962-1979			Δ P-Value		
IRA Spec.	NO	SE	DK	NO	SE	DK	NO	SE	DK	NO	SE	DK-NO	DK-SE	NO-SE
All	0.156	0.167	0.190	0.170	0.180	0.265	0.234	0.225	0.530	0.379	0.277	0.065	0.004	0.176
									(0.035)	(0.018)	(0.033)			
Son-Parent	0.242	0.245	0.225	0.222	0.233	0.280	0.241	0.235	0.360	0.085	0.014	0.000	0.000	0.736
									(0.036)	(0.024)	(0.061)			
Daughter-Parent	0.146	0.158	0.197	0.173	0.169	0.276	0.262	0.240	0.592	0.552	0.428	0.363	0.067	0.020
									(0.035)	(0.037)	(0.038)			
Son-Father	0.253	0.248	0.220	0.236	0.242	0.241	0.213	0.211	0.139	-	-	0.000	0.000	0.378
										0.160	0.224			
									(0.035)	(0.022)	(0.060)			
Son-Mother	0.068	0.080	0.098	0.089	0.101	0.198	0.150	0.155	0.619	0.324	0.307	0.000	0.007	0.723
									(0.026)	(0.026)	(0.041)			
Daughter-Father	0.137	0.139	0.175	0.144	0.152	0.216	0.195	0.192	0.342	0.325	0.239	0.778	0.570	0.421
									(0.031)	(0.032)	(0.034)			
Daughter-Mother	0.073	0.077	0.120	0.119	0.112	0.227	0.221	0.194	0.731	0.607	0.541	0.439	0.181	0.107
									(0.036)	(0.037)	(0.034)			

Note: Columns (1)-(7) report the IRA coefficients of separated regressions in the years 1951, 1962 and 1979 separately for Denmark, Norway and Sweden. Columns (8)-(10) report the coefficient of the fitted regression lines of country specific regressions of the IRA coefficient on a linear trend for the years 1962 to 1979. The trend coefficients and corresponding standard errors have been multiplied by 100 in order to avoid too many digits after the separator. Columns (11)-(13) report rounded p-values for the null hypothesis that the slopes for the respective countries (see column header) are equal. Robust standard errors are reported in parentheses.

3.C Calibrating Parameters in Model

Each set of country-year model parameters for trend decomposition are — loosely described — calibrated in the following steps:

1. If the year is the first year of observation for a given country, draw a random set of parameters. If the year is not the first year of observation, initialize the algorithm with the optimal set of parameters from the last year associated with the same country. These become the 'search parameters' until they are replaced.

2. Draw 100,000 parent-child pairs (the same in each year), and repeat the following procedure until there is a sufficiently close match between empirical rank associations and modeled rank associations[22]:

   (a) Compute skills and incomes for all individuals (father, mother, son and daughter) using the set of 'search parameters' along with randomly drawn values for x_{it}^{k} and ε_{it}^{k}.

   (b) Compute associations in income ranks between (i) fathers and sons, (ii) fathers and daughters, and (iii) mothers and sons, and (iv) mothers and daughters, while (v) matching the relationship between mother and father income ranks.

   (c) If the convergence criterion is not met, adjust the parameters using a customized variation of gradient descent. These now become the ¿search parameters'.

In the following set of graphs, we illustrate how the the implied empirical association between the two types of income in the (calibrated) simulated data compares to the empirical association between the same two incomes as observed in the data. These figures validate the quality of the calibration exercise.

[22] Or stop the algorithm early if it stops converging, i.e. it seems that a much better match cannot be achieved.

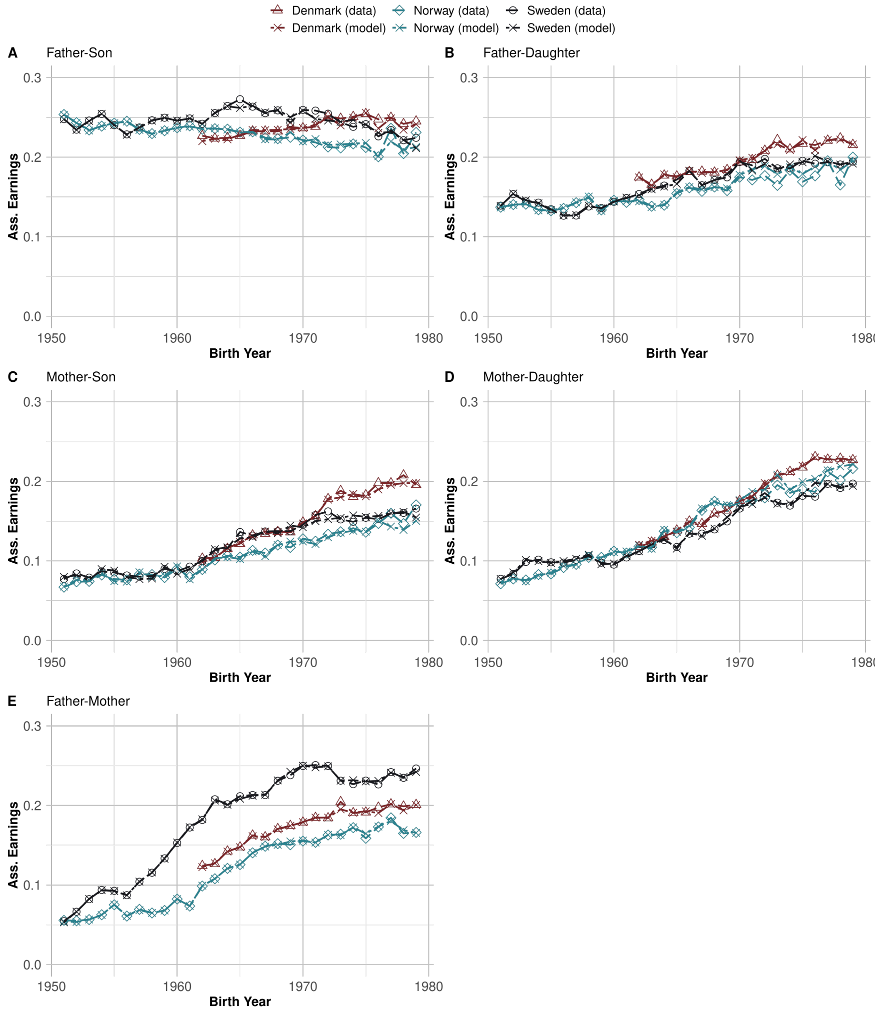


Figure 3.C1: Validation of Calibration Exercise.

Note: Each panel displays the empirical association between two incomes as observed in the data as well as the implied empirical association between the same two types of income in the simulated data as calibrated in the decomposition model.

Chapter 4

Wage Inequality, Selection and the Evolution of the Gender Earnings Gap in Sweden*

*This chapter is written together with Susan Niknami (Swedish Institute for Social Research) and Mårten Palme (Department of Economics, Stockholm University). Thanks to Anders Björklund, Anne Boschini, Hans Grönqvist, David Seim, David Strömberg and participants on seminars at the Department of Economics as well as the Institute for Social Research at Stockholm University for comments on earlier version of the paper. We are also very grateful to David Seim for providing us access to data.

4.1 Introduction

While economic differentials between men and women is one of the few areas where inequality has consistently decreased over the past decades in high income countries (see e.g. Blanchard and Rodrik, 2021) there is widespread disappointment with the pace at which it is happening (e.g. Goldin, 2014; Blau and Kahn, 1997, 2017; Olivetti and Petrongolo, 2008, 2016). To get an exhaustive picture of how the gender pay gap has evolved over time, it is however necessary to adjust for the most crucial movements in the labor market that have taken place in rich countries over this period, i.e. shifts in working hours, the widening of the wage distribution and changes in the gender composition of the work force.

In this paper, we measure the evolution of the gender pay gap in Sweden between 1968 and 2019, accounting for these key changes of the labor market. Sweden provides an interesting example for the study of changes in women's economic position as the rise in female labor force participation precedes the same development in most other industrialized countries. The empirical analysis is carried out in three steps. We start by measuring the unadjusted gender pay gap using observed weekly earnings. We then adjust for shifts in intensive margin labor supply by using observed hourly wages. Next, we use a percentile specific deflator, proposed by LaLonde and Topel (1992) to adjust for the great compression of the wage structure in the first era of the analysis (1960s-1990s) (see e.g. Hibbs, 1990) and the widening of the wage distribution in the second era of the analysis (1990s-2010s). Finally, in the third step, we account for non-random selection into the labor market, meaning that we estimate the gender gap in wage offers rather than observed wages.

Past literature on gender pay inequality has used three main methods to correct for potential selection: (i) parametric models based on the seminal paper of Heckman (1979); (ii) non-parametric bounds (following Manski, 1994, and Blundell et al., 2007); (iii) imputations using panel structures and observable human capital characteristics (as in e.g. Olivetti and Petrongolo, 2008, and Blau et al., 2021); (iv) sub-group analysis with high participation (following Chamberlain, 1986, and Heckman, 1990). However, as most recently pointed out by Blau et al. (2021), each of these methods has its limitations - ranging from strong identifying assumptions to difficulties in making conclusive inference due to wide bounds and unrepresentative samples. As a result, we implement all these methods to obtain a robust pattern of the evolution of the gender wage gap.

We use two main data sources. First, data from the 1968, 1974, 1981, 1991, 2000 and 2010 waves of the Swedish Level of Living surveys (SLLS). The original samples for these surveys consist of about 6,000 individuals in the ages 18-74, or about 0.1 percent of the Swedish population. Although the SLLS is a panel, a supplementary sample is made for every wave to ensure that each of the surveys have the properties of a random sample from the Swedish population corresponding to the year of the survey. Our motivation for using these data, rather than population register data, is that the SLLS includes detailed information from personal interviews on e.g. hours of work as well as on occupations and family situations, used in various stages of our empirical analysis.

Second, in order to extend the analysis to the more recent development, we also use data from the Wage Structure Statistics between 1995 and 2019. Like the SLLS data, it contains measures of hourly wage rates and information on hours of work. Unlike the SLLS, it is not a random sample, but contains information on the entire public sector as well as the private sector employees in companies with more than 500 employees. For smaller companies, the data set includes information from a random sample to ensure that the total sample, including weights, is representative for the entire Swedish labor force.

Our results show that the evolution of the gender wage gap appears very differently, depending on how one chooses to measure it.

1. Measuring the gender gap using observed weekly earnings (excluding zeros) gives a change from 66 percent in 1968 to 24 percent in 2010 – a 42 percentage point (64 percent) decrease in the gender gap over time. Until year 2019, we estimate that the gap has closed further, to a level of 20 percent. When using observed hourly wage instead, the narrowing of the gender gap is not as pronounced – a 13 percentage points (48 percent) decrease from 27 percent in 1968 to 14 percent in 2010. In 2019, the estimated wage gap in hourly wages is at 10 percent.

2. Correcting for changes in overall wage inequality has a strong impact. During the era of wage compression, between 1968 and 1991, our results show that the entire change in the gender wage gap, from 27 percent in 1968 to 19 percent in 1991, can be attributed to changes in overall wage inequality. For the era of increasing wage inequality, between 1991 and 2019, our results show that the change in the gender gap in observed wages from 19 to 11.1 percent (7.9 p.p.) is more than twice as large – from 27.5 to 14 percent (13.5 p.p.) - if one corrects for changes in the overall wage distribution.

3. Selection into employment has a big influence on the measured gender wage gap during the first half of the period under study – from 1968 to 1991 – since the great increase in female labor force participation in Sweden took place during that period. The 3.5 percentage points change in the gender gap in median distributional corrected wages between 1968 and 1991 – from 26.5 to 23 percent – would have been 16.5 – from 39 to 22.5 percent – when correcting for sample selection under the assumption of positive selection into the labor market within age and education level cells.

This paper relates to a very large literature on the evolution of the gender wage gap in various countries. The relation between the general wage inequality and the gender gap in earnings has been studies by Blau and Kahn (1996, 1997) and Edin and Richardsson (2002) among others. Blau and Kahn (1997) studies how increased general wage inequality in the US has affected the gender gap and Edin and Richardsson (2002) use Swedish data for the period 1968 to 1991 to study the effect of wage compression. The long period covered by our data in this study allows us to study the effects of both decreased *and* increased wage dispersion. In particular, we extend the analysis of Edin and Richardsson to also cover the period of increased wage dispersion after the 1990s on Swedish data.

Mulligan and Rubinstein (2008) shows that the increased general wage dispersion in the US starting in the 1980s led to that high skilled women to a larger extent entered the labor market. The changed selection of females into the labor market led, in turn, to a decrease in the gender gap in observed wages.[1] The implications of selections under study in this paper are driven by the opposite case. The decreased general wage dispersion in Sweden in the late 1960s and throughout 1970s, paired with increased labor force participation of primarily low skilled women generated, all else equal, an increase in the gender gap in observed wages. According to the cross-country analysis presented in Olivetti and Petrongolo (2008), positive selection into the labor market, and a negative bias from estimating the gender wage gap in observed wages, is more prevalent in most western industrialized countries. The empirical results, as well as the overview of previous evidence, provided in Blau et al. (2021) support this view.

We make several contributions to the literature. Most importantly, we show how each of the three most fundamental changes in the labor market that have taken place in all rich countries in recent decades, i.e. the changes in overall wage inequality and the increase in female labor force participation and hours of work, have affected income inequality between men and women. By using the same data for measuring all income concepts, we are able to assess the relative importance of each of the three changes, respectively. While past studies have tried to account for these movements separately, our study stresses the importance to account for these co-movements simultaneously – not at least since they overlap in time. We also contribute by documenting the long-term trend of the gender gap, a 42-year period, stretching from just before the dramatic rise in women's labor force participation until today. Last, we contribute by providing Swedish evidence on the development on the gender gap for each of the three measures separately. No previous study have shown the implications of both the great contraction of the wage distribution in Sweden beginning in the late 1960 *and* the increase in wage inequality starting in the 1990s. The same is true for the implications of the changes in labor force participation over the decades. Since a similar development have taken place in most industrial countries we believe that these evidences have external validity to the development in other economies and not only shed

[1] Blau and Beller (1988) obtains somewhat contradictory results for the same historical development.

light on an important historical development in Sweden.

The rest of the paper is organized as follows. Section 4.2 describes our data and sample restrictions. Sections 4.3, 4.4 and 4.5 describe methods and present results when adjusting the wage gap for female hours of work, wage dispersion and selection, respectively. Section 4.6 discusses the results and Section 4.7 concludes.

4.2 Data

We use two different data sets. The first one is compiled from six different waves of the Swedish Level of Living Survey (SLLS) collected in 1968, 1974, 1981, 1991, 2000 and 2010, respectively. The SLLS is a panel survey in the sense that all individuals are re-interviewed, although the sampling design assures that each independent survey is a representative sample for the Swedish population aged between 18 and 74.[2] The sample size is intended to be 0.1 percent of the Swedish population in the age group included in the survey, which means around 6,000 individuals in each wave. In the year of the surveys, the participants respond to detailed questions about their economic resources, hours of work, occupation and educational attainments.

We do a number of restrictions to our estimation sample. Table 4.1 shows how the sample sizes change when the restrictions are imposed. First, we only include individuals in working ages, and therefore restrict the sample to individuals between the ages of 20 and 64. Second, we exclude farmers, self-employed and workers in freelance occupations. We also exclude "assistants" in these occupations. For all waves, except the 1968 one, we extract this information from detailed questions on what respondents were doing the week before the survey: if they were employed full time or part time; self-employed or assistant to self-employed, farmers or involved in a family farm; freelance workers; doing household work; or unemployed searching for a job.

For the 1968 survey, individuals are classified as working or not

[2]The SLLS data has frequently been used in economics to study various aspects of the labor market (e.g. Blau and Kahn, 1996; Lindh and Ohlsson, 1996; Björklund and Jäntti, 1997; Johansson and Palme, 2002; Böhlmark and Lindquist, 2006 and Blomquist and Selin, 2010).

Table 4.1: SLLS Sample Size and the Sample Selection Process.

	1968	1974	1981	1991	2000	2010
Ages 20-64	4,551	4,402	4,295	4,297	4,394	3,587
Not farmer/self-emp./military	3,927	3,721	3,708	3,606	3,839	3,127
Not student	3,823	3,572	3,590	3,423	3,487	2,904
Observed wage	2,633	2,694	2,988	2,937	2,757	2,371
Missing wages of workers imputed	2,706	2,786	3,016	2,983	2,867	2,421
Not employed	1,117	786	574	440	620	483

Notes: Each row describes a step in the sample selection process, showing the remaining number of individuals. *Not student* represents the total sample for each survey wave. "Missing wages imputed" represents the working sample of observed wages. "Not employed" shows the number of individuals outside the labor force, in each sample year.

primarily based on whether they say they worked any days last week. The questions on activities other than work refers to the year before the year of the survey, i.e., in 1967, rather than week before the survey was done. It also contains information on number of weeks the respondents were involved in different activities. For most activities the answers are binned in eight categories: 52 weeks, 50-51 weeks, 40-49, 25-39, 10-24, 5-9, 3-4 or 1-2. We assign an individual to a particular activity if he or she devotes 10 or more weeks to the activity. Third, we exclude students who report that they work less than 20 hours per week. This gives our final sample of working and non-working individuals.

In each wave of the survey, a small fraction (maximum one percent) of individuals in the work force fail to report their earnings or wages. In order to separate these from the unemployed or non-employed, hourly wages are imputed using a Mincer-type linear prediction for these individuals.[3] The fifth row of Table 4.1 shows the number of

[3]The wage equation $Wage_{it} = Female_{it} + Education_{it} + Exp_{it} + Exp_{it}^2$ is

observed wages in each survey year, including imputed wages. This is the working sample of *observed wages* used in subsequent analysis. As a reference, the last row of Table 4.1 shows the yearly number of observations outside the workforce.

We use the measure of the hourly wage rate calculated by the SLLS surveys to be comparable across different waves of the survey and labelled *Gross hourly wage*. The variable is constructed from survey information on whether the respondent receive their earnings on the basis of hourly, weekly or monthly payments, their stated earnings and regular hours of work.

We also use measures of weekly earnings. This measure is constructed using each individual's stated length of a standard working week from the SLLS. This is multiplied with the hourly wage rate to obtain weekly earnings. In cases where weekly hours worked are missing for working individuals (maximum one percent of observations per survey year), the observation is dropped from the sample.

Table 4.B1 provides summary statistics for our sample, and Appendix Figure 4.B2 summarizes the age structure of the sample for each survey wave.

Our second main data set is obtained from merging individual information from the Longitudinal Integrated Database for Health Insurance and Labour Market Studies (LISA) with the Wage Structure Statistics (WSS) register for all years between 1995 and 2019. The LISA register mainly contains information from the Taxation Register, registers including transactions from the social security administration and the National Education Register, which includes information on educational attainments for all individuals living permanently in Sweden. By combining these two annual data sources we are able to construct a data set of repeated cross sections including the entire Swedish population aged 25-64 for each year between 1995 and 2019.

The main advantage with the Wage Structure Statistics data for our purposes, compared to the data from the Taxation register, is that it contains information on contracted hours of work as well as

estimated on all individuals i with an observed wage in year t. Wages in a given year are then predicted linearly for all in the workforce, and then transformed to logarithms according to: $\ln Wage_i = \ln Wage_i + 1$.

data on hourly wage rates.Wages are recorded as monthly pay in September or November of a given year, depending on the sector. We divide this measure with contracted weekly hours to construct hourly wages. This allows us to separate out the part of the gender wage gap that is, in a mechanical sense, attributed to gender differences in hours of work. The disadvantage with this data set compared to the LISA register is that it does not include the entire population. It contains, however, all individuals employed in organizations in the public sector as well as on all employees in companies with more that 500 employees. The data also includes a random sample of private companies with less than 500 employee, which includes roughly 50 percent of these firms.

We construct our main analysis sample by keeping all individuals with a wage record in the Wage Structure Survey, and a representative sample of individuals not participating in the labor force. We use the LISA population register to define participation as an indicator variable for having annual income in the tax records exceeding one basic amount (46,500 SEK in 2017).[4] Next, we calculate the share of the population with non-missing wages, i.e. the share of the working population covered by the WSS, and take a random sample of the non-working population of corresponding size. The final data set thus covers just over half of the Swedish working-age population also including a random sample of non-participants proportional in size to the share of non-participants calculated from the LISA register.

4.3 The Gender Gap in Earnings and Wages

A major source of earnings inequality between working men and women is differences in intensive margin labor supply. Figure 4.1 shows that the mean number of working hours per week among women is consistently lower than among men. Women who entered the labor force in the 60s and 70s did not generally work full time, and though the average length of women's work weeks has increased, it is in 2019 still shorter than the male average (35.8 and 38.2 hours per week, respectively). To assess the importance of these differences,

[4] Individuals with missing or zero wages but annual income exceeding one basic amount are excluded from the sample.

Table 4.2: Descriptive Statistics, Wage Structure Survey

	Males			Females		
	Wage	Population	Sample	Wage	Population	Sample
1995	18,429	2,294	1,033	15,054	2,229	1,378
1996	19,658	2,314	1,104	15,855	2,248	1,418
1997	20,522	2,332	1,143	16,587	2,265	1,441
1998	21,445	2,348	1,096	17,177	2,281	1,397
1999	22,424	2,365	1,127	18,033	2,298	1,432
2000	23,386	2,381	1,168	18,753	2,313	1,459
2001	24,491	2,395	1,152	19,644	2,326	1,451
2002	25,338	2,407	1,172	20,478	2,338	1,468
2003	26,101	2,414	1,179	21,269	2,346	1,481
2004	26,944	2,421	1,181	22,001	2,354	1,484
2005	27,797	2,431	1,180	22,677	2,364	1,486
2006	28,617	2,443	1,187	23,400	2,376	1,494
2007	29,523	2,453	1,177	23,971	2,384	1,491
2008	31,068	2,457	1,179	25,379	2,389	1,512
2009	31,938	2,462	1,181	26,370	2,394	1,511
2010	32,513	2,465	1,190	26,961	2,397	1,516
2011	33,378	2,469	1,131	27,703	2,401	1,481
2012	34,335	2,475	1,128	28,480	2,408	1,477
2013	35,071	2,489	1,133	29,359	2,421	1,476
2014	35,907	2,513	1,144	30,262	2,44	1,492
2015	36,607	2,542	1,157	31,123	2,464	1,508
2016	37,233	2,587	1,186	32,021	2,499	1,542
2017	37,807	2,624	1,199	32,832	2,532	1,558
2018	38,606	2,655	1,220	33,684	2,561	1,575
2019	39,508	2,655	1,212	34,609	2,561	1,568

Notes: “Wage” refers to average monthly wage. “Population” shows total population sample size, and “Sample” shows the Wage Structure Survey population plus our sample of non-workers (see description above).

we compare the gender gap in weekly wages to that in hourly wages.[5]

[5] Weekly hours may also have a negative effect on their wage rate per hour, if we believe that employers are unlikely to reward employees who are not able to

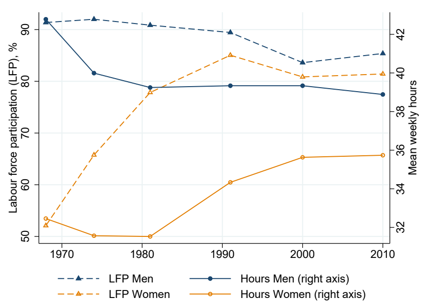
The following model is estimated separately for each survey year:

$$w_{it} = \alpha_t + \beta_t Female_i + e_{it}, \tag{4.1}$$

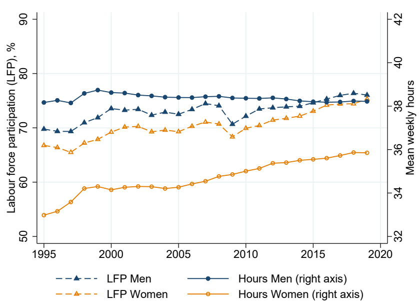
where w_{it} is log weekly or hourly wages, $Female$ is an indicator variable taking the value one for women, i denotes individuals in the work force and, finally, t represents the year. e_{it} is the i.i.d. error term. Thus, β_t measures the percentage difference in wages between men and women in the work force in year t.

The upper two panels of Figure 4.2 plots estimates on the SLLS data of the gender wage gap in weekly and hourly wages between 1968 and 2010 and the lower two panels show the corresponding estimates for the Wage Structure Statistics database for the period 1995 to 2019. The left panels show the estimates for the mean log gender wage gaps and the right hand side panels the corresponding estimates at the median.

Figure 4.1: Labor Force Participation and Hours Worked, by Gender.



(a) SLLS, 1968-2010



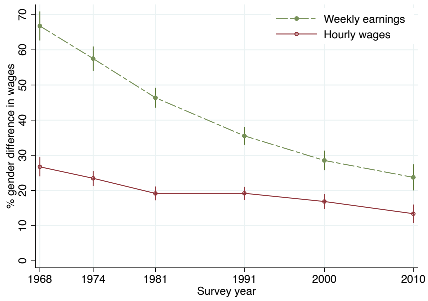
(b) WSS, 1995-2019

Notes: LFP (labor force participation) denotes extensive margin participation, i.e. the mean of an indicator variable equal to one for workers self-stated in the labour force. Hours denotes weekly hours conditional on employment, from survey responses to contracted number of hours in a normal work week. Panel a: Swedish Level of Living Survey, Panel b: Wage Structure Statistics.

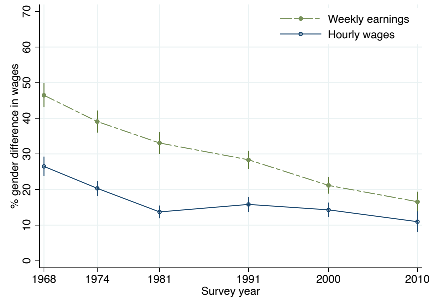
The SLLS estimates of log weekly pay show that the wage gap narrows over time at an almost linear rate, from around 70 percent in

work full time, or unwilling to hire such workers in the first place.

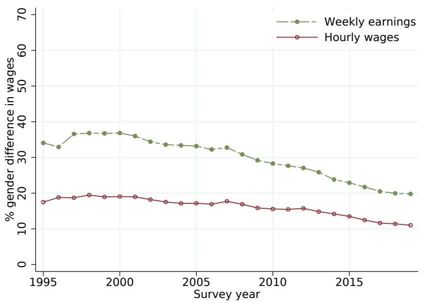
Figure 4.2: Observed Gender Gap in Weekly and Hourly Wages, at Mean and Median.



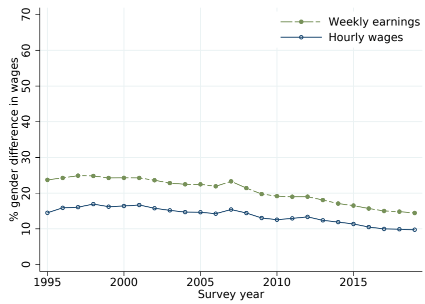
(a) SLLS Mean 1968-2010



(b) SLLS Median 1968-2010



(c) WSS Mean 1995-2019



(d) WSS Median 1995-2019

Notes: Coefficients from regression of log(wages) on female dummy, at mean and median. Missing wages for employed individuals are predicted out-of-sample with a wage regression, as described in Section 4.2.

1968 to 24 percent in 2010. In other words, the gender gap in weekly wages closes by more than 45 percentage points over the course of this time. The estimates from the Wage Structure Statistics data show that this trend continues after 2010, to 20 log points in 2019. Taken together, this implies that there is a 50 percentage point change over the entire period.

The gender gap at the median is consistently below that at the mean from 1968 and onward. This is a well known pattern of results observed in previous research on Swedish data (see e.g. Albrecht et al., 2003) indicating that the male wage distribution is more top heavy than that of females, and thus suggesting a “glass ceiling” in women’s wages. The measured change over the entire period is also much smaller for the median measure, both in the SLLS and the WSS results. However, the pattern of the changes are very similar between the median and the mean wage gap.

Working hours matter greatly for observed gender income differences. When we measure the gender gap in hourly wages, the gender gap is estimated at about 27 percent in 1968, 12 percent in 2010 and at 10 percent in 2019. The observed wage gap thus closes by 17 percentage points between 1968 and 2019 at the mean, and almost half of this change, 7.5 percentage points, occurs between 1968 and 1991. The results obtained from the measures obtained at the median gives overall a very similar picture.

The much smaller change in the wage gap measured in log hourly wages compared to log weekly wages (17 pp *versus* 50 pp) tells us that the largest part - about two thirds - of the change in the gender gap can be attributed to decreasing gender differences in hours of work.

To assess the potential effect of working part time on hourly wage rates, Appendix Table 4.B1 shows that splitting the sample into those who report working full time and those who work part time, respectively, the gender wage gap for part-time workers greatly exceeds that for full-time workers in 1968. In 1974, however, the gender wage gap is at the same level for both groups, and from then on, it is smaller for part-time than for full-time workers. This could indicate that the part-time wage penalty is greater for men than for women. Another explanation is that men who work part time are a negatively selected group, more so than women with part-time jobs. Historically, the reasons for working part-time are likely to differ between men and women. In 2010, the unconditional wage gap between part-time working men and women is estimated at zero.

All subsequent analysis will be done using hourly wages, since it reflects differences in the market price of labor.

4.4 Changes in the Overall Wage Distribution

4.4.1 Adjusting for the Overall Wage Distribution

The period covered in this study includes two eras with respect to changes in overall wage inequality. The first one, between 1968 and 1991, can be characterized as a period of major wage compression following the trade union strategy of centralized bargaining, the "solidarity wage policy" (see e.g. Hibbs, 1990) as well as the expansion of the public sector. Since women are over-represented in the lower part of the wage distribution, a more compressed wage structure will mechanically increase the relative earnings of women and decrease the gender wage gap, characterized as *swimming with the tide* by Edin and Richardsson (2002).

In the later era, 1991-2019, there is a reverse trend towards increased wage inequality, following more decentralized wage bargaining and a general trend towards a less compressed wage structure. Again, since women are over-represented at the lower end of the wage distribution, this development would lead to a mechanical increase in the gender wage gap, i.e. *swimming upstream* as coined by Blau and Kahn (1997). Appendix Table 4.B2 shows the development in wage inequality measured in the SLLS data as the coefficient of variation (standard deviation over mean), the 90/10 income ratio and the Gini coefficient.

To estimate the effect of changes in aggregate wage dispersion on the gender wage gap we have converted wages for each survey year to match the 1968 wage distribution. We use the *percentile deflator* proposed by LaLonde and Topel (1992), which reshapes the wage distribution for each year after 1968 to have the same first and second moments as the 1968 distribution, without imposing any parametric assumptions. For each percentile of the wage distribution in a given year, wages for those in the labor force are deflated by the wage growth in that percentile since 1968. For all workers, wages are recalculated according to:

$$\ln \widehat{W}_{ipt} = \ln W_{ipt}/\delta_{pt}, \tag{4.2}$$

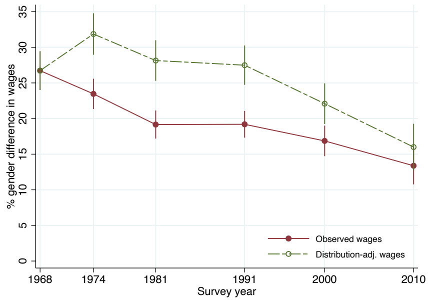
where δ_{pt} is the wage growth experienced by workers in percentile p between 1968 and year $t \in 1974, 1981, 1991, 2000, 2010$ for the SLLS data, and $t \in [1995, 2019]$ for the WSS data. The highest wages in each year (those above the 99th percentile) are winsorized at the maximum wage recorded in 1968. With this method, we make no assumptions about the observed or unobserved characteristics of an individual that places her or him in a given percentile.[6] The "percentile deflator" allows for between-individual wage differences to change, even within the age-education cells, and thus also reflects how wage offers depend on unobservables. Wages adjusted with this method are thus used going forward.

4.4.2 The Distribution-Adjusted Wage Gap

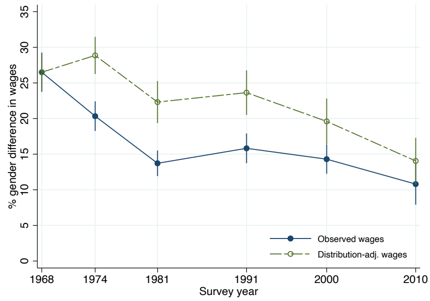
Figure 4.3 shows the gender wage gap where we have used the LaLonde-Topel percentile deflator to transform the overall wage distribution for each wave to correspond to that of 1968. As a point of reference, the unadjusted wage gaps are also included in the figures. The left panels shows the evolution of the gender wage gap at the mean and the right panel at the median. Again, the upper panels show the results 1968-2010 from the SLLS data and the lower ones the results 1995-2019 from the Wage Structure Statistics.

[6] To corroborate the results, we 1) calculate gender wage gaps in percentile rankings, and 2) implement a wage deflator based on an individual's level of education and age. Both of these produce similar estimates to the "percentile deflator", and results are available upon request.

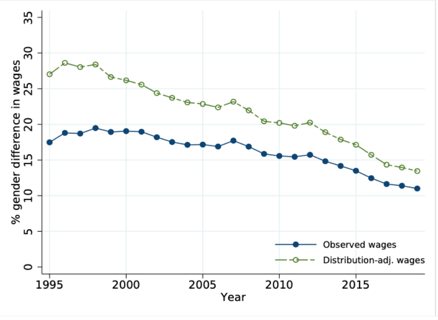
Figure 4.3: Wage Dispersion. Gender Wage Gap in Distribution-Adjusted Hourly Wages, at Mean and Median.



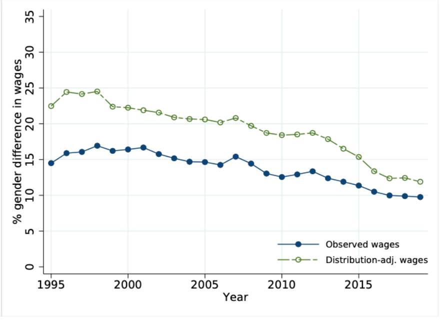
(a) SLLS Mean 1968-2010



(b) SLLS Median 1968-2010



(c) WSS Mean 1995-2019



(d) WSS Median 1995-2019

Notes: Yearly coefficients from regressions of log(wages) on female indicator variable, at mean and median. Distribution-adjusted wages are deflated with percentile method (LaLonde and Topel, 1992), meaning that they are adjusted according to the wage growth within a given percentile between 1968 and any given year.

The results in Figure 4.3 show, as expected, that the change in the wage gap between 1968 and 1991 is much smaller for the distributional adjusted series than the unadjusted ones. For the mean differentials, the results reveal that the entire drop in the gender wage gap from 27 to 20 log points can be attributed to changes in overall income inequality, rather than that the relative position of women in the wage distribution improved. For the median, the results are somewhat less extreme, although they show that 8 out of the 11 log point change in the gender wage - i.e., more than 70 percent - can be attributed to the more equal overall wage distribution.[7]

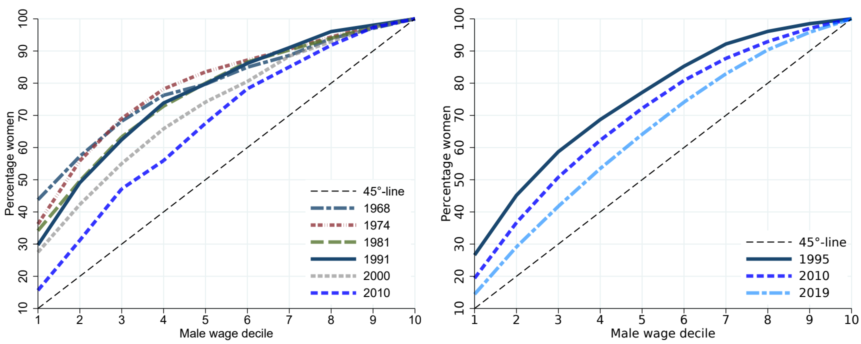
Figure 4.3 also shows that the convergence between the adjusted male and female wages is much faster than the unadjusted ones in the 1991 to 2010 period (1995 to 2019 in the WSS data). This is expected, since the overall wage dispersion increased during the first few years of this period. In fact, both the mean and the median results for the SLLS data suggest that the gender wage gap would have been five log points smaller if the overall wage distribution from 1991 would have been maintained - avoiding the *swimming upstream effect* (Blau and Kahn, 1997). This implies that the decrease in the mean gender wage gap over this period more than 60 percent larger in adjusted compared to unadjusted wages. For the median gender wage gap, the corresponding difference is even larger. Similar results can be inferred also for the WSS data, where the distribution-adjusted series at the mean and median decrease by 50 percent during 1995-2019, while the unadjusted series are reduced by only around 30 percent.

The fact that we find a larger decrease in the gender wage gap when we neutralize the effect of wage dispersion implies that women become less overrepresented in the bottom of the wage distribution over time. To further visualize the change in the relative position of female wages, Figure 4.4 shows the cumulative distribution of women's wages in the male wage distribution. In 1968, 45 percent of the employed women had wages that fell in the first decile of the male wage distribution. Over time, this changes at a modest rate and still in 1991 one third of all women are found in the bottom decile. The single largest change is observed between 2000 and 2010. In 2010, about 70 percent of women are paid a wage below the male median, and this number is reduced further to 60 percent when we look in the WSS data for 2019. However, this development does not seem to extend to the top of the wage distribution, where the share of women continues to be low throughout the sample period. Most of the changes seem to take place at the bottom half of the (male)

[7] This reinforces the findings of Edin and Richardsson (2002), that a large part of the closure in the gender wage gap over this time period can be attributed to changes in the overall wage distribution.

distribution, with the possible exception of the development between years 2010 and 2019, when a downward shift of the entire cumulative distribution is observed. Despite great changes, the male and female wage distributions differ substantially across all years in our study.

Figure 4.4: Cumulative Distribution of Female Wages, by Percentiles of Male Wages.



(a) SLLS 1968-2010 (b) WSS 1995-2019

Notes: The figure shows the share of female workers with wages in each decile of the male wage distribution. The distributions are calculated from the Swedish Level of Living Survey waves 1968-2010 in Panel a; from the Wage Structure Survey data 1995, 2010 and 2019 in Panel b.

4.5 The Selection-Corrected Wage Gap

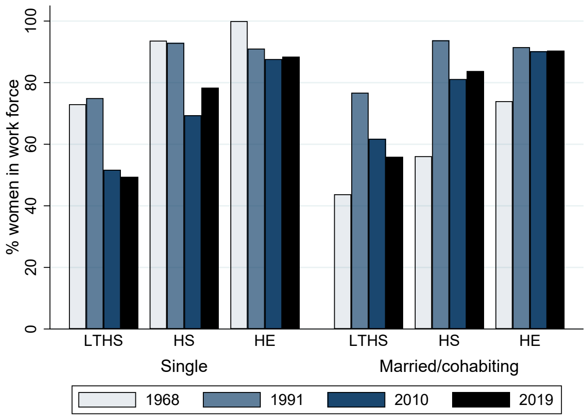
4.5.1 Adjusting for Selection on Observables

Reductions in the gender wage gap in industrialized countries are often partly attributed to female increased levels of accumulated human capital (see e.g. Goldin, 2014; Blau and Kahn, 2017). We will thus begin by exploring how the average age and level of education among women in Sweden have changed over our study period, and how these observed changes in workforce composition have affected the gender wage gap.

Figure 4.5 shows labor force participation in 1968, 1991, 2010 and 2019, separately for groups formed by marital status and level of education. For both single and married women, labor force participation in 1968 was higher among those with a high school degree and any higher education, than among those without a high school degree. For married women in particular, participation rates in 1968 depend on the level of education; married women with college education are almost twice as likely to be working, compared to women with only compulsory schooling. Between 1968 and 1991, employment increased the most in the less-educated groups among married women. The change between 1991 and 2010 marks a contraction in labor force participation, most prominent among less-educated groups. One explanation lies in the lower rates of participation among young adults. Between 2010 and 2019, participation rates seemingly increased among the high school educated population.[8]

[8]Note that participation rates for 2019 are calculated in a different sample (the WSS data), whereby this difference is to be interpreted with some caution.

Figure 4.5: Female Participation Rates 1968, 1991 and 2010, by Martial Status and Level of Education.



Notes: The figure shows percentage self-stated labor force participation (full time or part time) among women, calculated from SLLS samples 1968, 1991 and 2010 and from the WSS data for 2019, by marital status and highest achieved level of education. LTHS = less than high school, HS = high school degree, HE = higher education (college/post-secondary). Divorced and widowed individuals are excluded from this analysis.

To see this, Panel (a) of Appendix Figure 4.B2 shows the distributions of worker age for men and women separately for each survey year, while Panel (b) shows the age distribution of the full SLLS samples. Most notably, the workforce consists of a gradually smaller share of young individuals over time, and a progressively larger share of middle-aged workers. This change is especially pronounced among female workers.

We control for these changes by calculating the gender wage gap in residualized wages. The residuals are generated from a wage equation including age, education level and their interactions as controls.[9] The results are shown at the median in the gray dashed line in Appendix Figure 4.B3. Compared to the wage gap in distribution-corrected hourly wages, the total change from 1968 to 1991 in terms

[9]We regress log distribution-adjusted wages on a full set of age and education dummy variables and their interactions. Education is divided into three categories: compulsory schooling, high school degree and any higher education.

of residual wages is smaller; only 8 percentage points. In addition, the 2010 wage gap in residual wages is larger than that in observed (distributional corrected) wages: 17.9 versus 14.2 log points. This reflects the increased average level of education among employed women, relative to their male counterparts.

4.5.2 Adjusting for Selection on Unobservables

The core issue with selection into the labour force is a latent variable problem: we observe an individual's wage only when she has made the choice to work. To address this *selection bias*, we first estimate non-parametric Bounds on the distribution of wages, and calculate Bounds on the change in the gender wage gap over time.[10] We do this first with only observed wages of workers; then impute latent wages for non-workers based on observed individual characteristics, loosely following methods described in Olivetti and Petrongolo (2008), and calculate Bounds on the change in the wage gap with these. Finally, we estimate the Heckman selection model. Both methods will ultimately result in an estimate of the change in the gender gap in the distribution of *wage offers*, rather than the one consisting of observed wages.[11]

Bounds

We use the non-parametric bounds method first proposed by Manski (1991, 1994). Denote by $F(w|g)$ the gender-specific cumulative distribution function of wage offers, with $g \in \{male, female\}$. This can be expressed as an average of the wage distribution of employed and non-employed, weighted by the observed probability of working $p(g)$:

$$F(w|g) = F(w|g, E = 1)p(g) + F(w|g, E = 0)\left(1 - p(g)\right), \tag{4.3}$$

[10] See seminal work by Gronau (1974) and Heckman (1976) for a conceptual discussion.

[11] We use the term wage offer here as a population concept following Gronau (1974), reflecting the observed wage for workers, and the potential wage a non-working individual would have, were he or she employed. For non-workers, it is thus a theoretical construct, rather than an empirical observation.

where E is an indicator for employment taking the value 1 for being employed and 0 otherwise.

The *worst case bounds* for the wage distribution of each gender group are obtained by using the fact that the unobserved cumulative distribution function for the non-employed, $F(w|g, E = 0)$, is bounded between 0 and 1.

The *lower bound* is constructed by assuming that all those not in the labor force would earn a wage at the bottom of the (gender-specific) distribution, if they were participating. This means that we assume that those who participating in the labor force are positively selected from the population in terms of wage offers they receive. The lower bound is obtained by setting $F(w|g, E = 0)$ equal to 0 in equation (3), which implies that the last term in the equation cancels out.

Conversely, the *upper bound* is constructed by assuming the work force is negatively selected and wage offers of non-workers are in the top of the distribution. The upper bound for the wage distribution is then obtained by setting $F(w|g, E = 0)$ equal to 1 in equation (3) and the worst case bounds for the latent wage distribution is obtained by:

$$F(w|g, E = 1)p(g) \leq F(w|g) \leq F(w|g, E = 1)p(g) + (1 - p(g)) \,. \tag{4.4}$$

This expression allows us to obtain quantiles of the latent wage offer distribution $F(w|g, E = 0)$ from what we observe in the data ($F(w|g, E = 1)$ and $p(g)$). Solving the upper bound expression at a given quantile of the distribution q, yields a quantile of the observed distribution, $F(w|g, E = 1)$:

$$q(u) = F(w|g, E = 1)p(g) + (1 - p(g)) \,, \tag{4.5}$$

and the inverse function $F^{-1}(\cdot)$ gives a wage estimate of the upper bound. Correspondingly, $q(l) = F(w|g, E = 1)p(g)$ is solved for an estimate of the lower bound. In our application, we estimate Bounds on the median, i.e. for $q = 0.5$.

Denoting by $w^{q(u)}$ the upper bound to the gender-specific wage distribution and by $w^{q(l)}$ the lower bound, we define Bounds on the female-male difference D in wages at quantile q in year t to be:

$$w^{q(u)}_{t,female} - w^{q(l)}_{t,male} \leq D^q_t \leq w^{q(l)}_{t,female} - w^{q(l)}_{t,male}. \tag{4.6}$$

Note that we use the lower bound for male earnings for both the lower and upper Bounds on the gender wage gap. The reason for doing this, rather than combining the lower bound for female wages with the upper one for men, is that we believe that positive selection into the labor force is a reasonable assumption for the Swedish male population and therefore the lower wage bound is a good approximation of median latent wages. A lower bound to the change in the gender wage gap between years s and t, $\Delta D_{st}^{q(l)}$ is then calculated as the difference between the upper bound in year t minus the lower bound in year s:

$$\left(w_{t,female}^{q(l)} - w_{t,male}^{q(l)}\right) - \left(w_{s,female}^{q(u)} - w_{s,male}^{q(l)}\right). \tag{4.7}$$

The upper bound to the change over time, $\Delta D_{ts}^{q(l)}$, which is of main interest in this paper, is given by the lower bound to the gender wage gap in year t, minus the upper bound to the wage gap in year s:

$$\left(w_{t,female}^{q(u)} - w_{t,male}^{q(l)}\right) - \left(w_{s,female}^{q(l)} - w_{s,male}^{q(l)}\right). \tag{4.8}$$

We follow Blundell et al. (2007) and use the *median restriction* to obtain tighter bounds. This restriction asserts that the median of the observed wage distribution constitutes an upper bound for the median of the unobserved distribution. The Bounds on the distribution are then:

$$\begin{aligned} F(w|g, E = 1)p(g) + 0.5\,(1 - p(g)) \leq F(w|g) \leq \\ F(w|g, E = 1)p(g) + (1 - p(g)). \end{aligned} \tag{4.9}$$

The median restriction implies that although some non-participants may have high latent wages, the lower half of the distribution of those not participating are as a group negatively selected from the lower half of the overall population. For the entire male and female populations, this is likely a weak restriction. However, for some sub-groups, such as women with small children, one may argue that it would not necessarily apply. For example, if high-ability women on average have children later in life, selection into employment will be counteracted in certain age groups. This may be more of a concern in the earlier years in our study, when subsidized daycare was not readily available for small children.

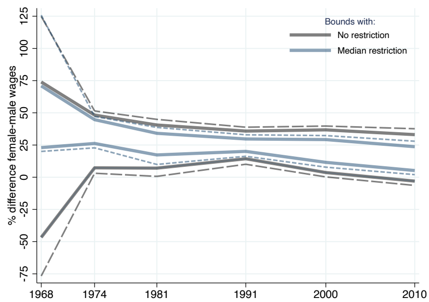
Bounding the change in the wage gap over time

Figure 4.6 shows the first set of results when we address also the problem of selection on unobservables into the labor market. Panels (a) and (c) shows the evolution of the worst-case bounds for the level of the gender wage gap adjusted for changes in the overall wage distribution 1968-2010 and 1995-2019, respectively. These panels also show the bounds when the median restriction is imposed. The dashed lines give 95 percent confidence intervals for each of the two sets of bounds. The results for the *change* in the gender wage gap shown in Panels (b) and (d) can be interpreted in different ways. The most conservative one is to compare the lower bound in 1968 with the upper bound in 2010. This procedure implicitly assumes that the selection into the labor force changes from being positive in the beginning of the period under study to being negative by end. This is of course an unlikely case and an overly cautious interpretation.

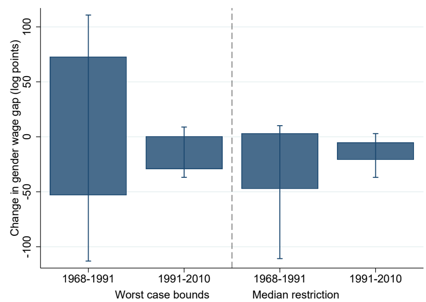
Panels (b) and (d) show the upper and lower bounds for the *change* in the gender wage gap, as suggested by equations (7) and (8), for the eras 1968-1991, 1991-2010, 1995-2010 and 2010-2019, respectively. The upper bounds for the wage gap are, as explained above, constructed from lower bounds for female wages, obtained under the assumption that females in the labor force are, as a group, positively selected from the population of females. The lower bounds are constructed from the upper bound for female median wages, assuming women in employment are negatively selected. The lower bound on male wages, based on the assumption that men are positively selected to participate in the labor force, are used for both the lower and the upper bounds for the gender wage gap in Figure 4.6.

The results in Panel (b) reveal that, due to the low female labor force participation in 1968, the worst-case bounds are not informative for the change over the 1968-1991 era. The change across these years ranges between a 50 log point increase and an equal-size decrease. Focusing on the right side of Panel (b), where we impose the median restriction, we see that the bounds for the change in the gender wage gap range from an increase of 2 and a decrease of 48 log points.

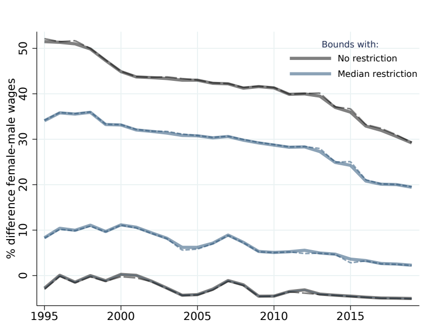
Figure 4.6: Bounds on the Median Gender Wage Gap.



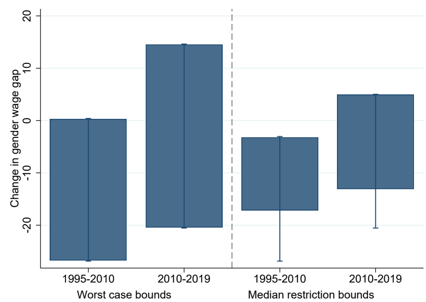
(a) SLLS Bounds for each survey wave.



(b) SLLS Change 1968-1991 and 1991-2010.



(c) WSS Bounds for each survey year.



(d) WSS Change 1995-2010 and 2010-2019.

Notes: Bounds on the median gender wage gap, calculated using the sample of men and women in the labor force and with distribution-adjusted wages. Panel (a): Bounds without restriction and with the median restriction (the median wage of non-workers is assumed not to be higher than the median wage among workers), calculated for each survey wave. Dashed lines represent bootstrapped 95 % confidence intervals. Upper bound to gender wage gap calculated as lower bound to female wages minus lower bound to male wages. Lower bound calculated as female upper bound minus male lower bound. Panel (b): Bounds on the change in the gender wage gap, between years 1968-1991 and 1991-2010. Bounds are calculated with the median restriction. Thin lines represent bootstrapped 95 % confidence intervals. **N.B. different scales on top and bottom graphs.**

The most interesting result shown in Panel (b) is that we cannot rule out a decrease in the gender wage gap of almost 50 log points

- even when disregarding the statistical error - as a result of the potential role of selection. Even when we impose the more plausible assumption that female workers were consistently positively selected between 1968 and 1991 (not shown in Panel b), the decrease in the gender wage gap is still 45 log points. Bounds on the change in the gender wage gap between 1991 and 2010 are much tighter. Again disregarding the statistical error, Panel (b) shows that these bounds range between 5 and 20 log points. Panel (d) shows that the corresponding result with the median restriction from the WSS data set (1995-2010) is very similar, with an estimated change of between 3 and 17 log points. Between 2010 and 2019, even the median restriction bounds are too wide to infer a certain reduction or increase of the selection-adjusted wage gap.

Combining bounds with imputations of wages

In order to tighten the bounds further, and enable more informative inference, we use two different strategies to impute unobserved wages. In the first one we use the panel structure of the data to predict whether or not the missing wage observation is above or below the sample median for males and females, respectively. We simply infer that the wages of an individual are below the gender-specific median in a particular year, if it is so in either of the two adjacent years. This method relies on the assumption that the position of an individual's wage with respect to the median wage remains the same across years. It would be possible to infer wages from more than one other survey wave. However, since this would require stronger assumptions about the propensity of individuals' wages to remain on the same side of the median wage over a longer time span, we have abstained from doing so.[12]

Columns 1, 2 and 3 in Table 4.B3 show the sample sizes, the number of observed wages and the number of observed wages after the panel imputations, respectively. As is evident from Column 3,

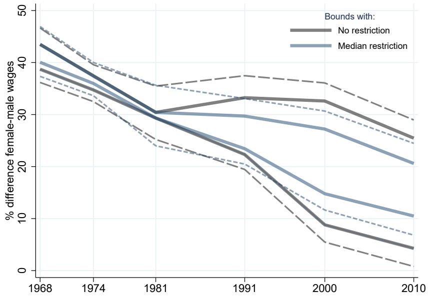
[12] Note that these panel imputations will address the concern about the median restriction discussed above, since high-ability women are likely to participate in the labor force in at least one of the survey years. It does not, however, allow us to infer wages for those never employed, which is about 30 percent of the SLLS sample in 1968, and about ten percent in 1991 and 2010.

the panel imputation method increases the sample size by about six to ten percent per wave, compared to the sample of workers.

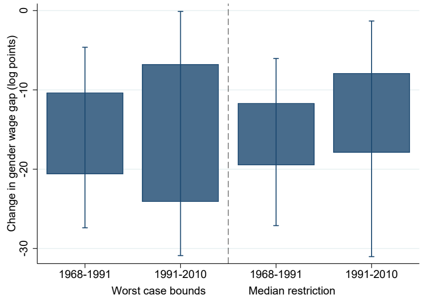
To increase the share of imputed wages, we complement the strategy described above by following Neal (2004) and Olivetti and Petrongolo (2008) in using human capital variables observable in the survey. The idea is that people with higher observed human capital are more likely to have above-median wage offers. For those with sufficiently high human capital, the risk of making a wrong prediction is likely to be quite low. However, lowering the threshold for above-median wage offer imputations increases the risk of making a prediction error. Since we are still able to use the non-parametric bounds, we are, however, not obligated to predict wage offers for all individuals in the sample. Setting the thresholds for a below or above median wage offer could thus be characterized as a consistency vs. efficiency trade-off, since setting the threshold for an above-median wage prediction very close to the below-median threshold would increase the number of observations, but also the number of false predictions, potentially causing inconsistent estimates of the wage gap.

There is no obvious way to find the optimal combination in this trade-off. We have chosen to apply a three-step procedure, which is described in detail in Appendix 4.A. In addition to the wage data obtained from the other years of the panel, we use two sets of human capital variables: educational attainments and work experience. We chose between the following threshold values for classifying observations: *Valid panel imputation*: v (wage in relation to median wage) $\in \{0.55, 0.6, 0.65, 0.7, 0.75, 0.8\}$; *High education*: h (years of schooling) $\in \{12, 13, 14, 15, 17, 18\}$; *Low education*: l (years of schooling)$\in \{7, 8, 9, 10, 11, 12\}$; *Experience:* e (years of work experience)$\in \{5, 10, 15\}$. In the first step, we calculate the in-sample mean squared error for each combination of threshold values, on the basis of the observations with observed wages, i.e. comparing actual wages and the imputed samples. In the second step, we rank each combination based on mean squared error and sample size, respectively. Finally, in the third step, we chose the threshold values producing the lowest average rank score.

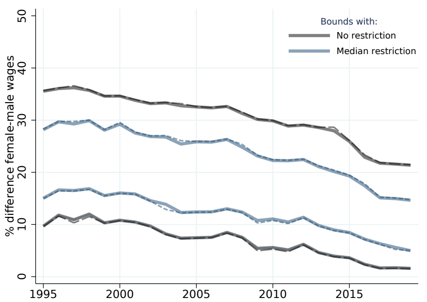
Figure 4.7: Bounds on the Median Gender Wage Gap, with Wage Imputations.



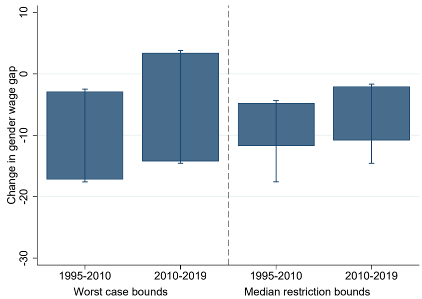
(a) SLLS Bounds for each survey wave.



(b) SLLS Change 1968-1991 and 1991-2010.



(c) WSS Bounds for each survey year.



(d) WSS Change 1995-2010 and 2010-2019.

Notes: Bounds on the median gender wage gap, calculated using the combined panel and human capital method described in Section 4.5.2. Sample sizes printed in Table 4.B3, Column 5. Panel a and c: Bounds without restriction and with the median restriction (the median wage of non-workers is assumed not to be higher than the median wage among workers), calculated for each survey wave. Dashed lines represent bootstrapped 95 % confidence intervals. Upper bound to gender wage gap calculated as lower bound to female wages minus lower bound to male wages. Lower bound calculated as female upper bound minus male lower bound. Panel b and d: Bounds on the change in the gender wage gap, between years (b) 1968-1991 and 1991-2010; (d) 1995-2010 and 2010-2019. Bounds are calculated with the median restriction. Thin lines represent bootstrapped 95 % confidence intervals. **N.B. different scales on top and bottom graphs.**

The procedure results in the following thresholds: $\{v = 0.6, l = 11, h = 18, e = 10\}$. Column 5 in Table 4.B3 shows the resulting number of observations, and Figure 4.7 shows bounds to the gender wage gap corresponding to those in Figure 4.6, but with these wage imputations. Again, Panels (a) and (c) show the evolution of the bounds for the wage gap with and without the median restriction, and Panels (b) and (d) show the bounds for the change in the gap.

Comparing the results in Figure 4.7 with those in Figure 4.6, it is obvious that the bounds are, as expected, much tighter. Panel (b) in Figure 4.7 shows that even the lower bound for the change in the gender wage gap now corresponds to a quite substantial 13 log points reduction in the gender wage gap for the period 1968 to 1991, and the upper bound is now almost 20 log points. From the early 90s until 2010, the gender wage gap is estimated to have shrunk by between 8 and 18 log points in the SLLS data; between 5 and 12 points in the WSS data. From 2010 to 2019, the change is found to be between 2 and 11 log points.

Bounds in groups with comparatively high female participation rates

An alternative way to obtain tighter bounds without imputing unobserved wages is to look at the change in the wage gap in population groups with high female labor force participation, under the assumption that the change to the wage gap between 1968 and 1991 is homogeneous between these and other, otherwise similar, demographic groups.[13] Following Blundell et al. (2007), we labels this the *additivity restriction*. We divide the sample into four groups based on level of education (high or low) and age (younger or older). Due to the changing nature of educational attainments at population level, we denote compulsory schooling “low” and everything beyond that “high” in the SLLS data 1968-1991, while high school completion is assigned to “low education” in the WSS data and SLLS data for 1991-2010.

Within these groups, we calculate the upper and lower bound to the change in the wage gap 1968-1991, for each subgroup formed by

[13] This method is used in Blundell et al. (2007) and also discussed in Blau et al. (2021), who attribute it originally to Chamberlain (1986) and Heckman (1990).

age in five-year intervals ($a \in A$), $\Delta D_{st}^{q(u)}(a)$ and $\Delta D_{st}^{q(l)}(a)$. Bounds for a given group g is set as the combination of the smallest upper bound and the largest lower bound among all ages within that larger age-education group:

$$\max_{a \in A} \Delta D_{st,g}^{q(l)}(a) \leq \Delta D_{st,g}^{q} \leq \min_{a \in A} \Delta D_{st,g}^{q(u)}(a).$$

For example, in group $g = 1$, formed by people with "low" education and aged between 20-44 in a given year, bounds are calculated in subgroups of individuals aged 20-24, 25-29, 30-34, 35-39 and 40-44. The tightest possible bounds for this group are found by combining the upper bound for ages 40-44 and the lower bound for ages 20-25.

Figure 4.8 show box plots with 95 percent confidence intervals for the results from our application of this procedure (using distribution-adjusted wages). Overall, the bounds are similar to those found using the imputation methods above, but less precise. Two features of the results stand out. First, the estimated bounds are generally much tighter for the the more educated in both age groups. This is expected, since the labor force participation rate for females is much higher among people with more education. However, the confidence intervals show that the precision of the estimates, due to small sample sizes, precludes conclusive inference from the results for most years. One exception is college educated older workers, where a statistically significant change between 1995 and 2010 of around 12 log points is found. In general, these bounds suggest that the change in the gender wage gap over time is larger for older workers, than for younger workers, although the bounds do not allow us to statistically distinguish estimates for most years and educational groups.

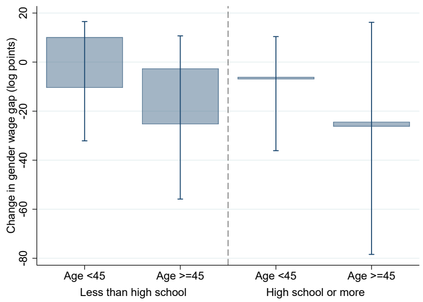
The Heckman selection model

To gain further understanding on the effect of selection of females to the labor force, we also estimate a fully parametric model based on the method proposed by Heckman (1976) to correct for bias from selective samples. We estimate the following wage equation:

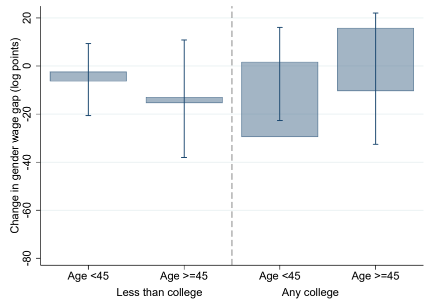
$$w_i = \alpha + \beta_1 educ_i + \beta_2 age_i + \beta_3 age_i^2 + \phi\lambda_i + \varepsilon_i, \tag{4.10}$$

where w is log hourly wage, $educ$ is years of schooling, age is age

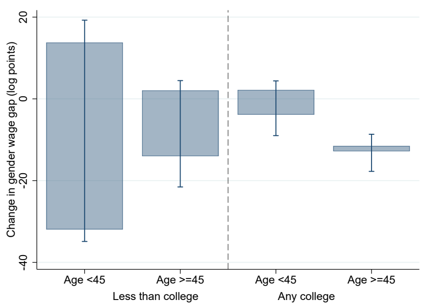
Figure 4.8: Bounds on the Change in the Gender Wage Gap of Workers, with Additivity.



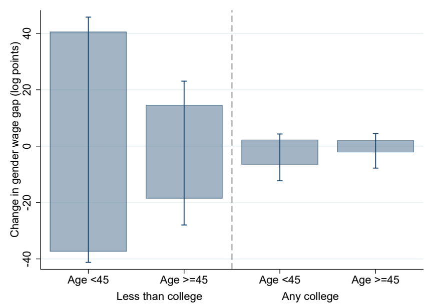
(a) SLLS 1968-1991.



(b) SLLS 1991-2010.



(c) WSS 1995-2010.



(d) WSS 2010-2019.

Notes: This figure shows Bounds on the change in the gender wage gap in distribution-adjusted wages, between years 1968-1991 and 1991-2010 for the SLLS data; for years 1995-2010 and 2010-2019 for the WSS data. Bounds are calculated with the median restriction (the median wage of non-workers is assumed not to be higher than the median wage among workers) and the *additivity restriction*. This states that the change in the wage gap over time within an age-education group is the same, and bounds can be obtained by combining the largest lower bound with the smallest upper bound among subsets of that group. Groups are described in text above.

at the time of the survey and λ is the inverse of Mill's ratio, i.e.,

$$\lambda_i = f(-y_i)/\left[1 - F\left(-y_i\right)\right]. \tag{4.11}$$

This model is estimated in two steps. In the first step, a probit model for participation in the labor force, we use an indicator variable for marital status, number of children aged between 0 and 6, number of children aged 6 to 18, total number of children and non-labor income as independent variables, in addition to the ones included in the wage equation above. The result from the first step is used to generate the inverse of Mill's ratio for the selection correction in the wage equation.

The results from the Heckman (1976) parametric selection model are shown in Table 4.3, and are confined to the female samples in the 1968 and 1974 waves of the SLLS. The reason for limiting the presentation to these years is that, as is revealed in Figure 4.6, female labor force participation after the 1974 survey is high enough, that selection has no major impact on the estimates of the gender wage gap. Table 4.3 shows significant positive selection into employment for both years under study. Adding the selection components (λ) to the observed average gender wage gaps (as presented in the left panel of Figure 4.2) of 0.267 in 1968 and 0.289 in 1974, gives wage gap estimates corrected for sample selection bias of 0.383 for 1968 and 0.399 in 1974. Reassuringly, these estimates are very similar to the ones reported in Figure 4.7 using our imputation strategies to correct for selection bias.

Table 4.3: Heckman 2-step Estimates of Selection Bias in SLLS 1968 and 1974 (Women).

	1968		1974	
	(1) Wage	(2) Participation	(3) Wage	(4) Participation
Education (years)	0.0726***	0.0984***	0.0599***	0.0944***
	(0.004)	(0.014)	(0.004)	(0.012)
Age	0.0379***	0.0975***	0.0544***	0.150***
	(0.006)	(0.021)	(0.007)	(0.023)
Age sq.	-0.000434***	-0.00150***	-0.000585***	-0.00210***
	(0.000)	(0.000)	(0.000)	(0.000)
1(Married)	-0.0265	-0.360**	-0.0429	0.265*
	(0.031)	(0.115)	(0.030)	(0.125)
1(Divorced)	0.00864	-0.0984	-0.0282	0.222
	(0.037)	(0.142)	(0.041)	(0.151)
Num. kids <=6 yrs		-0.727***		-0.515***
		(0.072)		(0.066)
Num. kids <18 yrs		-0.217*		-0.0740
		(0.087)		(0.119)
Total num. kids		0.100		-0.0756
		(0.081)		(0.114)
Non-labour inc.		-0.00000944***		-0.0000155***
		(0.000)		(0.000)
Constant	0.843***	-1.142**	0.438**	-2.316***
	(0.122)	(0.435)	(0.141)	(0.470)
Observations	2,004		1,910	
λ	0.116		0.110	
Prob>chi2	0.000		0.000	

Notes: Each column reports maximum likelihood estimates for the female SLLS sample; wage equations in (1) and (3), participation equations (self-reported working or not) in (2) and (4). Wages are log distribution-adjusted wages. Standard errors in parentheses. $\lambda = \phi \times \rho$. Prob>chi2 from LR of independent equations ($\rho = 0$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.6 Summary of Results and Discussion

The object of this paper is to assess the importance of changes in female labor supply, earnings inequality and selection into the labor force, on the evolution of the gender pay gap in Sweden since the late 1960s. Table 4.4, along with Figure 4.9, summarize our main findings.

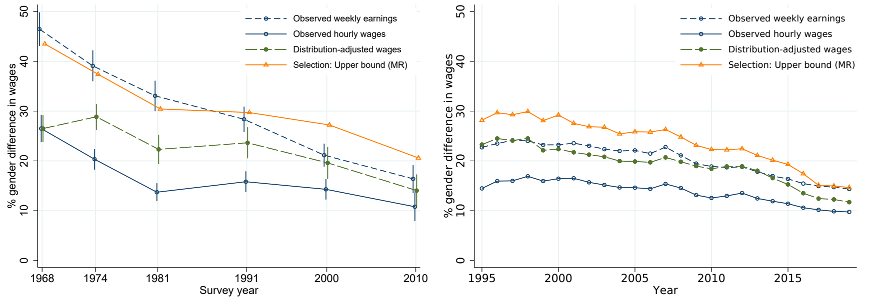
To study the effect of shifts in intensive margin labor supply, it is useful look at the difference between the first two columns in Table 4.4. Starting in 1968, we see that the gender gap in weekly earnings is 46.5 log points, compared to 26.5 in hourly wages. This means that about 43 percent of the gap in weekly earnings could be attributed to differences in hours of work, ignoring any secondary effects from restricting the hours of work to less than full time. The development until 2010 reveals a marked convergence between the gender gap in weekly earnings and hourly wages, implying a levelling out in hours of work between men and women. Between 1991 and 2010, when only marginal changes are discerned in the gap in hourly wages, weekly earnings differences declined as a result of increased length of the work week among women. However, still in 2019, almost 34 percent of the gender gap in weekly pay, at 15 percent, can be attributed to differences in hours of work — a result that is not surprising given the emphasis placed on intensive margin labor supply by recent literature on the gender pay gap (see e.g. Kleven et al., 2019a, 2019b; Goldin, 2014; Mas and Pallais, 2020).

Table 4.4: Estimated Change in the Median Gender Wage Gap, 1968-1991 and 1991-2010.

	Observed		Distribution adj.	Selection adj.	
	(1) Earnings	(2) Wages	(3) Distribution	(4) Panel	(5) Panel+HC
Change 1968-1991	-18.1	-10.7	-2.9	(1.8, -21.4)	(-12.2, -20.3)
Change 1991-2010	-11.8	-4.8	-9.5	(-6.7, -19.3)	(-8.0, -18.5)
Change 2010-2019	-4.6	-2.6	-4.7	(4.3, -10.9)	(-2.1, -10.6)
Gap 1968 (%)	46.5	26.5	26.5	(26.6, 43.7)	(40.5, 43.8)
SLLS Gap 2010 (%)	16.6	11.0	14.2	(9.1, 20.1)	(10.1, 20.1)
WSS Gap 2010 (%)	19.6	12.5	18.4	(6.1, 27.5)	(10.5, 22.0)
Gap 2019 (%)	15.0	9.9	13.8	(3.2, 19.5)	(4.7, 14.8)
Total change 1968-2019	-31.5	-16.6	-12.7	(-7.0, -40.5)	(-25.7, -39.1)

Notes: Columns: (1) Weekly earnings. (2) Hourly wages. (3) Distribution adjusted wages with percentile deflator. (4) Panel imputation from ± 1 survey wave (70 % in-sample accuracy threshold, age-education groups). (5) Combined imputation: panel (as in column 4) and human capital. In (4)-(5), the results refer to bounds with the median restriction instead of point estimates: (lower bound, upper bound). All gaps from median regression of log wage on female dummy.

Figure 4.9: Summary of Results for Gender Wage Gap at Median.



(a) SLLS (b) WSS

Notes: Coefficients from median regression of log(wage) on female dummy. Observed earnings and wages (in blue) have imputed values for missing observations. Distribution adjusted wages (in green) are adjusted with percentile deflator to match 1968 distribution. The selection result (in orange) uses percentile-deflated wages with panel and human capital imputations, and shows the upper bound (the difference between the male and female lower bounds) imposing the median restriction.

Next, to study the effects of changes in overall wage inequality, we compare Columns 2 and 3 in Table 4.4. We see that almost the entire reduction in the gender wage gap between 1968 and 1991 can be attributed to the compression in the wage distribution. However, as previously shown in Figure 4.3, the increase in wage inequality masks a substantial decrease in the gender wage gap between 1991 and 2019. The changes from 1991 to 2010 and from 2010 to 2019 are almost twice as large when taking the level of wage dispersion into account (4.8 vs. 9.5 log points; 2.6 vs. 4.7 log points). As expected, less wage dispersion leads to smaller gender differences in wages, and *vice versa*.

Among others, Mulligan and Rubinstein (2008) and Card and Hyslop (2018) discuss how wage inequality may have extensive-margin labor supply effects for women. Mulligan and Rubinstein (2008) argue that the increased wage inequality in the US in the 1970s and 80s provided incentives for high-ability women to enter the labor force, since wage offers in the wider wage distribution exceeded their reservation wages for labor market entry. Conversely, the wider wage distributions induced low-ability women to leave the labor force, as their wage offers fell below their reservation wages. Mulligan and Rubinstein (2008) show that the combination of these processes lead to a smaller gender gap in observed wages.

The wage compression following the solidarity wage policy in Sweden in the late 60s and throughout the 70s created stronger incentives for low-ability females to enter the labor force. This happened because the lower end of the distribution of wage offers shifted upwards, increasing the probability of receiving a wage offer exceeding reservation wages, and thus leading to labor market entry.[14] Our results from the Heckman selection model support the interpretation of positive female selection into the labor market in the 60s and 70s, and a smaller gender gap in observed wages than in wage offers. This corresponds to the results of Mulligan and Rubinstein (2008), although they study the effects of a more unequal wage distribution rather than a compression of the wage distribution.

Finally, Columns 4 and 5 in Table 4.4 summarize the development of the gender gap in hourly wages when we, in addition, have corrected for selection bias. As apparent from Section 5.2, we have used different methods to correct for selection bias and to avoid ambiguities caused by too many numbers, we here confine ourselves to present results from the panel imputation method, and the combined panel-human capital imputation method. However, these results are similar to the upper bounds estimates when we impose the median- and additivity restrictions, shown in Figure 4.8, and the results using the Heckman selection model.

Comparing Columns 3 and 4 in Table 4.4, the change in the gender gap from 1968 to 2019 is likely to be twice, or even three times, as large as that in observed wages corrected for changes in overall wage inequality only. This result is of course conditional on that we are willing to make the assumption of positive selection into the labor market and/or that we believe in the result on positive selection

[14]Previous research have shown that contemporary reforms introducing separate taxation of spouses and expanding public child care also contributed to the increase in female labor force participation. See e.g. Selin (2014) and Rosen (1997).

from the Heckman model.[15] This is different from findings for the US labor market in e.g. Blau and Kahn (2006) and Mulligan and Rubinstein (2008), who find ambiguous or no effects on the wage gap in 1970-1990 when correcting for selection bias. Since participation rates among women in Sweden were high by international standards already in the 1970s, direct comparisons with US results may however not be applicable.

4.7 Conclusions

In this study we have characterized the evolution of the gender pay gap in Sweden during a 51-year period, accounting for changes in hours of work, the wage distribution and in the gender composition of the work force. We have shown that this characterization differs in many respects from the gender pay gap reported in most public policy studies, which we think gives a partial, and in some respects misleading, measure of women's relative status at the labor market.

Three main conclusions can be obtained from this study. First, female intensive margin labor supply plays an important role for both the change in and the level of the gender gap in weekly earnings. Almost half of the change in the earnings gap 1968 to 2019 can be attributed to that employed women work more hours. Second, changes in overall wage inequality are important determinants for the evolution of the gender gap in observed wages. Almost the *entire* change in the gender gap in observed wages between 1968 and 1991 can be attributed to increased wage equality. Conversely, the more unequal wage distribution since the 1990s hides an improvement in the relative position of women in the labor market. Finally, positive selection into the labor force combined with a great increase in labor force participation hides a potentially large decrease in the gender gap in wage offers. Our upper bound estimates of the gender gap in wage offers in 1968 suggest that an analysis using observed wages

[15] Table 4.4 also shows that the Bounds on the selection-adjusted wage gap exceeds the gap for observed wages after 1991. A conceivable explanation to this development is a change in the relative age structure of men and women participating in the labor force. As is evident from the graphs in Appendix Figure 4.B2 there is a shift towards higher relative labor force participation of young men and a lower relative labor force participation of old men.

misses the major part of the change in wage offers until 1981. Taken together, the three measures show a continuous improvement of women's position at the labor market. Meanwhile, the remaining raw gap in 2019 is estimated at 9.9 log points in unadjusted hourly wages, reflecting the fact that substantial improvements of women's conditions are still needed for gender equality in pay.

While our results show that the inference on the evolution of the gender gap is highly dependent on the choice of wage concept, we remain agnostic about which we consider the most relevant. In the end it all boils down to what type of gender gap measure one is interested in studying. One may, for example, argue that the wage offer distribution is the most relevant measure if one wants to assess the evolution of gender discrimination at the labor market, while observed weekly earnings is more relevant if interest lies in within-household bargaining positions of men and women, respectively. Likewise, one could argue in favor of not correcting for changes in overall income inequality if one sees a smaller gender wage gap as a feature of a more equal overall wage distribution, while this is not the case if one focuses on the relative position of women in the wage distribution.

Bibliography

Albrecht, James, Anders Björklund, and Susan Vroman. 2003. "Is There a Glass Ceiling in Sweden?" *Journal of Labor Economics*, 21(1): 145–177.

Böhlmark, Anders, and Matthew Lindquist. 2006. "Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden." *Journal of Labor Economics*, 24(4): 879–900.

Björklund, Anders, and Markus Jäntti. 1997. "Intergenerational Income Mobility in Sweden Compared to the United States." *The American Economic Review*, 87(5): 1009–1018.

Blanchard, Olivier, and Dani Rodrik, ed. 2021. *Combating Inequality: Rethinking Government's Role.* Cambridge, MA:MIT Press.

Blau, Francine D., and Andrea H. Beller. 1988. "Trends in Earnings Differentials by Gender, 1971–1981." *ILR Review*, 41(4): 513–529.

Blau, Francine D., and Lawrence M. Kahn. 1996. "Wage Structure and Gender Earnings Differentials: An International Comparison." *Economica*, 63(250): S29–S62.

Blau, Francine D., and Lawrence M. Kahn. 1997. "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s." *Journal of Labor Economics*, 15(1): 1–42.

Blau, Francine D., and Lawrence M. Kahn. 2006. "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence." *ILR Review*, 60(1): 45–66.

Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature*, 55(3): 789–865.

Blau, Francine D, Lawrence M Kahn, Nikolai Boboshko, and Matthew L Comey. 2021. "The Impact of Selection into the Labor Force on the Gender Wage Gap." National Bureau of Economic Research Working Paper 28855.

Blomquist, Sören, and Håkan Selin. 2010. "Hourly Wage Rate and Taxable Labor Income Responsiveness to Changes in Marginal Tax Rates." *Journal of Public Economics*, 94(11-12): 878–889.

Blundell, Richard, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir. 2007. "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds." *Econometrica*, 75(2): 323–363.

Blundell, Richard, Antoine Bozio, and Guy Laroque. 2011. "Labor Supply and the Extensive Margin." *American Economic Review*, 101(3): 482–86.

Blundell, Richard, Howard Reed, and Thomas M. Stoker. 2003. "Interpreting Aggregate Wage Growth: The Role of Labor Market Participation." *American Economic Review*, 93(4): 1114–1131.

Card, David, and Dean R. Hyslop. 2021. "Female Earnings Inequality: The Changing Role of Family Characteristics and Its Effect on the Extensive and Intensive Margins." *Journal of Labor Economics*, 39(S1): S59–S106.

Chamberlain, Gary. 1986. "Asymptotic efficiency in semi-parametric models with censoring." *Journal of Econometrics*, 32(2): 189–218.

Edin, Per-Anders, and Katarina Richardson. 2002. "Swimming with the Tide: Solidary Wage Policy and the Gender Earnings Gap." *The Scandinavian Journal of Economics*, 104(1): 49–67.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 104(4): 1091–1119.

Gronau, Reuben. 1974. "Wage Comparisons-A Selectivity Bias." *Journal of Political Economy*, 82(6): 1119–1143.

Heckman, James. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." In *Annals of Economic and Social Measurement, Volume 5, number 4*. 475–492. National Bureau of Economic Research, Inc.

Heckman, James. 1990. "Varieties of Selection Bias." *The American Economic Review*, 80(2): 313–318.

Hibbs A, Douglas. 1990. "Wage Compression Under Solidarity Bargaining in Sweden." *FIEF RESEARCH REPORT NO. 30.*

Johansson, Per, and Mårten Palme. 2002. "Assessing the Effect of Public Policy on Worker Absenteeism." *Journal of Human Resources*, 37(2): 381–409.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and Gender Inequality: Evidence from Denmark." *American Economic Journal: Applied Economics*, 11(4): 181–209.

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. 2019. "Child Penalties across Countries: Evidence and Explanations." *AEA Papers and Proceedings*, 109: 122–26.

LaLonde, Robert J, and Robert H Topel. 1990. "The Assimilation of Immigrants in the U.S. Labor Markets." National Bureau of Economic Research Working Paper 3573.

Lindh, Thomas, and Henry Ohlsson. 1996. "Self-Employment and Windfall Gains: Evidence from the Swedish Lottery." *The Economic Journal*, 106(439): 1515–1526.

Manski, Charles F. 1991. "Identification of Endogenous Social Effects: The Reflection Problem." University of Wisconsin-Madison, Social Systems Research Institute SSRI Workshop Series 292712.

Manski, Charles F. 1994. "The selection problem." In *Advances in Econometrics: Sixth World Congress*. Vol. 1 of *Econometric Society Monographs*, , ed. Christopher A.Editor Sims, 143–170. Cambridge University Press.

Mas, Alexandre, and Amanda Pallais. 2020. "Alternative Work Arrangements." *Annual Review of Economics*, 12(1): 631–658.

Mulligan, Casey B., and Yona Rubinstein. 2008. "Selection, Investment, and Women's Relative Wages over Time." *The Quarterly Journal of Economics*, 123(3): 1061–1110.

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

Olivetti, Claudia, and Barbara Petrongolo. 2008. "Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps." *Journal of Labor Economics*, 26(4): 621–654.

Olivetti, Claudia, and Barbara Petrongolo. 2016. "The Evolution of Gender Gaps in Industrialized Countries." *Annual Review of Economics*, 8(1): 405–434.

Rosen, Sherwin. 1995. "Public Employment, Taxes and the Welfare State in Sweden." National Bureau of Economic Research Working Paper 5003.

Selin, Håkan. 2014. "The Rise in Female Employment and the Role of Tax Incentives. An Empirical Analysis of the Swedish Individual Tax Reform of 1971." *International Tax and Public Finance*, 21(5): 894–922.

4.A Wage Imputation Strategy

This section describes the methods used to select the exact procedure by which we infer the position of an individual's wage - either below or above - in relation to the survey year and gender-specific median wage. The purpose of this analysis is to understand which imputation strategy that best predicts the correct position in the wage distribution of the largest number of individuals, within the sample of workers.[16] We limit the evaluated sample to female workers, as these are our main focus in correcting for selection into the labor force.

Our evaluation criterion consists of three elements: post-imputation sample size, share of correct imputations (imputed and observed wage both above or both below the median wage) and difference between the predicted and "true" wage gap. We iterate over combinations of six different values of "correct panel imputation thresholds", six different "low" and six different "high" years of education- values, and three different levels of experience. These are the following:

Panel imputation thresholds: $v \in \{0.55, 0.6, 0.65, 0.7, 0.75, 0.8\}$
"High education": $h \in 12, 13, 14, 15, 17, 18$
"Low education": $l \in 7, 8, 9, 10, 11, 12$
Experience: $e \in 5, 10, 15$

For any given combination of these values, the following procedure is run:

- Separately by survey wave, drop a number of observed wages corresponding to the share of non-workers in female sample.
- On the subsample of missing female wages:
    1. Impute wage position relative to the median from plus/minus one survey wave for observations in age-education groups with v percent correct imputations (above/below median), as calculated within the sample of workers with an observed wage in either adjacent survey wave.
    2. For still missing observations: Impute a wage below median ($\hat{w} = 1$) if less than l years of schooling and less than

[16] The underlying assumption is thus that the same strategy best predict the counterfactual position of individual wages out-of-sample, among non-workers.

e years experience. Impute wage above median ($\hat{w} = 5$) if more than h years of schooling and more than e years of work experience.

3. Repeat 50 times, saving 1) share of correct imputations relative to the median, 2) predicted wage gap and 3) sample size after wage imputations.

- Calculate mean squared error between predicted and actual wage gap. Save this, the mean number of correct imputations, and the average number of imputed wages. Additionally, calculate the share of observations for whom the predicted and the observed positions of the wage with respect to the median are the same, i.e. both above or both below the median wage.

The result is N=$6 \times 6 \times 6 \times 3 = 648$ different imputation specifications to choose from. The MSE and sample size from each specification are plotted in Figure 4.A1, where the panels display results split by: (a) age-education group threshold value, (b) "low education" threshold, (c) "high education" threshold and (d) experience threshold. Dots in the lower right quadrant represent combinations that result in a high sample size and a low MSE between predicted and full-sample gender wage gaps.

More formally, we choose our preferred combination of imputation thresholds in the following way:

1. Calculate mean squared error for gender wage gap at median, mean number of imputations and share of correct imputations for each iteration (for each set of "threshold values").

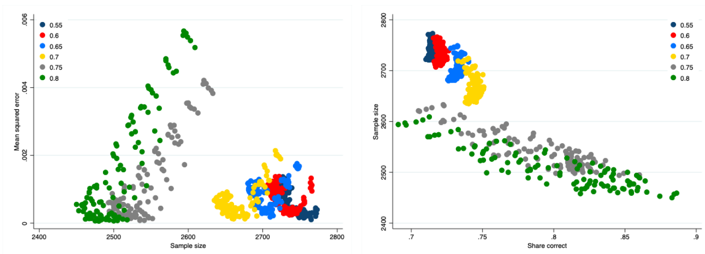
2. Rank observations, separately for 1968-1981 and 1991-2010: MSE low to high, imputation sample size and share correct imputations high to low.

3. Pick the set of threshold values that gives the lowest average rank.

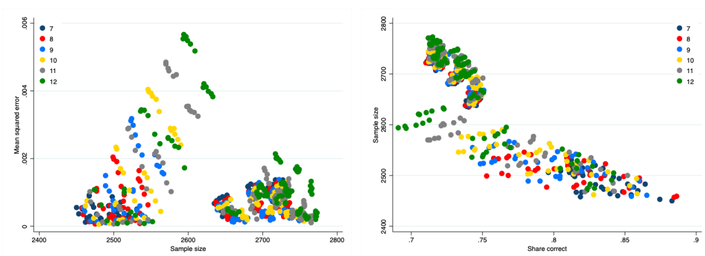
This gives the combination $\{v = 0.6, l = 11, h = 18, e = 10\}$. The resulting wage gap, sample size per survey wave and share of correct imputations relative to the median are displayed in Figures 4.A2 and 4.A3.

Figure 4.A1: Simulation results. Wage gap MSE, imputation sample size and share of correct imputations for each combination of criteria for wage imputed below or above median..

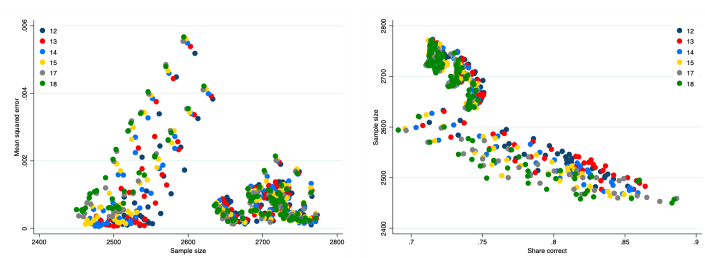
(a) Age-education group correct imputation share



(b) Low education



(c) High education



(d) Experience.

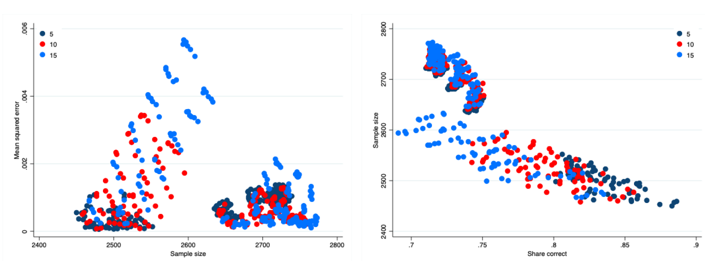
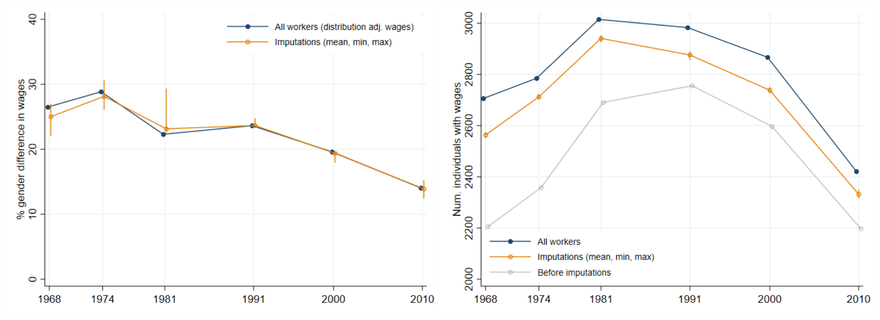


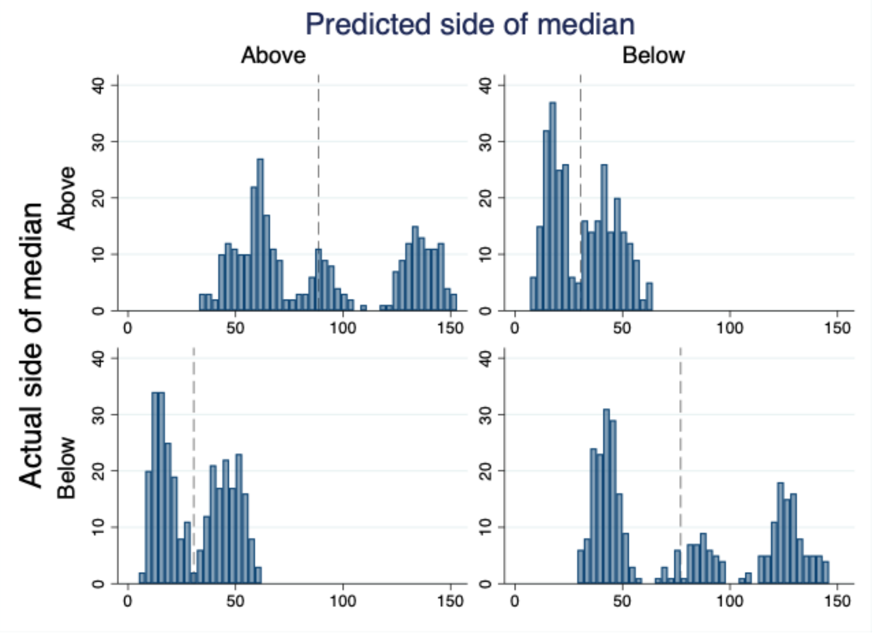
Figure 4.A2: Selected imputation strategy: Wage gap and sample sizes.

(a) Median Gender Wage Gap (b) Sample sizes



Notes: This figure evaluates the accuracy of the chosen wage imputation strategy, by showing its performance within the sample of workers. Panel a shows the median (distribution adjusted) gender wage gap for the sample of employed persons and for workers together with individuals with imputed wages after implementing the selected wage imputation strategy. Panel b shows sample sizes for 1) all workers, 2) after wage imputations according to the preferred strategy, and 3) when "dropping" wage observations at random among women, to match the year-specific participation rate (i.e. the step preceding in-sample imputations).

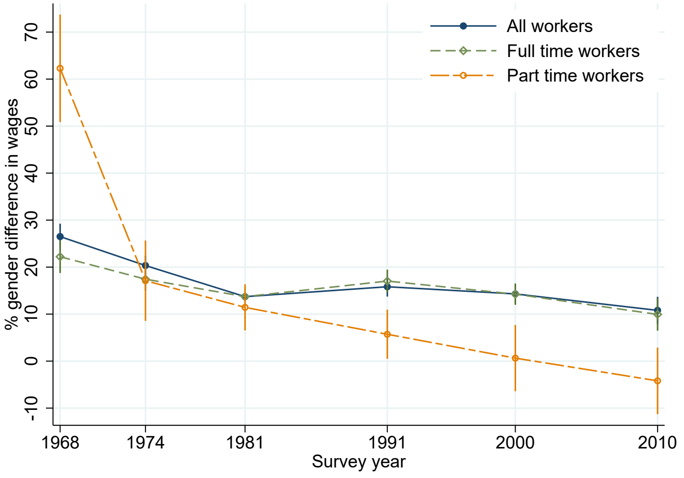
Figure 4.A3: Selected imputation strategy: Predicted vs. Real wage with respect to median wage.



Notes: This figure evaluates the accuracy of the chosen wage imputation strategy, by showing its performance within the sample of workers, in 50 random subsamples. Each iteration saves the number of observations either correctly imputed above the median, incorrectly imputed below the median, incorrectly imputed above the median, and correctly imputed below the median. Y-axis: frequency, x-axis: number of observations (iterations). The dashed line shows the mean over all subsamples.

4.B Additional Figures and Tables

Figure 4.B1: Gender Wage Gap for Full-time and Part-time Workers.

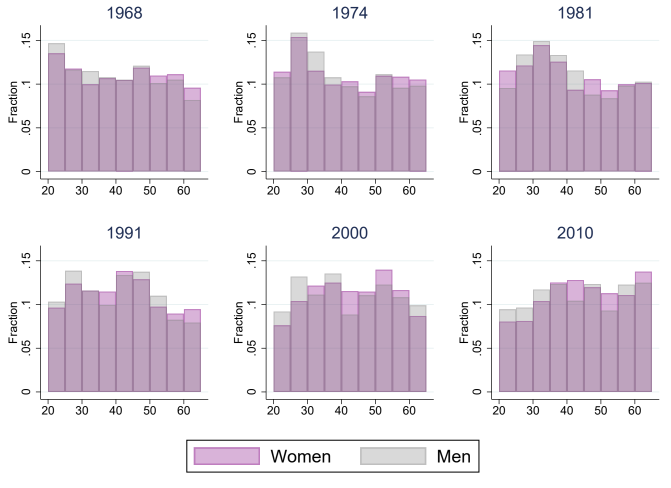


Notes: Coefficients from median regression of log(wage) on female dummy. Missing wages for working individuals imputed with wage equation.

Figure 4.B2: Age Structure of the Sample Population, by Survey Year.

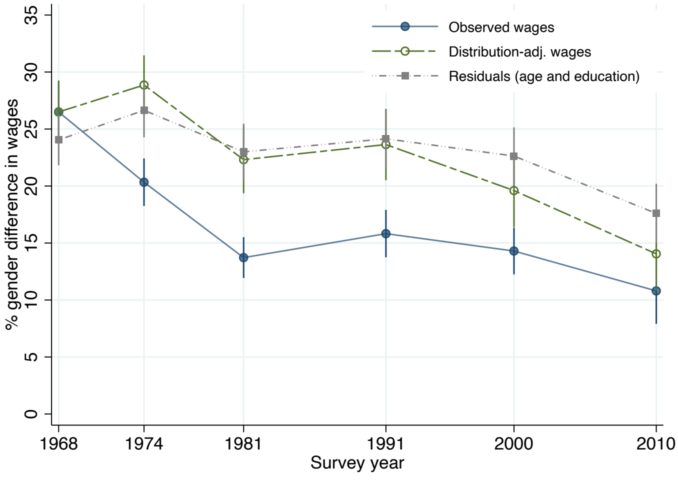
(a) Workers only

(b) Whole samples



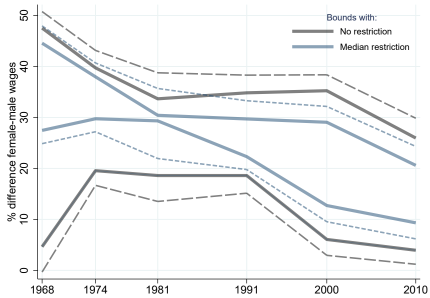
Notes: Figure shows histograms of age in five-year bins by survey year and gender. In (a), only workers are included (working status is inferred from survey responses). In (b) the whole main sample is included (workers and non-workers, subject to the sample restrictions presented above).

Figure 4.B3: Gender Wage Gap in Residuals, Controlling for Age and Education.

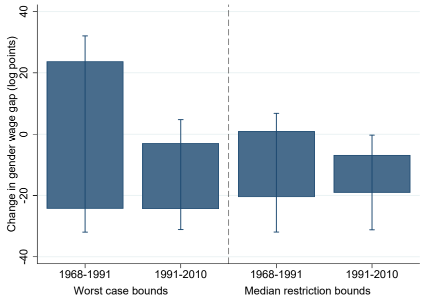


Notes: Coefficients from median regression of log(wage) on female dummy, in distribution-adjusted wages (deflated with percentile method, LaLonde and Topel (1992)). Missing wages for working individuals predicted out-of-sample with wage regression, as described in Section 4.2

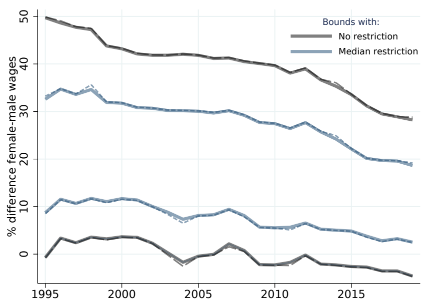
Figure 4.B4: Bounds on Change in the Gender Wage Gap with Panel Imputations.



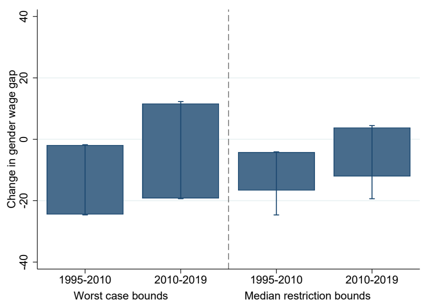
(a) SLLS Bounds on Evolution of Gender Wage Gap.



(b) SLLS Change in Gender Wage Gap 1968-1991 and 1991-2010.



(c) WSS Bounds on Evolution of Gender Wage Gap.



(d) WSS Change in Gender Wage Gap 1995-2010 and 2010-2019.

Notes: This figure shows Bounds on the change in the gender wage gap, between years 1968-1991 and 1991-2010. Bounds are calculated with the median restriction: the median wage of non-workers is assumed not to be higher than the median wage among workers. Panel imputations from plus/minus one survey wave for individuals belonging to age-education groups with above 55% correct in-sample imputations. Sample sizes printed in Table 4.B3, Column 4.

Table 4.B1: Summary Statistics, by Employment Status and Gender.

	1968	1974	1981	1991	2000	2010
Share not employed	0.292	0.220	0.160	0.129	0.178	0.166
Men	0.086	0.080	0.091	0.106	0.164	0.146
Women	0.479	0.342	0.222	0.150	0.191	0.186
	Labor supply by the employed					
Full time last week	0.809	0.795	0.731	0.778	0.805	0.808
Men	0.975	0.978	0.942	0.952	0.950	0.931
Women	0.544	0.573	0.511	0.610	0.658	0.678
Part time last week	0.184	0.200	0.268	0.222	0.195	0.192
Men	0.020	0.021	0.057	0.048	0.050	0.069
Women	0.445	0.418	0.489	0.390	0.342	0.322
	Activities of the non-employed					
Unemployed	0.042	0.076	0.103	0.251	0.348	0.395
Men	0.217	0.218	0.187	0.372	0.468	0.524
Women	0.014	0.047	0.071	0.172	0.246	0.293
Pension	0.158	0.236	0.511	0.513	0.441	0.495
Men	0.433	0.556	0.729	0.558	0.440	0.447
Women	0.113	0.171	0.431	0.483	0.441	0.532
	Marital status of the non-employed					
Married	0.816	0.756	0.711	0.636	0.629	0.560
Men	0.535	0.436	0.574	0.526	0.565	0.509
Women	0.863	0.821	0.762	0.708	0.682	0.600
Single	0.106	0.132	0.153	0.223	0.248	0.351
Men	0.382	0.383	0.310	0.370	0.332	0.406
Women	0.060	0.081	0.095	0.127	0.178	0.307
Divorced	0.078	0.111	0.136	0.141	0.123	0.089
Men	0.083	0.180	0.116	0.104	0.102	0.085
Women	0.077	0.096	0.143	0.165	0.139	0.093
	Years of schooling, by employment status					
Non-employed	7.756	8.118	8.087	9.378	11.029	12.287
Men	7.554	7.271	7.723	9.800	11.085	12.180
Women	7.790	8.291	8.221	9.109	10.982	12.372
Employed	8.634	9.722	10.514	11.551	12.654	13.809
Men	8.628	9.833	10.708	11.638	12.614	13.589
Women	8.645	9.587	10.311	11.468	12.694	14.038

Table 4.B2: Wage Statistics. SLLS Hourly Wages by Survey Year

	(1) 1968	(2) 1974	(3) 1981	(4) 1991	(5) 2000	(6) 2010
Mean wages	11.35	18.62	38.60	80.68	116.62	167.92
Median wages	10.02	17.32	35.00	74.00	104.00	148.47
CV all	0.85	0.68	0.58	0.53	0.70	0.66
CV conditional	0.47	0.37	0.35	0.34	0.47	0.44
CV cond. Women	0.41	0.32	0.28	0.25	0.31	0.33
CV cond. Men	0.46	0.36	0.36	0.35	0.53	0.49
90/10 ratio	2.47	2.03	1.93	1.90	2.00	2.14
Gini coefficient	0.22	0.18	0.16	0.16	0.18	0.19
Observations	2,706	2,786	3,016	2,983	2,867	2,421

Notes: Wages for all in each survey wave with non-missing wages. All statistics are conditional on working, except “CV all” and the Gini coefficient.

Table 4.B3: Sample Size after Wage Imputation Methods.

			Imputation method:		
	(1) All	(2) Wage	(3) Panel	(4) Panel (restr.)	(5) Panel+HC
Men					
1968	1,819	1,662	1,698	1,687	1,795
1991	1,638	1,465	1,539	1,531	1,564
2010	1,450	1,238	1,302	1,300	1,334
Women					
1968	2,004	1,044	1,284	1,247	1,930
1991	1,785	1,518	1,622	1,570	1,634
2010	1,453	1,183	1,258	1,247	1,296

Notes: Table reports yearly sample sizes with non-missing wages after each imputation method. Column samples: (1) Main SLLS sample. (2) Observed wage (i.e. workers). (3) Observed wage in the given year, or in either adjacent survey wave (plus/minus one survey wave). (4) Observed wages after imputations as in col. 3, but with imputations restricted as described in Appendix A. (5) Observed wages or imputations as in col. 4; then wage imputed below median if no high school degree and less than ten years of work experience, wage imputed above median if college degree and more than ten years of experience. In all imputation methods, wage is set to 1 if below median, 5 if above median.

Sammanfattning

Den här avhandlingen består av fyra fristående kapitel, som alla på olika sätt berör ämnet ekonomisk ojämlikhet. Kapitel ett och två handlar om hur individer som begått brott behandlas i det svenska rättsystemet. Brottslighet och ekonomisk utsatthet är ofta nära förknippade, och nationalekonomisk forskning har sedan länge beaktat att beslutet att begå en brottslig handling kan ses som ett rationellt beslut då individens uppskattade alternativkostnad för brottet är tillräckligt låg.

I det första kapitlet, **Youth Crime, Community Service and Labor Market Outcomes**, undersöker jag kopplingen mellan straff och arbetsmarknadsanknytning bland ungdomsbrottslingar. Påföljder för unga brottslingar i Sverige består nästan uteslutande av böter, ungdomsvård (individuellt anpassade vårdande insatser via socialtjänsten) eller ungdomstjänst (samhällstjänst för ungdomar). Dessa alternativ skiljer sig åt i kostnader för genomförande, men också i sin grad av rehabilitering — alltså i vilken utsträckning de hjälper ungdomsbrottslingar att integreras i samhället.

En deskriptiv analys visar att ungdomar som begår brott i betydligt högre utsträckning än den generella populationen av ungdomar mellan 15 och 17 år kommer från socioekonomiskt svaga familjer. Exempelvis bor de oftare med en ensamstående förälder, har föräldrar som är arbetslösa eller socialbidragstagare, och har en större benägenhet att ha hoppat av skolan i förtid. I teorin borde antisocialt beteende, såsom brottslighet, avhjälpas genom att individer känner sig inkluderade i majoritetssamhället, t ex genom en starkare anknytning till den formella arbetsmarknaden. Ungdomstjänst innebär kortare perioder av oavlönat arbete och liknar därför en sorts praktikplats. Skulle detta straff kunna vara ett sätt att åstadkomma en sådan anknytning?

Genom en effektutvärdering av 2007 års straffreform för unga (Prop. 2005/06:165) visar jag att ungdomar som gavs ungdomstjänst som ett alternativ till böter var mindre benägna att återfalla i brottslighet, medan ungdomar som gavs ungdomstjänst istället för ungdomsvård däremot löper högre risk att begå allvarliga brott och få fängelsestraff i framtiden. Dessutom varierar effekten av reformen beroende på socioekonomisk bakgrund och tidigare arbetslivserfarenhet. Domstolar i Sverige ska enligt lag ta hänsyn till individuella omständigheter vid bestämningen av rättsliga påföljder (Brottsbalk (1962:700), Kap 32). Resultaten i den här studien indikerar att allokeringen av straff till unga brottslingar skulle kunna göras mer skräddarsydd och därmed gynna den dömda, samtidigt som effektiviteten i rättssystemet skulle öka. Studien visar vidare att såväl straffvärdet som graden av vårdande insatser i ungdomspåföljder bör tas i beaktning för att minska återfall i brott och motivera tidigare dömda ungdomar att ta gymnasieexamen och delta aktivt på arbetsmarknaden.

Kapitel två handlar istället om själva domstolsprocessen, och om hur utfall i brottsmålsrättegångar kan skilja sig åt, beroende på vilka individer som slumpmässigt tilldelats att döma i målet. I **Identity in Court Decision-Making**, skriven tillsammans med Susan Niknami och Mårten Palme, visar vi att sociala grupptillhörigheter, eller "identitet", kan spela en avgörande roll för det straff som utmäts åt en tilltalad i brotssmål. I Sverige, precis som i många andra länder, tillämpas ett lekmannasystem vid rättegångar, där "vanligt folk" — nämndemän — deltar i beslutet om fällande dom och vilket straff som i så fall ska ges. Syftet är demokratisk representation i domstolar, och får till följd att brotssmisstänkta döms av sina jämlikar. Tidigare literatur inom socialpsykologi och beteendeekonomi har funnit en benägenhet hos människor att döma andra som i högre grad liknar oss själva mildare, ofta benämnt "in-gruppsfavorisering". Vilken effekt har i praktiken likhet mellan tilltalade och dömande parter för domstolsbeslut?

Vi använder oss av egeninsamlade data från brottsmålsprotokoll från Stockholms tingsrätt år 2000-2004, vilka vi har länkat samman med registerdata om demografiska uppgifter, utbildning och inkomster om den tilltalade och nämnemännen i ungefär 10 000 brotssmål. Genom detta kan vi skapa mått på likhet mellan tilltalad och nämndemän i sex olika observerbara egenskaper: kön, etnisk bakgrund, ålder (tre demografiska variabler) och utbildning, familjens disponibla inkomst och bostadsområde (tre socioekonomiska faktorer). Vi studerar effekten av större likhet mellan tilltalad och nämnden i termer av alla dessa sex egenskaper, och skapar vidare tre index utifrån dem: *demografisk identitet*, *socioekonomisk identitet* och totalen av alla sex (*"identitet"*). Eftersom nämndemän och domare är slumpmässigt anvisade till brottsmål, kan effekten på domslut av en högre andel liknande egenskaper mellan tilltalad och nämnden tolkas som kausal.

Resultaten visar att social identitet spelar en viktig roll för domslutens utformning: Nämndemän som slumpmässigt allokerats till en tilltalad som i högre grad liknar dem utdelar mildare straff. Särskilt finner vi att in-gruppsfavorisering uppstår när tilltalad och nämnd har liknande socioekonomisk bakgrund, vilket är en ny addering till literaturen, som tidigare mest kunnat studera gruppeffekter av kön och etnisk bakgrund. Vårt resultat tyder på att en tilltalad som ställs inför en nämnd där samtliga tre medlemmar har samma utbildningsnivå som hon själv löper 15 procent lägre risk att dömas till fängelse och får fängelsestraff som är i genomsnitt hälften så långa, jämfört med en tilltalad vars dom bestäms av en nämnd där samtliga har en annan utbildningsbakgrund. Vi finner alltså att även egenskaper som inte direkt går att observera om en främmande människa, såsom utbildning, leder till kännbara konsekvenser för individer i kontakt med rättssystemet. Det är inte otänkbart att liknande effekter skulle kunna uppstå i andra situationer där människor blir bedömda av samhällets representanter, såsom betygsättning i skolan och ansökningar till socialförsäkringssystemet.

Kapitel tre och fyra byter fokus till den historiska utvecklingen av kvinnligt arbetskraftsdeltagande. Under andra halvan av 1900-talet påbörjades en "revolution" på arbetsmarknaden, då kvinnor i allt högre utsträckning deltog i lönearbete på liknande villkor som män. I kapitel tre, **Intergenerational Mobility Trends and the Changing Role of Female Labor**, samskrivet med René Karadakic och Joachim Kahr Rasmussen, undersöker vi hur denna strukturella transformation av arbetsmarknaden har påverkat intergenerationell inkomströrlighet, alltså samvariationen i inkomster mellan barn och deras föräldrar. Detta mått används ofta för att studera i vilken utsträckning individers ekonomiska status bestäms av deras familjebakgrund och alltså i förlängningen om individer föds med liknande ekonomiska möjligheter i livet.

Vi använder registerdata från Sverige, Danmark och Norge för att studera hur den intergenerationella rörligheten har förändrats över tid, för individer födda mellan 1951 och 1979. Vi mäter rörligheten som *rangkorrelationer* (korrelationen mellan föräldrars inkomster rankade mellan noll och hundra och deras barns inkomstrank), där föräldrarnas inkomster definieras som medelvärdet mellan moderns och faderns inkomster. Denna beskrivning gör det genast uppenbart varför ökat kvinnligt arbetsutbud skulle påverka trenden i rörlighet: Över tid kommer kvinnors inkomster bättre reflektera deras *inkomstpotential*, alltså de färdigheter såsom utbildning och kognitiv förmåga som kan omsättas i lön på arbetsmarknaden. Given att en viss andel av dessa färdigheter överförs från mödrar till deras barn, kommer korrelationen mellan mödrars och barns inkomster att öka över tid, och den observerade rörligheten att minska. Denna utveckling uppstår oavsett om förändringen i arbetsutbud sker på den så kallade *extensiva marginalen* (kvinnors huvudsysselsättning byts från hushållsarbete till lönearbete) eller den *intensiva marginalen* (kvinnors veckoarbetstimmar ökar, deras yrke och plats i hierarkin på arbetsplatsen bestäms i högre grad av förmåga och mindre av t ex könsnormer eller diskriminering).

Våra resultat visar på precis denna utveckling. Den intergenerationella rörligheten minskade i hela Skandinavien mellan födelsekohorter från 1951 till 1979 (1962 till 1979 för Danmark), oavsett vilket inkomstmått vi använder och hur vi definierar föräldrars inkomster. Däremot är detta en utveckling som endast eller huvudsaklingen observeras mellan mödrar och söner och döttrar, och mellan fäder och döttrar — med andra ord i familjekonsellationer med minst en kvinna. I nästa steg styrker vi hypotesen att denna utveckling uppstår på grund av att kvinnligt arbete värderas på liknande grunder som manligt arbete. Detta görs först med hjälp av en modellbaserad dekomponering, och sedan empiriskt genom att vi mäter trenden i intergenerationell rörlighet i *ekonomisk status* (här definierat som inkomster, utbildning och yrke) istället för endast i inkomster. Dessa analyser tyder på att utvecklingen mot lägre social rörlighet inte ska tolkas som en samhällsutveckling där barns möjligheter i livet har blivit mindre likvärdiga, utan som en sidoeffekt av en socialt önskvärd utveckling mot en jämlikare arbetsmarknad för kvinnor.

Avslutningsvis bjuder kapitel fyra på en redogörelse för utvecklingen av löneskillnader mellan män och kvinnor på den svenska arbetsmarknaden under åren 1968 till 2019. Utgångspunkten för studien är att könsgapet i genomsnittliga timlöner mellan män och kvinnor har förändrats relativt lite över tid, samtidigt som kvinnors arbetsutbud har genomgått dramatiska förändringar. Under samma tidsperiod har dock en rad andra omvandlingar skett, som påverkar den uppmätta skillnaden i timlöner mellan män och kvinnor, och för att förstå hur lönegapet har utvecklats över tid krävs en analys som tar hänsyn till dessa influenser.

Vi tar fasta på tre strukturella skillnader över tid: Att kvinnor i högre utsträckning arbetar heltid, att *lönespridningen* (den absoluta skillnaden mellan höga och låga timlöner) har ökat, samt att arbetskraftens sammansättning har ändrats över tid ("selektion in i arbete"). För att kunna studera lönegapet i timlöner över en lång tidsperiod använder vi oss av data från Levnadsnivåundersökningarna (LNU), vilket är en panelundersökning som täcker 0.1 procent av befolkningen och har genomförts i sex omgångar mellan 1968 och 2010. Vi kompletterar detta med registerbaserad data från Lönestrukturstatistiken, för att kunna täcka in åren 1995-2019 i analysen. Huvudfokus i analysen är på könsgapet i medianlöner.

I ett första steg jämför vi utvecklingen av lönegapet i månadslöner med det i timlöner, för att förstå inverkan av fler arbetade timmar per vecka. Medan könsgapet i månadslöner minskar med 32 procentenheter mellan 1968 och 2019, så är motsvarande minskning i termer av timlöner endast 17 procentenheter. Detta visar alltså att den viktigaste faktorn för minskade löneskillnaden mellan män och kvinnor över tid har berott på kvinnors förändrade veckoarbetstimmar. Nästa steg är att förstå effekten av generalla förändringar i lönespridningen, vilken under vår tidsperiod först minskar drastiskt mellan 1968 och 1991, bland annat på grund av ökad fackligt inflytande, för att sedan öka igen från 1991 och framåt. Givet att kvinnor ofta är lägre betalda än män så innebär en komprimerad lönefördelning ett mindre könslönegap, allt annat lika, och vice versa. När vi korrigerar för dessa förändringar, finner vi att könslönegapet endast minskade med ett par procentenheter mellan 1968 och 1991, för att sedan minska i betydligt högre takt än vad icke-korrigerade löner visar.

Till sist visar vi, genom att använda oss av ett flertal olika metoder som föreslagits i tidigare literatur, att selektion in i arbete har haft en betydande inverkan på lönesgapets förändring mellan 1968 och 1991. Under antagandet att kvinnor i lönearbete på 60- och 70-talet var *positivt selekterade* (alltså hade egenskaper som gjorde att de tjänade relativt höga löner) och att selektionen därefter gradvis har avtagit allt eftersom den kvinnliga arbetskraften har utökats, finner vi att förändringen i könslönegapet hade varit upp till tre gånger så stor över denna tidsperiod, om arbetskraftens sammansättning varit oförändrad.

Dissertations in Economics (1975- 2022)

1975	Claes-Henric Siven	A study in the theory of inflation and unemployment
1976	Bo Axell	Prices under imperfect information: a theory of search market equilibrium
1976	Thomas Franzén Kerstin Lövgren Irma Rosenberg	Skatters och offentliga utgifters effekter på inkomstfördelningen: en teoretisk och empirisk studie
1977	Hans-Olof Hagén	Grafisk industri i omvandling: en produktionsteknisk studie
1977	Peter Svedberg	Foreign investment and trade policies in an international economy with transnational corporations
1978	Nils Bruzelius	The value of travel time: theory and measurement
1979	Alfred Kanis	Demand for factors of production: an interrelated model of Swedish mining & manufacturing industry
1979	Aleksander Markowski	A formal versus an unformal forecasting model: an investigation of the forecasting procedure of the Swedish National institute of economic research
1980	Lorenzo Brown	The technical representation of returns to scale on cost and production functions
1981	Dick Kling	Spridning av ny teknik: en analys av anpassningsprocesser i svensk industri
1981	Richard Murray	Kommunernas roll i den offentliga sektorn
1981	Åsa Sohlman	Education, labour market and human capital models: Swedish experiences and theoretical analyses
1981	Edward E. Palmer	Determination of personal consumption: theoretical foundations and empirical evidence from Sweden
1983	Mats Bohman	Effektivitetsproblem inom vatten- och avloppsområdet
1983	Hillman Soeria-Atmadja	Product guarantees and liability rules
1984	Dag Lindskog	Foreign disturbances and domestic reactions study of wage and exchange rate policy reactions in Denmark, Finland, Iceland and Sweden 1973-81
1985	Nils Gottfries	Essays on price determination and expectations
1986	Farouk Kobba	Foreign investment export promotion and economic development

1986	Hans Wijkander	Indirect corrections and disequilibrium pricing: studies in second-best policies
1986	Göran Östblom	Structural change in the Swedish economy: empirical and methodological studies of changes in input-output structures
1987	Ronny Norén	Comparative advantages revealed: Experiments with a quadratic programming model of Sweden
1988	Emil Ems	Economics of public information systems
1988	Sten Kjellman	International trade in steam-coal
1988	Christopher Sardelis	Local employment policy : a study on the properties of intermunicipal rivalry
1990	Hans Lind	Tanken bakom tänkta ekonomier: om forskningsstrategi i modern nationalekonomi
1991	Kari Lantto	Optimal deterrents to malingering: the role of incentives, attitudes and information costs in social insurance, especially sickness benefit and welfare
1991	Kjell Jansson	Efficient prices and quality in public transport
1991	Tekaligne Godana	The behavior and performance of public enterprises: a theoretical and empirical analysis
1992	Claes Berg	Optimal investment and option values under risk-aversion with empirical evidence from Swedish manufacturing
1992	No-Ho	Optimal tax mixes
1992	Mats Kinnwall	Target zones : theory and evidence
1993	Kurt Lundgren	Aspects of learning by doing and firm behavior
1993	Ann-Mari Sätre	The effect of the Soviet shortage economy on the environment and the use of natural resources
1995	Jonas Björnerstedt	Essays in evolutionary game theory
1996:1	Ariane Lambert-Mogiliansky	Essays on corruption
1996:2	Joakim Sonnegård	Essays on experimental economics: bargaining, auctions, reservation-price elicitation and political stock markets
1997:1	Jan-Eric Nilsson	Efficiency Implications of Innovations in Administering Transport Infrastructure
1998:1	Xiang Li	Essays on the political economy of central bank policy

1998:2	Pehr-Johan Norbäck	Multinational Firms, Technology and Location
1999:1	Gabriel C Oxenstierna	Market power in the Swedish banking oligopoly
1999:2	Kerstin Hallsten	Essays on the effects of monetary policy
1999:3	Tomas Forsfält	Timing options and taxation: essays on the economics of firm creation and tax evasion
2000:1	Björn Carlén	Studies in climate change policy: theory and experiments
2001:1	Sven-Olof Fridolfsson	Essays on endogenous merger theory
2001:2	Helen Jakobsson	Essays on concentration, border effects and R&D cooperation
2002:1	Matthew J. Lindquist	Essays on the dynamics of wage inequality
2002:2	Jesper Roine	The political economics of not paying taxes
2002:3	Anne D. Boschini	Three essays on the economics of institutions
2002:4	Helena Svaleryd	Essays in finance, trade and politics
2002:5	Mikael Priks	Corruption, rent seeking and efficient governance
2002:6	Anne-Sophie Crépin	Tackling the economics of ecosystems
2003:1	Yoshihiko Fukushima	Essays on employment policies
2003:2	Lena Nekby	Empirical studies on health insurance, employment of immigrants and the gender wage gap
2003:3	Nils Bohlin	Essays on urban wages, location and retail trade
2003:4	Dan Nyberg	Essays on exchange rate risk and uncertainty
2003:5	Tobias Lindqvist	Essays on mergers and financial markets
2003:6	Adam Jacobsson	War, drugs, and media: arenas of conflict
2004:1	Tobias Nilsson	Essays on Voting and Government Inefficiency

2004:2	John Ekberg	Essays in Empirical Labor Economics
2004:3	Petroulas Pavlos	International Capital Flows: Effects, Defects and Possibilities
2005:1	Carlos Razo Perez	Mergers, Congestion and Collusion: Essay on Merger Policy
2005:2	Jan Pettersson	Three Empirical Studies on Development: Democracy, the Resource Curse and Aid
2005:3	Anna Nilsson	Indirect effects of unemployment and low earnings: Crime and children's school performance
2005:4	Bo Larsson	Essays on Banking and Portfolio Choice
2005:5	Carl Wilkens	Auri sacra fames: Interest Rates – Prediction, Jumps and the Market Price of Risk
2006:1	Magnus Wiberg	Essays on the Political Economy of Protection and Industrial Location
2007:1	Alberto J. Naranjo	Drugonomics: Industrial Organization of Illegal Drug Markets
2007:2	Maria Jakobsson	Empirical Studies on Merger Policy and Collusive Behaviour
2007:3	Camilo von Greiff	Income Redistribution, Educational Choice and Growth
2007:4	Sara Åhlén	Firms, Employment and Distance: Essays on the Swedish Regional Economy
2008:1	Anders Åkerman	Essays on International Trade, Productivity and Firm Heterogeneity
2009:1	Li-Ju Chen	Essays on Female Policymakers and Policy Outcomes
2009:2	Johan Kiessling	Essays on technology adoption and political reform in developing countries
2009:3	Paolo Zagaglia	The Macroeconomics of the Term Structure of Interest Rates
2009:4	Lars M. Johansson	Studies of the relationship between aid and trade and the fiscal implications of emigration and HIV/AIDS interventions
2009:5	Tobias Heinrich	Essays on Growth Econometrics and Endogenous Information
2009:6	Marie Gartell	Educational Choice and Labor Market Outcomes: Essays in Empirical Labor Economics
2010:1	Shon Ferguson	Essays on Trade, Technology and the Organization of Firms

2010:2	Eva Skult	Studies on Saving under Uncertainty
2010:3	Hans Lindblad	Essays on Unemployment and Real Exchange Rates
2011:1	Gûlay Özcan	Essays on Labor Market Disparities and Discrimination
2011:2	Jaewon Kim	Trade, Unemployment and Real Exchange Rates
2011:3	Magnus Rödin	Gender, Ethnicity and Labor Market Disparities
2012:1	Karolina Holmberg	Empirical Essays in Macroeconomics and Finance
2012:2	Martin Olssson	Essays on Unemployment Protection, Private Equity and Spousal Behavior
2012:3	Nicholas Sheard	Regional Economics, Trade and Transport Infrastructure
2012:4	Gustav Engström	Essays on Economics Modelling of Climate Change
2012:5	Lisa Laun	Studies on Social Insurance, Income Taxation and Labor Supply
2013:1	Christian Odendahl	Parties, Majorities, Incumbencies
2013:2	Maria Cheung	Education, Gender and Media
2013:3	Emma von Essen	Understanding Unequal Outcomes
2013:4	Mathias Ekström	Cues, Conformity, and Choice Architecture
2013:5	Eric Sjöberg	Essays on Environmental Regulation, Management and Conflict
2013:6	Pedro Soares Brinca	Essays in Quantitative Macroeconomics
2013:7	Marc Sanctuary	Essays on trade and environment
2014:1	Linnea Wickström Östervall	Essays on Antibiotics Use: Nudges, Preferences & Welfare Benefits
2014:2	Wei Xiao	Migration, Crime and Search in Spatial Markets
2015:1	Johan Egebark	Taxes, Nudges, and Conformity: Essays in Labor and Behavioral Economics

2015:2	Manja Gärtner	Prosocial Behavior and Redistributive Preferences
2015:3	Sara Fogelberg Lövgren	Markets, Interventions and Externalities: Four Essays in Applied Economics
2016:1	Theodoros Rapanos	Essays on the Economics of Networks Under Incomplete Information
2017:1	Kiflu Gedefe Molla	Essays in International trade, exchange rates and prices
2017:2	Jürg Fausch	Essays on Financial Markets and the Macroeconomy
2017:3	Anders Österling	Housing Markets and Mortgage Finance
2017:4	Mengyi Cao	Labor, Trade and Finance: Essays in Applied Economics
2018:1	Evangelia Pateli	Essays on International Trade: Theory and Evidence on the Determinants and Implications of Firms' Import Behaviour
2018:2	Wei Si	Empirical Essays in Labor and Development Economics
2018:3	Jakob Almerud	Public Policy, Household Finance and the Macroeconomy
2018:4	Dany Kessel	School Choice, School Performance and School Segregation: Institutions and Design
2018:5	Daniel Knutsson	Public Health Programmes, Healthcare and Child Health
2018:6	Joakim Jansson	We are (not) anonymous: Essays on anonymity, discrimination and online hate
2018:7	Laurence Malafry	Inequality and Macroeconomic Policy: Essays on Climate, Immigration and Fiscal Intervention
2019:1	Elisabet Olme	Essays on Educational Choices and Integration
2019:2	Vanessa Sternbeck Fryxell	Essays on Interbank Markets
2019:3	Mathias Pronin	Essays in Macroeconomics and Political Economy
2020:1	Niklas Blomqvist	Essays on Labor Economics: The Role of Government in Labor Supply Choices
2020:2	Daniel Almén	Societal Impacts of Modern Conscription: Human Capital, Social Capital and Criminal Behaviour
2020:3	Louise Lorentzon	Empirical Essays on Public Policies: Social Insurances, Safety Nets, and Health Care

2020:4	Tamara Thornquist	Essays in economics: The impact of changes on the labor market induced by structural change, the adoption of a new computer-based technology and economic slowdowns on family formation, family fertility outcomes and new careers
2021:1	Nanna Fukushima	Essays on the Economics of the 1956 Clean Air Act
2021:2	Roza Khoban	Globalization and Development: The Impact of International Trade on Political and Social Institutions
2022:1	Ulrika Ahrsjö	Essays on Economic Disadvantage: Criminal Justice, Gender and Social Mobility

This thesis consists of four empirical essays that touch upon the subject of economic disadvantage. Chapters one and two concern the Swedish criminal justice system: how youth offenders are sanctioned and the role of identity in court outcomes. Chapter three examines how increased female labor force participation affects the intergenerational transmission of socioeconomic status, and chapter four provides an overview of the evolution of the gender wage gap when structural changes in the labor market are taken into account.

Ulrika Ahrsjö
... holds a BSc from Uppsala University, and a MSc from Stockholm University. In an alternate universe, this thesis would instead be about computational linguistics, or possibly about sourdough baking.

ISBN 978-91-7911-894-5
ISSN 1404-3491

Stockholm University

Department of Economics