Essays in Empirical Labour and Public Economics

Spencer Smith

A thesis submitted for the degree of
Doctor of Philosophy in Economics



Nuffield College
University of Oxford

Trinity Term 2019

Abstract

This thesis comprises three essays in empirical labour and public economics. Chapter 1, co-authored with Anne Brockmeyer, Marco Hernandez, and Stewart Kettle, concerns taxation and economic development. The majority of firms in developing countries are informal, yet even among firms registered with the government, tax filing rates are low. We argue that nonfiling of taxes among registered firms constitutes an important intermediate form of informality, which can be tackled cost-effectively. Using a randomised experiment in Costa Rica, we show that credible enforcement emails increased tax filing and payment rates among previously nonfiling firms. Emails that highlight third-party reports of a firm's transactions further increased compliance. The effect persisted over two years, and treated firms became more likely to report transactions with other firms, facilitating future tax enforcement.

Chapter 2 considers the wage effects of liability for employment discrimination. In both a standard competitive model and a monopsony model of the labour market, the net theoretical effect of liability for employment discrimination is ambiguous. Therefore, whether liability for employment discrimination increases the wages of protected workers and, if so, by how much, is an important empirical question. This essay proposes an empirical strategy to estimate the effects of liability for employment discrimination, using discontinuities in coverage of United States antidiscrimination law and in damages awardable in discrimination lawsuits based on firm size. Future work will aim to use confidential employer-employee matched data held by the U.S. Census Bureau. However, preliminary estimates using survey data suggest that coverage under federal antidiscrimination law increased relative wages for black men by 12 to 16 percent.

Chapter 3 uses data from the London Metropolitan Police Service to investigate discrimination in police searches and to estimate the effects of body cameras on discrimination. Prior to a body camera trial, black individuals were 2.5 times more likely to be searched than white or South Asian individuals, and men were 15.8 times more likely to be searched than women. These disparities cannot be completely explained by accurate statistical discrimination, as there were significant differences in search success rates between groups—differences consistent with preference-based or inaccurate statistical discrimination against young minority-ethnic men. Using variation in body camera use resulting from a one-year camera trial, I find suggestive but not conclusive evidence that cameras reduced these disparities somewhat, particularly the gender difference in search success rates.

Acknowledgements

I am grateful to my supervisors, Abi Adams and Brian Bell, for their guidance; to my co-authors, Anne Brockmeyer, Marco Hernandez, and Stewart Kettle, for their collaboration; to the Marshall Aid Commemoration Commission, Nuffield College, the World Bank, and the John M. Olin Center for Law, Economics, and Business at Harvard Law School, for financial and institutional support; to the Ministry of Finance and General Directorate of Taxation of Costa Rica; to the London Metropolitan Police Service; to Miguel Almunia, Pierre Bachas, Miriam Bruhn, Raul Félix Junquera-Varela, Louis Kaplow, Paul Klemperer, David McKenzie, Marina Mendes Tavares, Blanca Moreno-Dodson, Miguel Pecho, Simon Quinn, Steven Shavell, Gabriel Tourek, Bassirou Sarr, Joel Slemrod, Cass Sunstein, Crystal Yang, and seminar participants at the Copenhagen Workshop on Tax Evasion and Inequality, the Pacific Conference for Development Economics, the Centre for the Study of African Economies (CSAE) Annual Conference, the CSAE Research Workshop, the Oxford University Centre for Business Taxation, the Inter-American Development Bank, the National Tax Association Conference, the Harvard Law School Law and Economics Workshop, and the Harvard Law School Empirical Legal Studies Workshop, for helpful comments and discussions; to my family, for their unending support; to my friends, for their camaraderie; and to Kevin Carney, for filling the years spent working on this research with so much joy. This thesis would not have been possible without you.

Contents

Abstract iii

Acknowledgements iv

Introduction 1

1 Casting a Wider Tax Net: Experimental Evidence from Costa Rica 5
- 1.1 Introduction . . . 5
- 1.2 Context and Data . . . 10
  - 1.2.1 Income and Sales Tax . . . 10
  - 1.2.2 Third-Party Information . . . 13
  - 1.2.3 Data . . . 14
- 1.3 Experiment Design . . . 14
  - 1.3.1 Treatments . . . 15
  - 1.3.2 Timing and Follow-Up . . . 17
  - 1.3.3 Enforcement Context . . . 18
  - 1.3.4 Balance of Randomisation . . . 20
  - 1.3.5 Estimation . . . 21
- 1.4 Baseline Results . . . 21
  - 1.4.1 Income Tax Compliance . . . 22
  - 1.4.2 Other Compliance Outcomes . . . 27
  - 1.4.3 Heterogeneity . . . 32
- 1.5 Medium-Term Effects and Mechanisms . . . 36
  - 1.5.1 Persistence of Treatment Effects . . . 36
  - 1.5.2 Response to Repeated Interventions . . . 40
  - 1.5.3 Reporting Behaviour . . . 44
- 1.6 Cost-Benefit Analysis . . . 47
- 1.7 Conclusion . . . 49
- 1.8 References . . . 51
- 1.9 Appendix . . . 55
  - 1.9.1 A Simple Tax Filing Model . . . 55
  - 1.9.2 Potential Value of Additional Third-Party Information . . . 58
  - 1.9.3 Figures . . . 60
  - 1.9.4 Tables . . . 77

2 The Wage Effects of Liability for Employment Discrimination 91
2.1 Introduction . . . 91
2.2 Background . . . 94
2.2.1 Economic Models of Employment Discrimination . . . 94
2.2.2 Evidence of Employment Discrimination . . . 98
2.2.3 Liability for Employment Discrimination in the United States . 103
2.3 The Wage Effects of Liability for Employment Discrimination: Theory . 109
2.4 The Wage Effects of Liability for Employment Discrimination: Evidence 113
2.4.1 Existing Evidence . . . 113
2.4.2 Empirical Strategy . . . 116
2.4.3 Results . . . 136
2.5 Conclusion . . . 146
2.6 References . . . 149
2.7 Appendix . . . 155

3 Police Discrimination and Body Cameras: Evidence from London 161
3.1 Introduction . . . 161
3.2 Background . . . 164
3.3 Conceptual Framework . . . 166
3.4 Data . . . 170
3.5 Results: Searches and Search Outcomes . . . 172
3.6 Results: Body Cameras . . . 182
3.7 Conclusion . . . 187
3.8 References . . . 190
3.9 Appendix . . . 193

Conclusion 205

List of Figures

1.1 Tax Nonfiling Rates Across Countries . . . 12
1.2 Compliance by Information Coverage and Treatment Group, 2014 . . . 23
1.3 Heterogeneity by Third-Party Information . . . 35
1.4 Filing Response to Repeated Income Tax Interventions . . . 42
1.5 Reporting Behaviour of Tax Filers . . . 45
1.A1 Filing Rates Over Time . . . 60
1.A2 Email to Firms Covered by Third-Party Information . . . 61
1.A3 Email to Firms Not Covered by Third-Party Information . . . 62
1.A4 Impact of Follow-up Activities . . . 63
1.A5 Heterogeneity by Third-Party Information (Percentiles) . . . 64
1.A6 Self-Reported versus Third-Party Reported Sales . . . 65
1.A7 Self-Reported Sales versus Self-Reported Costs . . . 66
1.A8 Compliance by Information Coverage and Treatment Group, 2015 . . . 67
1.A9 Compliance by Information Coverage and Treatment Group, 2016 . . . 68
1.A10 Distribution of Third-Party Reports around Reporting Threshold . . . 69
1.A11 Filing Response to Sales Tax Interventions, 2015 . . . 70
1.A12 Filing Response to Sales Tax Interventions, 2016 . . . 71
1.A13 Filing Response to Multiple Repeated Interventions . . . 72
1.A14 Impact of Online Publication of Sales Tax Nonfilers . . . 73
1.A15 Availability and Use of Third-Party Information Across Countries . . . 74

2.1 Limits on Title VII Damages after the Civil Rights Act of 1991 . . . 106
2.2 Frequency of Reported Firm Sizes in NLSY79, 1994–2010 . . . 128
2.3 U.S. Firm Size Distribution, 1994–2000 & 2006 . . . 132
2.4 Wages and Title VII Coverage: Detail (Men) . . . 138
2.5 Wages and Title VII Liability: All Cutoffs (Men) . . . 144
2.A1 Wages and Title VII Coverage: Detail (Women) . . . 155
2.A2 Wages and Title VII Liability: All Cutoffs (Women) . . . 156

3.1 Search Grounds . . . 171
3.2 Ethnic Makeup of Searched Persons versus London Population . . . 174
3.3 Ethnic Makeup of Searched Persons by Search Grounds . . . 175
3.4 Gender Makeup of Searched Persons versus London Population . . . 176
3.5 Search Success Rates by Ethnic Appearance . . . 179
3.6 Search Success Rates by Gender . . . 180
3.7 Search Success Rates by Age . . . 181

3.8 Body Camera Trial Boroughs 184
3.9 Timing of Introduction of Body Cameras 186
3.A1 Police Stop and Search Form 193
3.A2 Gender Makeup of Searches by Search Grounds 194
3.A3 Search Success Rates by Ethnic Appearance and Age 195
3.A4 Search Success Rates by Gender and Age 196
3.A5 Search Success Rates by Ethnic Appearance and Gender 197

List of Tables

1.1 Impact on Income Tax Compliance . . . 25
1.2 Impact on Other Compliance Outcomes (Part 1) . . . 29
1.3 Impact on Other Compliance Outcomes (Part 2) . . . 31
1.4 Impact on Informative Declarations . . . 33
1.5 Medium-Term Impact on Compliance Outcomes . . . 37
1.6 Medium-Term Impact on Third-Party Reports . . . 39
1.A1 Summary of Filers and Nonfilers, Firm Characteristics . . . 75
1.A2 Summary of Filers and Nonfilers, Past Compliance . . . 76
1.A3 Experiment Timeline . . . 77
1.A4 Enforcement Activities, 2008–2015 . . . 78
1.A5 Balance of Randomisation, Firm Characteristics . . . 79
1.A6 Balance of Randomisation, Outcomes at Baseline . . . 80
1.A7 Impact on Income Tax Compliance, Firms with Third-Party Information (Full Results) . . . 81
1.A8 Impact on Income Tax Compliance, Firms without Third-Party Information (Full Results) . . . 82
1.A9 Impact on Income Tax Compliance, LATE . . . 83
1.A10 Impact on Other Compliance Outcomes, LATE . . . 84
1.A11 Medium-Term Impact on Compliance Outcomes, LATE . . . 85
1.A12 Medium-Term Impact on Third-Party Report Coverage, LATE . . . 86
1.A13 Heterogeneity of Treatment Effects by Firm Characteristics . . . 87
1.A14 Heterogeneity of Treatment Effects by Third-Party Information . . . 88
1.A15 Response to 2016 Income Tax Intervention . . . 89
1.A16 Impact on Reporting Behaviour of Tax Filers . . . 90

2.1 RD Estimates of the Effect of Title VII Coverage on Log Real Hourly Wages for Black Men Relative to White Men: Detail . . . 140
2.2 RD Estimates of the Effect of Title VII Liability on Log Real Hourly Wages for Black Men Relative to White Men: Upper Cutoffs . . . 145
2.A1 RD Estimates of the Effect of Title VII Coverage on Log Real Hourly Wages for Black Women Relative to White Women: Detail . . . 157
2.A2 RD Estimates of the Effect of Title VII Liability on Log Real Hourly Wages for Black Women Relative to White Women: Upper Cutoffs . . . 158
2.A3 Determinants of Log Real Hourly Wages: Men (First Stage) . . . 159
2.A4 Determinants of Log Real Hourly Wages: Women (First Stage) . . . 160

3.1 Search Outcomes . . . 172
3.2 Search Success Rates . . . 178
3.3 Effect of Body Cameras on Demographic Shares of Searches . . . 186
3.4 Effect of Body Cameras on Search Success Rates . . . 188
3.A1 Search Types . . . 198
3.A2 Searches by London Borough . . . 199
3.A3 Search Grounds: Detail . . . 200
3.A4 Ethnic Appearance of Searched Persons . . . 200
3.A5 Self-Defined Ethnicity of Searched Persons . . . 201
3.A6 Search Outcomes: Detail . . . 201
3.A7 Summary Statistics . . . 202
3.A8 Search Success Rates: Detail . . . 203
3.A9 Search Success Rates by Search Grounds . . . 204

Introduction

This thesis comprises three essays in empirical labour and public economics. Each essay relies on an experimental or quasi-experimental method—for example, a randomised controlled trial (Chapter 1), a regression discontinuity design (Chapter 2), or a difference-in-differences design (Chapter 3)—to evaluate a government activity or program. Two essays focus on discrimination: one on discrimination in labour markets (Chapter 2) and one on discrimination in police searches (Chapter 3).

Chapter 1, "Casting a Wider Tax Net: Experimental Evidence from Costa Rica," is co-authored with Anne Brockmeyer, Marco Hernandez, and Stewart Kettle. It is motivated by the finding that the majority of firms in developing countries are informal, yet even among firms registered with the government, tax filing rates are low. Firms that do not file taxes generate several important negative externalities, including loss of government revenue; horizontal inequities between firms, which can distort resource allocation and affect tax morale; and loss of third-party information on other firms' activities, which reduces the government's ability to enforce taxes and to design economic policy more generally. Despite this importance, nonfiling has received little attention in the literature, which has focused on tax registration of fully informal firms and misreporting among tax filers.

We argue that nonfiling of taxes among registered firms constitutes an important intermediate form of informality, which can be tackled cost-effectively. Using a randomised experiment in Costa Rica, we show that, for previously nonfiling firms, credible enforcement emails tripled the rate of income tax filing and more than doubled the rate

and amount of payment, relative to a control group that received no message. Among firms for which the government has information from third parties, these effects correspond to a 21 percentage point (p.p.) increase in the filing rate, a 3.4 p.p. increase in the payment rate, and a US $15 increase in the average payment amount. Emails that highlight third-party reports of a firm's transactions further increased compliance. The average payment return on the strongest email was US $18.

Increased income tax filing and payment rates due to the intervention persisted over two years, and the intervention had largely positive effects on other dimensions of compliance. For example, treated firms became more likely to report transactions with other firms, facilitating future tax enforcement. We calculate an overall cost-benefit ratio of 1:4, which leads us to conclude that the email intervention is a cost-effective way to "cast a wider tax net" by increasing compliance and by broadening the tax authority's information set for future tax enforcement.

Chapter 2, "The Wage Effects of Liability for Employment Discrimination," is a theoretical and empirical analysis of laws that prohibit discrimination in hiring, firing, and pay. In both a standard competitive model and a monopsony model of the labour market, the net theoretical effect of liability for employment discrimination is ambiguous. On the one hand, liability for discrimination in hiring is effectively an employment subsidy. On the other hand, liability for discrimination in firing is effectively an employment tax. The net effect of these two provisions is further complicated by an equal pay mandate. Therefore, whether liability for employment discrimination increases the relative wages of protected workers and, if so, by how much is an important empirical question.

Empirically, we know surprisingly little about the effects of liability for employment discrimination. This essay proposes an empirical strategy to estimate the effects of liability for employment discrimination using a distinctive feature of federal anti-discrimination law in the United States. After the Civil Rights Act of 1991, coverage

under Title VII of the Civil Rights Act of 1964 and damages awardable in discrimination lawsuits became discontinuous functions of firm size. The structure of liability for employment discrimination in the United States lends itself to a regression discontinuity (RD) design. This project aims to use confidential employer-employee matched data held by the U.S. Census Bureau to estimate the wage effects of liability for employment discrimination using an RD design, and I am working with the Census Bureau to gain access to the confidential data. In the meantime, this essay uses survey data to provide preliminary estimates of the effect of Title VII on black wages relative to white wages in the United States. A noisy estimate is that Title VII coverage increases relative wages for black men by 12 to 16 percent. Estimates of the effect of greater Title VII liability and estimates for women are mixed.

Chapter 3, "Police Discrimination and Body Cameras: Evidence from London," uses administrative data from the London Metropolitan Police Service to investigate discrimination in police searches and to estimate the effects of body cameras on discrimination. I find that, prior to a body camera trial, black individuals were 2.5 times more likely to be searched than white or South Asian (Indian or Pakistani) individuals. Ethnic disparities were greater for drug searches, for which black individuals were 2.9 times and South Asians 1.4 times more likely to be searched than whites. Men were 15.8 times more likely to be searched than women.

These disparities cannot be completely explained by accurate statistical discrimination, as there were significant differences in search success rates between ethnic, gender, and age groups. Considering all searches in the year prior to a body camera trial, search success rates were on average 0.7 p.p. lower for black persons and 2.0 p.p. lower for South Asians than for whites. However, after including a full set of control variables, the black-white difference falls to zero. The overall disparities between ethnic groups appear to be driven by drug searches (the most frequent kind), for which search success rates were 2.4 p.p. and 4.6 p.p. lower for black and South Asian persons,

respectively, as well as searches of young persons. In general, young persons (particularly those under 20 years old) experienced lower search success rates, and the age differences were most striking for South Asians and for drug searches. With respect to gender, the search success rate was on average 6–7 p.p. lower for men than for women. Overall, these differences are consistent with preference-based or inaccurate statistical discrimination against young minority-ethnic men.

Using variation in the timing and location of body camera use resulting from a one-year camera trial, I find suggestive but not conclusive evidence that cameras reduced these disparities somewhat, particularly the gender difference in search success rates. However, the estimated effects of body cameras on ethnic disparities are small. For activists and policymakers looking to address ethnic disparities in police searches, body cameras should not be viewed as a panacea.

The remainder of this document is organised into three chapters, each with a separate list of references and appendix. Figures and tables appear throughout. A short conclusion follows.

Chapter 1

Casting a Wider Tax Net: Experimental Evidence from Costa Rica

with Anne Brockmeyer, Marco Hernandez, and Stewart Kettle

1.1 Introduction

Low-income countries are characterised by low tax compliance (Gordon & Li 2009; Besley & Persson 2013), one key aspect of which is nonfiling. As Figure 1.1 shows, tax nonfiling rates are above 20 percent in half of the countries for which data are available, and are significantly higher in lower-income countries, reaching above 90 percent in some. This relationship holds across four major taxes. As a reflection of the importance of tax nonfiling, the International Monetary Fund's and World Bank's tax projects usually include activities to increase filing and monitor it as a key performance indicator (IMF 2015; Junquera-Varela et al. 2017). In Costa Rica, 25 percent of tax-registered firms and over 60 percent of firms that are unregistered but known to the tax authority through third-party reports do not file their income tax declaration.[1]

Firms that do not file taxes generate several important negative externalities. First, nonfiling leads to a loss of government revenue. We estimate that the tax authority in Costa Rica loses hundreds of thousands of dollars annually due to income

[1]We use the term "firm" to refer to both incorporated and unincorporated firms, the latter group encompassing predominantly self-employed professionals and other small businesses.

tax nonfiling. Second, selective nonfiling generates horizontal inequities between firms, which can distort resource allocation (Hsieh & Klenow 2009; Restuccia & Rogerson 2008) and affect tax morale (Luttmer & Singhal 2014). Third, firms that do not file deprive tax authorities of third-party information on other firms' activities, which reduces the government's ability to enforce taxes and to design economic policy more generally. Yet, despite its empirical importance, nonfiling has received little attention in the literature, which has focused on tax registration of fully informal firms (e.g., de Andrade, Bruhn, and McKenzie (2014)) and misreporting among tax filers (e.g., Carrillo, Pomeranz, and Singhal (2017)).

This paper argues that simple and highly cost-effective policies can increase tax-filing compliance. We evaluate a nationwide randomised trial conducted by the Costa Rican tax authority, which contacted 33,279 nonfiling firms by email and requested that they submit their income tax declaration for 2014. In addition to drawing on behavioural insights (Service et al. 2015), the messages contained various threats to deter non-compliers, mentioning the possibility of an audit, shop closure, or online publication of a list of nonfilers (similar to Slemrod, Blumenthal, and Christian (2001); Kleven et al. (2001); Pomeranz (2015); and Perez-Truglia and Troiano (2018)). These deterrents are strong but credible in our context; as discussed below, the tax authority implemented a personalised follow-up with phone calls and visits to a set of taxpayers that did not respond to the treatment.

The messages also leveraged third-party reports about firms' business activities, where available, including information that the tax authority collects through informative declarations presented by other firms, state institutions, and credit card companies. Within this subset of firms, the emails either highlighted the existence of third-party information about the taxpayer in question or provided specific examples of the taxpayer's third-party reported sales, such as the name of a client firm and the reported purchase amount. These examples make the presence of third-party reports credible and

salient without communicating total amounts, which could generate perverse responses (Carrillo, Pomeranz & Singhal 2017).

Using administrative data on filing and tax payments, we present four sets of results. First, the emails sent to nonfilers tripled their rate of income tax filing and more than doubled their rate and amount of payment, relative to a control group that received no message. Among firms with third-party information, these effects correspond to a 21 percentage point (p.p.) increase in the filing rate, a 3.4 p.p. increase in the payment rate, and a US$ 15 increase in the average payment amount. Listing specific examples of third-party reports about the firm had an additional positive effect on all outcomes. The average payment return on the strongest email was US$ 18.[2]

Second, we show that the intervention had largely positive effects on other dimensions of compliance. We find no reduction in sales tax filing rates or payments, and only a small increase in the deregistration rate (by 1–2 p.p.). Instead, we find that the intervention increased the rate of income tax filing and payment, both for the fiscal years prior to the intervention and in the two years following the intervention. For firms with third-party information, the effect persisted even for three years. In addition, the intervention induced firms to file 800 additional informative declarations, an effect that also persists in the year following the intervention. These new informative declarations improve future tax enforcement, as about 30 percent of them cover nonfilers or firms underreporting their sales.

Third, we study the mechanisms driving the increase in tax compliance. While part of this effect is driven by taxpayers updating their beliefs about enforcement, as evidenced by the medium-term persistence of the effects, part is also driven by a salience or reminder mechanism, which can be used repeatedly. Firms in our experiment, regardless of their treatment group and 2014 filing status, respond to repeated enforcement messages sent to nonfilers in 2016 and 2017. We observe a trend break in filing rates

[2]Figures are calculated using an exchange rate of 574 Costa Rican *colones* (CRC) per US dollar.

at the exact calendar date on which messages are sent, even among firms contacted for the fourth or fifth time. Together, the medium-term and dynamic effects suggest that frequently contacting non-compliant firms can be a sustainable policy.

Finally, we show that the treatment leads to an improvement in reporting behaviour conditional on filing among the self-employed, a group with notoriously low tax compliance rates in almost all settings. Self-employed individuals contacted by the tax authorities report higher sales and tax liabilities and make higher tax payments, as compared with self-employed filers who were not contacted. Bunching evidence suggests that this is due to the mentioning of third-party information and a deduction rule which discourages offsetting an increase in self-reported sales with an increase in self-reported costs, showing that a clever policy design can avert evasion shifting.

We calculate an overall cost-benefit ratio of 1:4, which leads us to conclude that the email intervention is a cost-effective way to "cast a wider tax net": by increasing compliance in current, future, and even past periods, and by broadening the tax authority's information set for future tax enforcement. We discuss under what conditions the intervention is welfare-improving in the presence of filing costs, and review alternative policies that would avoid filing costs.

This paper contributes to three strands of literature. First, this experiment adds to the literature on taxpayer communication designed to increase tax compliance, including Dwenger et al. (2016), Hallsworth et al. (2017), and other studies reviewed in Hallsworth (2014). Most importantly, we extend the literature by using rich administrative data to measure a variety of outcomes. Contrary to compliance crowd-out, we find positive impacts on compliance in prior tax periods and on filing of informative declarations, a persistence of the main treatment effects in the medium term, and continued responsiveness to repeated interventions. We use a cost effective delivery method—emails[3]—and maximise message impact by combining different message ele-

[3]See Ortega and Scartascini (2015) on the larger impact of emails compared to letters.

ments that proved successful in other contexts, generating large effects compared with most other studies. In addition, our study is one of few in this literature to focus on firms as opposed to individuals, and the first (with Kettle et al. (2016)) to focus on tax filing.[4] Indeed, most studies in this literature have focused on correct reporting of liabilities or on the payment of already assessed liabilities.

Second, we contribute to the literature on taxation and development, as surveyed in Besley & Persson (2013), which analyses how tax capacity grows along the development path. While most recent contributions to this literature focus on misreporting on the intensive margin (Best et al. 2015; Pomeranz 2015; Naritomi 2019; Carrillo, Pomeranz & Singhal 2017), our study highlights the importance of compliance gaps on the extensive margin. The role of third-party information in enhancing tax compliance features prominently in the literature (Kleven et al. 2011; Kleven, Kreiner & Saez 2016), and this information has been shown to enhance tax compliance on the intensive margin for the value-added tax (Pomeranz 2015; Naritomi 2019), but results for the corporate income tax are less clear-cut. Carrillo, Pomeranz, and Singhal (2017) and Slemrod et al. (2017) show that the use of third-party information in enforcement leads to evasion shifting to the cost margin, but Brockmeyer & Hernandez (2018) document sharp increases in reported tax liabilities when firms in Costa Rica become subject to third-party reporting for the first time.[5] This paper is the first to focus on the role of third-party reporting in enhancing extensive-margin compliance, showing that third-party information is simultaneously a tool to identify nonfiling firms, a tool to strengthen enforcement interventions, and a by-product of better filing compliance.

Finally, this paper contributes to the literature on firm formalisation, as reviewed in Bruhn & McKenzie (2014) and de Andrade, Bruhn, and McKenzie (2014). Previ-

[4]Kettle et al. (2016) find much smaller treatment effects than this study. Hallsworth (2014) refers to three other papers that considered filing behaviour, but all of them focused on individual taxpayers.

[5]Brockmeyer & Hernandez (2018) also study the impact of tax withholding on business sales, finding that a doubling of the average withholding rate increased sales tax payment by treated firms by over 20 percent.

ous studies find that providing information, reducing registration costs, or simplifying regulation is relatively ineffective in encouraging formalisation. Only enforcement or large payments (e.g., one month of profits) significantly increase registration rates. We suggest that encouraging regular tax filing among firms that are tax-registered but do not file (regularly) is a more cost-effective way of casting a wider tax net. As nonfilers are firms that chose to register for taxes, their perceived benefit from tax filing likely exceeds that of fully informal firms that chose not to register. In addition, the tax authority already has these firms' contact information, and (in some cases) also third-party reports about business activities that can be leveraged to estimate outstanding tax liabilities.

The rest of this paper is organised as follows. Section 1.2 describes the context and data; Section 1.3 presents the experimental design; Sections 1.4, 1.5, and 1.6 present the baseline results, medium-term results and mechanisms, and a cost-benefit analysis; and Section 1.7 concludes.

1.2 Context and Data

1.2.1 Income and Sales Tax

Total tax revenues in Costa Rica represented 13.3 percent of GDP in 2016, of which 33 percent were due to the sales tax and 34 percent to the income tax (Costa Rica Ministry of Finance 2016). The income tax is levied according to different schedules for corporations and unincorporated firms, i.e., self-employed individuals. Corporations face an average tax rate on profits of 10 percent, 20 percent, or 30 percent, depending on their revenue level. The self-employed face a marginal tax rate on profits between 0 percent and 25 percent, depending on their profit level. The self-employed have lower effective tax rates on average, and the self-employed below the exemption threshold file

their income tax declaration without incurring a liability. All firms in our sample are required to file their income tax declaration by December 15.[6]

Firms selling manufactured goods and certain service providers, such as hotels and restaurants, are liable for the monthly sales tax. This tax is effectively a narrowly based VAT, providing credits for taxed inputs. Approximately 20 percent of income tax-liable firms are liable for the sales tax.

Instead of paying the regular income tax and sales tax, retailers in certain sectors and below certain size thresholds[7] can opt into a simplified regime, in which firms file quarterly and pay tax on inputs at sector-specific rates that vary from 3 percent to 9.8 percent. During the period we study, approximately 30,000 firms filed their taxes under the simplified regime, while 360,000 firms filed their income taxes under the regular regime. Firms can opt into or out of the simplified regime by submitting a D140 switch form, or deregister by submitting a D141 deregistration form.

Figure 1.1 shows the share of tax-registered firms that did not file a tax declaration in Costa Rica in 2014 (square and triangle markers). For the income tax, among firms without third-party information, this share was 25 percent of corporations and 19 percent of the self-employed.[8] It was lower but still substantial at 17 percent of firms liable for the sales tax. This proportion has been stable over the last five years, and it is based on a tax register that is updated on a regular basis. It is possible that the tax register contains some inactive firms. However, even among firms covered by third-party information, i.e., shown to be economically active, 12 percent did not file income tax for 2014. Appendix Tables 1.A1 and 1.A2 compare the characteristics and compliance

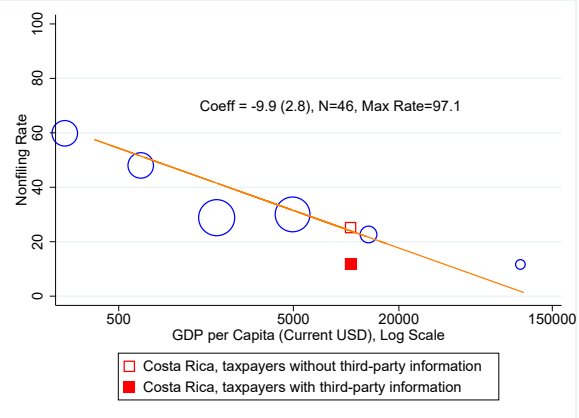
[6]The fiscal year is October 1 to September 30. Firms have to make three quarterly advance tax payments for the income tax if their previous year's liability or their average liability over the last three years is non-zero, with the maximum of these two amounts determining the level of advance payments.

[7]In particular, those with annual purchases below 150 base salaries, net assets below 350 base salaries, or with fewer than six employees.

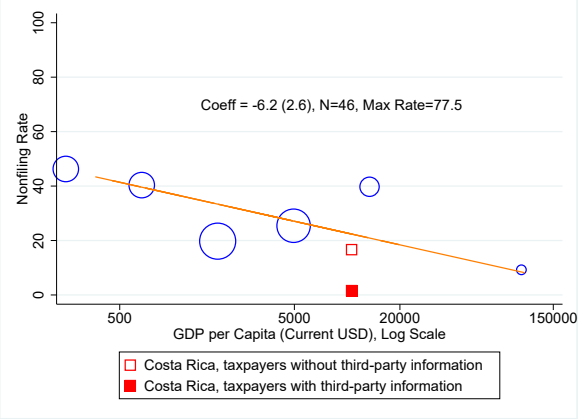
[8]The corresponding figures in Brockmeyer and Hernandez (2018) are comparable, albeit lower, as they use an algorithm based on tax declarations and registration records to construct the tax register for different years but do not directly observe the tax register. The figures in this paper are the Costa Rican tax authority's own estimates based on all available data.

Figure 1.1: Tax Nonfiling Rates Across Countries

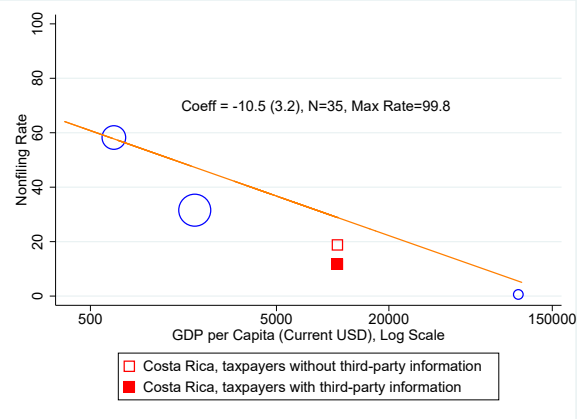
A: Corporate Income Tax



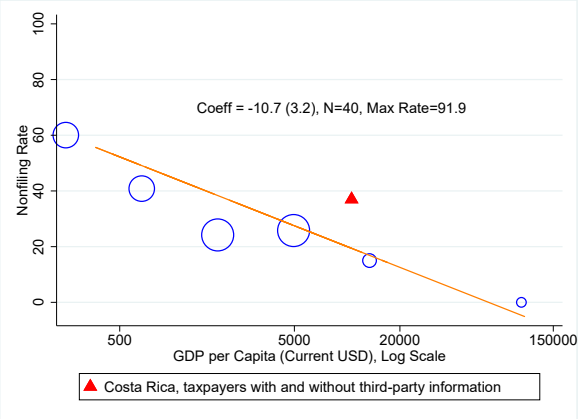
B: Value-Added Tax



C: Personal Income Tax, Self-Employed



D: Personal Income Tax, Employer Report



Note: These figures show the correlation between the nonfiling rate and log GDP per capita across countries, for different taxes. To avoid disclosing country-specific nonfiling rates, the data are binned by log GDP per capita (bin size is 0.5 log points). The marker size reflects the relative number of observations per bin. The red square/triangle markers represent Costa Rica, where we measure nonfiling rates directly using the tax authority's list of nonfilers in 2014 (based on the tax register and the income and sales tax declarations submitted). The round markers represent other countries. The regression coefficients (standard errors in parentheses) are based on the raw data. The filing rate data is from the Tax Administration Diagnostic Assessment Tool (TADAT), and from the 2016 report by CIAT (Centro Interamericano de Administrative Tributarios), "Revenue Administration in Latin America and the Caribbean." TADAT is an assessment tool that is implemented globally by development partners and technical assistance providers including the International Monetary Fund (IMF) and the World Bank. The GDP data is from the World Development Indicators. The nonfiling rate is defined as one minus the share of returns filed on time in the TADAT data, and as the share of returns not submitted in the CIAT data. The negative correlation is also present when considering only nonfiling or only one minus on-time filing. Appendix Figure 1.A1 shows that filing rates are also increasing over time.

behaviour of filers and nonfilers. While nonfilers are smaller than the average filer, they are actually larger than filers with below-median sales, and most nonfilers have filed income tax in the past. Thus, encouraging the compliance of this valuable segment of taxpayers appears worthwhile from a revenue and horizontal equity perspective.

1.2.2 Third-Party Information

To enforce taxes, the Costa Rican tax authority makes use of third-party information from firms, state institutions, and credit/debit card companies. An informant submits one informative declaration for each customer or provider, specifying their tax identification number, the transaction amount, the tax withheld (if applicable), and the income or transaction type (in general categories).[9]

Firms have to report in form D151 all transactions with those clients and suppliers with whom the annual amount of transactions reaches at least CRC 2.5 million (US$ 4,400). The payment of rent, commissions, professional service fees, or interests must be reported for transaction partners with whom the annual transaction amount reaches CRC 50,000 (US$ 90). These transactions must be reported by both the seller and the buyer.

State institutions report all purchases from private firms (form D150). They also withhold tax at a rate of 2 percent of the transaction amount, which is remitted to the tax authority as an advance payment on the income tax. Credit or debit card companies report all card sales by affiliated businesses (form D153) and withhold tax at a firm-specific rate between 0 percent and 6 percent on card sales. The withheld amount is remitted to the tax authority as advance payment on the sales tax.[10]

[9]Taxpayers are not provided with the informative declarations about their transactions. Given the structure of reporting requirements, firms should be aware of the existence of third-party reports about them, but firms with unsophisticated accounting systems may not be able to accurately estimate the amount of third-party reports.

[10]The filing deadline for third-party reports by firms and state institutions is the same as the filing deadline for the income tax, December 15. Credit and debit card reports are filed monthly. See Brockmeyer and Hernandez (2018) for further details on the withholding schemes.

State institutions and card companies are considered to be highly compliant with their reporting obligation, while firms are only partially compliant with their obligation to report transactions. The tax authority uses all informative declarations and customs information to cross-check income tax returns. Nonfilers are thus identified and taxpayers with strong discrepancies between third-party information and self-assessed tax declarations are selected for (desk) audits.

1.2.3 Data

We use income and sales tax returns of corporations and the self-employed in Costa Rica during 2010–2017. This amounts to about 360,000 income tax returns per year and 68,000 sales tax returns per month. The data allow us to measure compliance prior to the experiment, estimate the effect of the experiment on filing for fiscal year 2014, and estimate the medium-term impact on filing in 2015 and 2016. The data includes all line items from the tax declarations and tax payment receipts. The fact that all declarations and payment receipts carry a time-stamp and indicate the corresponding fiscal period allows us to precisely capture compliance at different points in time. We match the firms' declarations with third-party reports submitted by other firms, state institutions, and card companies. About 60 percent of firms that filed an income tax return in 2014 were reported by at least one client or supplier, 9 percent were reported by a card company, and 6 percent by a state institution.

1.3 Experiment Design

Our study relies on a randomised communication experiment implemented by the tax authority in Costa Rica. Appendix Table 1.A3 summarises the experiment design. The target population included 115,000 firms that were registered with the tax authority but had not filed their income tax declaration for 2014 by February 15, 2015,

two months after the regular filing deadline. Of these nonfilers, the experiment targeted 49,757 firms that had an email address on file.[11] The experiment was divided into two sub-experiments, targeting firms that were covered by at least one third-party report ($N = 12,515$), i.e., definitely economically active, and those that were not covered by any third-party report ($N = 37,242$).

1.3.1 Treatments

Firms covered by third-party reports were randomly assigned to one of three groups: a control group and two treatment groups. The control group received no email message. The treatment groups received an email from the tax authority (shown in Appendix Figure 1.A2) requesting them to file an income tax declaration for 2014. The message content was similar to past communication campaigns. It emphasised that nonfiling is a serious offence, that offenders could be audited or subject to business closure as stipulated by law, and that the names of nonfilers might be published online. Additionally, the message integrated findings from behavioural economics to strengthen the message impact.[12] The content of the message was fact-based and credible, given the enforcement context discussed below.

As the tax authority routinely uses third-party information in its enforcement activities and mentions this to taxpayers, the baseline treatment message (T1) also stated:

[11] While our results do not have full external validity for nonfilers without an email address on file, they are nonetheless relevant for this sample. Firms without email addresses were only slightly less likely to have filed in 2013 or to file in 2014, compared to our experimental control group, and exhibit no differences in reporting or payment behaviour conditional on filing.

[12] The message used simplified text (Service et al. 2015; Dwenger et al. 2016), featured a clear call to action in red capital letters—"Please file your income tax return in the next 10 days" (Gabaix & Laibson 2006), provided a web link to the tax form (Service et al. 2015), and was personalised (Haynes et al. 2013). The message also presented the social norm—"8 out of 10 [tax-registered] Costa Ricans have filed their 2014 income tax return" (Wenzel 2005; Hallsworth et al. 2017; Del Carpio 2013).

Baseline Treatment (T1), firms with third-party information:

We have third-party information confirming that you or your client performed activities in 2014 which require you to pay taxes.

The information treatment (T2) further strengthened this statement by providing firms with specific examples of third-party reports:

Information Treatment (T2), firms with third-party information:

We have third-party information confirming that you or your client performed activities in 2014 which require you to pay taxes. From third-party reports (D150, D151 and D153), we know about your operations, for example:

- *Sales of at least XXX reported by COMPANY,*
- *Sales from credit/debit cards of at least YYY reported by BANK,*
- *Sales or contracts with state entities of at least ZZZ.*

For each type of third-party information held by the tax authority (reports from other firms, state institutions, and credit/debit card companies), we listed the transaction amount and trading partner from the largest reported transaction. This demonstrates the tax authority's possession of third-party reports and thus makes the message highly credible. In addition, providing examples rather than the total amount of third-party reported transactions allows us to apply this treatment to firms with small third-party reports without giving away how much or how little the tax authority knows. Finally, the design of the treatment allows us to test whether firms over- or underestimate the amount of information that the tax authority holds about them.

Firms without third-party information were also assigned randomly to either a no-message control group or one of two treatment groups. Treated firms received an email that was nearly identical to the email in the first sub-experiment (Appendix Figure 1.A3). The only difference was the paragraph on third-party information. The

first treatment group received an email with no mention of third-party information whatsoever:

> **Baseline Treatment (T1), firms without third-party information:**
>
> *[No mention of third-party information.]*

The second treatment condition highlighted the tax authority's use of third-party information, but it did not claim that the tax authority held third-party reports regarding the firm's activities:

> **Information Treatment (T2), firms without third-party information:**
>
> *The tax authority uses information from third-party returns (D150, D151 and D153) to identify economic activity and sources of income.*

This treatment thus does not claim that the tax authority has third-party information about the taxpayer in question, but it may encourage compliance among firms who are uncertain about whether they have been reported by a supplier or client.

1.3.2 Timing and Follow-Up

Emails were sent to firms with third-party information starting March 4, 2015, and to firms without third-party information starting April 7, 2015. The list of nonfilers was extracted on February 15, 2015, and it was not possible to update it afterwards. Some firms in the experiment sample had thus already filed by the time the emails were sent, which allows us to demonstrate parallel filing-rate trends in the treatment and control groups prior to the experiment.

To ensure the credibility of the enforcement threat in our intervention, the tax authority subjected a significant share of the taxpayers that did not file by July 15, 2015,

to follow-up activities. These started with a phone call by a tax officer and proceeded (in the case of continued non-compliance) according to a specified sequence of steps that could lead to an audit. Cases were assigned to individual tax officers and dealt with between July 2015 and July 2016. Unlike the experimental emails, these follow-up activities were not randomly assigned, but instead targeted remaining nonfilers (and thus disproportionately the control group) and large firms with third-party information: 15 percent of remaining nonfilers with third-party information were contacted, 50 percent of which were in the control group (which represents only 33 percent of the sample).[13] The vast majority of firms subject to follow-up took action upon contact by the tax officer. Among the remaining non-compliers, at least 20 firms were subjected to an official enforcement procedure, five were temporarily closed, and four were audited.

The threat of online publication was also credible, as the authorities published the names of 6,788 sales tax nonfilers in August 2015, 266 of whom were also income tax nonfilers and part of our experiment; and published the names of 13,025 income tax nonfilers for the period 2011–2015 on March 10, 2016, 1,509 of whom were in our experiment.[14]

The claim that third-party information is used in enforcement was supported in 2015 by desk audits targeted at filers exhibiting discrepancies between self-reported and third-party reported sales (130 of these desk audits targeted firms in our experiment sample), and desk audits for firms with a net liability of zero (1,289 of which targeted firms in our sample).

1.3.3 Enforcement Context

The credibility of our intervention is further supported by the general enforcement environment, in particular, the fact that the tax authority conducts a variety of

[13] Among remaining nonfilers without third-party information, 85 were contacted.

[14] We examine the impact of online publication in footnote 43 and Appendix Figure 1.A14.

enforcement interventions between the extremes of an audit and a simple email. Examples of these interventions for Costa Rica during 2008–2015 are listed in Appendix Table 1.A4. Every year, the tax authority conducts approximately 700 audits, targeting taxpayers with large estimated tax liabilities or suspicious behaviour, including nonfilers (less than 10 percent of audits are for large taxpayers), and forces several hundred temporary firm closures, about one-third of which target nonfilers. The threat of an audit or shop closure is routinely used in communications with taxpayers. Although the likelihood of an audit or shop closure for any particular firm is low, the monetary and reputational costs can be substantial.[15]

Less invasive enforcement activities are usually conducted on a larger scale and can cover several thousand taxpayers per year. These activities include desk audits, invitations of taxpayers for interviews in the tax office, controls of the provision of receipts, follow-up on complaints about taxpayers, and phone calls and messaging campaigns to reduce nonfiling for the income tax, sales tax, and informative declarations. The list of taxpayers in arrears is usually published online. There is thus a continuum of enforcement activities, and the specific type of activities and criteria for taxpayer selection change from year to year in a way that is not predictable for taxpayers. While interventions usually target particular compliance dimensions (e.g., nonfiling for a specific tax and fiscal period), they are also used to verify other compliance margins and can thus be viewed as general enforcement interventions.

Consistent with this, in a 2014 tax compliance cost survey (UN 2014), 9 percent of taxpayers in Costa Rica reported having been subject to some kind of control activity by the tax authority in the previous 12 months.[16] Importantly, this is not specific

[15]Focus group discussions suggest that audits generate reputational costs to a firm if the firm's neighbours, suppliers, or clients learn about the audit and interpret it as a sign of illegal behaviour. The likelihood of this is larger for shop closures, which are advertised on a shop's entrance through an irremovable sign posted by the tax authority. There is abundant empirical evidence on the impact of alleged misconduct (and publication thereof) on firm performance via a reputational channel (Hanlon & Slemrod 2009; Karpoff & Lott 1993; Bø, Slemrod & Thoresen 2015).

[16]Of these taxpayers, 19 percent took more than two hours to comply with the control activity, suggesting that these activities often represent more than just a phone call.

to Costa Rica. Other tax administrations in Latin America conduct a similar array of enforcement activities, and in a sample of 18 lower-income countries around the world, 28 percent of taxpayers on average reported having been subject to some kind of tax control activity in the last year.[17] We thus believe that our experimental treatment accurately represents the general enforcement probability, which is the structural parameter underpinning the treatment effect. By providing several examples of enforcement activities without stating precise probabilities, the email conveys the idea that the tax authority conducts an array of enforcement activities and that the likelihood of being subject to one of these is non-negligible for a taxpayer.

1.3.4 Balance of Randomisation

Given the random assignment of treatment status, the treatment and control groups do not significantly differ in terms of their characteristics and compliance outcomes measured at baseline (Appendix Tables 1.A5 and 1.A6). When jointly testing the significance of the difference between the three experiment groups, we fail to detect a statistical difference between the groups at the 10 percent level of significance for 20 out of 22 firm characteristics, and for 18 out of 20 compliance outcomes.

Among firms with (without) third-party information, 14 percent (26 percent) are corporations, 61 percent (66 percent) use a legal representative, and about 50 percent are located in the capital city. The larger share of self-employed among the nonfilers is consistent with the general view that the self-employed are particularly prone to tax evasion. As for compliance in the pre-experiment year, 33 percent (17 percent) filed income tax, 5 percent (1 percent) made a payment, and 19 percent (5 percent) filed a third-party report. These firms are thus truly at the margin of compliance. Regarding

[17] The minimum, maximum and median values are 4.4 percent, 60.0 percent and 27.5 percent, respectively. We focused on small firms where possible and note that the mean was even higher among informal firms. The data is from tax compliance cost surveys conducted by the World Bank in Albania, Armenia, Bangladesh, Bosnia and Herzegovina, Colombia, Ethiopia, Georgia, India (Rajasthan), Kenya, Kyrgyz Republic, Nepal, Peru, Tajikistan, Uganda, Ukraine, Uzbekistan and Yemen (Coolidge 2012).

compliance at baseline, we find that 6 percent (1 percent) filed an income tax declaration and 1 percent (0.04 percent) made a payment after the extraction of the nonfiler list and before the start of the experiment. As would be expected given the short window between the extraction of the nonfilers list and the experiment start date, hardly any firms deregistered.

1.3.5 Estimation

In all main specifications reported in this paper, we use ordinary least squares (OLS) to estimate the treatment effect on outcome y_i for firm i as

$$y_i = \alpha + \beta_1 \cdot T1_i + \beta_2 \cdot T2_i + \gamma X_i + \varepsilon_i, \tag{1.1}$$

where $T1_i$ and $T2_i$ indicate treatment 1 and 2, respectively, for firm i; X_i is a vector of covariates; and ε_i is an error term.[18] Outcomes are measured at 15 weeks after the experimental interventions, unless stated otherwise. In addition to testing the hypotheses that β_1 and β_2 are significantly different from zero, we report the p-value from a (Wald) test of the hypothesis that β_1 and β_2 are equal. To increase power, we control for a set of firm characteristics measured prior to the experiment.

1.4 Baseline Results

This section presents the baseline results of our empirical analysis, considering the treatment impact on income tax compliance and on other compliance outcomes, as well as heterogeneity in the treatment effect.

[18] We compute (Huber-White) standard errors that are robust to within-cluster correlation, as randomisation was conducted by clusters of the primary email address. The results are also robust to conducting the estimation using the sample of one-firm clusters only.

1.4.1 Income Tax Compliance

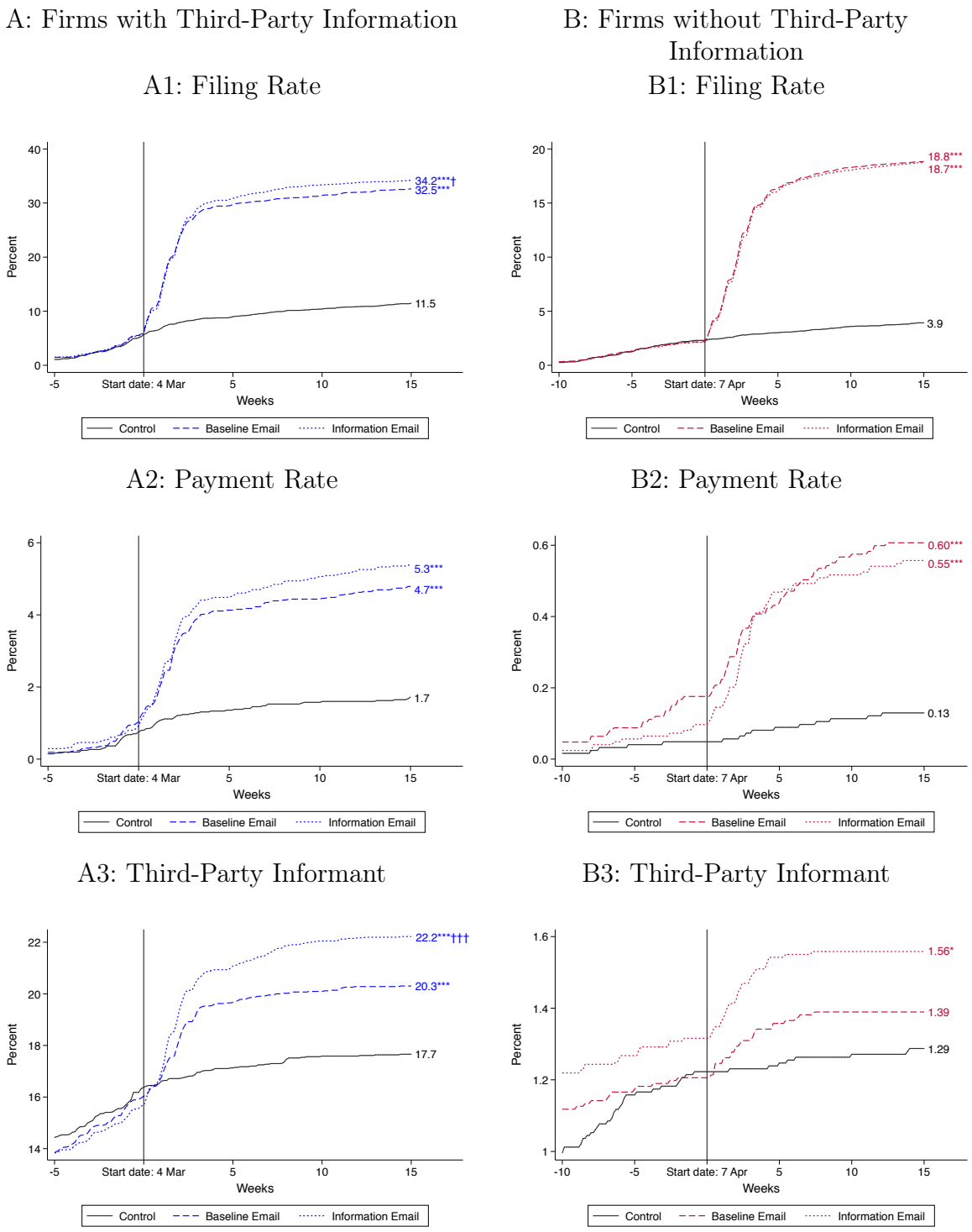
We start by analysing the impact of the emails non-parametrically. Figure 1.2 shows the rate of income tax filing and payment over time by treatment status, with the start of the experiment indicated by a vertical line. The left (right) column is for firms with (without) third-party information. While pre-intervention trends in the treatment and control groups were nearly identical for all outcomes, a positive treatment effect on filing and payment emerged sharply at the start of the experiment. This effect stabilised by about five weeks after the experiment start date, and remained approximately constant during the next ten weeks. This confirms that the emails generated additional tax payments rather than just bring forward payments that firms would have made anyway.[19] By 15 weeks after the start of the intervention, the filing rate for information-covered firms reached 32.5 percent for those sent the baseline email, and 34.2 percent for those sent the email with examples of third-party information, relative to 11.5 percent for the control group.[20] The difference between the two treatment effects is statistically significant at the 10 percent level. The payment rate was 4.7 percent for those sent the baseline email, 5.3 percent for those sent the email with examples of third-party information, and 1.7 percent for the control group.

For firms not covered by third-party information, the shape of the filing and payment response relative to the control group is similar, except that there is no difference between the two treatments. Emphasising the use of third-party information thus did not enhance compliance among firms not covered by an information trail. This suggests that either most of these firms did not expect the tax authority to possess any third-party reports about them, or their response to the potential use of third-party information is small compared to the large response to the message itself. By 15 weeks

[19]The absence of a trend break in the control group suggests treatment spillovers were minor or absent.

[20]The fact that over 60 percent of the treatment group did not file in response to the email might partially be attributable to the fact that some taxpayers did not receive and open the email. It is unfortunately not possible to identify these taxpayers.

Figure 1.2: Compliance by Information Coverage and Treatment Group, 2014



Note: These figures show the share of firms filing income tax (row 1), paying income tax (row 2), and filing a third-party informative declaration (D151) about a client or supplier (row 3), all for fiscal year 2014. Column A corresponds to firms with third-party information and column B corresponds to firms without third-party information. The vertical line in each figure indicates the experiment start date. The black solid line corresponds to the control group and the blue/red dashed/dotted lines correspond to the baseline treatment and information treatment, respectively, for the two different subsamples. The numbers indicate the mean for each outcome and treatment group at 15 weeks after the start of the experiment. Stars indicate a significant difference compared with the control group and come from regressions that include controls (as in Table 1.1). Significance levels are noted as per convention: * $p < .10$, ** $p < .05$, *** $p < .01$. Daggers indicate the significance level of the difference between the two treatments.

after the intervention, the proportion of treated firms filing an income tax declaration reached 19 percent, relative to 3.9 percent for the control group. The rate of payment was below 1 percent, as most firms without third-party information declare a net liability of zero, but the payment rate was still significantly higher for the treatment group.[21]

To consider a larger number of outcomes and control for covariates, we report OLS estimates of the treatment effect at 15 weeks in Table 1.1. Panel A presents estimates for firms with third-party information. The first three columns report extensive-margin responses in income tax filing, in reporting a positive net liability, and making a payment. We estimate that the baseline email increased the probability of filing by 21.3 p.p. The effect of the information email, at 23.0 p.p., is significantly greater than the effect of the baseline email. The two emails also increased the probability that a firm reported a positive net liability (by about 5 p.p.) and made a positive payment (by about 3.4 p.p.).[22] The fourth and fifth columns show that the emails also increased the average payment amount, which is partly driven by the greater number of payers, and partly by larger payments conditional on making a payment.[23] The baseline email increased income tax payments by CRC 8,353 (US$ 15) per taxpayer, on average, while the information email increased payments by CRC 10,441 (US$ 18).[24] The information email had a larger impact on all outcomes than the baseline email, but the difference is statistically significant at conventional levels only for filing, which is presumably the most precisely measured outcome.[25]

[21] In panels B2 and B3, it appears that the treatment and control groups are not equal in the pre-period, but this is a result of the small scale of the vertical axis. There are no statistically significant differences between the groups prior to the intervention (see Appendix Table 1.A6).

[22] Note that the payment rate is lower than the share of firms reporting a positive liability, as some firms make quarterly advance payments or are subject to withholding at amounts that fully cover their liability, and others underpay or pay with delay.

[23] The dependent variable in columns 4 and 5 is 0 for firms who do not file and for filers who do not make any payment. We consider only payments made by the taxpayer with her annual declaration, and ignore the quarterly advance payments during the year. If a nonfiler has outstanding advance payments, those would be made with the annual declaration, and are thus considered in our estimation.

[24] We winsorize payment amounts at the top 0.1 percent to reduce the influence of outliers.

[25] Filing is measured accurately if a filer's tax declaration appears in our database, while line items

Table 1.1: Impact on Income Tax Compliance

	OLS				
	(1)	(2)	(3)	(4)	(5)
	Filed (percent)	Positive net liability (percent)	Positive payment (percent)	Log payment	Payment (CRC)
A: Firms with Third-Party Information					
T1: Baseline email	21.12	4.44	3.07	0.3408	8352.9
	(0.87)	(0.48)	(0.38)	(0.0428)	(2449.8)
T2: Information email	22.98	5.19	3.73	0.4088	10441.2
	(0.89)	(0.49)	(0.40)	(0.0444)	(2823.7)
Other controls	Yes	Yes	Yes	Yes	Yes
Control group avg.	11.48	3.39	1.72	0.1940	5015.8
T1=T2 p-value (Wald test)	0.074	0.189	0.159	0.195	0.514
Observations	12,515	12,515	12,515	12,515	12,515
B: Firms without Third-Party Information					
T1: Baseline email	15.05	0.65	0.47	0.0496	265.1
	(0.41)	(0.09)	(0.08)	(0.0083)	(59.4)
T2: Information email	14.93	0.59	0.43	0.0411	168.9
	(0.40)	(0.09)	(0.07)	(0.0075)	(51.5)
Other controls	Yes	Yes	Yes	Yes	Yes
Control group avg.	3.93	0.18	0.13	0.0142	96.2
T1=T2 p-value (Wald test)	0.809	0.565	0.649	0.397	0.161
Observations	37,242	37,242	37,242	37,242	37,242

Note: Each column in this table reports estimates from an OLS regression run at the taxpayer level. Robust standard errors clustered by email address (to account for the fact that a few taxpayers share the same email address) are in parentheses. Panel A is for the sample of firms for which the tax authority has third-party information on their economic activities. Panel B is for the sample of firms without third-party information. The dependent variables, as indicated by the column headings, are indicator variables for whether the taxpayer filed income tax for 2014, reported a positive net liability, and paid income tax (considering only final payments made with the declaration and not advance payments that may have been made earlier) for 2014 (columns 1–3, respectively), log tax payment (column 4), and tax payment in Costa Rican colones (CRC) (column 5). The dependent variables in columns 4 and 5 are zero for firms who do not file and for filers who do not make any payment. All outcomes are measured 15 weeks after the start of the experiment. Payment amounts are winsorized at the top 0.1 percent to reduce the influence of outliers. All regressions include as independent variables two indicator variables for the baseline email treatment (T1) and for the information email treatment (T2), which were randomly assigned within the samples, and the following other indicator variables (which are balanced across treatment groups): whether the taxpayer is a corporation, has a legal representative, has a second email address, is located in the capital city, has third-party information above 2.5 million CRC, has third-party information above 6 million CRC (the administratively used thresholds), made advance tax payments for the income tax in 2014, filed income tax in 2013, reported a positive net liability for the income tax in 2013, made an income tax payment in 2013, and was a third-party informant in 2013. The bottom rows in each panel report the control group average of the outcome variable and a p-value from a Wald test evaluating whether the coefficients on the two treatment variables are different from each other. Appendix Tables 1.A7–1.A8 display the full regression results, including the coefficients on the control variables.

Panel B shows the analogous results for firms without third-party information. The emails increased the probability of income tax filing by 15 p.p.; they also increased the probability of reporting a positive net liability and of making a payment, but the magnitude of the effect is small (0.6 p.p.). The treatment increased the payment amount by CRC 217 on average (US$ 0.38).[26]

In sum, we find that the emails significantly improved income tax compliance by nonfilers. For firms covered by third-party information, the impact of emails that listed specific examples of information known to the tax authority was larger. All treatment effects estimated here are similar to or exceed those from other communication experiments (Castro & Scartascini 2015; Del Carpio 2013; Fellner, Sausgruber & Traxler 2013; Kettle et al. 2016),[27] which suggests that combining different message features that proved to be impactful individually (deterrence, behavioural design, use of third-party information) can generate a larger impact overall.

Appendix Figure 1.A4 examines the impact of follow-up activities, which were non-randomly targeted at remaining nonfilers starting in July 2015. These activities, combined with the continued gradual increase in the filing rate between July and December 2015, lead the filing rate among firms with third-party information to reach over 40 percent in December 2015, as shown in Panel A. Panel B provides evidence of a clear trend break after July 15 in the filing rate of the targeted priority group firms, with no trend break occurring among non-targeted firms. Consistent with the personalised and sequential nature of this campaign, the trend break is less sharp than in the baseline intervention. Importantly, the follow-up targeted the control group disproportionately,

on the declaration are measured accurately only if the declaration appears in our database and if the taxpayer made no mistakes when filling it out.

[26] Appendix Table 1.A9 shows that the treatment effects for all payment outcomes are significantly higher when restricting the sample to firms that are more likely to be required to make a tax payment (corporations and firms with a positive tax liability or a positive income tax payment in 2013).

[27] In Fellner, Sausgruber, and Traxler (2013), who consider compliance with TV license fees, the proportional effect is larger than ours—7.7 percent of letter recipients start to pay the fee within 50 days of the experiment versus 0.8 percent in the no-letter control group—though the absolute effect is smaller.

and thus lead to a slight narrowing of the filing gap between the treatment and control group, as evidenced in Panel C.

1.4.2 Other Compliance Outcomes

As the treatment had a strong impact on the targeted compliance outcomes, notably income tax filing and payment, it is appropriate to consider spillovers to other compliance outcomes. Indeed, taxpayers might perceive the intervention as a general increase in enforcement, and hence improve compliance with other tax filing obligations. Alternatively, they might perceive the intervention as pertaining to a specific tax only, and increase compliance with the targeted tax but compensate for lost income by reducing compliance elsewhere. To shed light on this, we use rich administrative data on firms' filing of informative declarations, sales tax declarations, sales tax payments, deregistration, and switches to the simplified regime, all of these representing compliance outcomes that were not directly targeted by the intervention. We also consider income tax compliance in 2013, which may be affected by the intervention (targeting nonfilers for 2014), as a non-negligible share of firms file or pay their taxes with substantial delay (documented in Brockmeyer and Hernandez (2018)). This is to our knowledge the first study to examine the impact of enforcement on such a large variety of compliance measures.

Tables 1.2 and 1.3 show the treatment impact on the above-mentioned outcomes, for firms with and without third-party information in Panels A and B respectively. Contrary to the possibility of a crowd-out effect, columns 1–3 in Table 1.2 find no evidence that the treatment has an impact on compliance with the sales tax, the most important tax paid by firms. The treatment affects neither the number of months in which a firm files sales tax, nor the number of months it pays sales tax, nor the total sales tax amount. We also confirm that there is no evidence of a crowd-out effect in the sample of firms liable for the sales tax (i.e., having filed sales tax at some point prior

to our experiment, Appendix Table 1.A10).

Columns 5–6 in Table 1.2 show that the emails increased firms' likelihood of deregistering with the tax authority. This is consistent with the fact that deregistration generates a hassle cost (visiting the tax office[28]), and there are effectively no fines for remaining registered but economically inactive, so that firms are unlikely to voluntarily deregister when ceasing activities. For firms with third-party information, treatment increased the deregistration rate from 0.9 percent to about 2 percent, and large amounts of third-party information were reassuringly negatively correlated with deregistration. For firms not covered by third-party information, treatment increased the deregistration rate from 1.2 percent to about 3 percent. These firms were also marginally more likely to deregister only from the sales tax or switch to the simplified tax regime, but this concerns less than 0.5 percent of the sample. The smaller effect on information-covered firms is consistent with these firms' reported economic activity in the fiscal year in question.

When considering the firms that deregistered and those that did not, pooling the two experiments, it appears that the deregistrants were firms striving to be compliant on paper, without transferring any revenue to the tax authority. Among deregistrants, 61 percent filed an income tax declaration for 2014 and 52 percent did so for 2013, versus 21 percent and 26 percent, respectively, for firms that remained in the tax register. Yet the mean reported liability was more than an order of magnitude lower for deregistrants compared to firms remaining in the tax register: CRC 40,101 (US$ 70) versus CRC 656,409 (US$ 1,144) in 2013. While some deregistrants continued their business activities informally (103 firms were third-party reported in 2015), they would have been unlikely to pay more taxes. It thus appears that the deregistrations reduced the number of taxpayers to be managed by the administration without affecting tax

[28]The possibility for online deregistration was abolished as firms allegedly exploited it to register with the aim of obtaining a book of official receipts, only to deregister immediately afterwards.

Table 1.2: Impact on Other Compliance Outcomes (Part 1)

	OLS					
	(1)	(2)	(3)	(4)	(5)	(6)
	No. months decl. sales tax	No. months paid sales tax	Sales tax payment (CRC)	Deregistered (sales tax) (percent)	Deregistered (percent)	Switched to simplified regime (percent)
	A: Firms with Third-Party Information					
T1: Baseline email	-0.0393	-0.0231	-9751.5	0.02	1.15	0.05
	(0.0504)	(0.0205)	(7965.0)	(0.10)	(0.27)	(0.05)
T2: Information email	-0.0418	-0.0231	-5568.2	-0.05	1.11	0.05
	(0.0493)	(0.0203)	(8331.9)	(0.10)	(0.27)	(0.05)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Control group avg.	0.6371	0.1330	29095.0	0.22	0.92	0.02
T1=T2 p-value (Wald test)	0.960	0.999	0.563	0.481	0.885	0.982
Observations	12,515	12,515	12,515	12,515	12,515	12,515
	B: Firms without Third-Party Information					
T: Any email	0.0004	0.0062	26.9	0.07	2.03	0.04
	(0.0162)	(0.0035)	(76.9)	(0.06)	(0.15)	(0.02)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Control group avg.	0.2547	0.0159	357.5	0.32	1.22	0.02
Observations	37,242	37,242	37,242	37,242	37,242	37,242

Note: This table is identical to Table 1.1, except for the outcome variables, which are indicated by the column headings. For Panel B, the regressions include only a single treatment indicator (combining the baseline and the information treatment groups), in addition to the control variables. The dependent variables, as indicated by the column headings, are the number of months in 2015 during which the taxpayer declared sales tax (column 1), the number of months in 2015 for which the taxpayer paid sales tax (2), the total sales tax payment for 2015 in CRC (3), and indicator variables for whether the taxpayer deregistered from the sales tax (4), deregistered completely with the tax authority (5), or switched to the simplified tax regime available to small firms (6). All outcome variables are measured at 15 weeks after the start of the experiment and refer to compliance for fiscal year 2014, unless otherwise noted. Payment amounts are zero for taxpayers who did not file or did not make any payment. Sales tax payment amounts are winsorized at the top 0.1 percent to reduce the influence of outliers.

collection.[29]

Continuing to Table 1.3, columns 1–2 show that the emails improved income tax compliance for the previous tax year (2013), by significantly increasing the probability of (late) filing and payment. This is true even though the emails specifically mentioned the requirement to file the 2014 income tax return. Emails to information-covered firms increased firms' likelihood of filing an income tax declaration by 2–3 p.p., compared to the control group's average of 35 percent. The impact on firms not covered by third-party information was even larger—about 5 p.p. compared to the control group's average of 18.9 percent. In both samples, treated firms were also slightly more likely to make a payment for the income tax in 2013.[30]

Finally and importantly, columns 3–5 consider the treatment impact of filing third-party reports. Panel A shows that treated firms were marginally more likely to file an information report about their supplier, and significantly more likely to file a report about their client. This can be explained by the fact that, for firms filing an income tax declaration, filing an informative declaration generates only a small hassle cost, but no monetary cost, and prevents a possible fine for non-compliance with the reporting obligation. As the emails with additional third-party information made the existence of third-party reports salient and provided information about the nature of these reports, it is not surprising that this treatment generated a response twice as large as the baseline treatment (a 5.1 p.p. versus a 2.8 p.p. increase in the reporting rate). The impact on information reporting of clients is also present and significant, though smaller in magnitude, among firms not covered by third-party information. For both samples, we confirm in the bottom panels of Figure 1.2 that the pre-intervention trends in treatment and control groups were similar, and that the effect emerged sharply at

[29]To prevent deregistrations of active firms, in 2016 the tax authority introduced a new rule according to which all nonfilers that request deregistration have to be visited by a tax official.

[30]The treatment effect on income tax payment for 2013 is significant when focusing on the sample most susceptible to respond, the self-employed and firms that had filed income tax at some point before (Appendix Table 1.A10). More generally, Table 1.A10 shows that the results discussed in this section are qualitatively identical, with larger point estimates, in the sub-samples most likely to respond.

Table 1.3: Impact on Other Compliance Outcomes (Part 2)

	OLS				
	(1) Filed income tax for 2013 (percent)	(2) Paid income tax for 2013 (percent)	(3) Presented informat-ive declaration (percent)	(4) Reported supplier (percent)	(5) Reported client (percent)
	A: Firms with Third-Party Information				
T1: Baseline email	3.21	0.19	2.76	0.36	2.64
	(0.41)	(0.15)	(0.78)	(0.59)	(0.71)
T2: Information email	2.28	0.23	5.14	0.97	4.44
	(0.39)	(0.16)	(0.83)	(0.60)	(0.76)
Other controls	Yes	Yes	Yes	Yes	Yes
Control group avg.	35.03	5.71	17.66	9.28	12.79
T1=T2 p-value (Wald test)	0.048	0.837	0.006	0.305	0.024
Observations	12,515	12,515	12,515	12,515	12,515
	B: Firms without Third-Party Information				
T: Any email	4.94	0.09	0.20	0.02	0.19
	(0.20)	(0.03)	(0.13)	(0.11)	(0.08)
Other controls	Yes	Yes	Yes	Yes	Yes
Control group avg.	18.92	1.09	1.29	0.94	0.44
Observations	37,242	37,242	37,242	37,242	37,242

Note: This table is the continuation of Table 1.2. Everything is identical, except the outcome variables, which here are indicator variables for whether the taxpayer filed income tax for 2013 (1), paid income tax for 2013 (2), presented an informative declaration (third-party report) about a client or supplier for 2014 (3), presented an informative declaration about a supplier for 2014 (4), and presented an informative declaration about a client in 2014 (5).

the time the intervention started. A targeted (income tax) enforcement intervention thus led to an expansion of the tax authority's information set for future enforcement of the income tax and other taxes.

Table 1.4 examines the importance of the additional third-party reports in more detail, showing regressions of the number of reports filed on a treatment indicator and control variables. Columns 1 and 2 suggest that the treatment lead to an increase of 0.1 in the number of third-party reports filed, or approximately 820 new reports in total, 70 percent of which are reports about clients. Columns 3–5 show that slightly more than half of the new reports are unmatched by a corresponding report in the other direction, i.e., covering a transaction link previously unobserved by the tax authority; about 4 percent of the new reports cover nonfilers; and 30 percent of them cover under-reporters, i.e., firms that underreport their sales compared to third-party reports. As the tax authority uses third-party reports to identify and follow-up on under-reporters (e.g., through desk audits which were found to increase under-reporters' reported tax liability and tax payment (Brockmeyer & Hernandez 2018)), this additional third-party information could be of substantial value to tax enforcement.[31]

In sum, our analysis of a diverse set of compliance outcomes detects positive treatment effects on compliance with third-party reporting and past income tax obligations, and a small increase in deregistration rates, mostly reflecting the exit of firms with very low reported liabilities.

1.4.3 Heterogeneity

Since communication campaigns such as the one analysed here generate costs to both the tax authority and to the taxpayer (which we examine in Section 1.6), they might need to be targeted. This section therefore studies heterogeneity in taxpayer

[31] Appendix subsection 1.9.2 proposes a back-of-the-envelope estimation of the value of the additional third-party reports, suggesting that they could increase the tax revenue return of the intervention by up to 20 percent.

Table 1.4: Impact on Informative Declarations

	(1) No. of third-party reports filed	(2) No. of reports about clients	(3) No. of reports about suppliers	(4) No. of unmatched reports	(5) No. of reports about nonfilers	(6) No. of reports about under-reporters
T: Any email	0.098	0.070	0.027	0.051	0.004	0.030
	(0.024)	(0.012)	(0.017)	(0.016)	(0.002)	(0.008)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Control group avg.	0.101	0.046	0.055	0.038	0.003	0.028
Observations	12,515	12,515	12,515	12,515	12,515	12,515

Note: Each column in this table reports estimates from an OLS regression run at the taxpayer level. Robust standard errors clustered by email address (to account for the fact that a few taxpayers share the same email address) are in parentheses. The dependent variables, indicated by the column headings, are the number of informative declarations (third-party reports) a taxpayer filed (column 1), the number of reports filed about clients (2), the number of reports filed about suppliers (3), the number of unmatched reports (4), the number of reports about nonfilers (5), and the number of reports about under-reporters (6). The table considers third-party reports for fiscal year 2014, filed after experimental messages were sent. Unmatched third-party reports (column 4) are reports for which there is not a corresponding report in the other direction for the same client-supplier pair and transaction amount (rounding to the nearest 5,000 CRC (about US$ 10)—the results are robust to rounding to the nearest 10,000 CRC or not rounding at all). Reports about nonfilers (column 5) are reports filed before the subject filed an income tax declaration. Reports about under-reporters (column 6) are reports about subjects with total amounts of third-party reported sales (taking into account the latest report) that exceed the amounts of self-reported sales from their income tax declarations. The independent variable is an indicator variable for whether the taxpayer was in one of our two experimental treatment groups. The control variables are the same as in Table 1.1.

responsiveness to the intervention. Table 1.1 already showed that firms with third-party information display larger p.p. increases in filing and payment rates than firms without third-party information. Table 1.A13 considers heterogeneity by firm type, location in the capital city, and past compliance record, separately for firms with third-party information in Panel A and for firms without such information in Panel B. Corporations exhibit a weaker filing rate response but a stronger payment response compared to the self-employed (columns 1 and 6). There is less heterogeneity along this dimension among firms without third-party information. Firm location in the capital city is associated with a stronger treatment effect in only one out of the four estimations (columns 2 and 7). Past compliance in the form of filing and payment of the income tax in 2013, and filing of sales tax, is associated with a larger treatment effect (with the exception of sales tax compliance among firms with third-party information).

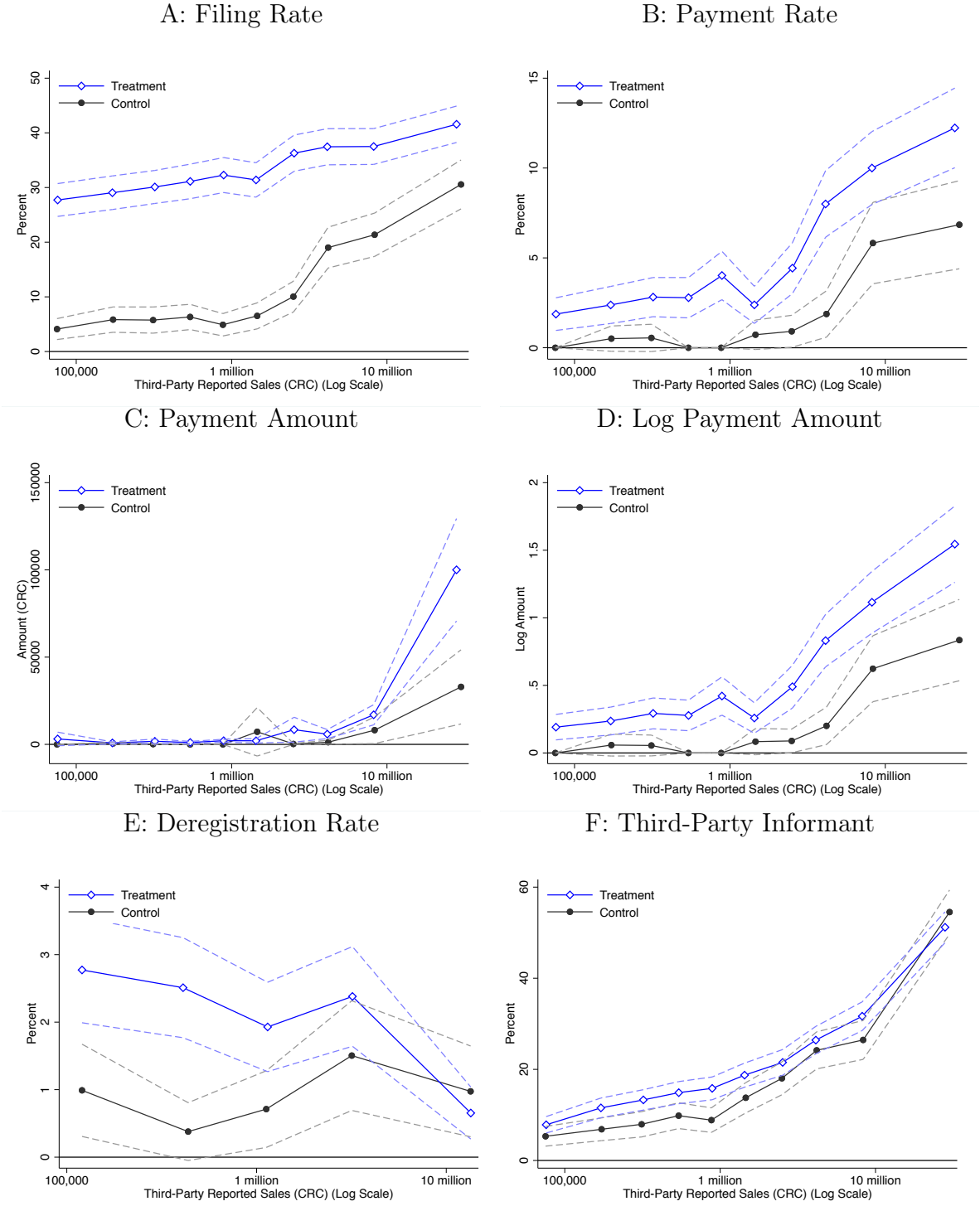
We then zoom in on the sample of firms with third-party information and examine treatment effects by firm size. Figure 1.3 plots the main outcomes, for the treatment and control group, by percentiles of third-party reported sales. We use deciles for most outcomes and quintiles for deregistration.[32] The treatment effect on filing is driven by smaller firms, as larger firms are more likely to declare even when in the control group (Panel A). The effect on payment rate and amount, on the other hand, is driven by larger firms (Panels B–D). This is consistent with heterogeneity by firm type discussed above, as larger firms are more likely to be corporations, and with differences in the two tax schedules, as the self-employed can file without making a payment if below the exemption threshold.[33]

When targeting a communication intervention, it is thus advisable for the tax authority to take into account a firm's third-party reported sales, its firm type, and its

[32]The results are similar when using deciles of the maximum of self-reported sales in year $t-1$ (or the most recent year available) or third-party reported sales in year t. Table 1.A14 confirms the heterogeneity results in regression form, interacting the treatment dummies with different indicators for high third-party reported sales.

[33]The deregistration response is driven by smaller firms, as would be expected, while the effect on submitting an informative declaration does not differ by firm size (Panels E–F).

Figure 1.3: Heterogeneity by Third-Party Information



Note: This figure shows the share of nonfilers that filed an income tax declaration (Panel A), the share that made an income tax payment (Panel B), and payment amounts (Panels C and D) for 2014, the share that deregistered (Panel E), and the share that filed an informative declaration (Panel F) at 15 weeks after the start of the experiment, among firms with third-party information (experiment 1). Payment amounts are winsorized at the top 0.1 percent to reduce the influence of outliers. The values are displayed by decile/quintile of log third-party reported sales. The blue hollow markers are for the pooled treatment group and the black solid markers are for the control group. For the 4th and 5th decile in the control group, not one observation made any payment, so the average payment rate is zero, and so is the standard deviation. Estimates are similar when calculated by bins of the maximum of self-reported sales in year $t-1$ (or the most recent year available) or third-party reported sales in year t.

past compliance record.

1.5 Medium-Term Effects and Mechanisms

This section examines the medium-term persistence of treatment effects, the impact of repeated compliance interventions, and the treatment effect on reporting behaviour conditional on filing.

1.5.1 Persistence of Treatment Effects

Having shown that the treatment had a large impact on contemporaneous income tax compliance and other compliance outcomes, we now examine whether these effects persisted over time. In Table 1.5 we pool the two treatments, as the coefficients do not differ significantly by treatment, and control for whether or a not a firm was subject to follow-up.[34] For firms with (without) third-party information, the email messages increased the income tax filing rate one year later by 8 p.p. (7.2 p.p.), over a control group average of 30 percent (15 percent).[35] The emails also generated a statistically significant increase in the payment rates one year later by 1.3 p.p. (0.3 p.p.). Finally, treated firms were more likely to file informative declarations in 2015. The emails generated a 2.3 p.p. (0.4 p.p.) increase in the third-party reporting rate, thus further expanding the tax authority's enforcement capacity. While the medium-term effects on third-party reporting dissipate after two years, a significant impact on income tax filing and payment persists for three years. In 2016, treated firms with (without) third-party information were 2.6 p.p. (1.5 p.p.) more likely to file their income tax declaration, and 1.6 p.p.

[34]The tax authority also conducts other communication campaigns for the filing of sales tax and informative declarations, and the targeting of those would likely have been either orthogonal to our treatment assignment or disproportionately focused on the control group.

[35]The fact that the medium-term impact on filing is smaller than the short-term impact for 2014 is likely because of the higher control group average in the following years, which is due to a small degree of mean reversion and the follow-up activities started by the tax authority in July 2015, examined in Appendix Figure 1.A4 and Section 1.4.1.

Table 1.5: Medium-Term Impact on Compliance Outcomes

	By filing deadline					
	(1) Filed income tax (percent)	(2) Paid income tax (percent)	(3) Reported client (percent)	(4) Reported supplier (percent)	(5) No. of clients reported	(6) No. of suppliers reported
	A: Firms with Third-Party Information					
T effect in 2015: Any 2014 email	8.02	1.26	2.34	0.71	0.0209	0.0039
	(0.92)	(0.42)	(0.63)	(0.52)	(0.0056)	(0.0044)
Control group avg. 2015	30.83	4.67	11.20	7.89	0.0930	0.0613
T effect in 2016: Any 2014 email	2.59	1.58	1.08	0.30	0.0102	0.0002
	(0.96)	(0.50)	(0.70)	(0.56)	(0.0063)	(0.0047)
Control group avg. 2016	35.07	6.27	13.78	9.08	0.1165	0.0702
T effect in 2017: Any 2014 email	2.03	0.97				
	(0.95)	(0.53)				
Control group avg. 2017	34.20	7.73				
Observations	12,515	12,515	12,515	12,515	12,515	12,515
	B: Firms without Third-Party Information					
T effect in 2015: Any 2014 email	7.20	0.25	0.40	0.31	0.0034	0.0022
	(0.42)	(0.12)	(0.17)	(0.15)	(0.0016)	(0.0013)
Control group avg. 2015	15.34	1.14	2.34	1.69	0.0209	0.0150
T effect in 2016: Any 2014 email	1.52	0.30	0.15	0.26	0.0006	0.0015
	(0.44)	(0.15)	(0.21)	(0.18)	(0.0020)	(0.0016)
Control group avg. 2016	19.08	1.89	3.94	2.60	0.0347	0.0214
T effect in 2017: Any 2014 email	0.15	0.14				
	(0.45)	(0.19)				
Control group avg. 2017	19.50	3.00				
Observations	37,242	37,242	37,242	37,242	37,242	37,242

Note: Each cell in this table reports estimates from an OLS regression run at the taxpayer level. Robust standard errors clustered by email address (to account for the fact that a few taxpayers share the same email address) are in parentheses. Panel A is for the sample of firms for which the tax authority has third-party information on their economic activities. Panel B is for the sample of firms without third-party information. The dependent variables, as indicated by the column headings, are indicator variables for whether the taxpayer filed income tax for a particular year (column 1), paid income tax (2), presented an informative declaration about a client (3), presented an informative declaration about a supplier (4), the number of clients reported (5), and the number of suppliers reported (6). Within each panel, these outcomes are measured either in fiscal year 2015 (top cells), 2016 (middle), or 2017 (bottom). All regressions include as an independent variable an indicator for whether the firm was in one of the two treatment groups of our experimental intervention for nonfilers in 2014. Each cell reports the coefficient on this treatment variable and its standard error. The bottom row in each cell reports the control group average of the outcome variable. All regressions include the same control variables as in Table 1.1. The regressions for Panel A additionally control for whether the taxpayer was subject to the follow-up campaign in 2015, and an interaction between this follow-up indicator variable and the treatment indicator. The regressions for Panel B additionally include the follow-up indicator. The number of treated firms subject to follow-up for Panel B is small, which is why we do not control for treatment and follow-up interaction. The control group average represents the average for firms in the control group not subject to follow up.

(0.3 p.p.) more likely to make a payment than firms in the control group for 2014.[36] In 2017, treated firms with third-party information still exhibited a 2.0 p.p. higher filing rate and a 1.0 p.p. higher payment rate than 2014 control firms. The persistence of treatment effects for several years is unparalleled in enforcement-letter interventions. As the 2017 effects for firms with third-party information are only slightly smaller than the 2016 effects, these treatment effects might even persist for longer than three years.

In contrast to the positive impact on firms' medium-term tax compliance and third-party reporting behaviour, we do not detect any effect of the emails on firms' propensity to be subject to third-party reporting in a later year. Table 1.6 displays the treatment effect on firms' likelihood of being reported by state institutions, private sector clients or suppliers, or card processing companies respectively in 2015 and 2016. Treated firms were no more or less likely to be subject to these reports, even if they received the stronger information email, which provided them with examples of third-party information held by the tax authority.[37] Thus, the treatment does not seem to have distorted firms' trade networks or coverage by third-party reports.[38]

Overall, these findings suggest that one-time enforcement messages can have a positive impact on compliance in the medium term, including compliance with information reporting requirements. The persistence of the main effects shows that the messages lead some firms to update their beliefs regarding the tax authority's enforcement capacity, i.e., the capacity to identify and follow up on nonfilers, and that the update was persistent over time, hence confirming the messages' credibility.[39] Consistent with this

[36] Appendix Figures 1.A8 and 1.A9 show the medium-term effects graphically.

[37] The result holds when considering the two treatments separately and among sub-samples of firms more susceptible to respond to the treatment (Appendix Table 1.A12). The result is consistent with the absence of a behavioural response (missing mass) above the mandatory third-party reporting threshold, and the fact that a large share of transactions below the threshold are voluntarily declared (Appendix Figure 1.A10). This strengthens the power of third-party information as a compliance tool.

[38] We examined potential spillovers effects on firms' trading partners, as identified by the third-party reports, but did not find any spillovers.

[39] It is possible that the treatment allowed firms to gain new information about the tax filing procedure, which would have reduced the cost of future tax filing, but this is unlikely as tax filing in Costa Rica is very simple and all the necessary information is available online.

Table 1.6: Medium-Term Impact on Third-Party Reports

	(1) Reported by state institution (percent)	(2) Reported by private client or supplier (percent)	(3) Reported by card company (percent)
T effect in 2015: Any 2014 email	-0.13	-0.42	0.23
	(0.36)	(0.94)	(0.56)
Control group avg. 2015	3.36	67.50	8.44
T effect in 2016: Any 2014 email		-1.31	
		(1.00)	
Control group avg. 2016		59.56	
Observations	12,515	12,515	12,515

Note: Each cell in this table reports estimates from an OLS regression run at the taxpayer level. Robust standard errors clustered by email address (to account for the fact that a few taxpayers share the same email address) are in parentheses. This table covers only firms for which the tax authority has third-party information on their economic activities. The dependent variables, as indicated by the column headings, are indicator variables for whether the taxpayer was reported by a state institution (1), by a private client or supplier (2), or by a credit/debit card company (3). These outcomes are measured either in fiscal year 2015 (top cells) or 2016 (bottom cells). All regressions include as an independent variable an indicator for whether the firm was in one of the two treatment groups of our experimental intervention for nonfilers in 2014. Each cell reports the coefficient on this treatment variable and its standard error. The bottom row in each cell reports the control group average of the outcome variable. All regressions include the same control variables as in Table 1.1, and additionally control for whether the taxpayer was subject to the follow-up campaign in 2015, and an interaction between this follow-up indicator variable and the treatment indicator. The control group average represents the average for firms in the control group not subject to follow up.

interpretation, Appendix Table 1.A11 shows that the medium-term impacts are larger among firms that were less compliant prior to the intervention, and hence more likely to update their enforcement perceptions through the treatment.

The persistence of the treatment effect also sheds some light on the size of potential real effects. By reducing (compliant) firms' after-tax profits, the treatment might have lowered investment and firm growth. This would reduce firms' future profits, and hence their likelihood of complying with taxes, if tax compliance is partly a function of firm profits. The fact that this potential negative effect on medium-term compliance is overcome through the experimental treatment and updating of enforcement perceptions suggests that either the real effects were small, or profitability is only weakly related to tax filing propensity.[40]

1.5.2 Response to Repeated Interventions

A potential concern with our intervention, as with any enforcement threat, is that it may work differently as a one-off intervention than as a frequently used policy tool. On the one hand, firms that revised upward their beliefs about enforcement in response to the treatment message may be less likely to respond to additional messages once their beliefs are closer to the true (or suggested) enforcement probability. On the other hand, treated taxpayers which did not file and were not punished for their non-compliance might revise downward their beliefs about enforcement. These changes in beliefs should be stronger for firms in the treatment group, but could also affect control firms that interact with and learn from treated firms or from the media. Both channels would lower and potentially even eliminate the impact of repeated interventions.[41]

[40] It is also conceivable that the real effect is positive, for instance if better tax compliance allows firms to trade with more formalised and productive firms.

[41] While taxpayers may update their beliefs about enforcement probabilities, it is unclear whether beliefs should ultimately be in equilibrium. As discussed in Section 1.3.3, the enforcement strategy is multi-dimensional and changes in arbitrary ways over time, so firms would need high-frequency observations on all enforcement parameters to form accurate beliefs. This is unlikely, as tax enforcement is not widely discussed online or in the media. Thus, while beliefs may correspond to the true en-

While pinning down the exact process through which taxpayers update enforcement beliefs after being treated is beyond the scope of this paper, we provide evidence that taxpayers repeatedly respond to enforcement messages. Our analysis relies on the fact that the tax authority conducted repeated (non-randomised) interventions following up on nonfilers during 2015–2017. In particular, the tax authority emailed income tax nonfilers for fiscal year 2015 on April 5, 2016; and income tax nonfilers for fiscal year 2016 on January 18, 2017.[42] The message content differed slightly across interventions, but was similar to our experimental treatment. Figure 1.4 examines the filing response to these interventions. As in Figure 1.2, we plot the filing rate among contacted nonfilers around the time of the intervention, distinguishing the experimental treatment group (solid line) and control group (dotted line). Given the two aforementioned channels through which firms may update their beliefs, we further distinguish firms that remained income tax nonfilers for 2014 (Panels A1 and B1) and firms that ultimately filed income tax for 2014 (Panels A2 and B2).

The key finding from these figures is that all groups still respond to the new interventions in 2016 and 2017, sharply at the time the messages were sent.[43] There is some evidence of decreasing marginal returns for the interventions, as the filing-rate increases achieved by the 2016 and especially by the 2017 intervention are not as high as those achieved by our earlier experimental intervention. Comparing the magnitudes of the response across groups is complicated by the fact that the treatment is no longer

forcement probability on average, individual taxpayers' beliefs are likely to be widely dispersed around this average. Beyond that, it is not necessarily the case that the average taxpayer belief equals the true enforcement probability, even in equilibrium. There is ample evidence that humans overestimate small probabilities (Kahneman & Tversky 1979), and specific survey evidence showing that taxpayers consistently overestimate the audit probability (Erard & Feinstein 1994; Scholz & Pinney 1995). Behavioural models, such as the availability heuristic, can help explain this.

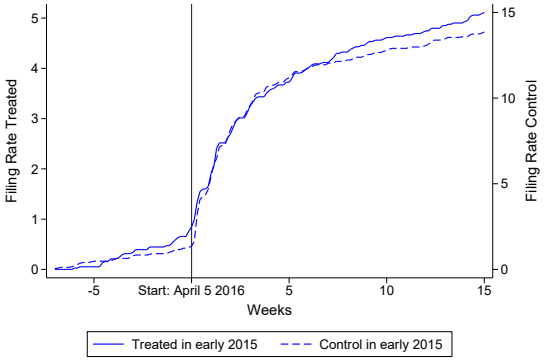
[42] Additional interventions conducted for the sales tax in 2015 and 2016 are presented in Appendix Figures 1.A11 and 1.A12. Our analysis focuses on filing responses, but payment responses behave similarly.

[43] The August 2015 sales tax intervention also provides direct evidence that merely publishing nonfilers' names online increases the filing rate. On August 10, 2015, the tax authority published the names of 6,788 sales tax nonfilers, 1,366 of which were also income tax nonfilers for 2014. As Appendix Figure 1.A14 shows, the filing rates for both taxes were constant before the intervention and sharply increased once the names were published.

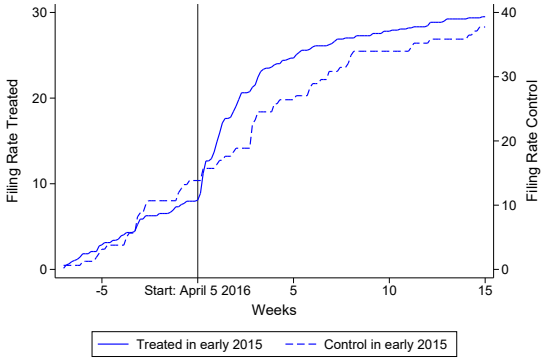
Figure 1.4: Filing Response to Repeated Income Tax Interventions

A: Response to 2016 Income Tax Intervention

A1: Remaining 2014 Nonfilers

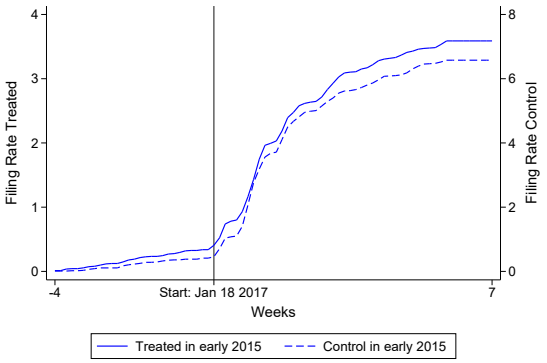


A2: 2014 Filers

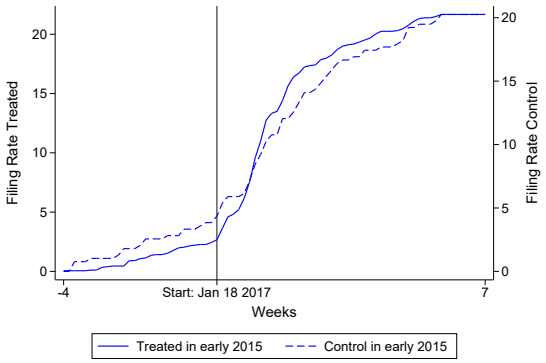


B: Response to 2017 Income Tax Intervention

B1: Remaining 2014 Nonfilers

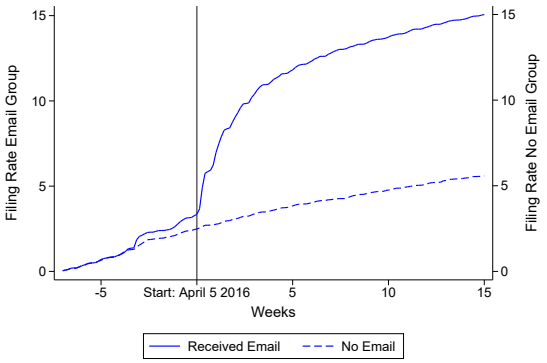


B2: 2014 Filers

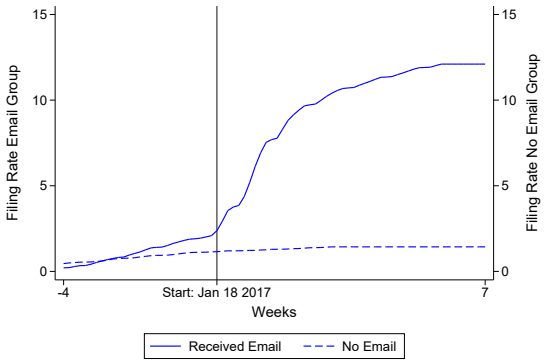


C: Comparing Firms That Did and Did Not Receive an Email

C1: 2016 Income Tax Intervention



C2: 2017 Income Tax Intervention



Note: These figures show the tax filing rate around two enforcement messaging campaigns, as indicated by the row titles, which used enforcement emails similar to the ones used in our experiment. Panels A and B focus on firms in our experimental sample, distinguishing firms that remained nonfilers for the income tax in 2014 (left side), and firms that ultimately did file income tax for 2014 (right side). The black vertical line in each figure marks the date on which messages were sent. Panel C looks at the full sample of nonfilers considered for the 2016 and 2017 interventions, comparing firms that did receive an email and firms that ultimately did not receive an email.

randomly assigned within the target sample for these new interventions. For instance, the lower overall filing rate among firms treated in our experiment might be due to the fact that treated firms which nonetheless become nonfilers again are particularly persistent nonfilers. However, two qualitative observations can be made.

First, the sharpness of the response at the time of the intervention is actually stronger among firms that remained nonfilers for 2014 and might thus have revised their enforcement beliefs downwards (panels A1 and B1). Yet the filing rate in this sample remains below 15 percent in all cases, regardless of experimental treatment status. This suggests that a concern that the experiment might have lowered these firms' responsiveness to future enforcement interventions is of second order importance, as these firms are very unlikely to file anyway. Second, among 2014 filers, the average response rate is similar to the response rate in our experiment (pooling firms with and without third-party information), with filing rates reaching up to 30 percent among previously treated firms and up to 40 percent among previous control firms. Both the sharpness of the response, and the proportional effect size, compared to the pre-intervention filing rate, are larger among the experimental treatment group. There is thus no evidence that our treatment reduced firms' responsiveness to future enforcement interventions.[44]

Panels C1 and C2 compare firms that were targeted by the email campaigns in 2016 and 2017 with firms that had also been identified by the tax authority as nonfilers for the relevant fiscal period, but which did not receive an email (likely due to administrative capacity constraints). We detect no change in filing rates among this group at the time of the intervention, confirming that the filing rate increase observed among targeted firms is an effect of the email.

In conclusion, while some firms update their beliefs about enforcement permanently as discussed in Section 1.5.1, for other firms (those that respond repeatedly to

[44] Appendix Table 1.A15 shows regression estimates of firms' responsiveness to the 2016 income tax intervention, controlling for 2015 treatment and filing status and the follow-up activities started in July 2015. For firms that were subject to follow-up activities, the results rule out any potential backfiring of the treatment (i.e., reduced responsiveness to future interventions).

enforcement messages) the treatment effect seems to work through a salience or reminder channel, which can be activated repeatedly. The effectiveness of this channel is still evident even among firms that have been contacted four or five times for nonfiling behaviour, and which still respond to enforcement messages, as Figure 1.A13 shows.[45]

1.5.3 Reporting Behaviour

In addition to raising the filing and hence payment rate, the treatment was designed to improve reporting behaviour and hence tax payment amounts conditional on filing, by highlighting the use of third-party information in enforcement. To analyse changes in reporting behaviour, we focus on the sample of filers with third-party information and examine to what extent firms match self-reported sales to third-party reported sales, and self-reported costs to self-reported sales. Figure 1.5 summarises the results.[46] Panels A1 and B1 show, for each firm type and treatment group, the fraction of filers that underreport, overreport, or match self-reported sales to third-party reported sales. There is no difference in filing behaviour between the baseline treatment and the control group. However, firms filing in response to the information treatment were significantly less likely to underreport sales and significantly more likely to match third-party reports than other firms. This effect is found among both corporations and the self-employed. As the effect is not present in the baseline treatment group, it must be driven by the specific nature of the treatment, i.e., the salient mentioning of third-party information, rather than the selection effect (the fact that filers in the treatment group have different characteristics than filers in the control group). Panels A2 and B2 show that the increase in reported sales in the information treatment group was

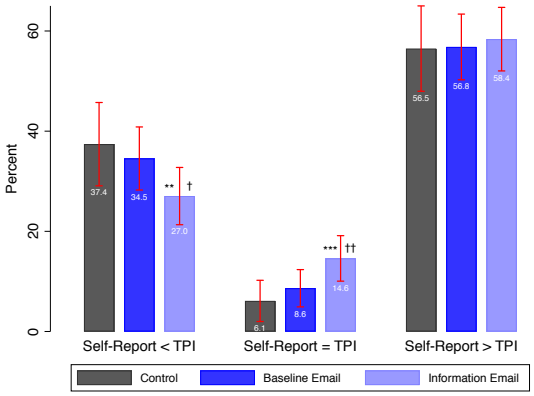
[45] It is also possible that periodic messaging or contact with the tax authority is necessary for firms to maintain constant enforcement beliefs. In fact, given the numerous and repeated nonfiling interventions that the tax authority conducts, one might consider the response to our experiment as an equilibrium response. Besides, the mere fact that the tax authority conducts repeated interventions to follow up on nonfilers could be considered *prima facie* evidence that these interventions have an impact.

[46] Appendix Figures 1.A6 and 1.A7 present more finely-grained bunching graphs, similar to those found in Slemrod et al. (2017).

Figure 1.5: Reporting Behaviour of Tax Filers

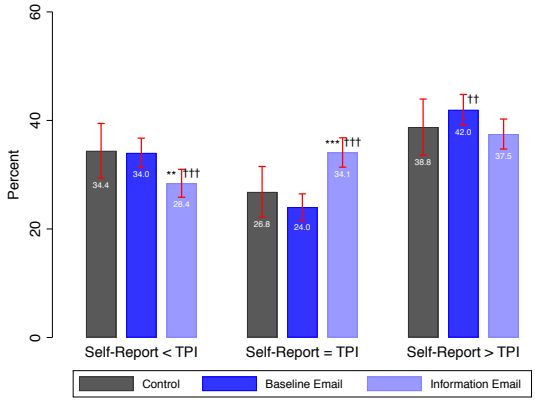
A: Corporations

A1: Self-Reported versus Third-Party-Reported Sales

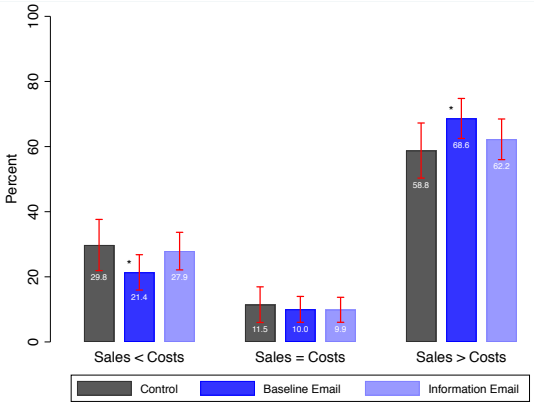


B: Self-Employed

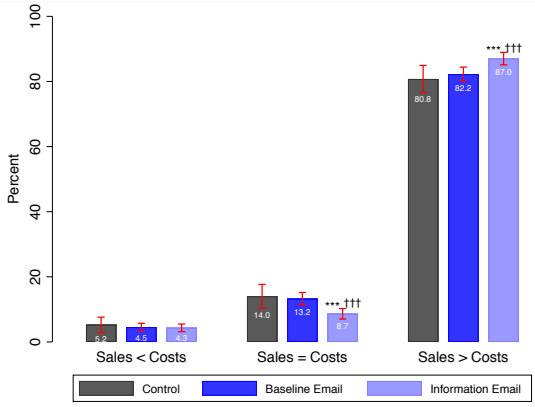
B1: Self-Reported versus Third-Party-Reported Sales



A2: Self-Reported Sales versus Self-Reported Costs



B2: Self-Reported Sales versus Self-Reported Costs



Note: These figures show the share of firms with self-reported sales (recorded on firms' income tax declarations) less than, equal to, or greater than third-party reported sales (row 1) or self-reported costs (row 2). The figures are for firms with third-party information (TPI) that filed an income tax declaration within fifteen weeks of the experiment start. Error bars indicate 95 percent confidence intervals. Stars indicate a significant difference from the control group (at the .10, .05, and .01 levels) and daggers (†) indicate a significant difference between the two treatments.

not offset by an increase in reported costs. Among corporations, there are hardly any significant differences between treatment groups. Among the self-employed, firms in the information treatment were in fact less likely than control firms to match sales to costs and more likely to report sales higher than costs.

This result is surprising, given that previous studies found firms to nearly fully offset reported sales increases with cost increases (Carrillo, Pomeranz & Singhal 2017; Slemrod et al. 2017). It can be explained by the presence of an enforcement notch. The self-employed in Costa Rica are allowed to deduct 25 percent of their sales as costs without proof of receipts, but are required to maintain auditable receipts of all costs if they deduct costs in excess of 25 percent of sales.[47] Consequently, the self-employed bunch either at a costs-to-sales ratio of 0.25 or 1, whereas corporations bunch at a costs-to-sales ratio of 1. The enforcement threat lead self-employed in the information treatment group to reduce bunching at a costs-to-sales ratio of 1 and increase bunching at 0.25, to use the unenforced 25 percent deduction, as shown in Appendix Figure 1.A7. The treatment effectively pushed firms to trade cost deductions for reduced enforcement. Our results show that a clever design of deduction rules combined with credible enforcement can prevent firms from offsetting increases in reported sales.

Together, these results confirm that the treatment not only increased government revenue due to increased filing rates, but also improved reporting behaviour among the self-employed, who declare larger tax bases and make larger payments in response to the mentioning of third-party information (Appendix Table 1.A16). The presence of an enforcement-free minimum deduction prevents firms from offsetting sales increases with cost increases.

[47] This is similar in spirit to an alternative minimum tax or presumptive tax (Best et al. 2015), schemes which are common around the world and used to enhance tax compliance and reduce compliance costs. Spain has a deduction rule similar to Costa Rica's.

1.6 Cost-Benefit Analysis

To examine under what conditions governments should conduct interventions like the one presented in this paper, we analyse in this section the intervention's direct costs and benefits, its welfare implications, and indirect benefits such as deterrence. The primary cost to the tax authority is the human resource cost of sending the personalised emails and responding to taxpayer inquiries. The sending of the emails was executed in seven and a half days by 32 tax officers paid at about CRC 36,700 (US$ 64) a day. We assume that each officer spent an additional five days answering taxpayers' inquiries. In addition, we account for the cost of the follow-up activities, considering that a tax officer deals with about four cases per day, and that 876 treated firms with third-party information and 47 firms without third-party information were subject to follow-up. This means the cost per email was on average 695.6 CRC (US$ 1.20), and for firms without third-party information it was one-fourth of the cost for firms with third-party information, due to the higher number of follow-up cases in the latter group.[48]

The benefits are measured by the increase in tax payment from treated firms. Among firms with third-party information, the baseline email increased firms' income tax payments by CRC 8,353 (US$ 15), on average, 15 weeks after the start of the experiment, while an email that lists specific examples of information known to the tax authority increased payment by CRC 10,441 (US$ 18). For firms without third-party information the email increased payment by CRC 217 (US$ 0.38). With this conservative approach, we find that the intervention was highly cost-effective, with an overall cost-benefit ratio of about 1:4. As the cost-benefit ratio increases to 1:6 among firms with third-party information, a tax authority with limited resources might consider targeting

[48]To draw the most conservative conclusion possible, we take the tax officer's time cost into account, although other studies implicitly assume that the opportunity cost of time for the tax officers and hence the cost of the entire intervention is zero. When we attribute the total cost to the group of firms with and without third-party information respectively, we take into account the number of messages sent for each group, the number of filers that should be proportional to the number of inquiries, and the number of follow-up cases.

the intervention at these firms.

The benefit of focusing on firms with third-party information is reinforced when taking into account firms' filing costs, as we do in the Appendix with a simple tax filing model based on Keen and Slemrod (2017). Ignoring filing costs and penalty costs, our intervention is welfare improving if the marginal value of public funds is larger than 1.33. With filing costs, our intervention is welfare improving if the filing cost is below 5 percent of additional tax payments. The likelihood of this being true is highest for firms that have an employee responsible for tax accounting, so that the marginal cost of filing a tax form is close to zero, assuming the employee receives a fixed wage and is not working at full capacity.

Although the aggregate revenue gains generated by this intervention are small, as in most other communication experiments, the intervention serves the broader purpose of sustaining voluntary compliance, as evidenced by the fact that treated firms improve their tax filing and payment behaviour also in future tax periods. In addition, the intervention and follow-up activities allowed the tax authority to update taxpayer contact information and collect other missing tax declarations. Potentially most importantly, the intervention enhanced compliance with information reporting requirements, which facilitates future tax enforcement. Beyond revenue considerations, the intervention improved horizontal equity of taxation by enhancing compliance among relatively small firms, levelling the playing field between full and partial tax compliers, and could thus improve production efficiency, tax morale, and the perception of fairness of the tax system. The new information generated through firms' self-reports and third-party reports gives the government a broader view of the economy and hence a better basis for policy design.

1.7 Conclusion

This paper has argued that nonfiling among tax-registered firms constitutes an important and under-researched compliance gap in lower-income countries that can be addressed cost-effectively. We evaluate a randomised enforcement campaign in which the tax authority in Costa Rica requested nonfiling firms by email to file their income tax declaration. We find that the enforcement emails increase the tax payment rate by 3.4 p.p. and the payment amount by US$ 15 among firms covered by third-party information — treatment effects that further increase to US$ 18 when emails specifically mention examples of third-party reports. The treatment also increases firms' propensity to provide information reports about other firms, thus facilitating future tax enforcement. The main treatment effects persist for at least two years, and for three years among firms with third-party information. In addition, firms continue to be highly responsive to similar enforcement interventions by the tax authority in future fiscal years.

The intervention is highly cost-effective, and potentially welfare enhancing if taxpayers' filing costs are sufficiently small. From the tax authority side, the costs of the intervention could be further reduced by automating the identification of nonfilers and the personalised delivery of messages with the taxpayer's name and third-party information. To avoid taxpayer filing costs, tax authorities may consider alternative policies such as tax withholding at source (Brockmeyer & Hernandez 2018), registration fees or flat-rate presumptive taxes, or automatically filled tax returns that merely need to be approved by taxpayers.

The intervention examined in this paper could likely be implemented in other developing countries, including poorer countries. Indeed, Figure 1.A15 shows that the vast majority of countries have access to third-party information that allows them to identify nonfiling firms and target communication campaigns (bars 1–4). Yet, while the information is available, most countries do not make optimal use of it in their enforcement activities (bars 5–8). Studying what impedes the use of available third-

party information is an important avenue for future research, which could focus on administrative capacity as well as political economy constraints.

1.8 References

Besley, Timothy, and Torsten Persson. 2013. "Taxation and Development." In *Handbook of Public Economics*, vol. 5, edited by Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, 51–110. Amsterdam: North-Holland.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem. 2015. "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan." *Journal of Political Economy* 123 (6): 1311–55.

Bø, Erlend E., Joel Slemrod, and Thor O. Thoresen. 2015. "Taxes on the Internet: Deterrence Effects of Public Disclosure." *American Economic Journal: Economic Policy* 7, no. 1 (February): 36–62.

Brockmeyer, Anne, and Marco Hernandez. 2018. "Taxation, Information, and Withholding: Evidence from Costa Rica." Unpublished manuscript.

Bruhn, Miriam, and David McKenzie. 2014. "Entry Regulation and Formalization of Microenterprises in Developing Countries." *World Bank Research Observer* 29, no. 2 (August): 186–201.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal. 2017. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement." *American Economic Journal: Applied Economics* 9, no. 2 (April): 144–64.

Castro, Lucio, and Carlos Scartascini. 2015. "Tax Compliance and Enforcement in the Pampas: Evidence from a Field Experiment." *Journal of Economic Behavior and Organization* 116:65–82.

Coolidge, Jacqueline. 2012. "Findings of Tax Compliance Cost Surveys in Developing Countries." *eJournal of Tax Research* 10 (2): 250–87.

Costa Rica Ministry of Finance. 2016. "Detalle de los principales ingresos del Gobierno Central." Accessed June 23, 2019. https://bit.ly/2SE1sZK.

de Andrade, Gustavo Henrique, Miriam Bruhn, and David McKenzie. 2014. "A Helping Hand or the Long Arm of the Law? Experimental Evidence on What Governments Can Do to Formalize Firms." *World Bank Economic Review* 30 (1): 24–54.

Del Carpio, Lucia. 2013. "Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru." https://bit.ly/2K4Q6tX.

Dwenger, Nadja, Henrik Kleven, Imran Rasul, and Johannes Rincke. 2016. "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany." *American Economic Journal: Economic Policy* 8, no. 3 (August):

203–32.

Erard, Brian, and Jonathan S. Feinstein. 1994. "The Role of Moral Sentiment and Audit Perceptions in Tax Compliance." *Public Finance* 49 (Supplement): 70–89.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler. 2013. "Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information." *Journal of the European Economic Association* 11, no. 3 (June): 634–60.

Gabaix, Xavier, and David Laibson. 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics* 121, no. 2 (May): 505–40.

Gordon, Roger, and Wei Li. 2009. "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation." *Journal of Public Economics* 93, no. 7–8 (August): 855–66.

Hallsworth, Michael. 2014. "The Use of Field Experiments to Increase Tax Compliance." *Oxford Review of Economic Policy* 30 (4): 658–79.

Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev. 2017. "The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance." *Journal of Public Economics* 148 (April): 14–31.

Hanlon, Michelle, and Joel Slemrod. 2009. "What Does Tax Aggressiveness Signal? Evidence from Stock Price Reactions to News about Tax Shelter Involvement." *Journal of Public Economics* 93, no. 1–2 (February): 126–41.

Haynes, Laura C., Donald P. Green, Rory Gallagher, Peter John, and David J. Torgerson. 2013. "Collection of Delinquent Fines: An Adaptive Randomized Trial to Assess the Effectiveness of Alternative Text Messages." *Journal of Policy Analysis and Management* 32 (4): 718–30.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124, no. 4 (November): 1403–48.

International Monetary Fund (IMF). 2015. *Current Challenges in Revenue Mobilization: Improving Tax Compliance*. International Monetary Fund. Washington, DC, April.

Junquera-Varela, Raul Felix, Marijn Verhoeven, Gangadhar P. Shukla, Bernard Haven, Rajul Awasthi, and Blanca Moreno-Dodson. 2017. *Strengthening Domestic Resource Mobilization: Moving from Theory to Practice in Low- and Middle-Income Countries*. Washington, DC: World Bank.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decisions under Risk." *Econometrica* 47, no. 2 (March): 263–92.

Karpoff, Jonathan M., and John R. Lott, Jr. 1993. "The Reputational Penalty Firms Bear from Committing Criminal Fraud." *Journal of Law and Economics* 36, no. 2 (October): 757–802.

Keen, Michael, and Joel Slemrod. 2017. "Optimal Tax Administration." *Journal of Public Economics* 152 (August): 133–42.

Kettle, Stewart, Marco Hernandez, Simon Ruda, and Michael Sanders. 2016. "Behavioral Interventions in Tax Compliance: Evidence from Guatemala." World Bank Policy Research Working Paper No. 7690.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica* 79, no. 3 (May): 651–92.

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez. 2016. "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries." *Economica* 83, no. 330 (April): 219–46.

Luttmer, Erzo F.P., and Monica Singhal. 2014. "Tax Morale." *Journal of Economic Perspectives* 28 (4): 149–68.

Naritomi, Joana. 2019. "Consumers as Tax Auditors." *American Economic Review* 109, no. 9 (September): 3031–72.

Ortega, Daniel, and Carlos Scartascini. 2015. "Don't Blame the Messenger: A Field Experiment on Delivery Methods for Increasing Tax Compliance." Development Bank of Latin America Working Paper No. 2015/09.

Perez-Truglia, Ricardo, and Ugo Troiano. 2018. "Shaming Tax Delinquents: Evidence from a Field Experiment in the United States." *Journal of Public Economics* 167 (November): 120–37.

Pomeranz, Dina. 2015. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." *American Economic Review* 105, no. 8 (August): 2539–69.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Plants." *Review of Economic Dynamics* 11, no. 4 (October): 707–20.

Scholz, John T., and Neil Pinney. 1995. "Duty, Fear, and Tax Compliance: The Heuristic Basis of Citizenship Behavior." *American Journal of Political Science* 39, no. 2 (May): 490–512.

Service, Owain, Michael Hallsworth, David Halpern, Felicity Algate, Rory Gallagher, Sam Nguyen, Simon Ruda, et al. 2015. *EAST: Four Simple Ways to Apply Behavioral Insights.* London: Behavioural Insights Team.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian. 2001. "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota." *Journal of Public Economics* 79, no. 3 (March): 455–83.

Slemrod, Joel, Brett Collins, Jeffrey L. Hoopes, Daniel Reck, and Michael Sebastiani. 2017. "Does Credit-Card Information Reporting Improve Small-Business Tax Compliance?" *Journal of Public Economics* 149 (May): 1–19.

United Nations (UN). 2014. *Measuring Tax Transaction Costs in Small and Medium Enterprises*. New York: United Nations.

Wenzel, Michael. 2005. "Misperceptions of Social Norms about Tax Compliance: From Theory to Intervention." *Journal of Economic Psychology* 26, no. 6 (December): 862–83.

1.9 Appendix

1.9.1 A Simple Tax Filing Model

1.9.1.1 Model Set-up and Taxpayer Behavior

This section presents a simple tax filing model based on Keen and Slemrod (2017). Consider a taxpayer i who earns an income (gross profit) y_i, with $y_i \sim G(y)$, incurs a tax filing cost f, and is taxed at a rate t on income above a threshold $T \geq 0$. All taxpayers are required to file a tax declaration. This also applies to taxpayers below the exemption threshold, as they may provide valuable information to the tax authority. A taxpayer who does not file is inspected with probability $p(a)$ where $p_a > 0$, a being the administrative enforcement level,[49] and inspected nonfilers suffer a penalty θ proportional to their income y_i. The nonfiling penalty θy_i may represent the monetary and reputational cost of a temporary shop closure, a bribe to avoid a shop closure or fine, or the fear of an inspection. The implicit assumption that the government collects no revenue from inspected nonfilers is close to reality.[50] To further simplify the exposition, we assume that taxpayers are risk-neutral and that y_i is the income detectable by the tax authority. Filers thus always declare y_i and we can ignore the intensive margin of compliance.

Assuming $f > p\theta T > tT$, $\exists\ \bar{y} \equiv \frac{f - tT}{p\theta - t} > T$ such that taxpayers file if $y_i > \bar{y}$. The government thus has to set (or create the perception that) $p(a)\theta > t$ to ensure that at least a fraction of taxpayers file voluntarily. The filing propensity then increases in the nonfiling penalty and the probability of detection, and decreases in the tax rate, filing cost, and exemption threshold. Heterogeneity in filing costs f could yield a situation in

[49] Without affecting the core results, one can bring the model closer to the data by assuming that larger firms are more likely to be inspected (and hence more likely to file), $p_y > 0$. In this case, it would also be reasonable to assume $p_{ya} > 0$.

[50] Allowing the government to instead collect a fine from nonfilers would complicate the model, but, provided some parameter restrictions that ensure that the government would prefer firms to file voluntarily rather than pay the fine, the model would still yield the same results.

which there are nonfilers and filers among firms both above and below the exemption threshold, as is empirically the case.[51]

1.9.1.2 Welfare Considerations

Given a continuum of taxpayers of mass one, the share of filers is $1 - G(\bar{y}) = 1 - G(f, t, \theta, T, a)$. We assume that all taxpayers derive a value $v(g)$ from the public good g that the government provides. The government budget constraint is $g + a = R = t \int_{\bar{y}}^{\infty} max(y - T, 0)dy$. Given this constraint, the government maximises social welfare

$$W = \int y \cdot dy - R - (1 - G(\bar{y})) \cdot f - p\theta \int_0^{\bar{y}} y \cdot dy + v\left(R - a\right).$$

A discrete investment in administrative enforcement Δa increases $p(a)$ by Δp from p_0 to p_1, lowers $\bar{y}$ from $\bar{y}_0$ to $\bar{y}_1$, and hence increases tax filing and payment, assuming $\bar{y}_0 > T$. The filers have to pay the filing cost and those with $y_i > \bar{y}_1$ have to pay the tax, but they no longer face the risk of the nonfiling penalty. The remaining nonfilers, in contrast, face a larger expected penalty, which generates no benefit to the government. The intervention Δa is welfare enhancing if it generates a sufficiently large gross revenue increase $\Delta R = t \int_{\bar{y}_1}^{\bar{y}_0} max(y - T, 0)dy$:

$$\Delta W = W_1 - W_0 = -\Delta R - \int_{\bar{y}_1}^{\bar{y}_0} (f - p_0\theta y)dy - \int_0^{\bar{y}_1} \Delta p\theta y \cdot dy + v'\Delta R - v'\Delta a > 0.$$

1.9.1.3 Extensions

We consider two extensions relevant to our empirical application:

- **Filers with $y < T$ provide valuable information to tax authority**: If filers below the exemption threshold provide information, e.g., about transactions with

[51] The higher initial nonfiling rate among corporations compared to the self-employed in Costa Rica, despite the absence of an exempt income amount for corporations, could be explained by the higher average income levels and a negative correlation between firm size and filing costs.

clients and suppliers, to the tax authority, and this information reduces the cost of future tax enforcement a or increases the detectable income level y, even an intervention that pushes $\bar{y}$ (further) below T could generate positive welfare gains. This is especially likely if filing costs fall discontinuously at $y_i = T$, because firms that file but do not make a payment avoid an additional errand (going to the bank).

- **Deregistration**: Firms with high filing costs might decide to deregister in response to an increase in enforcement. Assume that a deregistered yet still active firm faces a punishment (detection probability and fine) $q(a) \cdot \phi y < p(a) \cdot \theta y$ and incurs a loss k_i due to lost access to some public services. A firm deregisters in response to an increase in p (which leaves q unaffected) if its filing cost f is sufficiently high and its valuation for the public good k_i is sufficiently low (compared to firm size), i.e.,

$$(1 - p_0\theta)y_i > (1 - q\phi)y_i - k_i > (1 - t)y_i - f_i$$

and

$$(1 - q\phi)y_i - k_i > (1 - p_1\theta)y_i > (1 - t)y_i - f_i.$$

1.9.2 Potential Value of Additional Third-Party Information

To calculate the potential value of additional third-party information generated by the intervention, we consider the following. The tax authority uses third-party reports in desk audits targeted at taxpayers who reported sales lower than third-party reported sales or costs. Third-party reported sales are more conventionally used as a lower bound for the sales the taxpayer should self-report, but cost reports could conceivably play a similar role. In our case, the experiment generates new reports about treated firms' suppliers (which are sales reports for the reportee) and reports about clients (which are cost reports for the reportee). We thus calculate the additional value of both types of reports separately.

We assume that, when contacted by the tax authority and informed about the new third-party information, firms increase reported sales by the amount of the new third-party report, but compensate by increasing reported costs so as to keep the reported profit rate constant. This means the reported profit rate on the newly declared sales is the same as the initially reported profit rate. This is consistent with the results from desk audits reported in Brockmeyer and Hernandez (2018) and similar to the assumption made in that paper in the calculation of tax revenue lost due to under-reporting.

In our calculation, we use the estimates from Table 1.4, as well as estimates from regressions in which we separately consider unmatched reports about suppliers and unmatched reports about clients as outcome variables (results available upon request). To be conservative, we consider only unmatched reports about under-reporters (as defined in the footnote to Table 1.4) as valuable, and ignore the new reports about nonfilers.

We also use the average transaction amount reported in third-party reports about clients and suppliers (which does not significantly differ by treatment status conditional on filing a third-party report) and the average reported profit rate among

our experiment firms' clients and suppliers that are subject of unmatched reports,[52] and we assume that firms' income is such that they pay a marginal tax rate of 15%.

- Value of new reports about clients: 0.008 [number of new unmatched reports about under-reporters] * 12,515 * 2/3 [treated sample] * 11500000 [avg. transaction amount] * 0.13 [avg. profit rate] * 15% [hypothesised marginal tax rate] = 17 million CRC.

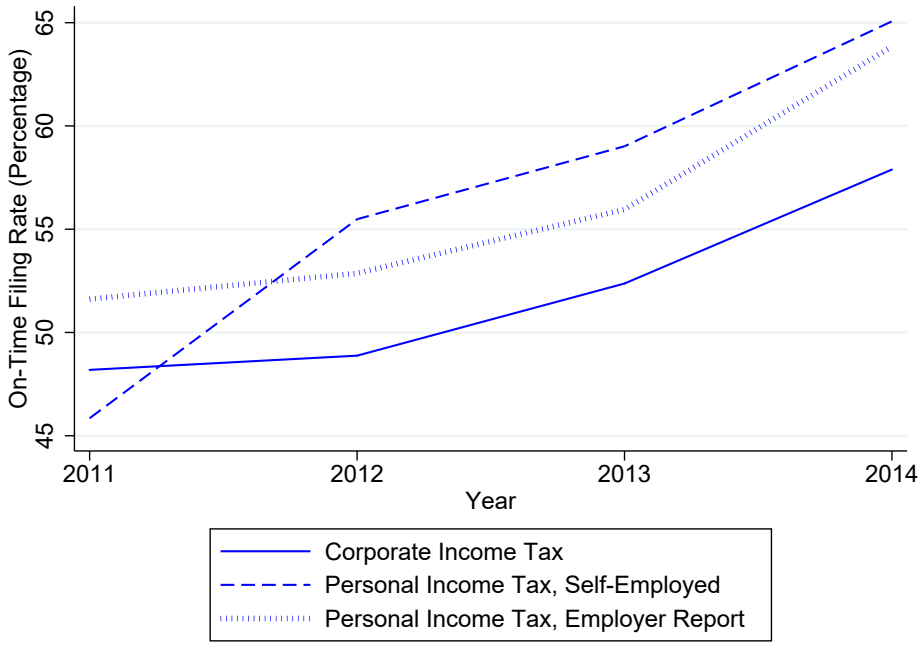
- Value of new reports about suppliers: 0.019 [number of new unmatched reports about under-reporters] * 12,515 * 2/3 [treated sample] * 3291500 [avg. transaction amount] * 0.51 [avg. profit rate] * 15% [hypothesised marginal tax rate] = 10 million CRC.

These figures amount to 20% and 12% of the total return of the experimental treatment. As the increase in unmatched reports about misreporting suppliers was positive but not statistically significant, we retain the conservative result that the generation of new third-party reports could increase the return of the experiment by up to 20%, under the assumptions laid out in this section.

[52] These results are similar when using the profits rate of all experiment firms' suppliers or clients respectively, and winsorizing the profit rate distribution in different ways.

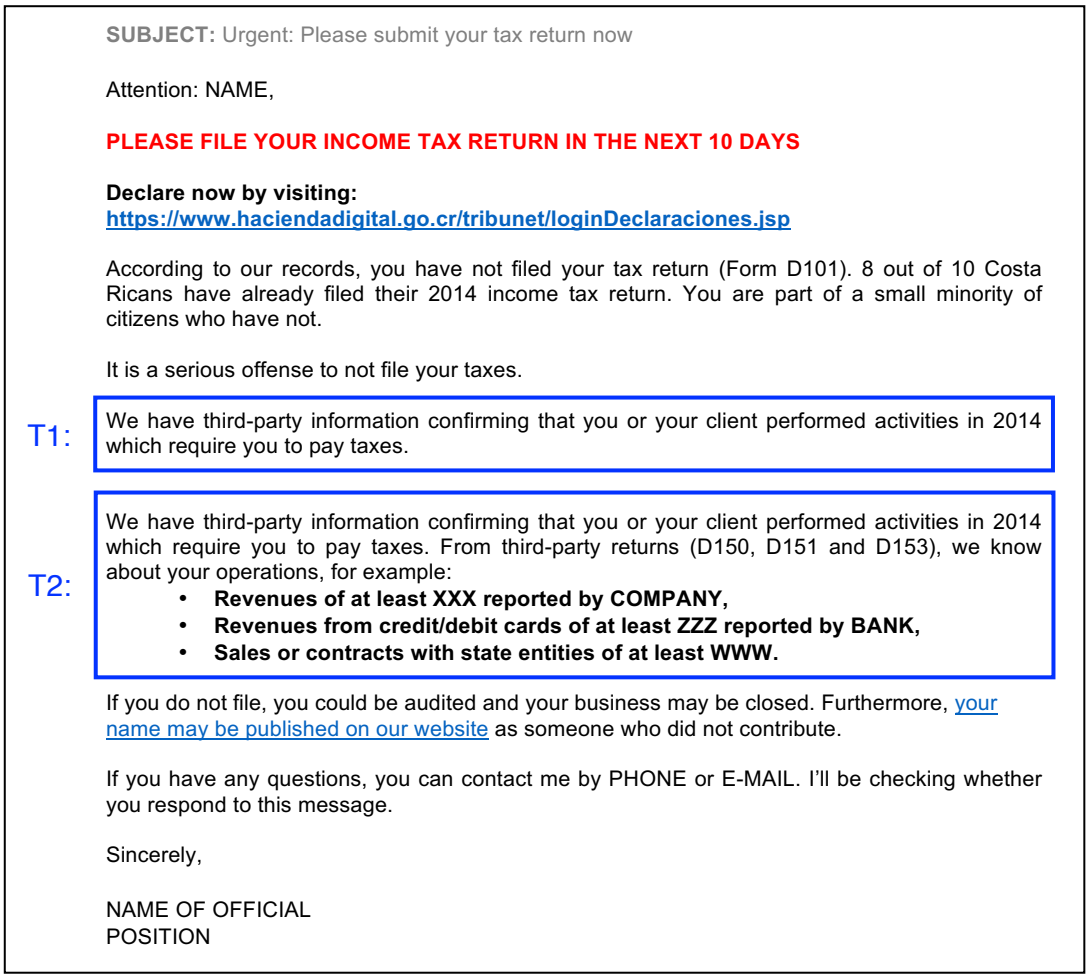
1.9.3 Figures

Figure 1.A1: Filing Rates Over Time



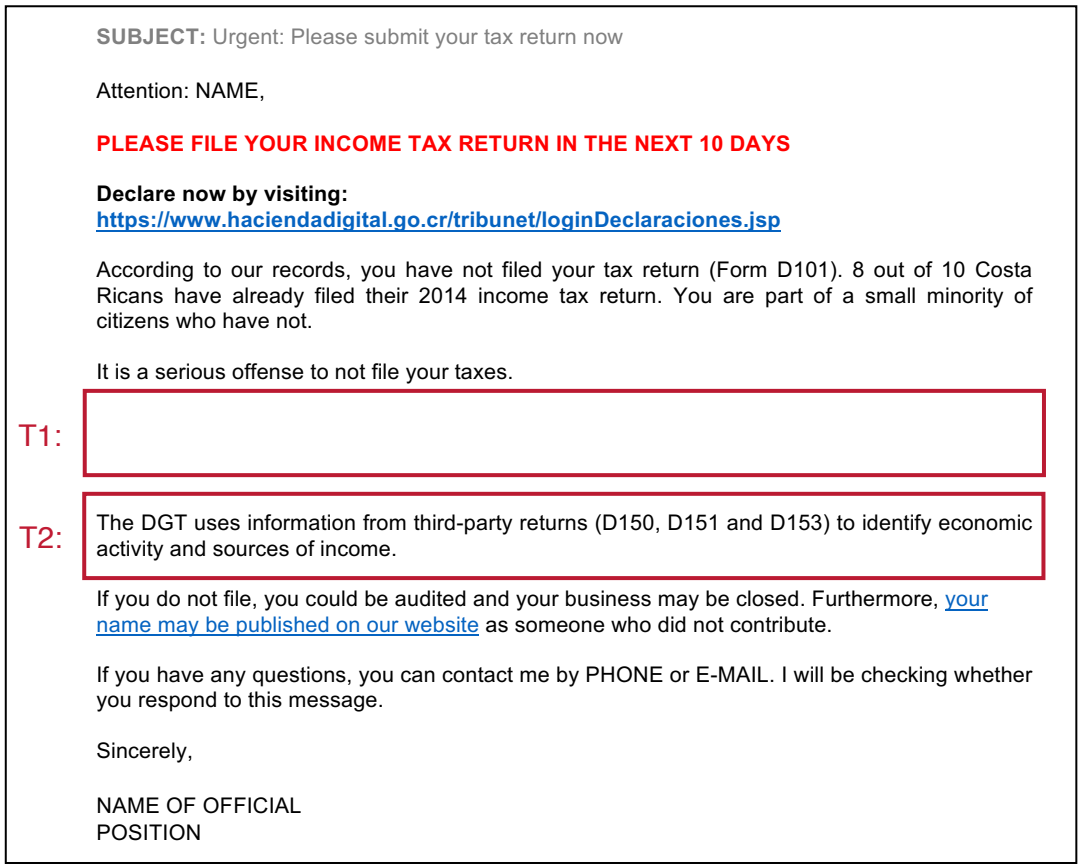
Note: This figure displays the rate of on-time filing over time, for a balanced sample of 15 countries (the identities of which cannot be disclosed), as measured by the IMF-led RA-FIT survey (Revenue Administration Fiscal Information Tool). The filing rates are self-reported by the relevant tax authorities.

Figure 1.A2: Email to Firms Covered by Third-Party Information



Note: This figure displays the treatment email used among firms with third-party information. Firms in the study received an email with either treatment 1 (T1, baseline email) or treatment 2 (T2, information email) shown. The mention of "your client" is added to make the message suitable for being sent directly to the firm or the firm's legal representative.

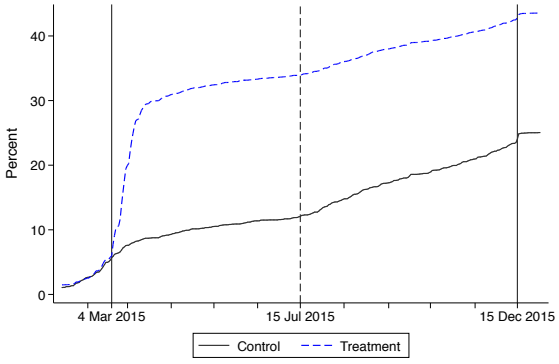
Figure 1.A3: Email to Firms Not Covered by Third-Party Information



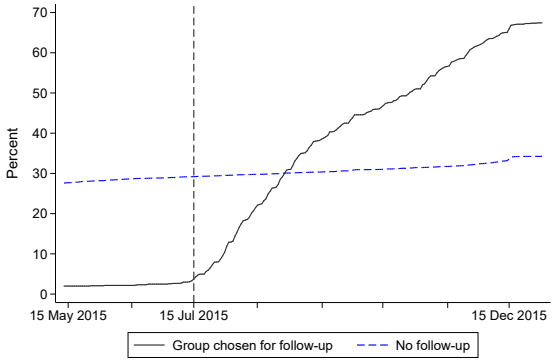
Note: This figure displays the treatment email used among firms without third-party information. Firms in the study received an email with either treatment 1 (T1, baseline email) or treatment 2 (T2, information email) shown. The mention of "your client" is added to make the message suitable for being sent directly to the firm or the firm's legal representative.

Figure 1.A4: Impact of Follow-up Activities

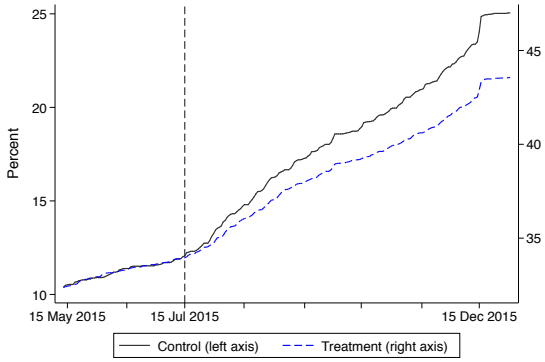
A: Income Tax Filing Behaviour, 2015



B: Impact of Follow-up on Filing Rate by Assignment to Follow-up

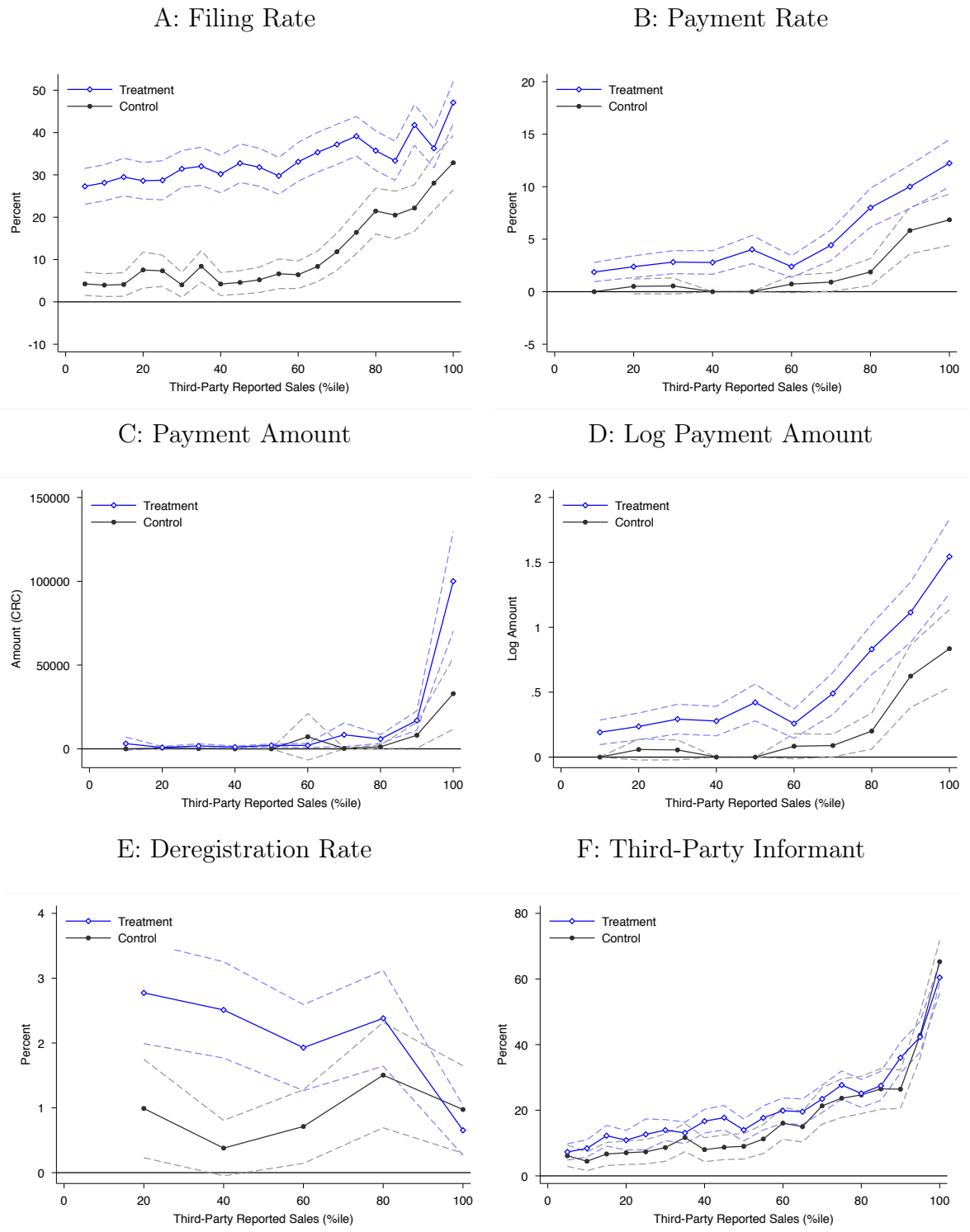


C: Impact of Follow-up on Filing Rate by Initial Treatment Group



Note: This figure displays the impact of follow-up activities conducted by the tax authority after July 15, 2015, targeting remaining nonfilers for the income tax in 2014. Targeted firms were mostly firms with large amounts of third-party information, and were disproportionately in the control group. Tax officers were assigned to follow up on these taxpayers according to a detailed protocol, starting with a phone call, then a written communication or a visit, with the possibility of referring a case to audit in case of continued non-compliance. The control and treatment groups referred to in the graphs are our experimental control and treatment groups.

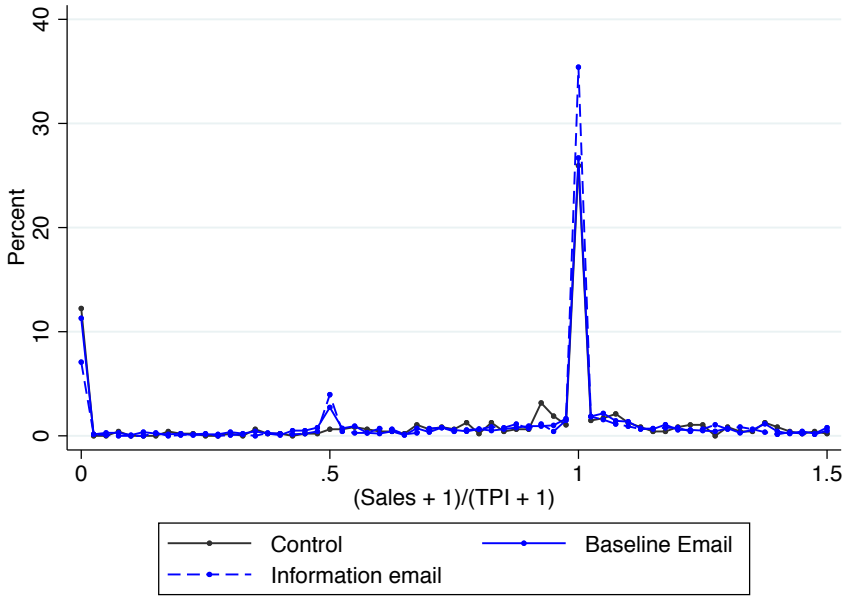
Figure 1.A5: Heterogeneity by Third-Party Information (Percentiles)



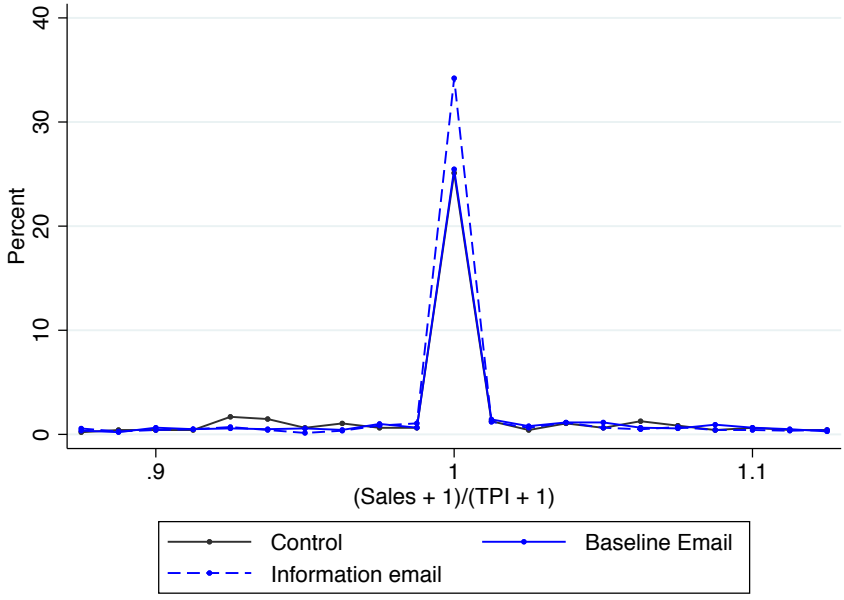
Note: This figure displays heterogeneity in the main treatment effects. It is identical to Figure 1.3, except that the horizontal axis here displays the relevant percentile of third-party reported-sales rather than the level of (log) sales at the relevant percentile, and that Panels A and F use ventiles instead of deciles.

Figure 1.A6: Self-Reported versus Third-Party Reported Sales

A: Full Distribution



B: Zoom at 1

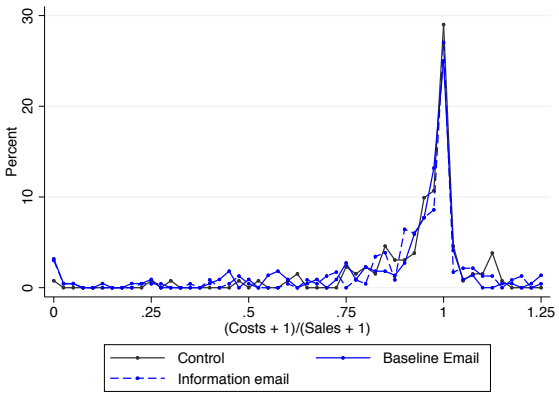


Note: These figures show the distribution of firms based on the ratio of self-reported sales to third-party reported sales, for firms with third-party information that filed an income tax declaration within fifteen weeks of the experiment start. Figure 1.5 displays the results of statistical tests for whether firms' likelihood to match self-reported sales to third-party reported sales differs significantly across treatment groups.

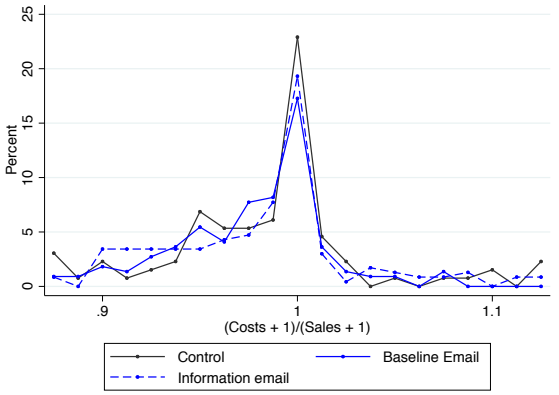
Figure 1.A7: Self-Reported Sales versus Self-Reported Costs

A: Corporations

A1: Full Distribution

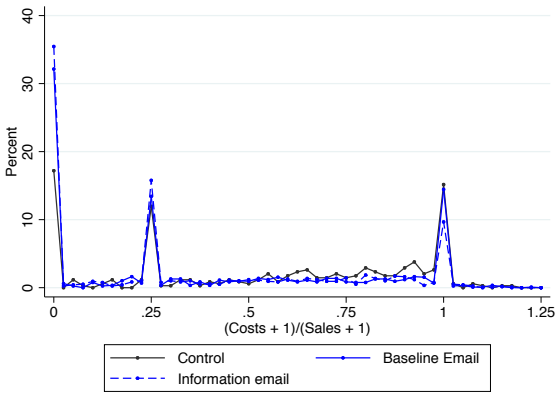


A2: Zoom at 1

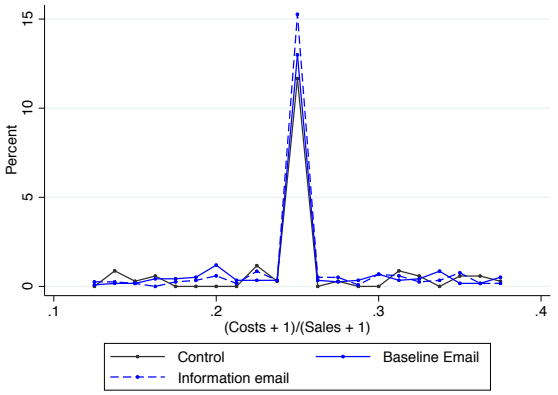


B: Self-Employed

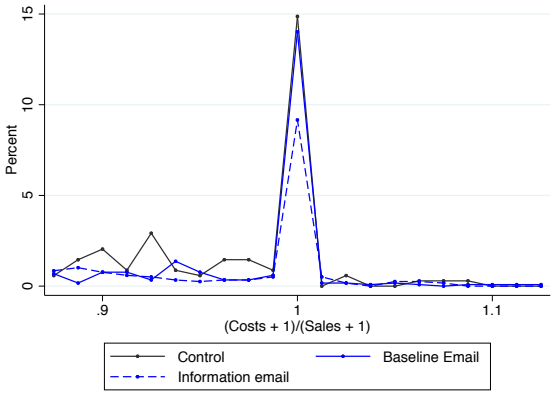
B1: Full Distribution



B2: Zoom at 0.25



B3: Zoom at 1



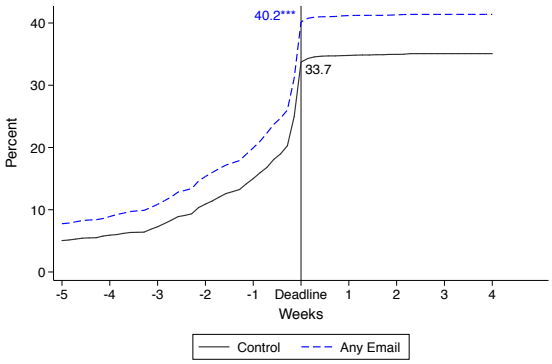
Note: These figures show the distribution of firms based on the ratio of self-reported sales to self-reported costs, for firms with third-party information that filed an income tax declaration within fifteen weeks of the experiment start. Figure 1.5 displays the results of statistical tests for whether firms' likelihood to match self-reported sales to self-reported costs differs significantly across treatment groups.

Figure 1.A8: Compliance by Information Coverage and Treatment Group, 2015

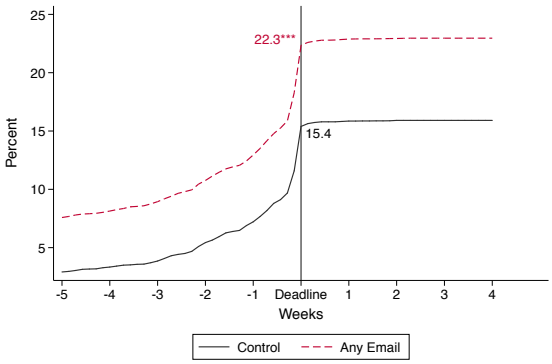
A: Firms with Third-Party Information

B: Firms without Third-Party Information

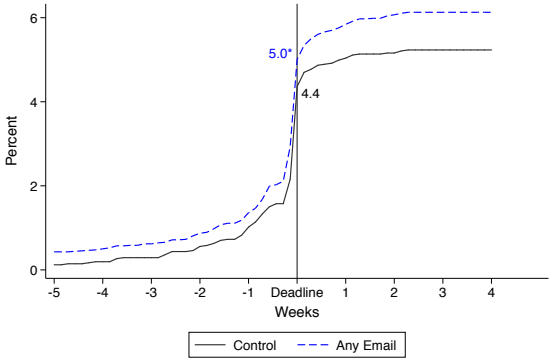
A1: Filing Rate



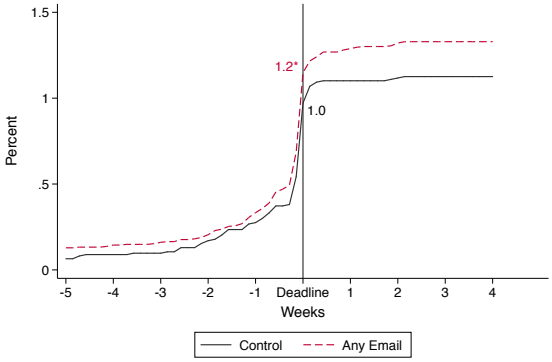
B1: Filing Rate



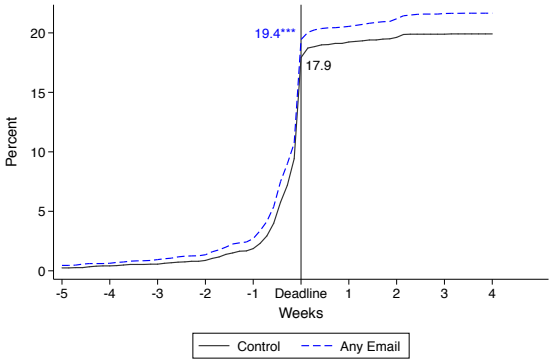
A2: Payment Rate



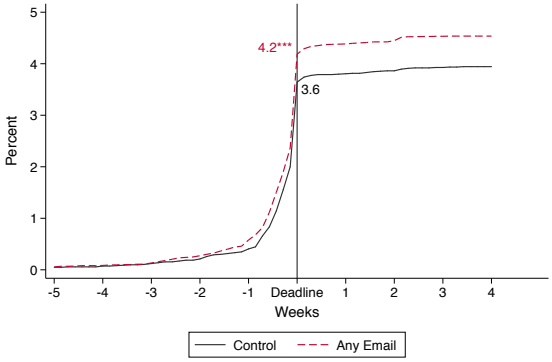
B2: Payment Rate



A3: Third-Party Informant



B3: Third-Party Informant

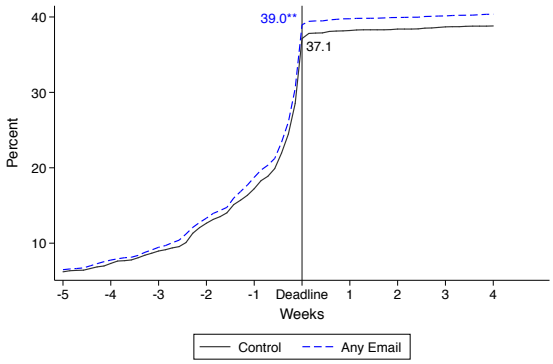


Note: These figures show the share of firms filing income tax (row 1), paying income tax (row 2), and filing a third-party informative declaration (D151) about a client or supplier (row 3), all for fiscal year 2015. Column A corresponds to firms with third-party information and column B corresponds to firms without third-party information. The vertical line in each figure indicates the filing deadline. The black solid line corresponds to the control group and the blue/red dashed lines correspond to the treatment group for the two different subsamples, as explained in Table 1.A3. The numbers indicate the mean for each outcome and treatment group at the deadline. Stars indicate a significant difference compared to the control group and come from regressions that include controls. Significance levels are noted as per convention: * $p < .10$, ** $p < .05$, *** $p < .01$.

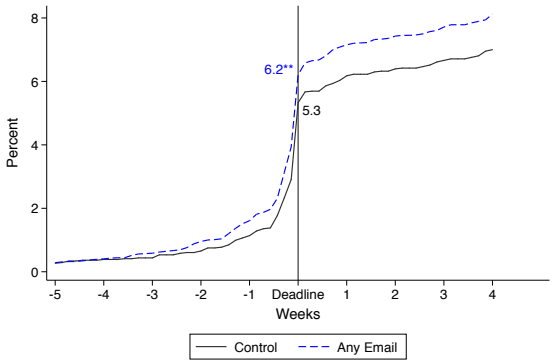
Figure 1.A9: Compliance by Information Coverage and Treatment Group, 2016

A: Firms with Third-Party Information

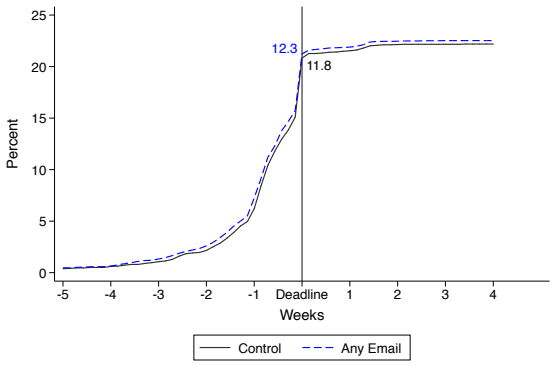
A1: Filing Rate



A2: Payment Rate

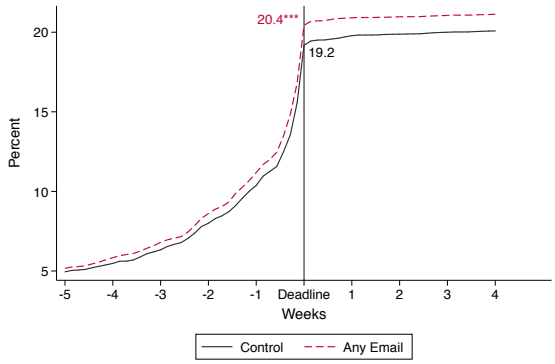


A3: Third-Party Informant

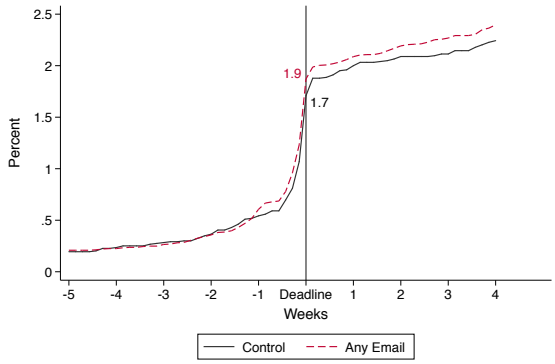


B: Firms without Third-Party Information

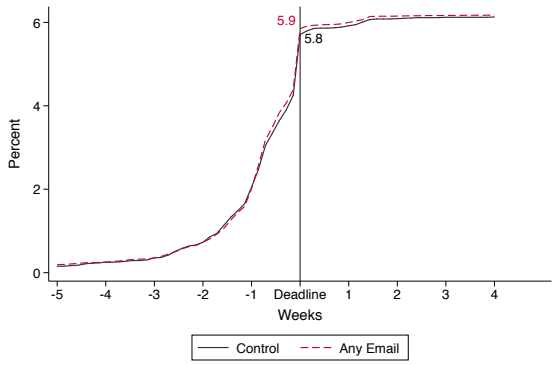
B1: Filing Rate



B2: Payment Rate

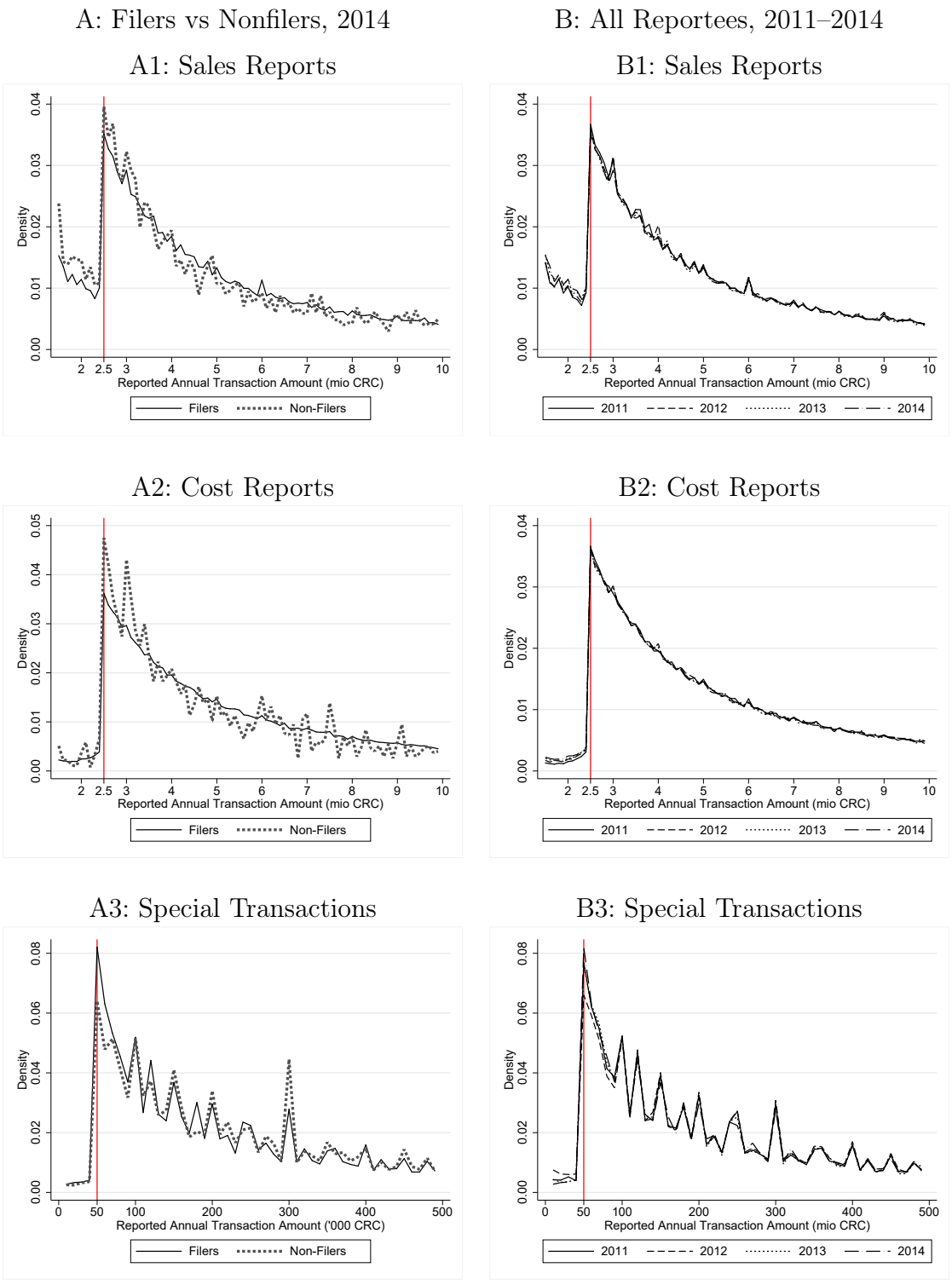


B3: Third-Party Informant



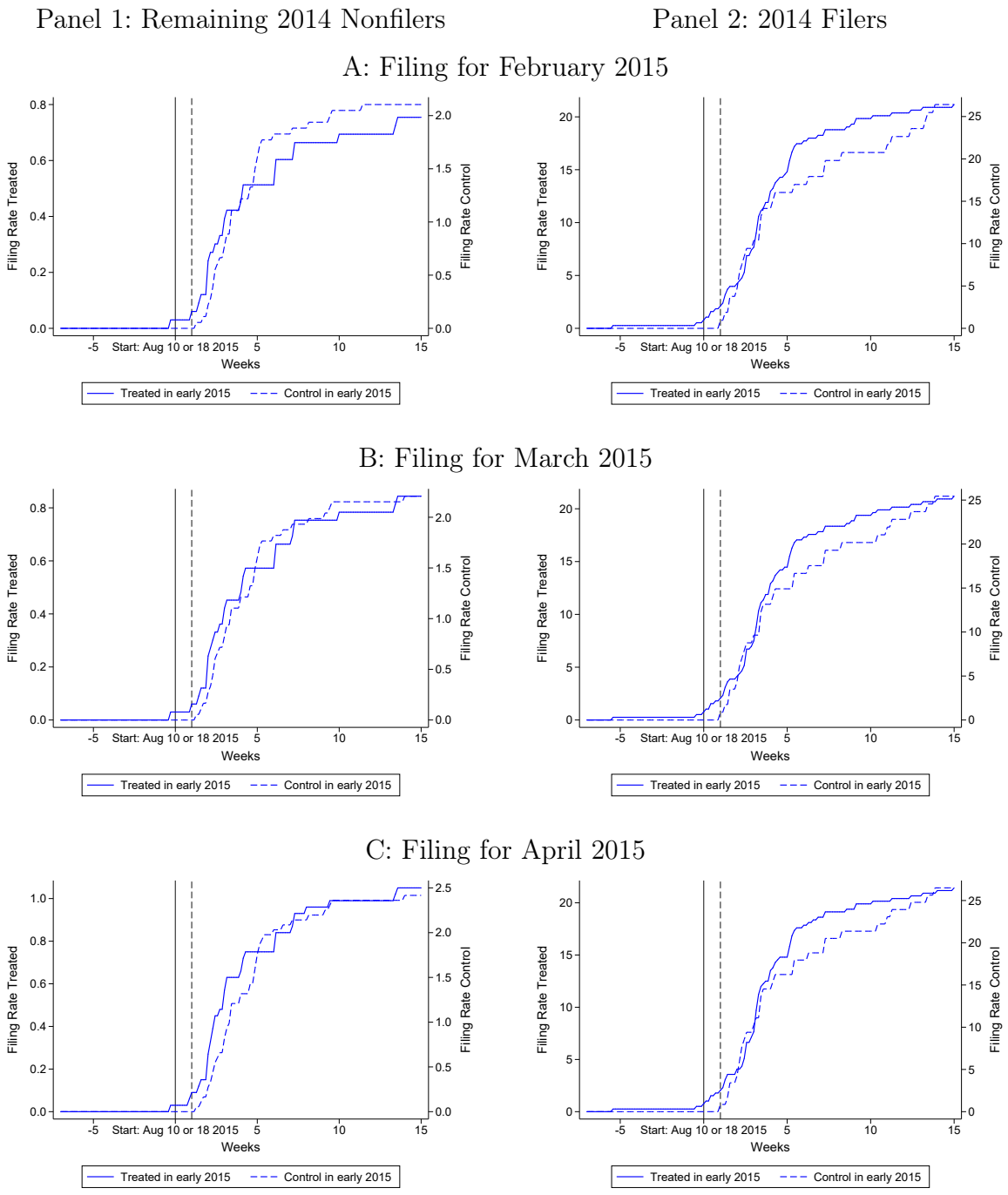
Note: This figure is identical to Figure 1.A8, but considers compliance outcomes for fiscal year 2016.

Figure 1.A10: Distribution of Third-Party Reports around Reporting Threshold



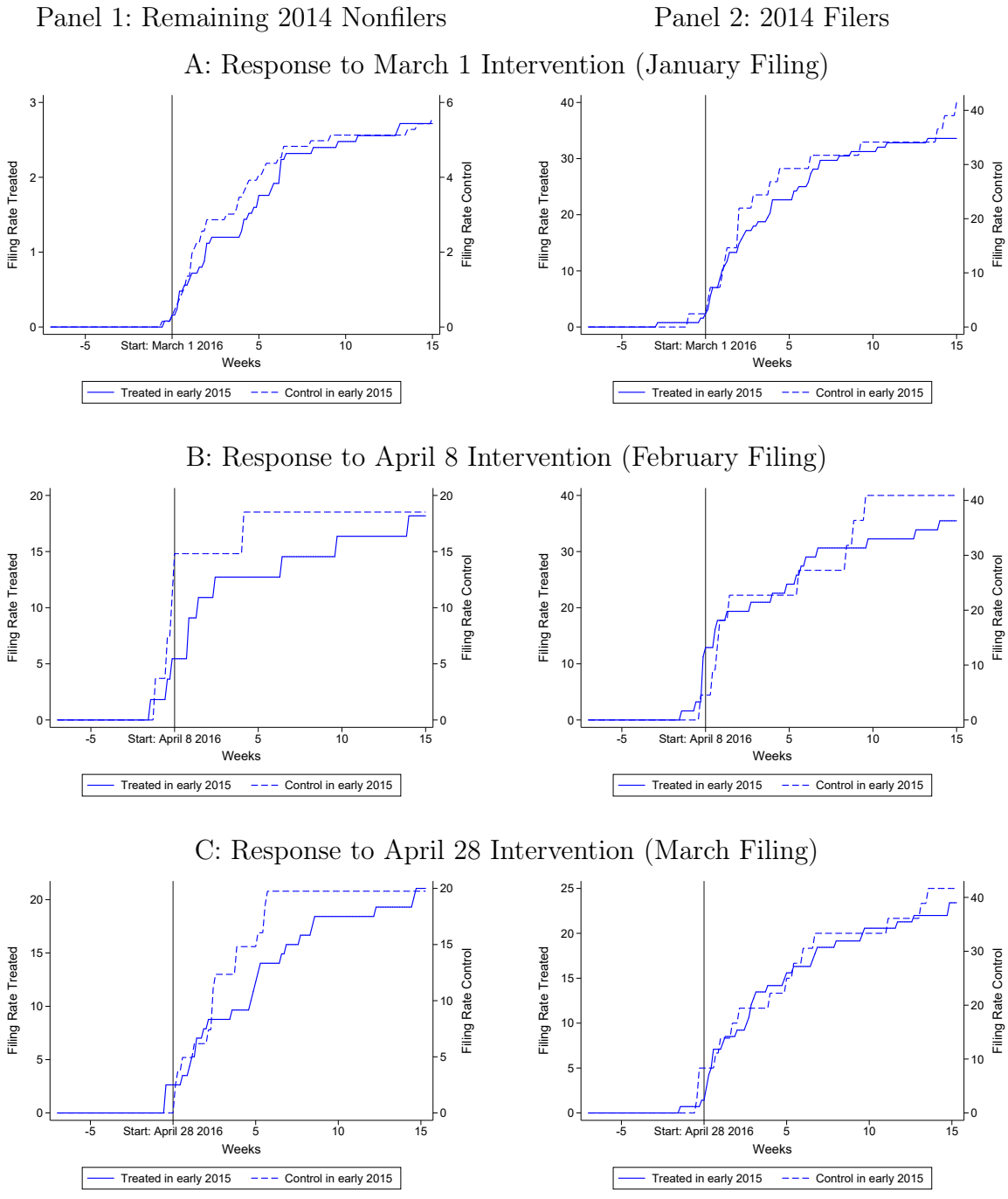
Note: These figures show the density distribution of third-party reports of sales (row 1), costs (row 2), and special transaction reports (row 3) from the D151 firm-to-firm transaction reports. An observation is a reporter-reportee-year. For sales and costs, the reporting requirement threshold is an annual transaction amount of CRC 2.5 million (US$ 4418); for special transactions (rent, interest, commissions, and professional services), it is CRC 50,000 (US$ 88). The panels in column A compare filers to nonfilers for the year 2014 (experiment sample), and the panels in column B compare the years 2011–2014. The reporting requirement threshold is marked by a red horizontal line.

Figure 1.A11: Filing Response to Sales Tax Interventions, 2015



Note: These figures show the sales tax filing rate for the months of February, March, and April 2015, as measured later in 2015. Nonfilers for the sales tax for these months (and for the month of May, as per the results displayed in Figure 1.4) received an enforcement email from the tax authority (similar to the one used in our experiment) on August 10 or August 18, 2015, as indicated by the vertical lines. The figures focus on firms in our experimental sample, distinguishing firms that remained nonfilers for the income tax in 2014 (Panel 1, left side), and firms that ultimately did file income tax for 2014 (Panel 2, right side).

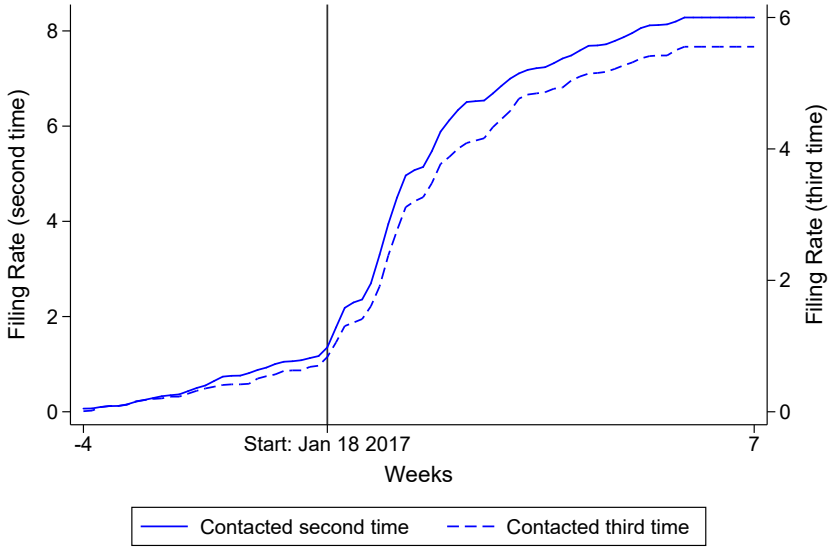
Figure 1.A12: Filing Response to Sales Tax Interventions, 2016



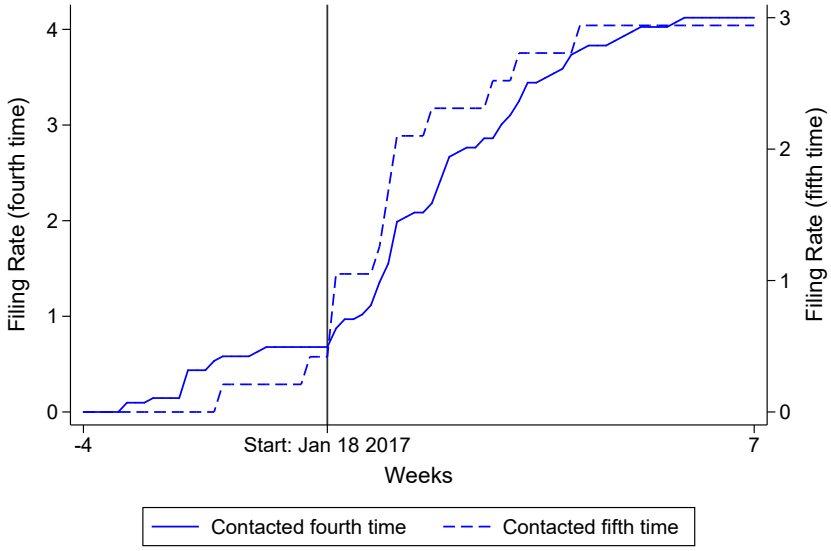
Note: These figures show the sales tax filing rate for the months of January, February and March 2016, as measured around different dates in March and April 2016 at which nonfilers for the sales tax for the respective months received an enforcement email from the tax authority (similar to the one used in our experiment). The figure is otherwise constructed in the same way as Figure 1.A11.

Figure 1.A13: Filing Response to Multiple Repeated Interventions

A: Taxpayers Contacted Two or Three Times

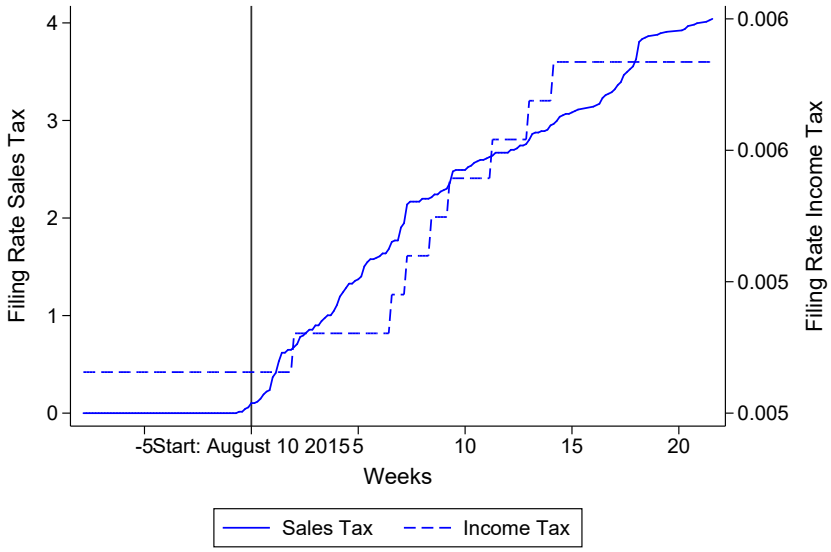


B: Taxpayers Contacted Four or Five Times



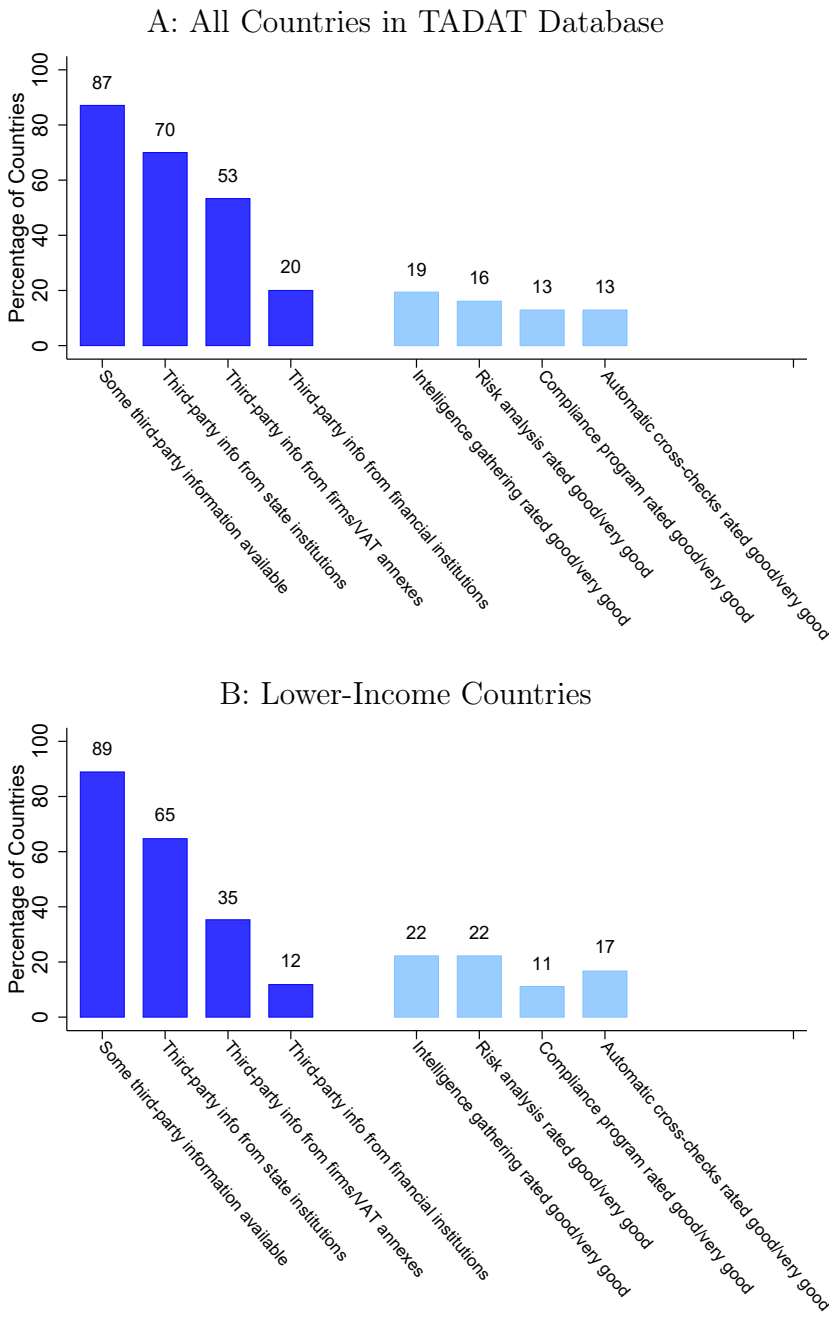
Note: These figures show the rate of income tax filing for fiscal year 2016, around January 18, 2017, when nonfilers for the income tax for 2016 received an enforcement email from the tax authority (similar to the one used in our experiment). The different series distinguish taxpayers for whom this enforcement email was the second, third, fourth, or fifth time they received such an email for nonfiling behaviour. In constructing these groups, we take into account the 2016 and 2017 income tax nonfiling campaigns, and the 2015 and 2016 sales tax nonfiling campaigns, as discussed in Section 1.5.2.

Figure 1.A14: Impact of Online Publication of Sales Tax Nonfilers



Note: This figure shows the rate of filing for the income tax for 2014 and the sales tax for May 2015, among firms whose names were published online on August 10, 2015. The majority of these firms were also sales tax nonfilers for February–April 2015, and 1,366 of them were also income tax nonfilers for 2014 and part of our experimental sample (265 among the firms without third-party information, one with third-party information, and 1,100 among a separate SMS experiment). Among these firms whose names were published online, 412 ultimately filed sales tax for May 2015 and 61 filed income tax for 2014.

Figure 1.A15: Availability and Use of Third-Party Information Across Countries



Note: This figure displays results from an analysis of 31 country reports completed under the Tax Administration Diagnostic Assessment Tool (TADAT). TADAT is an assessment tool implemented globally by development partners and technical assistance providers including the IMF and World Bank. Bars 1–4 are coded based on a text analysis of the report, and reflect the share of countries that have access to any kind of third-party information (bar 1), and to specific types of third-party information (bars 2–4). Third-party information from state institutions includes customs and procurement information. Bars 5–8 display the share of countries receiving a grade of A or B (instead of C or D) for TADAT performance indicators P2-3-1, P2-3-2, P2-4, and P6-14, respectively. Grades are given by an external assessment team after a visit to the country and are backed by evidence. Panel B focuses on the subsample of low and lower-middle income countries according to the current World Bank classification. The results are similar when focusing on low-income countries only. The remaining countries are almost exclusively upper middle-income countries, not high-income countries.

Table 1.A1: Summary of Filers and Nonfilers, Firm Characteristics

	(1) 2014 On-Time Filers	(2) 2014 On-Time Filers (Below Median Sales)	(3) 2014 On-Time Filers (Below Median Sales if Positive)	(4) 2014 Nonfilers	(5) 2014 Nonfilers (Above Median Sales)	(6) 2014 Nonfilers (with TPI)	(7) 2014 Nonfilers (without TPI)
Sales	40.50	1.253	3.250	5.497	10.99	12.23	3.235
	(0.208)	(0.00375)	(0.00656)	(0.193)	(0.383)	(0.606)	(0.156)
Costs	37.46	0.885	2.050	5.251	10.46	11.56	3.132
	(0.203)	(0.0207)	(0.0230)	(0.194)	(0.384)	(0.596)	(0.162)
Profit rate	0.538	0.605	1.041	0.203	0.408	0.314	0.165
	(0.00177)	(0.00299)	(0.00367)	(0.00343)	(0.00656)	(0.00815)	(0.00366)
TPI reportee in 2014	0.599	0.376	0.652	0.252	0.441	1	0
	(0.000808)	(0.00113)	(0.00124)	(0.00195)	(0.00315)	(0)	(0)
Third-party reported sales	19.59	0.719	1.831	2.043	4.082	8.096	0.00905
	(0.0968)	(0.00309)	(0.00605)	(0.0774)	(0.154)	(0.301)	(0.00244)
Corporation	0.327	0.225	0.146	0.234	0.206	0.140	0.265
	(0.000774)	(0.000974)	(0.000923)	(0.00190)	(0.00257)	(0.00310)	(0.00229)
San José	0.420	0.421	0.401	0.490	0.492	0.489	0.491
	(0.000818)	(0.00116)	(0.00129)	(0.00224)	(0.00317)	(0.00447)	(0.00259)
Observations	367,488	183,744	146,577	49,774	24,887	12,515	37,259

Note: This table displays summary statistics for firms that filed income tax in 2014 by the filing deadline versus firms that did not and that were part of our experiment sample (i.e., had an email address on file). The rows display means, with standard errors in parentheses. Amounts are in million CRC. Columns 2 and 3 restrict the sample to below or above median sales. For nonfilers, sales in the most recent period is the maximum of sales reported in the most recent available tax declaration and the most recent third-part report (as far back as 2010). For filers, we use the maximum of the most recent sales report and the current third-party report. The results are robust to alternative sales definitions.

Table 1.A2: Summary of Filers and Nonfilers, Past Compliance

	(1) 2014 On-Time Filers	(2) 2014 On-Time Filers (Below Median Sales)	(3) 2014 On-Time Filers (Below Median Sales if Positive)	(4) 2014 Nonfilers	(5) 2014 Nonfilers (Above Median Sales)	(6) 2014 Nonfilers (with TPI)	(7) 2014 Nonfilers (without TPI)
Year First Registered	2007.4	2007.6	2007.7	2008.2	2008.1	2008.4	2008.1
	(0.00404)	(0.00614)	(0.00704)	(0.0126)	(0.0172)	(0.0274)	(0.0141)
No. of Years Declared 2010-2013	2.956	2.685	2.623	0.904	1.359	1.059	0.851
	(0.00234)	(0.00354)	(0.00405)	(0.00549)	(0.00825)	(0.0119)	(0.00613)
Most Recent Gross Liability Conditional on Declaring	0.355	0.00684	0.0245	0.0603	0.0816	0.127	0.0345
	(0.00202)	(0.0000753)	(0.000188)	(0.00301)	(0.00406)	(0.00852)	(0.00253)
				[22,819]	[16,868]	[6,394]	[16,425]
Most Recent Payment Conditional on Declaring	0.104	0.00380	0.0125	0.0189	0.0255	0.0319	0.0138
	(0.000614)	(0.0000577)	(0.000121)	(0.000995)	(0.00134)	(0.00248)	(0.000988)
				[22,819]	[16,868]	[6,394]	[16,425]
No. of Years Presented D151 Report 2010-2013	1.729	0.956	1.317	0.324	0.609	0.503	0.263
	(0.00273)	(0.00313)	(0.00393)	(0.00345)	(0.00622)	(0.00871)	(0.00351)
No. of Years Covered by Third-Party Report 2010-2013	1.936	1.305	1.726	0.802	1.487	1.389	0.605
	(0.00268)	(0.00338)	(0.00406)	(0.00513)	(0.00773)	(0.0124)	(0.00504)
Observations	367,488	183,744	146,577	49,774	24,887	12,515	37,259

Note: This table displays summary statistics for firms that filed income tax in 2014 by the filing deadline versus firms that did not and that were part of our experiment sample (i.e., had an email address on file). The rows display means, with standard errors in parentheses. Where different from the column total, the number of observations is displayed in brackets. Amounts are in million CRC. Columns 2 and 3 restrict the sample to below or above median sales. For nonfilers, sales in the most recent period is the maximum of sales reported in the most recent available tax declaration and the most recent third-party report (as far back as 2010). For filers, we use the maximum of the most recent sales report and the current third-party report. The results are robust to alternative sales definitions.

1.9.4 Tables

Table 1.A3: Experiment Timeline

Extraction of nonfilers list ($N = 49,757$)	February 15
Emails to taxpayers with third-party information ($N = 12,515$) No message control group T1: Baseline email T2: Information email (mentions amount of largest third-party information report)	March 4–10
Emails to taxpayers without third-party information ($N = 37,242$) No message control group T1: Baseline email T2: Information email (emphasises tax authority's use of third-party information)	April 7–23

Note: This table presents the experiment timing, as discussed in Section 1.3. The treatment messages are displayed in Figures 1.A2 and 1.A3. To determine which firms were covered by third-party information, the tax authority considered all reports by other firms, by state institutions, and by credit/debit card companies. A small number of firms in the study shared the same primary email address, either because they had a common owner or shared a common legal representative. For this reason, firms were grouped into clusters based on their primary email address and randomisation took place at the cluster level. Throughout our analysis, we use standard errors that are robust to within-cluster correlation, and the results are robust to limiting the sample to clusters with only one firm (98% of clusters). The mean number of firms per cluster is 1.04 and the median is 1.

Table 1.A4: Enforcement Activities, 2008–2015

	Number of taxpayers contacted							
Type of activity	2008	2009	2010	2011	2012	2013	2014	2015
Full audits	825	897	817	829	895	847	880	662
Shop closures*	n/a	n/a	120	43	75	166	293	167
Invitation of "suspicious taxpayers" for interviews				680	2,293			
Control of provision of receipts	994	1,415	1,319				1,226	1,141
Control of mega events		117	102	30	25			12
Preventive controls (undefined)								559
Checks of discrepancies (desk audit)	2,864	2,470	2,400			5,150	6,197	7,099
Checks on suppliers to state institutions		748						
Checks on importers		386						
Follow-up on complaints about taxpayers (visits to taxpayer address)	865	1,608	969				5,522	8,130
Messages to nonfilers for informative declaration	2,304	2,924	3,344			2,130	1,190	1,024
Messages to nonfilers for income and sales tax	4,318	4,284	4,788	17,500	4,586	9,220	9,685	7,963
Publication of sales tax nonfilers								7,104
Of which, also income tax nonfilers								1,376

Note: Authors' compilation based on tax authority's annual enforcement plans and their evaluations. This table contains a non-exhaustive list of documented enforcement activities conducted by the extensive control unit. It does not include interventions conducted by the units for payment arrears and taxpayer services. "Checks" and "follow-ups" follow a detailed protocol underpinned by the tax administration regulations. They usually begin with a phone call, which (in case of no response by the taxpayer) is followed by an official notification with justification, a personal visit, a shop closure or administrative determination of amount to pay (as per a formula and all available information), and finally the transmission of the case to payment arrears (unless the amount owed is below a certain threshold) or to the audit department. The delay between each of these steps is specified by the regulation. Tax declarations presented or rectified in this process are checked by the assigned tax officer for consistency with third-party information and sector-specific average profit margins. "Messages" are done by email or SMS. Given data availability and the desire to have comparable information across years and types of activities, the figures in this table refer to planned activities (except for control of mega events where only realised numbers were available). The execution rate varies between 70% and 150% and is on average close to the planned rate. The activities presented in this list are similar to control activities found in other low- and middle-income countries (e.g., Dominican Republic, Guatemala, and Senegal).

*Out of the shop closures, about one-half to one-third are for nonfiling, the remainder for non-emission of receipt or non-payment of declared taxes.

Table 1.A5: Balance of Randomisation, Firm Characteristics

	Firms with Third-Party Information (Experiment 1)					Firms without Third-Party Information (Experiment 2)				
	(1) Control	(2) D1: Control - Baseline email	(3) D2: Control - Information email	(4) D1=D2=0 F-statistic	(5) D1=D2=0 p-value	(6) Control	(7) D1: Control - Baseline email	(8) D2: Control - Information email	(9) D1=D2=0 F-statistic	(10) D1=D2=0 p-value
Corporation	0.142	-0.00276	-0.00360	0.113	0.893	0.262	0.0111	-0.00209	1.434	0.238
	(0.00561)	(0.00789)	(0.00794)			(0.00445)	(0.00786)	(0.00655)		
Has legal representative	0.607	-0.00215	0.00354	0.126	0.882	0.659	0.00915	-0.00114	1.240	0.289
	(0.00808)	(0.0114)	(0.0115)			(0.00469)	(0.00694)	(0.00690)		
Has second email	0.0562	0.00131	0.00292	0.162	0.851	0.0542	-0.00112	0.00207	0.602	0.547
	(0.00359)	(0.00506)	(0.00514)			(0.00205)	(0.00290)	(0.00294)		
Located in San José	0.490	-0.00202	0.000361	0.0256	0.975	0.488	0.00408	0.00366	0.197	0.821
	(0.00795)	(0.0113)	(0.0114)			(0.00484)	(0.00753)	(0.00699)		
Was contacted early	0.117	0.00554	-0.00215	0.612	0.543	0.00202	0.000453	0.000723	0.715	0.489
	(0.00503)	(0.00716)	(0.00713)			(0.000405)	(0.000601)	(0.000621)		
TPI > 2.5 million CRC	0.346	-0.00485	-0.0188	1.585	0.205					
	(0.00794)	(0.0111)	(0.0110)							
TPI > 6 million CRC	0.182	0.00774	-0.00626	1.131	0.323					
	(0.00692)	(0.00976)	(0.00932)							
Made advance payment in 2014	0.0233	0.00139	0.00000564	0.114	0.892	0.00462	0.000738	-0.000818	1.674	0.188
	(0.00235)	(0.00334)	(0.00332)			(0.000671)	(0.000936)	(0.000869)		
Filed in 2013	0.329	0.00716	-0.00477	0.660	0.517	0.174	-0.00974	-0.00740	2.035	0.131
	(0.00745)	(0.0104)	(0.0106)			(0.00355)	(0.00506)	(0.00498)		
Reported net liability > 0 in 2013	0.0853	0.000409	0.00608	0.579	0.561	0.0156	0.000752	0.0000480	0.135	0.873
	(0.00441)	(0.00620)	(0.00630)			(0.00112)	(0.00161)	(0.00158)		
Made payment in 2013	0.0533	-0.00165	0.0000129	0.0763	0.927	0.0106	0.000339	-0.000103	0.0604	0.941
	(0.00356)	(0.00494)	(0.00500)			(0.000923)	(0.00133)	(0.00130)		
Was third-party informant in 2013	0.190	-0.00707	-0.0116	0.875	0.417	0.0457	-0.000366	-0.00155	0.182	0.834
	(0.00634)	(0.00875)	(0.00881)			(0.00190)	(0.00270)	(0.00269)		
Observations	4,128	4,260	4,127			12,350	12,516	12,376		

Note: This table shows the balance of randomisation in terms of firm characteristics, as measured before the experiment start. The rows correspond to the different variables. Columns 1–5 (6–10) correspond to the sample of firms with (without) third-party information, i.e., experiment 1 (2). Column 1 (6) displays the mean for the control group, columns 2 and 3 (7 and 8) show the mean difference between the control group and treatment groups 1 and 2, respectively, column 4 (9) reports an F-statistic from a test of the hypothesis that the two treatment groups are jointly equal to the control group, and column 5 (10) provides the corresponding p-value. Robust standard errors clustered by email address are in parentheses. TPI stands for third-party information (third-party reported sales), meaning the sum of sales reported by clients (D151), state institutions (D150), and credit/debit card companies (D153). The cutoffs of 2.5 million and 6 million CRC correspond to priority group designations used by the tax authority. When treated firms had two email addresses on file, the tax authority sent the same email to both. We ensure that the proportion of firms with two email addresses is balanced across control and treatment groups, and we control for second email addresses when we estimate treatment effects. As one regional tax office deviated from the experiment design and contacted firms prior to the start of the experiment date through phone calls and emails (different from the experimental emails presented in the previous sections), we also consider whether the occurrence of such early communication is balanced across treatment groups.

Table 1.A6: Balance of Randomisation, Outcomes at Baseline

	Firms with Third-Party Information (Experiment 1)					Firms without Third-Party Information (Experiment 2)				
	(1) Control	(2) D1: Control - Baseline email	(3) D2: Control - Information email	(4) D1=D2=0 F-statistic	(5) D1=D2=0 p-value	(6) Control	(7) D1: Control - Baseline email	(8) D2: Control - Information email	(9) D1=D2=0 F-statistic	(10) D1=D2=0 p-value
Filed	0.0567	0.00411	0.00195	0.298	0.742	0.0131	0.000545	0.000296	0.0686	0.934
	(0.00370)	(0.00532)	(0.00522)			(0.00102)	(0.00148)	(0.00152)		
Had positive net liability	0.0167	0.00324	0.00243	0.667	0.513	0.000648	0.000950	0.000483	2.420	0.0889
	(0.00202)	(0.00294)	(0.00293)			(0.000229)	(0.000453)	(0.000379)		
Made payment	0.00751	0.00282	0.00170	0.985	0.373	0.000405	0.000474	0.000161	0.976	0.377
	(0.00134)	(0.00205)	(0.00200)			(0.000181)	(0.000340)	(0.000280)		
Log payment	0.0857	0.0343	0.0185	1.056	0.348					
	(0.0154)	(0.0238)	(0.0229)							
Was third-party informant	0.164	-0.00343	-0.00699	0.342	0.710	0.0117	0.000165	0.00103	0.273	0.761
	(0.00586)	(0.00818)	(0.00844)			(0.000967)	(0.00138)	(0.00146)		
No. months decl. sales tax	0.619	-0.0396	-0.0607	0.692	0.501	0.243	-0.0249	0.00361	1.431	0.239
	(0.0381)	(0.0524)	(0.0521)			(0.0131)	(0.0183)	(0.0188)		
No. months paid sales tax	0.126	-0.0227	-0.0264	0.940	0.390	0.0148	0.00643	0.00441	1.347	0.260
	(0.0153)	(0.0204)	(0.0204)			(0.00248)	(0.00421)	(0.00393)		
Sales tax payment	108900.9	-75006.4	-64056.2	0.716	0.489	736.7	1416.5	245.1	0.444	0.641
	(62816.2)	(64536.4)	(65884.1)			(208.0)	(1728.3)	(483.3)		
Deregistered	0.00388	-0.00106	0.000000939	0.486	0.615	0.00372	-0.000768	-0.00130	1.705	0.182
	(0.00102)	(0.00131)	(0.00141)			(0.000548)	(0.000731)	(0.000704)		
Deregistered (sales tax)	0.000727	-0.0000225	-0.000242	0.132	0.876	0.000891	-0.000491	-0.000487	1.355	0.258
	(0.000420)	(0.000584)	(0.000542)			(0.000268)	(0.000322)	(0.000324)		
Switched to simplified tax regime	0.000242	-0.000242	-0.000242	1.000	0.317					
	(0.000242)	(0.000242)	(0.000242)							
Observations	4,128	4,260	4,127			12,350	12,516	12,376		

Note: This table shows the balance of randomisation in terms of outcomes, as measured by the day before the experiment start. The rows correspond to the different variables for fiscal year 2014. The number of months that a taxpayer filed and paid sales tax, and the sales tax payment are calculated over July 2013 until June 2014. Columns 1–5 (6–10) correspond to the sample of firms with (without) third-party information, i.e., sub-experiment 1 (2). Column 1 (6) displays the mean for the control group, columns 2 and 3 (7 and 8) show the mean difference between the control group and treatment groups 1 and 2 respectively, column 4 (9) reports an F-statistic from a test of the hypothesis that the two treatment groups are jointly equal to the control group, and column 5 (10) provides the corresponding p-value. Robust standard errors clustered by email address are in parentheses.

Table 1.A7: Impact on Income Tax Compliance, Firms with Third-Party Information (Full Results)

	OLS				
	(1) Filed	(2) Positive net liability	(3) Positive payment	(4) Log payment	(5) Payment
T1: Baseline email	0.211	0.0444	0.0307	0.341	8352.9
	(0.00873)	(0.00480)	(0.00377)	(0.0428)	(2449.8)
T2: Information email	0.230	0.0519	0.0373	0.409	10441.2
	(0.00889)	(0.00494)	(0.00397)	(0.0444)	(2823.7)
Corporation	-0.00417	0.0443	0.00724	0.116	11589.0
	(0.0157)	(0.0116)	(0.00846)	(0.0997)	(7276.9)
Has legal representative	-0.000636	-0.00119	0.00263	0.0412	2819.4
	(0.00809)	(0.00422)	(0.00341)	(0.0376)	(2083.4)
Has second email	0.0500	-0.00527	0.0128	0.169	13804.4
	(0.0220)	(0.0168)	(0.0124)	(0.151)	(13278.5)
Located in San José	0.00844	-0.00224	0.00369	0.0567	8170.2
	(0.00770)	(0.00424)	(0.00343)	(0.0385)	(2331.9)
TPI > 2.5 million CRC	0.0593	0.0384	0.0286	0.297	-1035.9
	(0.0113)	(0.00632)	(0.00538)	(0.0574)	(1846.5)
TPI > 6 million CRC	0.00816	0.0717	0.0230	0.335	28358.5
	(0.0150)	(0.0104)	(0.00824)	(0.0932)	(5297.4)
Made advance payment for 2014	0.143	0.126	0.114	1.303	67205.6
	(0.0315)	(0.0292)	(0.0260)	(0.305)	(27609.5)
Filed in 2013	0.104	-0.000137	0.00274	0.000274	-7209.5
	(0.0104)	(0.00558)	(0.00455)	(0.0509)	(2385.0)
Reported net liability > 0 in 2013	-0.0677	0.117	-0.0155	-0.213	-17826.7
	(0.0250)	(0.0212)	(0.0114)	(0.126)	(4711.1)
Made payment in 2013	0.107	0.0453	0.137	1.679	96446.0
	(0.0305)	(0.0275)	(0.0189)	(0.219)	(16718.6)
Was third-party informant in 2013	0.0504	0.00722	-0.00905	-0.0744	8708.7
	(0.0133)	(0.00832)	(0.00664)	(0.0754)	(4035.9)
Constant	0.0397	-0.0133	-0.00983	-0.136	-12384.2
	(0.00809)	(0.00436)	(0.00352)	(0.0396)	(2780.4)
Control group avg.	0.115	0.0339	0.0172	0.194	5015.8
T1=T2 p-value (Wald test)	0.074	0.189	0.159	0.195	0.514
Observations	12,515	12,515	12,515	12,515	12,515

Note: This table displays estimates of the treatment effect on tax filing and payment, controlling for firm characteristics. It is identical to Table 1.1, Panel A, in the paper, but displays the coefficients on all control variables, as noted in the table rows. TPI stands for third-party information (third-party reported sales), meaning the sum of sales reported by clients (D151), state institutions (D150), and credit/debit card companies (D153). The cutoffs of 2.5 million and 6 million CRC correspond to the priority group designations used by the tax authority.

Table 1.A8: Impact on Income Tax Compliance, Firms without Third-Party Information (Full Results)

	OLS				
	(1) Filed	(2) Positive net liability	(3) Positive payment	(4) Log payment	(5) Payment
T1: Baseline email	0.151	0.00651	0.00475	0.0496	265.1
	(0.00410)	(0.000897)	(0.000773)	(0.00829)	(59.38)
T2: Information email	0.149	0.00587	0.00431	0.0411	168.9
	(0.00405)	(0.000862)	(0.000740)	(0.00754)	(51.54)
Corporation	-0.0129	0.00164	0.00140	0.0167	195.9
	(0.00496)	(0.00113)	(0.000974)	(0.0104)	(81.09)
Has legal representative	-0.000571	0.000727	-0.0000674	0.00297	87.05
	(0.00407)	(0.000811)	(0.000723)	(0.00722)	(43.00)
Has second email	0.0281	0.00245	0.00296	0.0330	196.9
	(0.00893)	(0.00252)	(0.00224)	(0.0246)	(191.9)
Located in San José	-0.00133	-0.000892	-0.00101	-0.0111	-68.24
	(0.00369)	(0.000799)	(0.000693)	(0.00722)	(50.74)
Made advance payment for 2014	0.0976	0.0476	0.0127	0.153	1627.1
	(0.0344)	(0.0203)	(0.0141)	(0.156)	(1335.7)
Filed in 2013	0.132	0.00144	0.000816	0.00563	-22.34
	(0.00652)	(0.00143)	(0.00124)	(0.0130)	(86.75)
Reported net liability > 0 in 2013	-0.00941	0.0528	-0.00400	-0.0477	-564.1
	(0.0299)	(0.0167)	(0.00689)	(0.0779)	(648.3)
Made payment in 2013	0.0435	0.0440	0.0749	0.801	5627.5
	(0.0365)	(0.0228)	(0.0160)	(0.176)	(1459.9)
Was third-party informant in 2013	0.0164	0.00274	0.000828	0.0162	148.8
	(0.0121)	(0.00410)	(0.00328)	(0.0359)	(288.9)
Constant	0.0177	-0.000710	0.000336	0.00134	-51.26
	(0.00363)	(0.000755)	(0.000619)	(0.00647)	(45.24)
Control group avg.	0.0393	0.00178	0.00130	0.0142	96.15
T1=T2 p-value (Wald test)	0.809	0.565	0.649	0.397	0.161
Observations	37,242	37,242	37,242	37,242	37,242

Note: This table displays estimates of the treatment effect on tax filing and payment, controlling for firm characteristics. It is identical to Table 1.1, Panel B, in the paper, but displays the coefficients on all control variables, as noted in the table rows.

Table 1.A9: Impact on Income Tax Compliance, LATE

Subsample:	Corporations				Positive liability in 2013				Paid in 2013			
Outcome:	(1) Positive net liability	(2) Positive payment	(3) Log payment	(4) Payment	(5) Positive net liability	(6) Positive payment	(7) Log payment	(8) Payment	(9) Positive net liability	(10) Positive payment	(11) Log payment	(12) Payment
A: Firms with Third-Party Information												
T1: Baseline email	0.0905	0.0481	0.592	32776.1	0.146	0.0893	1.062	51713.2	0.170	0.124	1.506	76880.7
	(0.0209)	(0.0145)	(0.175)	(14505.2)	(0.0309)	(0.0235)	(0.278)	(22835.9)	(0.0402)	(0.0348)	(0.413)	(36706.4)
T2: Information email	0.0901	0.0707	0.879	48928.3	0.174	0.110	1.308	76168.4	0.238	0.175	2.087	112442.3
	(0.0209)	(0.0154)	(0.185)	(16336.6)	(0.0306)	(0.0235)	(0.281)	(27139.4)	(0.0408)	(0.0360)	(0.431)	(43345.4)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control group avg.	0.113	0.0427	0.498	18322.9	0.176	0.0767	0.894	34273.1	0.168	0.105	1.218	52755.0
T1=T2 p-value (Wald test)	0.987	0.185	0.165	0.373	0.392	0.440	0.430	0.386	0.128	0.212	0.231	0.440
Observations	1,750	1,750	1,750	1,750	1,094	1,094	1,094	1,094	660	660	660	660
B: Firms without Third-Party Information												
T1: Baseline email	0.0105	0.00752	0.0837	611.5	0.108	0.0507	0.518	2711.1	0.134	0.0704	0.708	3236.8
	(0.00217)	(0.00184)	(0.0205)	(171.8)	(0.0268)	(0.0202)	(0.219)	(1728.6)	(0.0350)	(0.0293)	(0.316)	(2508.6)
T2: Information email	0.00754	0.00558	0.0555	312.9	0.0880	0.0520	0.538	3584.3	0.0975	0.0636	0.637	3826.9
	(0.00200)	(0.00169)	(0.0178)	(141.0)	(0.0257)	(0.0207)	(0.222)	(1791.0)	(0.0334)	(0.0305)	(0.326)	(2699.1)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control group avg.	0.00278	0.00185	0.0208	159.2	0.0259	0.0155	0.176	1322.8	0.0305	0.0305	0.348	2871.5
T1=T2 p-value (Wald test)	0.260	0.392	0.249	0.133	0.563	0.961	0.943	0.700	0.407	0.857	0.861	0.858
Observations	9,875	9,875	9,875	9,875	592	592	592	592	398	398	398	398

Note: This table displays OLS estimates of the treatment effect on income tax payment. It reproduces columns 2–5 of Table 1.1 for different subsamples, as noted in the column titles.

Table 1.A10: Impact on Other Compliance Outcomes, LATE

Subsample:	Previously filed sales tax				Small or inactive firm[1]		Likely to file for 2013[2]		Self-Employed	Likely to file informative declaration[3]		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Outcome:	No. months decl. sales tax	No. months paid sales tax	Sales tax payment	Deregistered (sales tax)	Deregistered	Switched to simplified regime	Filed income tax for 2013	Paid income tax for 2013	Paid income tax for 2013	Presented informat-ive declaration	Reported supplier	Reported client
A: Firms with Third-Party Information, Treatments Separated												
T1: Baseline email	-0.0653	-0.111	-51874.0	0.00339	0.0211	0.000612	0.0704	0.00130	0.00234	0.0334	0.00925	0.0384
	(0.278)	(0.163)	(65860.6)	(0.00802)	(0.00452)	(0.000615)	(0.0123)	(0.00423)	(0.00150)	(0.0166)	(0.0130)	(0.0149)
T2: Information email	-0.324	-0.171	-41774.1	-0.00322	0.0164	0.000580	0.0627	0.00656	0.00271	0.0922	0.0196	0.0866
	(0.278)	(0.158)	(70312.8)	(0.00719)	(0.00408)	(0.000578)	(0.0122)	(0.00508)	(0.00153)	(0.0210)	(0.0147)	(0.0190)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Control group avg.	5.567	1.161	257583.4	0.0129	0.00613	0	0.0298	0.00777	0.0409	0.0439	0.0317	0.0293
T1=T2 p-value (Wald test)	0.370	0.706	0.873	0.413	0.385	0.970	0.613	0.304	0.831	0.009	0.499	0.022
Observations	1,385	1,385	1,385	1,385	5,016	5,016	2,325	2,325	10,765	1,204	1,204	1,204
B: Firms with Third-Party Information, Treatments Pooled												
T: Any email	-0.192	-0.140	-46944.7	0.000166	0.0187	0.000596	0.0667	0.00386	0.00252	0.0616	0.0142	0.0616
	(0.238)	(0.139)	(60281.3)	(0.00647)	(0.00334)	(0.000422)	(0.00967)	(0.00390)	(0.00124)	(0.0151)	(0.0115)	(0.0134)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Control group avg.	5.567	1.161	257583.4	0.0129	0.00613	0	0.0298	0.00777	0.0409	0.0439	0.0317	0.0293
Observations	1,385	1,385	1,385	1,385	5,016	5,016	2,325	2,325	10,765	1,204	1,204	1,204
C: Firms without Third-Party Information												
T: Any email	0.0737	0.0578	173.5	0.00696	0.0203	0.000429	0.0763	0.00192	0.000357	-0.00203	0.00478	-0.00682
	(0.117)	(0.0325)	(699.5)	(0.00467)	(0.00160)	(0.000223)	(0.00415)	(0.000512)	(0.000275)	(0.0192)	(0.0180)	(0.00683)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Control group avg.	2.382	0.146	3249.5	0.0169	0.0109	0.000197	0.0173	0.000270	0.00834	0.0377	0.0314	0.00629
Observations	3,794	3,794	3,794	3,794	30,797	30,797	11,090	11,090	27,367	465	465	465

Note: This table displays OLS estimates of the treatment effect on a variety of compliance outcomes. It reproduces Tables 1.2 and 1.3 for different subsamples, as noted in the column titles.

[1] A firm is considered small or inactive if it had third-party reported sales less than CRC 2.5 million and it did not file an income tax declaration for 2013.

[2] A firm is considered likely to file for 2013 if it filed an income or sales tax declaration at any time in the three years before 2013 or it made an advance payment but did not file for 2013.

[3] A firm is considered likely to file an informative declaration if it filed one in a previous year, it was the subject of a firm-to-firm informative declaration by the start of the experiment, and it had not yet filed one for the current year. (The reason that there are some observations who satisfy this condition for Panel C is that although firms were separated into two groups—those with and without TPI—for this experiment, a modest number of non-TPI firms (20%) were the subjects of firm-to-firm declarations after this separation. The fraction of non-TPI firms that were covered by a firm-to-firm informative declaration by the start of the experiment is balanced across treatment and control groups.)

Table 1.A11: Medium-Term Impact on Compliance Outcomes, LATE

	Did not file income tax in 2013	No liability or payment in 2013	Did not file informative declaration in 2013			
	(1)	(2)	(3)	(4)	(5)	(6)
	Filed income tax	Paid income tax	Reported client	Reported supplier	No. of clients reported	No. of suppliers reported
	A: Firms with Third-Party Information					
T effect in 2015: Any 2014 email	0.095	0.011	0.023	0.007	0.023	0.006
	(0.011)	(0.004)	(0.006)	(0.005)	(0.006)	(0.004)
Control group avg. 2015	0.256	0.035	0.091	0.059	0.078	0.045
T effect in 2016: Any 2014 email	0.027	0.015	0.003	0.000	0.005	0.000
	(0.011)	(0.004)	(0.007)	(0.006)	(0.007)	(0.005)
Control group avg. 2016	0.304	0.043	0.126	0.079	0.109	0.060
Observations	8,267	11,417	10,210	10,210	10,210	10,210
	B: Firms without Third-Party Information					
T effect in 2015: Any 2014 email	0.069	0.001	0.005	0.003	0.004	0.002
	(0.004)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Control group avg. 2015	0.115	0.009	0.021	0.013	0.019	0.012
T effect in 2016: Any 2014 email	0.018	0.002	0.002	0.002	0.001	0.001
	(0.005)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
Control group avg. 2016	0.151	0.016	0.037	0.023	0.033	0.019
Observations	30,825	36,646	35,565	35,565	35,565	35,565

Note: This table displays OLS estimates of the treatment effect on tax compliance in the medium-term. It reproduces Table 1.5 for different subsamples, as noted in the column titles.

Table 1.A12: Medium-Term Impact on Third-Party Report Coverage, LATE

	Previously reported by third party[1]			Below median TPI			Unlikely to file informative declaration[2]		
	(1) Reported by state institution	(2) Reported by private client or supplier	(3) Reported by card company	(4) Reported by state institution	(5) Reported by private client or supplier	(6) Reported by card company	(7) Reported by state institution	(8) Reported by private client or supplier	(9) Reported by card company
	Firms with Third-Party Information								
T effect in 2015: Any 2014 email	0.001	-0.005	0.010	-0.004	-0.017	0.006	-0.002	-0.007	0.007
	(0.030)	(0.009)	(0.034)	(0.004)	(0.013)	(0.008)	(0.004)	(0.011)	(0.007)
Control group avg. 2015	0.308	0.697	0.549	0.020	0.631	0.092	0.033	0.654	0.091
T effect in 2016: Any 2014 email		-0.013			-0.020			-0.013	
		(0.010)			(0.013)			(0.012)	
Control group avg. 2016		0.618			0.551			0.563	
Observations	1,279	12,298	1,104	6,689	6,689	6,689	8,267	8,267	8,267

Note: This table displays OLS estimates of the treatment effect on tax compliance in the medium-term. It reproduces Table 1.6 for different subsamples, as noted in the column titles.

[1]For columns 1–3, each regression is limited to firms that were the subject of the relevant third party report, i.e., firms that were previously reported by a state institution for column 1, by a private client or supplier for column 2, and by a card company for column 3.

[2]A firm is considered unlikely to file an informative declaration if it did not file one (before the experiment began) for any year between 2010–2014.

Table 1.A13: Heterogeneity of Treatment Effects by Firm Characteristics

	Outcome: Declared Income Tax for 2014					Outcome: Paid Income Tax for 2014				
	(1) Corporation	(2) HQ in San José	(3) Filed in 2013	(4) Paid in 2013	(5) Declared sales tax	(6) Corporation	(7) HQ in San José	(8) Filed in 2013	(9) Paid in 2013	(10) Declared sales tax
	A: Firms with Third-Party Information									
Any message	0.230	0.222	0.205	0.216	0.224	0.030	0.035	0.028	0.027	0.032
	(0.007)	(0.010)	(0.008)	(0.007)	(0.007)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Characteristic	0.036	0.012	0.073	0.047	0.039	-0.014	0.005	-0.010	0.052	-0.038
	(0.020)	(0.010)	(0.013)	(0.038)	(0.026)	(0.010)	(0.004)	(0.006)	(0.023)	(0.009)
Any message × characteristic	-0.066	-0.004	0.045	0.091	-0.060	0.031	-0.002	0.019	0.126	0.030
	(0.023)	(0.014)	(0.016)	(0.038)	(0.033)	(0.012)	(0.006)	(0.007)	(0.029)	(0.014)
Characteristic avg.	0.140	0.489	0.330	0.0527	0.0637	0.140	0.489	0.330	0.0527	0.0637
Observations	12,515	12,515	12,515	12,515	12,515	12,515	12,515	12,515	12,515	12,515
	B: Firms without Third-Party Information									
Any message	0.147	0.143	0.129	0.148	0.148	0.004	0.005	0.004	0.004	0.004
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Characteristic	-0.020	-0.011	0.050	-0.062	0.000	-0.000	-0.001	-0.001	0.030	0.001
	(0.004)	(0.004)	(0.007)	(0.039)	(0.017)	(0.001)	(0.001)	(0.002)	(0.018)	(0.003)
Any message × characteristic	0.010	0.015	0.124	0.156	0.070	0.003	-0.000	0.003	0.067	0.007
	(0.007)	(0.006)	(0.010)	(0.040)	(0.026)	(0.001)	(0.001)	(0.002)	(0.025)	(0.006)
Characteristic avg.	0.265	0.491	0.168	0.0107	0.0251	0.265	0.491	0.168	0.0107	0.0251
Observations	37,242	37,242	37,242	37,242	37,242	37,242	37,242	37,242	37,242	37,242

Note: This table displays OLS estimates of the treatment effect on income tax filing and payment (measured 15 weeks after the experiment start), with an interaction between the treatment indicator and different characteristics, as indicated by the column titles. The rows display the coefficients on the treatment indicator (pooling the two treatment groups), the coefficient on the characteristic, and the interaction between the treatment and the characteristic. Robust standard errors clustered by email address are in parentheses. The regressions use the same controls as in Table 1.A7.

Table 1.A14: Heterogeneity of Treatment Effects by Third-Party Information

	Outcome: Filed Income Tax for 2014				Outcome: Paid Income Tax for 2014			
	(1) Above median TPI	(2) TPI > 2.5 million CRC	(3) TPI > 6 million CRC	(4) No. of D151 reports by clients	(5) Above median TPI	(6) TPI > 2.5 million CRC	(7) TPI > 6 million CRC	(8) No. of D151 reports by clients
Any message	0.247	0.247	0.241	0.146	0.025	0.024	0.031	0.018
	(0.009)	(0.008)	(0.007)	(0.016)	(0.003)	(0.003)	(0.003)	(0.006)
Characteristic	0.047	0.111	0.083	-0.020	-0.011	0.009	0.010	-0.002
	(0.012)	(0.014)	(0.020)	(0.013)	(0.005)	(0.006)	(0.010)	(0.005)
Any message × characteristic	-0.053	-0.079	-0.112	0.096	0.018	0.029	0.018	0.016
	(0.014)	(0.016)	(0.022)	(0.018)	(0.006)	(0.008)	(0.012)	(0.007)
Characteristic avg.	0.500	0.338	0.183	1.443	0.500	0.338	0.183	1.443
Observations	12,515	12,515	12,515	12,515	12,515	12,515	12,515	12,515

Note: This table displays OLS estimates of the treatment effect on income tax filing and payment (measured 15 weeks after the experiment start), with an interaction between the treatment indicator and different characteristics, as indicated by the column titles. The rows display the coefficients on the treatment indicator (pooling the two treatment groups), the coefficient on the characteristic, and the interaction between the treatment and the characteristic. Robust standard errors clustered by email address are in parentheses. The regressions use the same controls as in Table 1.A7, and use only the sample of firms with third-party information. TPI stands for third-party information (third-party reported sales), meaning the sum of sales reported by clients (D151), state institutions (D150), and credit/debit card companies (D153). The cutoffs of 2.5 million and 6 million CRC correspond to the priority group designations used by the tax authority.

Table 1.A15: Response to 2016 Income Tax Intervention

	Outcome: Filed Income Tax After April 2016 Intervention					
	(1)	(2)	(3)	(4)	(5)	(6)
Any message	-0.0360	-0.0567	-0.0607	-0.0358	-0.0366	-0.0622
	(0.00515)	(0.00506)	(0.00504)	(0.00515)	(0.00517)	(0.00506)
Filed Income Tax for 2014		0.214	0.158			0.167
		(0.0107)	(0.0276)			(0.0307)
Any message × Filed			0.0667			0.0638
			(0.0297)			(0.0328)
Follow-up				0.0756	0.0546	0.0336
				(0.0229)	(0.0370)	(0.0427)
Any message × Follow-up					0.0332	0.0865
					(0.0438)	(0.0520)
Any message × Filed × Follow-up						-0.109
						(0.0944)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,405	19,405	19,405	19,405	19,405	19,405

Note: Extending the discussion in subsection 1.5.2, this table presents an OLS regression analysis of taxpayers' responsiveness to repeated enforcement interventions. The regressions focus on taxpayers in our experiment that become nonfilers again for the income tax in 2015 and received enforcement messages reminding them to file in April 2016. The regressions allow us to control for firm characteristics and examine the role of the follow-up activities in July 2015, which were intended to ensure the credibility of our experimental intervention. We regress an indicator for whether a taxpayer filed income tax for 2015 in response to receiving the enforcement email in April 2016 on indicators for our experimental treatment, whether the firm ultimately declared income tax for 2014, whether it was subject to a follow-up, and interactions of these indicators, controlling for firm characteristics as in our main results tables. We do not include the interaction between filing for 2014 and follow-up, because almost all firms that were subject to follow-up ultimately filed their declaration for 2014.

Interpretation of results: Given the selection of the sample, having been treated in our experiment is negatively associated with responsiveness to the new intervention (column 1). However, this is true only for firms that did not file in response to the treatment. Firms that filed for 2014, both treated and non-treated, are actually more likely to respond to the new intervention than firms previously in the control group that did not end up filing (columns 2–3). In addition, the follow-up intervention "neutralises" any potential perverse effect of our treatment. Firms that were subject to a follow-up are generally more responsive to the new intervention (column 4). Among 2014 nonfilers, those that were treated and followed up on are not less likely to respond to the new intervention than firms that were not treated at all in the experiment; and filers that were treated and followed up on are marginally more responsive (column 6).

Table 1.A16: Impact on Reporting Behaviour of Tax Filers

	(1) Log Sales	(2) Log costs	(3) Log profits	(4) Log tax liability	(5) Log payment	(6) Positive profits	(7) Positive tax liability	(8) Positive payment
				A: Corporations				
T: Any email	0.815	0.321	0.116	0.141	0.619	0.0153	0.0190	0.0421
	(0.542)	(0.542)	(0.677)	(0.581)	(0.478)	(0.0454)	(0.0451)	(0.0396)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control group avg.	15.61	15.73	8.458	7.020	2.232	0.573	0.550	0.191
Observations	584	584	584	584	584	584	584	584
				B: Self-Employed				
T1: Baseline email	0.256	-0.504	0.514	0.565	0.437	0.0359	0.0512	0.0391
	(0.318)	(0.386)	(0.341)	(0.255)	(0.222)	(0.0240)	(0.0224)	(0.0199)
T2: Information email	0.845	-0.387	1.102	0.633	0.450	0.0837	0.0585	0.0432
	(0.308)	(0.383)	(0.336)	(0.248)	(0.221)	(0.0236)	(0.0219)	(0.0199)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control group avg.	13.12	10.44	11.48	2.669	1.484	0.808	0.236	0.134
T1=T2 p-value (Wald test)	0.001	0.652	0.004	0.677	0.929	0.001	0.626	0.762
Observations	2,692	2,692	2,692	2,692	2,692	2,692	2,692	2,692

Note: This table displays OLS estimates of the treatment effect on reporting behaviour conditional on filing, for firms with third-party information. The outcome variables, mentioned in the column titles, correspond to line items on the 2014 income tax declarations. The table considers only firms that filed an income tax declaration, and it uses the same controls as in Table 1.A7. For Panel A, the treatment groups are pooled into one binary treatment variable, as there is not a significant difference between the two treatment effects. Robust standard errors clustered by email address are in parentheses.

Chapter 2

The Wage Effects of Liability for Employment Discrimination

2.1 Introduction

What is the effect of liability for employment discrimination? Laws that prohibit discrimination are meant to improve employment outcomes for protected individuals, but do these laws have their intended effect? We know surprisingly little about the answer to this question. In the United States, liability for employment discrimination has expanded at the state and local level, and there are proposals to amend the Civil Rights Act of 1964 to include sexual orientation and gender identity as protected classes.[1] We might want to know: how would this affect protected workers? Would it increase their relative earnings and, if so, by how much?

One thing is clear: labour market outcomes continue to differ by social group in the United States. Consider race: according to the Bureau of Labor Statistics, in 2017, median weekly earnings of black men working full-time were approximately three-quarters those of white men. Some of the difference in wages between white and black workers in the United States can be explained by factors such as education, skills, or occupation, but recent evidence suggests that racial prejudice and discrimination con-

[1] In fact, recent rulings by federal courts and the Equal Employment Opportunity Commission (EEOC) extended protections established by the Civil Rights Act of 1964, which prohibits discrimination "because of sex," to transgender individuals.

tinue to play a role. Fryer, Pager, and Spenkuch (2013) argue that market discrimination explains at least one-third of the black-white wage gap, and experimental evidence demonstrates that many employers harbour racial prejudice (Bertrand & Mullainathan 2004; Pager, Western & Bonikowski 2009).

What can be done? Investments in education and skills could improve labour market outcomes for minority workers, but this will not eliminate prejudice or discrimination by employers, at least not in the short run. If minority Americans earn lower wages partly due to unfair treatment, is there scope for additional liability for employment discrimination?

In a standard competitive model of the labour market, the net theoretical effect of a law that prohibits discrimination in hiring, firing, and pay is ambiguous (Acemoglu & Angrist 2001). As I show in this essay, the same is true in a model of equilibrium search and matching with costly search—the net effect depends on the relative size of hiring and firing costs introduced by the law, as well as the strength of its equal pay provision. The lack of consensus regarding the likely effect of liability for employment discrimination demonstrates the importance of empirical work on the topic.

This essay proposes an empirical strategy to estimate the effect of liability for employment discrimination on relative wages using a special feature of federal anti-discrimination law in the United States. Title VII of the Civil Rights Act of 1964 prohibits discrimination in employment on the basis of race, religion, sex, or national origin. It applies to firms with 15 or more employees, and it is enforced by the Equal Employment Opportunity Commission (EEOC). The Civil Rights Act of 1991 amended Title VII to provide explicit legal remedies for victims of intentional discrimination. Under the law, plaintiffs in discrimination lawsuits can recover punitive and emotional damages in court, but these damages are subject to limits that vary discontinuously according to firm size. Using a regression discontinuity (RD) design, this essay provides preliminary estimates of the effect of Title VII coverage on black relative wages in the

United States, and it also considers the effect of discontinuities in Title VII liability (that is, maximum damages awardable in cases of intentional discrimination) established by the Civil Rights Act of 1991.

A regression discontinuity design is invalid in this context if workers sort into firms covered by the law or firms manipulate their employment in order to avoid coverage. Recent evidence from France suggests that size-dependent regulations do affect the distribution of firm employment (Garicano, Lelarge & Van Reenen 2016; Gourio & Roys 2014). Using administrative data on U.S. businesses, this essay fits a power law to the distribution of firm employment and tests whether firms bunch just below thresholds in coverage. I find no evidence that firms alter their employment in response to federal antidiscrimination law in the United States, despite heaps near treatment thresholds in the data set used, which can be attributed to rounding on the part of survey respondents. This essay proceeds with a sharp regression discontinuity design, adjusting for heaping in the assignment variable, which is valid as long as workers cannot *precisely* sort into covered firms.

A noisy estimate is that Title VII coverage increases relative wages for black men by 12 to 16 percent, but we cannot reject the hypothesis that the effect is zero. This is consistent across several specifications that account for measurement error in the assignment variable and non-random data heaping. A point estimate of the effect on wages for black women relative to white women is smaller, likely because white women also receive some protections under the law. There is no evidence that higher Title VII liability, that is, higher damages allowable in discrimination lawsuits for larger firms, affects black relative wages. However, the data around upper thresholds are especially noisy.

Overall, this evidence suggests that Title VII coverage may increase wages for black men relative to white men, but higher liability in Title VII discrimination lawsuits likely does not. This makes sense if the expected cost of discrimination increases

substantially once a firm is covered under Title VII (e.g., in terms of EEOC investigations or expected legal fees from litigation) but higher damages awardable in court do not add much to this cost. Future work will aim to use confidential matched employer-employee data held by the U.S. Census Bureau to obtain a more precise measure of firm employment and expand the sample size, which would improve the precision of these estimates and allow for interactions with liability imposed by states.

2.2 Background

This section begins by describing several economic models of employment discrimination, including discrimination based on animus, or "taste-based" discrimination (Becker 1957); discrimination in the context of costly search (Black 1995; Lang, Manove & Dickens 2005); and statistical discrimination (Phelps 1972; Arrow 1973; Aigner & Cain 1977). The focus is on discrimination by employers, but there are also models of customer discrimination (Borjas & Bronars 1989) and discrimination by employees (Chiswick 1973). Next, this section reviews evidence of employment discrimination. Wage regressions, field and quasi-experiments, and estimates from structural models provide considerable evidence of employment discrimination on the basis of race and gender, and some lend support to specific theoretical models of discrimination. Finally, this section describes liability for employment discrimination in the United States, paying particular attention to the Civil Rights Act of 1991 and its amendments to Title VII of the Civil Rights Act of 1964.

2.2.1 Economic Models of Employment Discrimination

Economic models of employment discrimination began with the seminal work of Becker (1957). The Becker model of employer tastes for discrimination was applied and extended several times, notably in settings with costly search (Black 1995; Lang, Manove

& Dickens 2005). Phelps (1972) and Arrow (1973) introduced models of statistical discrimination, which were also embedded into search and matching models of the labour market. For a more detailed overview of theoretical models of discrimination, see Lang and Lehmann (2012).

2.2.1.1 Employer Animus

Becker (1957) models discrimination based on employer animus. Suppose there are two groups of individuals, a majority group and a minority group. (Much of this section refers to a majority and minority group for convenience, but these could be any social groups. They need not constitute a majority or minority of the population.) Further suppose that the two groups have identical productive characteristics. In this case, any difference in wages must result from differential treatment, not pre-market characteristics such as education or skills. Suppose the labour market is perfectly competitive, so that employers face constant wage rates for the two groups, but suppose prejudiced employers incur disutility from hiring minority workers. These prejudiced employers have discriminatory tastes.

If employers maximise profits less the monetary cost of disutility from hiring minority workers, an employer who is prejudiced will view minority workers as if their wage is the market wage plus this cost. That is, minority workers must "compensate" prejudiced employers by accepting a lower wage rate for any given level of productivity. Workers will segregate into firms based on the discriminatory tastes of owners. If there are enough prejudiced employers, so that demand for minority workers when their wages equal majority workers' is below supply, a wage gap emerges. If, on the other hand, the number of prejudiced employers is sufficiently small, there is no wage gap, even though some prejudiced employers exist. That is, discrimination on average does not imply discrimination at the margin.

With free entry or constant returns to scale, discrimination cannot exist in equi-

librium in this model. In the long run, competitive markets eliminate firms with discriminatory tastes. Majority and minority workers are perfect substitutes, yet when the majority wage rate exceeds the minority rate, employers with discriminatory tastes are equally willing to hire them. These employers hire too few minority workers (up to the point where the value marginal product of labour equals the wage plus the disutility value).

2.2.1.2 Equilibrium Search and Matching

Search costs can generate a long-run equilibrium in which minority workers receive lower wages, even when there are a large number of non-discriminating firms, as shown by Black (1995). Imagine discrimination in an equilibrium search model where a fraction of firms refuses to hire minority workers. A worker samples a firm at a cost, the firm makes a wage offer, and the worker either accepts or samples another firm. Matches generate identical productivity for firms, but workers get a match-specific utility. Firms have some monopsony power, which allows them to pay workers less than their marginal product.

In equilibrium, unprejudiced firms hire minority workers, but they offer these workers lower wages. This is true as long as there exists at least one prejudiced firm. The presence of prejudiced firms increases minority workers' search costs, lowering their reservation utility and therefore the wages offered to them by "unprejudiced" but monopsonistic employers. This is a departure from the Becker model, which assumes a frictionless labour market. With costly search, the existence of just one prejudiced employer guarantees minority wages will be lower in equilibrium, even if minority workers are equally productive. Entry of unprejudiced employers reduces but does not eliminate discrimination.

Lang, Manove, and Dickens (2005) developed a model of discrimination with costly search and wage posting. If posted wage offers are binding and cannot be condi-

tioned on group status, employers cannot pay minority workers less, thereby ruling out the market antidiscrimination mechanism in Becker (1957). When majority and minority workers apply for the same job, a discriminating employer will choose a majority worker, which discourages minority workers from applying to jobs with high wage offers. Workers segregate, and minority workers earn less in equilibrium, even if prejudicial attitudes or productivity differences are small.

2.2.1.3 Statistical Discrimination

Given the conclusions of the Becker model, it is not surprising that others sought a different model to explain discrimination in the labour market. Arrow (1972) commented that the Becker model "predicts the absence of the phenomenon it was designed to explain." Models of statistical discrimination began with Phelps (1972) and Arrow (1973). In these models, employers observe an imperfect signal of worker productivity and offer a wage based on expected productivity given the signal and given the group membership of the worker. If groups have different average levels of productivity, workers with equal signals but different group memberships will receive different wage offers. This is true even though the productivity signal is unbiased. Note that this does not result from animus on the part of the employer, but rather the employer's optimal response to a signal extraction problem. Also, although some workers face discrimination in the sense that they earn less than their marginal product, others from the same group earn more than their marginal product, so that expected productivity equals true productivity, on average. The key insight of this model is that two workers with the same observed characteristics could earn different wages depending on their group status.

Aigner and Cain (1977) extended this model to consider cases where signal-to-noise ratios differ between groups. Even if the groups' productivities are identically distributed, they will earn different wages given the same signal when employers are risk-

averse. As with differences in mean productivity, this hurts some group members but helps others within the same group. Average wages equal average group productivities. Oettinger (1996) extended this work to a dynamic setting with two periods. In that case, workers with noisy signals are worse off in the second period since workers with precise signals are more able to switch firms and earn higher wages from a better match. This dynamic model predicts no wage gap at labour force entry but one that emerges over time. However, Altonji and Blank (1999) question the usefulness of models that assume some difference in signal quality without empirical evidence showing that these signals differ in practice.

2.2.1.4 Other Models

Becker's taste-based model of discrimination and models of statistical discrimination have been influential, but other models exist. For example, Goldberg (1982) reformulated the Becker model in terms of nepotism, that is, favouritism toward members of your own social group, rather than animus toward others. Goldberg's model predicts that discriminating employers and wage gaps survive in the long run. Akerlof (1985) showed that discrimination can persist in a model of traders with discriminatory social customs in which it is costly to lose trading partners. Similarly, McAdams (1995) developed a model of "status production," in which social groups discriminate against others and individual deviators lose intra-group status. Other models incorporate sorting (Epstein 1992) or occupational exclusion (Johnson & Stafford 1998). More recently, Fryer and Jackson (2008) showed that optimal categorization in cognition can generate bias that resembles statistical discrimination toward a minority group.

2.2.2 Evidence of Employment Discrimination

There is considerable evidence of employment discrimination. Some of this evidence is reviewed here, in part because the concepts and methods used to measure

discrimination are important to the analysis of liability for employment discrimination in this essay, and in part because the existence of discrimination motivates the question this essay tries to answer.

This essay focuses on differences in wages between black and white individuals, as does much of the literature. Darity and Mason (1998), Altonji and Blank (1999), and Lang and Lehmann (2012) offer thorough reviews of the empirical findings in this literature. Charles and Guryan (2011) describe ongoing challenges, such as data availability and racial classifications. Each of these reviews discusses implications for theoretical models of discrimination, some of which are mentioned here.

2.2.2.1 Wage Regressions

A substantial amount of empirical work relies on Mincer earnings regressions to estimate discrimination in the labour market. It is well-established that in the United States, black workers earn less than white workers, but how much of this difference can be attributed to discrimination, rather than pre-market characteristics such as education, skills, or ability? The answer, for some, is to condition on observables and conclude that the residual wage gap represents discrimination. This approach risks omitted variable bias, but it is popular nonetheless. If unobservables vary systematically by race, the approach is incorrect.

Wage regressions of this kind offer different estimates of the extent of discrimination. Early work suggested as much as 50–60 percent of the raw wage gap between black and white men was "unexplained" by observable characters and could be due to discrimination (Corcoran & Duncan 1979; Reimers 1983).

In an influential paper, Neal and Johnson (1996) used data from the National Longitudinal Survey of Youth 1979 (NLSY79) to show that racial wage gaps almost entirely disappear (or reverse, in the case of women) after controlling for performance on a cognitive test taken before labour market entry, called the Armed Forces Qualification

Test (AFQT), which they argued is a racially unbiased measure of skill. They concluded that differences in earnings can be accounted for by differences in pre-market skills, not discrimination. Carneiro, Heckman, and Masterov (2005) argued further that cognitive test scores primarily reflect schooling, so efforts to bridge education and skills gaps are more likely to promote racial equality than additional antidiscrimination efforts or affirmative action.

Lang and Manove (2011) challenged these findings. They showed that conditional on AFQT, black individuals get more education than whites, which they claim is consistent with a model of statistical discrimination and educational sorting, since the existence of statistical discrimination encourages black workers to signal ability through education (more so than whites). They argued that one should control for education in addition to AFQT when comparing earnings for black and white workers, in which case, a large wage gap re-emerges. In the absence of discrimination, controlling for AFQT only should result in higher wages for blacks since conditional on AFQT, they are more highly educated, and a market without discrimination would reward this education.

Fryer, Pager, and Spenkuch (2013) used data on wage offers (accepted and unaccepted) and past wages of unemployed workers in New Jersey to establish a lower bound on the effect of discrimination in the labour market, which they estimate to be one-third of the black-white wage gap. Their estimate is a lower bound if we assume that conditional on previous wages (which reflect the market valuation of cognitive and non-cognitive skills and other characteristics related to marginal product), black individuals are at least as productive as whites. If wage differentials due to discrimination accumulate with experience, then their lower bound does not bind, that is, differential treatment by race explains even more of the gap.

Fryer, Pager, and Spenkuch (2013) made several other empirical findings, which they argue are consistent with their model of statistical discrimination in the context of search and matching in the labour market. They found that black workers have lower

reservation wages than whites (consistent with other studies), as the labour market provides them with less insurance, and within firms, black workers have higher returns to tenure. The latter is consistent with a model of statistical discrimination and employer learning since, conditional on productivity, black workers on average receive lower wage offers than whites, and firms pay their workers according to their marginal product, so within a firm, conditional on productivity, black individuals should have higher returns to tenure. The black-white wage gap could increase with age or experience across firms, however, since conditional on productivity, in the case of a separation, black workers have lower reservation wages than whites.

Much of the empirical work in this area focuses on the black-white wage gap for men, but a substantial gap exists for women, once you account for black-white differences in female labour force participation (Neal 2004). And there is of course a large body of empirical work on differences in pay between men and women, in general (see Blau and Kahn (2000) for a review).

2.2.2.2 Audit Studies

A growing body of work uses field and quasi-experiments to estimate the extent of employment discrimination. These studies offer convincing evidence of race and gender discrimination in hiring.

Bertrand and Mullainathan (2004) randomly assigned "black-sounding" and "white-sounding" names to identical fictitious resumes, which they sent to help-wanted advertisers. Job callback rates substantially differed for the two groups, with white names receiving 50 percent more callbacks. Pager, Western, and Bonikowski (2009) used a similar approach and found that white felons fared about as well as non-criminal black and Hispanic applicants. However, their study also involved in-person interviews with actors who might have consciously or unconsciously affected the probability of receiving a job offer by behaving in a way in order to achieve the outcome they expected.

Also, these studies do not reveal market discrimination in the Becker sense, since they describe the average, not marginal, employer.

Audit studies and quasi-experiments point to discrimination based on gender, too. Riach and Rich (2006) found occupation-specific gender discrimination in hiring using matched pairs of written applications, and Goldin and Rouse (2001) uncovered gender bias in paid orchestras using blind and non-blind auditions.

2.2.2.3 Structural Models

Several authors use structural models to estimate wage discrimination on the basis of race. Eckstein and Wolpin (1999) developed and estimated a search-matching model that accounts for the fact that the distribution of unobserved wage offers may not correspond to the distribution of accepted wages. They found that racial differences in unobserved wage offers exceeded differences in accepted wages, and they used their model to estimate an upper bound on the extent to which discrimination explains the gap in wage offers. They found that discrimination could explain the entire gap. Bowlus and Eckstein (2002) used an equilibrium search model to account for differences in unobserved productivity. They found that the productivity level of black workers is only 3.3 percent lower than that of whites, but slightly more than half of firms have some disutility value to hiring black employees, and this value is nearly one-third of the white productivity level. A drawback of structural models such as these is that they require restrictive assumptions.

Wage regressions, audit studies, and structural models provide considerable evidence of employment discrimination, particularly with respect to race. Some of this work lends support to specific theories of discrimination. For example, Fryer, Pager, and Spenkuch (2013) uncovered empirical evidence consistent with their model of statistical discrimination in the context of search and matching in the labour market, as

described above. On the other hand, Altonji and Pierret (2001) showed that if firms statistically discriminate, the relationship between wages and easily observable characteristics, such as education or race, should fall over time with the same employer, whereas the relationship between wages and hard-to-observe characteristics that may be related to productivity, such as AFQT score or father's education, should rise, as the employer learns about worker productivity. They found evidence of this phenomenon for education but not for race.

Charles and Guryan (2008) offered empirical evidence in favour of Becker's theory of discrimination. Recall from the Becker model that it is the discriminatory tastes of the marginal, not average, employer that determine the extent of market discrimination. Using questions about racial prejudice from the General Social Survey (GSS) and data on wages from the Current Population Survey (CPS), they showed that racial differences in wages in the United States are closely related to the prejudice of the "marginal" person in a state, defined as the pth percentile of the racial prejudice distribution of the state's white population, where p is the fraction of the state population that is black. This corroborates Becker's theory of discrimination, as do their findings that the share of black members of the workforce is negatively correlated with the racial wage gap and that the left tail of the racial prejudice distribution matters more for the wage gap than the right tail. They estimated that discrimination explains roughly one-quarter of the black-white wage gap.

2.2.3 Liability for Employment Discrimination in the United States

In the United States, a number of laws penalise employment discrimination on the basis of race, gender, and other social identities. These laws operate at the federal, state, and local levels, and they sometimes overlap. This section provides an overview of antidiscrimination laws in the United States, paying particular attention to Title VII

of the Civil Rights Act of 1964, as amended by the Civil Rights Act of 1991, which is at the centre of the analysis in this essay. The description below focuses on provisions related to race and employment.

2.2.3.1 Title VII of the Civil Rights Act of 1964

The Civil Rights Act of 1964 was a major piece of legislation enacted during the civil rights movement in the United States. The Act made it illegal to discriminate on the basis of race, colour, religion, national origin, or sex. In particular, Title VII of the Civil Rights Act of 1964 outlawed discrimination of this kind by private employers.[2] Initially, this prohibition applied to firms with 25 or more employees, but the Equal Employment Opportunity Act of 1972 expanded coverage of Title VII to include firms with 15 to 24 employees.[3]

Title VII prohibits discrimination in hiring, firing, pay, or conditions of employment. It is enforced by the Equal Employment Opportunity Commission (EEOC), which investigates claims of discrimination. The law requires an employee complainant to file a charge of discrimination with the EEOC before they can file a discrimination lawsuit (except in the case of a violation of the Equal Pay Act, described later). The EEOC offers to resolve the dispute using mediation, but if mediation fails, it conducts an investigation. Following an investigation, the EEOC will determine whether there was a violation of the law, and if so, it will try to reach a settlement with the employer. If the EEOC finds no violation of the law, or if its attempt to settle with the employer fails, it will notify the complainant of their right to sue. In the latter case, that is, when the EEOC finds a violation but the parties do not reach a settlement, the EEOC often files a lawsuit in federal court on the complainant's behalf. If the EEOC decides not to litigate, or if its original investigation found no violation of the law, an individual can

[2] 42 U.S.C. § 2000e-2(a)(1).

[3] Specifically, the law covers private employers with "fifteen or more employees for each working day in each of twenty or more calendar weeks in the current or preceding calendar year." 42 U.S.C. § 2000e(b).

sue privately.

Between 1997 and 2017, lawsuits filed by the EEOC with Title VII claims have numbered between 46 and 341 per year, with annual monetary benefits between $15 million and $159 million (EEOC 2017). This does not include private lawsuits brought against employers. The number of charges filed with the EEOC is far greater, with 30,000 per year, on average, for race discrimination alone. Annual reports from the EEOC's Office of General Counsel reveal that lawsuits related to harassment or discharge are more frequent than ones related to pay or terms of employment. However, even the possibility or threat of a lawsuit for unequal pay might affect firms' wage-setting behaviour.

Title VII also bans discriminatory practices by employment agencies and labour unions, and it protects individuals from retaliation if they file a charge of discrimination or engage in a discrimination investigation or lawsuit.

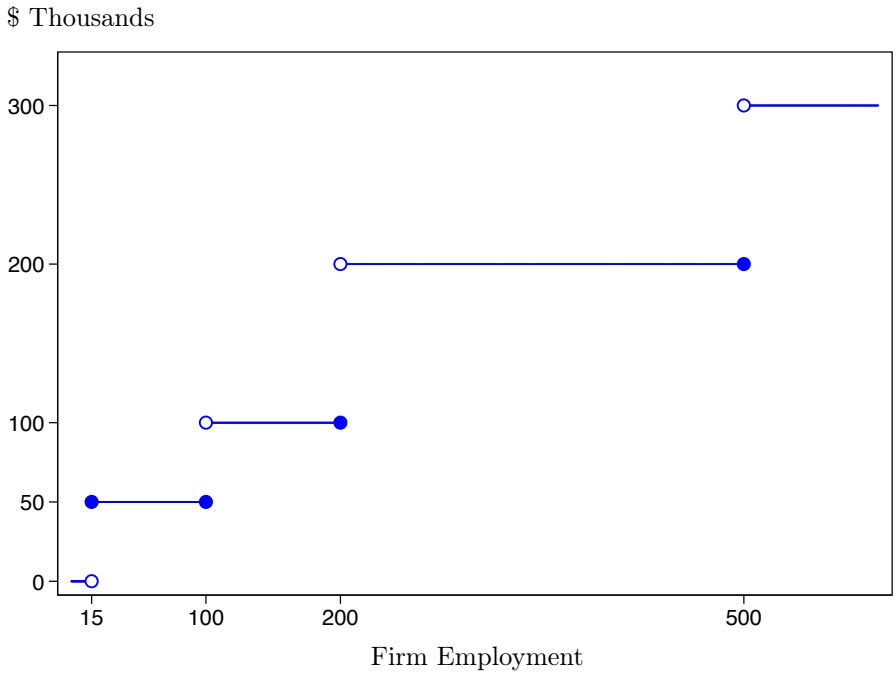
2.2.3.2 The Civil Rights Act of 1991

The Civil Rights Act of 1991 clarified the right of Title VII complainants to seek damages in court, but it established limits on those damages. Following the Civil Rights Act of 1991, plaintiffs could recover punitive and emotional distress damages in cases of intentional discrimination, and they were allowed a trial by jury. Before, they could only seek back pay and reinstatement (where applicable) and a trial by the court.

The Civil Rights Act of 1991 placed limits on punitive and compensatory damages recoverable in Title VII lawsuits, which vary according to the employment size of the firm.[4] Figure 2.1 displays the structure of these limits. Private employers with fewer than 15 employees are not covered under Title VII. This means they cannot be sued for discrimination (under Title VII) in court and they do not answer to the EEOC for charges of Title VII discrimination. Firms with 15 to 100 employees are liable for up to

[4] 42 U.S.C. § 1981(b).

Figure 2.1: Limits on Title VII Damages after the Civil Rights Act of 1991



$50,000 in punitive or compensatory damages if they are found guilty of discrimination under Title VII. This amount increases to $100,000, $200,000, and $300,000 for firms with 101 to 200, 201 to 500, and more than 500 employees, respectively. Thus, the limits on Title VII damages are a step function of firm size, as Figure 2.1 shows, with 15 employees as the cutoff for Title VII coverage and 101, 201, and 501 employees as cutoffs for higher Title VII liability under the Civil Rights Act of 1991.

I say the Civil Rights Act of 1991 "clarified" the right of Title VII plaintiffs to seek damages in court because a series of U.S. Supreme Court cases established and subsequently withdrew this right for victims of race discrimination in employment before passage of the Civil Rights Act of 1991. After the earlier Civil Rights Act of 1964, plaintiffs in discrimination lawsuits could seek remedies such as back pay and reinstatement but not punitive or compensatory damages. They were not allowed trial by jury but only trial by the court. Then, in 1976, the Supreme Court ruled that cases of race discrimination could be brought under Section 1981, a Civil War–era law, and

that damages could be sought without limit.[5] The Court later narrowed its position in *Patterson v. McLean Credit Union* in 1989, when it ruled that damages for some forms of race discrimination in employment were not recoverable under Section 1981, which required only that black persons have the same right to make and enforce contracts as white persons.[6] The Civil Rights Act of 1991 settled the matter, allowing for punitive and compensatory damages in all cases of intentional discrimination under Title VII (not just race discrimination like Section 1981).

2.2.3.3 Other Federal Laws

The Civil Rights Act of 1964 outlawed discrimination in other areas, such as voter registration, public facilities provided by state and local governments, the work of government agencies who receive federal funds, and entry to hotels, restaurants, and theatres. The Act also promoted desegregation of schools and strengthened the U.S. Commission on Civil Rights, which investigates civil rights issues to this day.

A number of other federal laws address the issue of discrimination in the United States. For example, the Equal Pay Act of 1963 made it illegal to pay men and women unequal wages for equal work in the same workplace. This law covers all employers. The Age Discrimination in Employment Act of 1967 prohibits discrimination against workers aged 40 years or older because of their age, but it applies only to employers with 20 or more employees. Finally, the Americans with Disabilities Act of 1990, which covers employers with 15 or more employees, outlaws discrimination based on disability, and it requires covered employers to provide reasonable accommodation to workers with disabilities.

[5] Runyon v. McCrary, 427 U.S. 160 (1976).
[6] 491 U.S. 164 (1989).

2.2.3.4 State Laws

Most states have antidiscrimination laws, and even some municipalities supplement federal and state law with their own antidiscrimination ordinances. State antidiscrimination laws, often referred to as Fair Employment Practice laws, sometimes allow complaining parties to file charges of discrimination with state agencies. Usually they can file these charges with the federal EEOC simultaneously. States have their own procedures for resolving disputes, such as resolution through a state agency, a lawsuit in state court, or both.

To get a sense of how state antidiscrimination laws vary, and to better understand how they might interact with liability for employment discrimination under federal law, consider the current antidiscrimination laws of six large states (which together account for approximately 40 percent of the U.S. population): California, Texas, New York, Florida, Illinois, and Georgia.

The California Fair Employment and Housing Act (FEHA) outlaws employment discrimination on the basis of race, covers all public employers and private employers with 5 or more employees, and requires filing a complaint with the Department of Fair Employment and Housing, which can take administrative action to resolve a dispute. The law allows for lawsuits in state court, where employees can recover unlimited monetary damages, including back pay, emotional distress damages, punitive damages, and attorney's fees and costs. The Texas Labor Code also prohibits racial discrimination in employment, covers all public employers and private employers with 15 or more employees, and allows for lawsuits in state court. However, it limits the sum of compensatory and punitive damages in state court in precisely the same way that federal law does, with all of the same cutoffs and dollar values.

In New York, state law prohibits employment discrimination on the basis of race, covers employers with 4 or more employees, and empowers the New York Division of Human Rights to investigate and resolve cases. It allows for lawsuits in state court,

in which complainants can recover lost income damages and compensatory damages (without limit), but no punitive damages or attorney's fees. In Florida and Illinois, state law prohibiting employment discrimination on the basis of race applies to employers with 15 or more employees, just like federal law, and it places no limits on compensatory damages awardable in state court (except in the case of public employers in Florida). Before 2008, Illinois residents had to resolve discrimination cases through a state agency. They could not file a lawsuit in state court.

Some states have weak or no antidiscrimination provisions at all. For example, Georgia law prohibits employment discrimination on the basis of race by state agencies only. Private and other public employers are not covered, and the state does not have a general antidiscrimination statute. Alabama and Mississippi have no state law prohibiting racial discrimination by private employers.

From these examples, we can see that state antidiscrimination laws vary in terms of coverage, procedure, and remedies such as punitive or compensatory damages. Many of them overlap with (or mimic) federal law. This has implications for measuring the effects of Title VII and the Civil Rights Act of 1991, as explained later.

2.3 The Wage Effects of Liability for Employment Discrimination: Theory

In a standard competitive model of the labour market, the net theoretical effect on wages of liability for discrimination in hiring, firing, and pay is ambiguous (Acemoglu & Angrist 2001). On the one hand, liability for discrimination in hiring might lead firms to hire more minority workers, increasing their wages. That is, the threat of a hiring discrimination lawsuit is effectively an employment subsidy. On the other hand, liability for discrimination in firing might lead firms to hire fewer minority workers, decreasing their wages. That is, the threat of a wrongful termination lawsuit is effectively an employment tax. Finally, the effect of an equal pay provision depends on the strength

of its enforcement, as well as its interaction with hiring and firing costs.

Recently, increasing attention has been paid to labour market monopsony—that is, buyer power among employers. This section shows that in a standard monopsony model, the net theoretical effect of liability for employment discrimination remains ambiguous. Therefore, whether liability for employment discrimination increases the wages of workers it is designed to help, and by how much, is an empirical question.

Consider a model of discrimination in the context of equilibrium search and matching, with costly search, as in Black (1995). In this model, some firms refuse to hire minority workers. Firms that do employ minority workers have monopsony power, and they exploit the reservation wages of minorities, which are lower because search is costly and minorities have fewer firms to choose from. Thus, minority workers earn lower wages in equilibrium as long as there exists at least one prejudiced firm who refuses to hire them.

Suppose workers sample firms at a cost, firms make wage offers, and workers either accept or sample another firm, following Black (1995) and Albrecht and Jovanovic (1986). Firms have monopsony power as a consequence of the cost of search. Suppose there are two types of workers, $i \in B, W$, but matches generate identical productivity for firms, v. Workers get a wage plus a match-specific utility, $z \sim F(z)$. Assume $F(z)$ is strictly log concave (e.g., uniform, normal), so that the inverse hazard, $g(z) = [1 - F(z)]/f(z)$ is strictly decreasing.

Following Acemoglu and Angrist (2001), suppose firms face the threat of administrative or legal action, which creates the following costs:

$$h_i = \text{expected cost of not hiring worker of type } i$$

$$f_i = \text{expected cost of firing worker of type } i$$

Each of these incorporates the probability that a worker sues or that an administrative agency, such as the EEOC, takes action; the expected cost of such action,

including damages and legal fees; and, in the case of firing, the chance that a firm will want to fire a worker. Assume $h_W = 0$ and $f_W = 0$. (Acemoglu and Angrist (2001) allow for the possibility that non-disabled workers claim to be disabled and sue, but this is unlikely in the case of race discrimination.) Further assume that these costs operate as a pure tax, that is, workers do not recoup the cost of administrative or legal action, following Acemoglu and Angrist (2001). This is sensible if enforcement involves payments to third-parties. If, instead, they operate as a transfer from firms to workers, firing costs may not affect employment (Lazear 1990).

The optimal behaviour of the worker is to follow a reservation utility strategy. Let $\bar{u}_i$ be the reservation utility. This is the level of utility such that a worker is indifferent between accepting a job offer or continuing to search. The worker accepts if $w_i + z \geq \bar{u}_i$. Black (1995) shows that minority workers have lower reservation utilities if there exists at least one prejudiced employer that refuses to hire minorities. 'Unprejudiced' but monopsonistic employers exploit these lower reservation utilities and offer equally-productive minority workers lower wages in equilibrium absent employment protection.

Suppose firms make wage offers to maximise expected profits per applicant, this time incorporating the expected cost of firing or not hiring:

$$\max_{w_i} \pi_i = [1 - F(\bar{u}_i - w_i)](v - w_i - f_i + h_i) \tag{2.1}$$

The necessary condition for profit maximization is:

$$v - w_i - f_i + h_i - g(\bar{u}_i - w_i) = 0 \tag{2.2}$$

And the second-order condition

$$-1 + g'(\bar{u}_i - w_i) < 0$$

is satisfied since $g'(\cdot) < 0$. Let $\widehat{w}_i$ be the value of w_i that satisfies (2.2) for $i \in B, W$. (Recall $f_W = 0$ and $h_W = 0$, so we expect $w_B \neq w_W$ even if $\bar{u}_B = \bar{u}_W$, such as the case when there are no discriminating employers.) Then, the wage offered to black workers is:

$$w_B^* = \max\{\widehat{w}_B, \eta\widehat{w}_W\}, \tag{2.3}$$

where $\eta \in [0, 1]$ indicates the strength of the equal pay provision, using notation from Acemoglu and Angrist (2001).

How does w_B^* respond to f_B, h_B and η? From equation 2.2, we have:

$$-1 < \frac{\partial \widehat{w}_B}{\partial f_B} = \frac{1}{g'(\cdot) - 1} < 0$$

and

$$0 < \frac{\partial \widehat{w}_B}{\partial h_B} = \frac{-1}{g'(\cdot) - 1} < 1$$

since $g'(\cdot) < 0$, so that

$$\partial w_B^* / \partial f_B \leq 0,$$

$$\partial w_B^* / \partial h_B \geq 0 \text{ and}$$

$$\partial w_B^* / \eta \geq 0.$$

That is, if the equal pay provision does not bind, an increase in firing costs decreases offered wages, while the threat of a lawsuit from not hiring acts as a hiring subsidy (but the change in wage does not fully offset the change in expected costs from firing or not hiring). The net effect will depend on the relative size of firing and (non-)hiring costs, as well as the strength of the equal pay provision.

If employment protection increases (decreases) the wage offers made to minority workers, it should also increase (decrease) their share of employment at covered firms.

This is roughly the same theoretical prediction as Acemoglu and Angrist (2001), but the mechanism is different: instead of firing costs lowering demand for minority workers, causing the equilibrium wage (taken exogenously by the firm) to fall, here, firing costs enter directly into wage offers made by monopsonistic employers.

In this model, liability for employment discrimination that causes firms that previously refused to employ minority workers to do so would increase the wages of minority workers by lowering their search costs and therefore their reservation wages. This would lead to an across-the-board increase in minority wages, not just at firms covered by the legislation. However, the expected cost of firing and (not) hiring minority workers, together with the equal pay provision—to the extent it binds—should create pay differences based on coverage, unless the provisions exactly offset one another.

2.4 The Wage Effects of Liability for Employment Discrimination: Evidence

If the net theoretical effects of liability for employment discrimination are ambiguous, as the previous section suggests, understanding these effects becomes an important empirical question. This section has three principal goals: to describe existing evidence of the effects of liability for employment discrimination (and to discuss potential shortcomings of this evidence); to propose a new empirical strategy to measure these effects using the structure of liability for employment discrimination following the Civil Rights Act of 1991; and to present preliminary results, pending a proposal to the U.S. Census Bureau to access confidential data.

2.4.1 Existing Evidence

A number of empirical studies have aimed to estimate the effect of antidiscrimination laws in the United States. For an overview, see Donohue (2007). A significant

challenge is to isolate the effect of laws that have broad coverage. Have antidiscrimination laws improved economic outcomes for black Americans since the 1960s and, if so, by how much? To be sure, a number of factors have contributed to black economic progress, including migration and changes in the overall economy. But what has been the specific effect of liability for discrimination?

Freeman (1973) made an early attempt to measure the effect of federal antidiscrimination law using time series methods. He found that EEOC expenditures were closely related to improvements in the relative income and occupational position of black workers, suggesting that antidiscrimination enforcement had some success. Butler and Heckman (1977) offered a different explanation: expanded transfer programs, which encouraged low-skilled black men to leave the workforce, caused a selection effect that explains this "improvement" in relative wages. Smith and Welch (1989) also cast doubt on the role of federal antidiscrimination law, noting that black earnings had been increasing since the 1940s, likely because of better education and migration out of the South.

Donohue and Heckman (1991) argue that federal antidiscrimination laws, such as Title VII, were instrumental to black economic progress, which they describe as episodic rather than continuous. They show that the greatest advances in the economic status of black Americans came after World War II and in the 1960s and the 1970s, and that much of the progress in the latter period was due to advances in civil rights policies at the federal level. In other periods, black progress was more or less continuous and, for the most part, it could be attributed to traditional explanations such as migration or increased education. See Card and Krueger (1992) for evidence on school quality and black-white relative earnings.

Recall that in 1972, Title VII coverage was extended to include firms with 15–24 employees. Chay (1998) uses this change in coverage and existing variation in state fair employment laws to establish treatment and control groups across industries and re-

gions. He documents shifts in black employment and earnings in industries most affected by the change and concludes that coverage under antidiscrimination law positively affected labour market outcomes for black Americans. Carrington, McCue, and Pierce (2000) also consider employer size and show that after 1964, black workers moved to large employers (for whom Title VII applied), and this movement accounted for about 15 percent of the convergence in the black-white wage gap over 1965 to 1980.

Hahn, Todd, and van der Klaauw (1999) use a regression discontinuity design to estimate the effect of Title VII coverage on minority share of employment. They use data from the 1979–1993 waves of the NLSY79 to compute implied estimates of minority employment by firm size, and they exploit the 15-employee coverage threshold to nonparametrically estimate the effect of Title VII. Using this method, the authors conclude that Title VII likely increased minority employment in covered firms, but their results are sensitive to excluding observations at the threshold and to the choice of bandwidth. The authors do not account for measurement error in firm size in the NLSY79, which is an important feature of the data, as shown in this essay. They do not estimate the effect of Title VII on relative wages, and they do not consider higher cutoffs in Title VII liability established by the Civil Rights Act of 1991.

Oyer and Schaefer (2000) examine the effect of the Civil Rights Act of 1991 using data from the 1987–1993 Surveys of Income and Program Participation (SIPP). They show that in the period following the enactment of the law, firms moved away from individual firings toward mass layoffs. Black workers were less likely to be fired following the law, but overall rates of involuntary displacement were largely unaffected. Oyer and Schaefer argue that this is consistent with new litigation possibilities firms faced after the Civil Rights Act of 1991. In another study, Oyer and Schaefer (2002) show that prior to the Civil Rights Act of 1991, black and women shares of employment had been growing in industries where these groups were least represented, but that this trend ended around the time of Act. They conclude that the law's "sorting effect," that

is, the disincentive to hire protected workers, outweighed its "quota effect." Donohue (2007) argues that much of Oyer and Schaefer's work is complicated by the Supreme Court rulings leading up to the passage of the Civil Rights Act of 1991, which prevent clean "before and after" comparisons.

In a related context, Acemoglu and Angrist (2001) investigate the labour-market effects of the Americans with Disabilities Act (ADA), which banned discrimination in hiring, firing, and pay of disabled workers. Using data from the Current Population Survey, they find disabled employment declined following enactment of the ADA. Employment fell more in medium-size firms, as small firms were exempt (and large firms may have been able to absorb the cost of accommodation). There was no clear evidence of an offsetting change in relative wages, suggesting the equal pay provision of the ADA had some effect. Acemoglu and Angrist suggest this may have exacerbated the decline in disabled employment.

The primary limitations of these studies are: (1) they tend rely on before-and-after comparisons; (2) they ignore the structure of Title VII liability following the Civil Rights Act of 1991; and (3) they are somewhat out of date. The aim of this essay is to propose a new empirical strategy to measure the wage effects of liability for employment discrimination, which is to exploit all four cutoffs in liability established by the Civil Rights Act of 1991, and to provide updated estimates using recent data.

2.4.2 Empirical Strategy

To estimate the wage effects of Title VII coverage and liability under the Civil Rights Act of 1991, this essay proposes a sharp regression discontinuity design with firm employment as the assignment (or "running") variable. Firm employment determines treatment in terms of overall coverage and legal liability under federal antidiscrimination law. For now, this essay will focus on black wages relative to white wages in the United States. This subsection describes the statistical model used and its requirements for

proper identification.

2.4.2.1 Statistical Model

The regression discontinuity design was introduced by Thistlethwaite and Campbell (1960), although it became much more popular in recent years. For an overview of regression discontinuity designs in economics, see Imbens and Lemieux (2008) and Lee and Lemieux (2010). The advantages of a regression discontinuity design are that it can be justified under relatively weak assumptions, and it has strong internal validity. The design can be thought of as a local randomised experiment, and estimated treatment effects have a causal interpretation (Lee 2008). On the other hand, its external validity is potentially weak. The treatment effect is identified at the assignment cutoff, not globally.

This essay proposes using a "sharp" regression discontinuity design, as the assignment of Title VII coverage and of different levels of legal penalties under the Civil Rights Act of 1991 are deterministic (and discontinuous) functions of firm size.[7]

Potential outcomes framework. Consider a potential outcomes framework. Let W_{ij} denote the real hourly wage of person i from social group j. Consider the pair of "potential" outcomes $W_{ij}(0)$ and $W_{ij}(1)$, where the latter denotes real hourly wages in the presence of liability for employment discrimination. The unit-level causal treatment effect of the law is

$$W_{ij}(1) - W_{ij}(0).$$

Obviously, we cannot observe both $W_{ij}(0)$ and $W_{ij}(1)$ for the same person.

Sharp regression discontinuity design. Let the binary treatment variable T_{ijk} be

[7]For an overview of "fuzzy" regression discontinuity designs, in which the probability of treatment increases discontinuously as a function of an assignment variable, but not from zero to one, see Lee and Lemieux (2010).

a function of a continuous assignment variable X_{ij} such that

$$T_{ijk} = \begin{cases} 0 & \text{if } X_{ij} < c_k \\ 1 & \text{if } X_{ij} \geq c_k \end{cases}$$

where individuals with $X_{ij} \geq c_k$ receive treatment and those with $X_{ij} < c_k$ do not. In this case, T_{ijk} denotes treatment under $k = 4$ levels of liability, X_{ij} is firm employment, and $\{c_k\}_{k=1}^{4}$ are cutoffs for Title VII coverage and for higher damages following the Civil Rights Act of 1991. A sharp regression discontinuity design estimates

$$\tau_{jk} = \lim_{\varepsilon \downarrow 0} \mathbb{E}[W_{ij} | X_{ij} = c_k + \varepsilon] - \lim_{\varepsilon \uparrow 0} \mathbb{E}[W_{ij} | X_{ij} = c_k + \varepsilon], \tag{2.4}$$

which equals the average causal treatment effect at the cutoff c_k,

$$\mathbb{E}[W_{ij}(1) - W_{ij}(0) | X_{ij} = c_k], \tag{2.5}$$

if we assume $\mathbb{E}[W_{ij}(0)|X_{ij} = x]$ and $\mathbb{E}[W_{ij}(1)|X_{ij} = x]$ are continuous in x. This is the condition for identification.[8] I will elaborate on what this condition means shortly. Since we want to understand the effect of liability for employment discrimination on black wages relative to white wages, the object of interest in this essay is $\tau_{bk} - \tau_{wk}$, where $j \in \{b, w\}$ denotes race.

2.4.2.2 Data

To estimate the effect of Title VII and the Civil Rights Act of 1991 on black relative wages in the United States, this essay uses data from the National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY79 is a nationally representative sample of men and women who were between 14 and 22 years old when they were first surveyed

[8] Actually, this is stronger than what is required for identification, which is that these functions are continuous at the cutoff c_k.

in 1979. The original sample contained 12,686 respondents; 9,964 remain eligible for interview today. There are roughly equal numbers of men and women in the NLSY79. In total, there have been 24 rounds of surveying between 1979 and 2010, with surveys taking place annually until 1994 and biannually after.

The NLSY79 has three subsamples. One is a cross-sectional sample of the non-institutionalised civilian population in the United States, another is a supplemental sample of either Hispanic, black, or economically disadvantaged non-Hispanic, non-black persons, and a third is a sample of persons serving in the U.S. military. By 1990, much of the military sample and the economically disadvantaged non-Hispanic, non-black sample were excluded from the survey. Observations from all three subsamples are used in this analysis of the Civil Rights Act of 1991.

Variable definitions. From the NLSY79, this essay uses data on hourly wages, employer size, employer history, race, gender, age, education, Armed Forces Qualification Test (AFQT) score, and urban residence.

Hourly wages are reported by respondents, or calculated if the respondent uses a time unit for earnings other than per hour (e.g., per day, per week, or per year). Nominal wages are converted to real wages using the Personal Consumption Expenditures: Chain-Type Price Index (PCECTPI) with 1987 as the base year, following Altonji and Pierret (2001) and Fryer, Pager, and Spenkuch (2013).[9] Employer size is granular, rather than in bins, which is essential for the empirical strategy. Respondents are asked to report the number of employees at the location where they work. They are also asked whether they work for a private company, government, or non-profit organization, or if they are self-employed or work for a family business. Beginning in 1996, respondents were assigned unique longitudinal employer identification numbers, allowing me to link employers through survey years.

[9]In much of this essay, it does not matter whether nominal or real wages are used, since the focus is on black-white relative wages and most regressions control for year effects. However, for sample restrictions, and for figures that display wages across multiple years, it makes sense to use real wages.

At the start of the panel, interviewers assigned respondents to one of three racial or ethnic cohorts: Hispanic, black, or non-Hispanic and non-black. This is obviously an enormous simplification of social identity based on skin colour or ancestry, not least because a person could identify as both Hispanic and black. In 2002, questions were added to the NLSY79 to offer respondents more options for their racial or ethnic identity. Limiting the sample to observations from 2002 and after ignores a substantial amount of data collected shortly after the Civil Rights Act of 1991. For this reason, this essay uses the original cohort classifications to define race, and it (like others) refers to the non-Hispanic, non-black category as "white." For an overview of taxonomical issues related to race in studies of discrimination in economics, see Charles and Guryan (2011).

Every respondent is classified as male or female. Age at interview date is calculated using birth dates, and education is measured in terms of years of schooling and top-coded at 20. The Armed Forces Qualification Test (AFQT) was administered in 1980 to more than 90 percent of the NLSY79 cohort. The test, which includes four sections of the broader Armed Services Vocational Aptitude Battery (ASVAB), was used for enlistment screening and job assignment in the U.S. military. AFQT scores have been used as a proxy for ability or pre-market skills (O'Neill 1990; Neal & Johnson 1996; Lang & Manove 2011). AFQT scores are standardised by age at the time the test was given. Finally, year-specific sampling weights account for selection and nonresponse.

Sample restrictions. In 1993 and 1994, the NLSY79 underwent significant changes, including the exclusive use of Computer-Assisted Personal Interviewing (CAPI) and changes to sections on employment in order to match the Current Population Survey (CPS), which underwent a major revision in January 1994. For this reason, the analysis in this essay is limited to data from 1994 to 2010. This amounts to 9 waves—one every other year.

I further restrict attention to individuals who work for a private company or the government, who report the number of employees at the location where they work,

and who report that their employer has only one location. This limits the sample to employees of single-establishment firms. Single-establishment firms may respond differently to antidiscrimination law, since they are less likely to have a human resources department, for example. Still, in order to correctly identify which persons are covered under different thresholds from the Civil Rights Act of 1991, the analysis must exclude persons who report that their employer has multiple locations. Subjects were asked how many employees there are at the place where they work, corresponding to employment at the establishment level, rather than at the firm level. They were then asked whether their employer has multiple locations, and, if so, whether the total number of employees was more or less than 1,000. Since Title VII applies to firms, not establishments, and since the empirical strategy requires a granular measure of firm size, this analysis relies on individuals who report they work for single-establishment firms and ignores others. Finally, observations with real wages below $2 or above $100 per hour are discarded. In total, these restrictions result in a final sample of 13,563 observations on 5,084 individuals.

Sources of variation. Data from 1994 to 2010 are pooled to increase the sample size, which requires computing standard errors that are robust to clustering. After applying the sample restrictions described above, much of the variation in the data is "between" variation, that is, across individuals, rather than within variation across time. There are, on average, 2.7 observations per individual in the restricted sample, and the distribution is skewed, with 35.7 percent of individuals with only one observation across all 9 waves. The median number of observations per individual is 2. Individuals get dropped in other waves for a variety of reasons, such as not completing an interview or reporting that their employer has multiple locations.

This raises an important question: how often do respondents report that their employer has one location but report in another year that the same employer has multiple locations? We can answer this using the unique longitudinal employer identification

numbers in the NLSY. About 27 percent of the observations in the restricted sample have this property. Looking at specific cases, it is hard to determine whether these represent actual changes in an employer's number of establishments or errors in reporting. As a robustness check, estimates that ignore observations when individuals report multiple locations in one wave, a single location in the next, and multiple locations right after do not differ.

To summarise, the analysis that follows uses a restricted sample of 13,563 observations on 5,084 individuals. Of these observations, 7,330 (54 percent) are male and 6,233 (46 percent) are female. Black, Hispanic, and white cohorts account for 3,615 (26.7 percent), 2,782 (20.5 percent) and 7,166 (52.8 percent) observations, respectively. The raw black-white wage gap in the restricted sample is –0.331 log points, or about 28 percent, for men, and for women, it is –0.186 log points, or 17 percent.[10] If we apply only the restriction that real wages are between $2 and $100 per hour, the gaps are –0.363 and –0.162 log points for men and women, respectively. These raw gaps are in line with other work that uses the NLSY79, such as Neal and Johnson (1996), Neal (2004), and Fryer, Pager, and Spenkuch (2013). They are slightly lower than estimates that use NJ data on unemployed workers in Fryer, Pager, and Spenkuch (2013). Using data from the Current Population Survey (CPS), I estimate that the black-white log hourly wage gap for the same age group over the same time period is –.214 for men and –.118 for women.

If you thought of this essay as an analysis of firms, we might be concerned with size-biased sampling. That is, since the NLSY79 surveys individuals rather than establishments or firms, it is more likely to include workers from large firms. Indeed, you can see this if you compare the cumulative distribution of firm size in the NLSY79 with official data. However, since this essay uses a regression discontinuity design to obtain estimates of the effect of liability for employment discrimination, size-biased sampling

[10]Like most "raw" racial wage gaps in the literature, these estimates condition on age.

does not pose a significant problem. This point and the distribution of reported firm sizes in the NLSY79 are examined in more detail in the next section.

2.4.2.3 Estimation

Local linear regression. To estimate τ_{jk}, one can use standard nonparametric methods. However, we want to estimate the conditional expectation functions in (2.4) evaluated at a specific point, which is also a boundary point. For this reason, Hahn, Todd, and van der Klaauw (2001) recommend local linear regression or local polynomial regression, as standard kernel regression estimates will be biased. This essay uses local linear regression, which still requires choosing a bandwidth and kernel weighting function.

Many regression discontinuity designs use a triangular kernel, since it is optimal for estimating boundary points using local linear regression (Fan & Gijbels 1996). However, there are several advantages to using a rectangular (uniform) kernel. While the triangular kernel places greater weight on observations near the boundary, the rectangular kernel weights all observations (within the bandwidth) equally. Lee and Lemieux (2010) point out that the choice of kernel does not substantially affect estimates, in practice. They argue that a more transparent way of placing additional weight on observations close to the cutoff is simply to shrink the bandwidth.

A principal advantage of the rectangular kernel is that it is easy to interpret. Estimating τ_{jk} using local linear regression and a rectangular kernel is equivalent to estimating the linear model

$$W_{ij} = \alpha_{jl} + \tau_{jk}T_{ijk} + \beta_{jl}(X_{ij} - c_k) + (\beta_{jr} - \beta_{jl})T_{ijk}(X_{ij} - c_k) + \upsilon_{ij}, \qquad (2.6)$$

$$c_k - h \leq X_{ij} \leq c_k + h,$$

where υ_{ij} is a stochastic error term with mean zero, h is the bandwidth, and subscripts

l and r designate left and right of the cutoff, c_k. The equation above pools data to the left and to the right of the cutoff and allow for different slopes on either side, but one could just as easily estimate α_{jl} and α_{jr} in separate regressions,

$$W_{ij} = \alpha_{jl} + \beta_{jl}(X_{ij} - c_k) + \upsilon_{ij},$$

$$c_k - h \leq X_{ij} < c_k,$$

and

$$W_{ij} = \alpha_{jr} + \beta_{jr}(X_{ij} - c_k) + \upsilon_{ij},$$

$$c_k \leq X_{ijk} \leq c_k + h,$$

to compute an estimate of the treatment effect, τ_{jk}. Note that $\tau_{jk} = \alpha_{jr} - \alpha_{jl}$, since

$$\alpha_{jr} = \lim_{\varepsilon \downarrow 0} \mathbb{E}[W_{ij}|X_{ij} = c_k + \varepsilon]$$

and

$$\alpha_{jl} = \lim_{\varepsilon \uparrow 0} \mathbb{E}[W_{ij}|X_{ij} = c_k + \varepsilon].$$

Pooling the data, as in (2.6), and centring X_{ij} around c_k allows me to easily compute the standard error of $\widehat{\tau}_{jk}$, which is just the standard error of the estimated coefficient on T_{ijk} in (2.6). This is an additional benefit of using a rectangular kernel.

In the linear model (2.6), the condition for identification becomes that $\mathbb{E}[\upsilon_{ij}|X_{ij} = x]$ is continuous in x (specifically, at the point $x = c_k$). If this condition is satisfied, the sharp regression discontinuity design yields a consistent estimate of the treatment effect. Note that we are not required to make W_{ij} linear in X_{ij}. Instead, wages could be some other function of firm employment on either side of the cutoff, which could differ by social group, $f_{jl}(X_{ij} - c_k)$ and $f_{jr}(X_{ij} - c_k)$, allowing for higher order terms in X_{ij}.[11]

[11]This is one way in which estimation is nonparametric. This method does not take a stance on

To estimate $\tau_{bk} - \tau_{wk}$, we can estimate τ_{bk} and τ_{wk} separately and compute the difference, or we can combine the linear models characterised by (2.6) for black and white individuals (adding indicator variables for race and interaction terms to allow for different slopes and intercepts for each race on either side of the cutoff) and estimate $\tau_{bk} - \tau_{wk}$ directly. Like before, the latter approach has the advantage of straightforward computation of standard errors, and, again, this is not possible without using a rectangular kernel. If $\widehat{\tau}_{bk}$ and $\widehat{\tau}_{wk}$ are computed separately, the standard error of $\widehat{\tau}_{bk} - \widehat{\tau}_{wk}$ could be obtained using the bootstrap, however.

This essay estimates $\tau_{bk} - \tau_{wk}$ by combining (2.6) for black and white individuals, as described above, but doing this separately for men and women. The paper uses local linear regression, a rectangular kernel, population sample weights, and the bandwidths described below. To conduct inference, I compute standard errors that are robust to heteroscedasticity and to clustering on individuals (since the analysis pools multiple waves of the NLSY79).

Choice of bandwidth. The choice of bandwidth involves a trade-off between bias and precision. Using a larger range of data increases precision but also bias, as the linear or polynomial specification becomes a worse approximation (Lee & Lemieux 2010). The bandwidth should tend to zero as the number of observations tends to infinity, and Hahn, Todd, and van der Klaauw (2001) show that the optimal bandwidth is proportional to $N^{-1/5}$, but shrinking the bandwidth at a slightly faster rate ($h \propto N^{-\delta}$ with $1/5 < \delta < 2/5$) ensures that the usual formula for least squares standard errors is valid.

These details are important for the asymptotic theory behind inference, but in practice, one cannot increase the sample size and shrink the bandwidth to zero. Instead, there are several options for choosing an appropriate bandwidth. Ludwig and

the actual functional form of W_{ij}; it only uses local linear or polynomial approximations to estimate its conditional expectation at the cutoff. Although, as we will see, large bandwidths end up being necessary because of the data. In this case, the assumption of local linearity could be thought of as restrictive or result in estimation that is "parametric" in some sense.

Miller (2005) propose a "leave-one-out" cross-validation procedure, and Imbens and Kalyanaraman (2012) offer another data-driven rule. Calonico, Cattaneo, and Titiunik (2014) argue that the bandwidths resulting from these methods are too large, and they offer their own alternative. All three of these approaches suggest bandwidths in this case that meet or exceed the maximum bandwidth allowable without mixing different treatments. This is especially true of the higher cutoffs. For example, the maximum (symmetric) bandwidth around the 101-employee cutoff from the Civil Rights Act of 1991 is 86. This ensures that firms with 14 or fewer employees are not included, as these firms are not covered under Title VII at all. With this in mind, I construct bandwidths $h/2$, h and $2h$, where $2h$ equals the maximum allowable bandwidth without crossing another treatment threshold. Including estimates using bandwidths equal to half and double some standard bandwidth, h, is common practice, and it demonstrates how sensitive estimates are to the choice of bandwidth. The large "optimal" bandwidths raise concerns about the specification or about the data, as discussed later.

Covariates and two-stage estimation. One could directly include a vector of covariates, Z_{ij}, in the local linear regression described by (2.6). This does not affect the consistency of the estimator for τ_{jk}, but it can improve precision, and it may eliminate some bias, especially when one uses observations far from the cutoff value, as is the case here (Imbens & Lemieux 2008). An alternative would be to regress wages on the set of covariates, form residuals using predicted values, and use the residuals as the dependent variable in (2.6). Lee and Lemieux (2010) suggest this approach as a supplement to directly including the covariates, although "residualising" the data in this way can raise standard errors in some cases.

I estimate three specifications: (i) a baseline local linear specification, (ii) a specification which includes a set of covariates, and (iii) a two-stage estimation procedure which regresses wages on the covariates, forms residuals, and re-estimates the baseline linear specification using residualised wages as the dependent variable. Specifications

(ii) and (iii) do not allow the coefficients on additional covariates to differ by race or by side of the cutoffs, but they are estimated separately for men and women. The covariates included are: age, age squared, urban residence, schooling, AFQT score, AFQT score squared, and indicator variables for years (year fixed effects). In the two-stage estimation procedure, the sample is not limited to workers who report that their employer has only one location until the second stage. That is, in forming residuals, all valid real wage observations (between \$2 and \$100 per hour) are used.

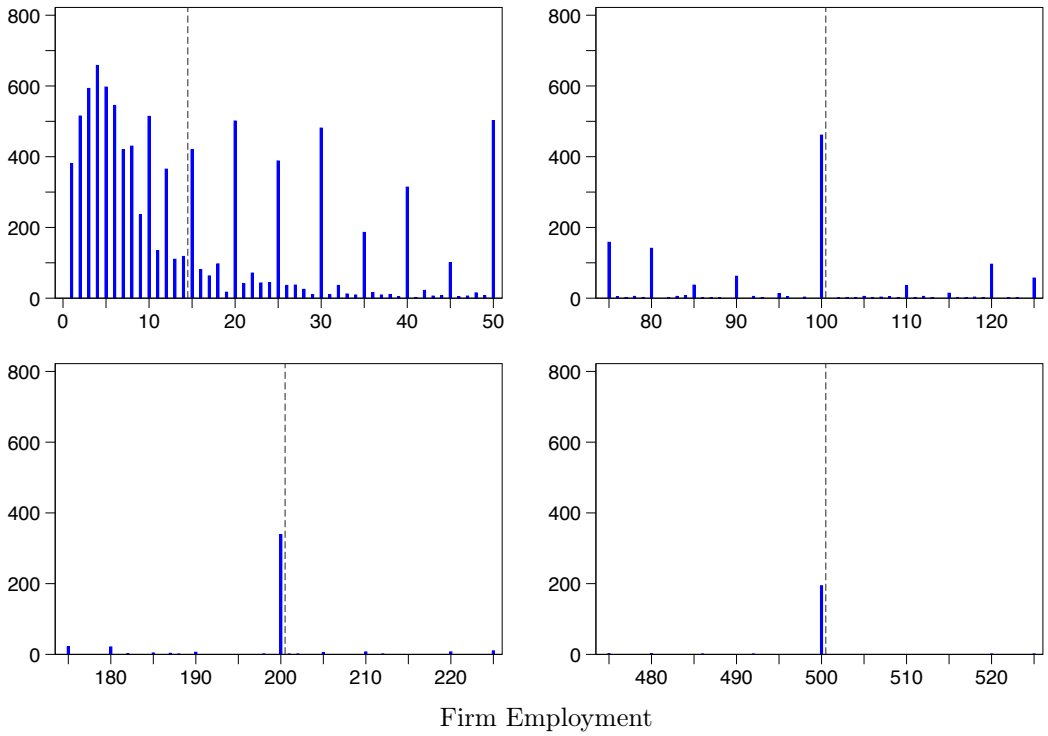
2.4.2.4 Identification

A sufficient condition for identification is that $\mathbb{E}[W_{ij}(0)|X_{ij} = x]$ and $\mathbb{E}[W_{ij}(1)|X_{ij} = x]$ are continuous in x. In the linear model without covariates, an equivalent condition is that $\mathbb{E}[\upsilon_{ij}|X_{ij} = x]$ is continuous in x. If economic agents precisely manipulate the assignment variable X_{ij} in response to treatment, or if observed or unobserved determinants of real wages jump discontinuously at the cutoff, the regression discontinuity design is invalid. In particular, if other laws or regulations that affect wages, such as state antidiscrimination laws, change discontinuously at the cutoff, we cannot identify the effect of the federal law.

In many other regression discontinuity design contexts, agents exert some degree of control over the assignment variable, but "precise" manipulation is not possible. For example, a student can increase their test score by exerting effort, but if the student cannot precisely choose their score, a regression discontinuity design with test score as the assignment variable could still be valid. In this study, agents obviously exert precise control over the assignment variable: firm employment. If firm owners choose their number of employees to select into, or—more likely—out of, treatment, identification fails.

I examine evidence from the NLSY79 data set and from official data on firm employment to determine whether precise manipulation of the assignment variable

Figure 2.2: Frequency of Reported Firm Sizes in NLSY79, 1994–2010



Note: Sample restricted as described in Section 2.4.2.2. In particular, respondents limited to those who report that their employer has only one location (that is, employees of single-establishment firms). Dashed lines indicate cutoffs in Title VII coverage and liability following the Civil Rights Act of 1991.

threatens the regression discontinuity design. This reveals measurement error and nonrandom heaping in firm employment in the NLSY79. Using recent work from the literature on regression discontinuity designs, I explain the consequences for the identification strategy, and I consider several alternatives for estimation.

Distribution of firm employment. To check whether the assignment variable, firm employment, varies smoothly through the cutoffs in Title VII coverage and liability under the Civil Rights Act of 1991, its distribution in the NLSY79 data set is shown in Figure 2.2. The figure displays the frequency of reported firm sizes (in employees) in the 1994–2010 waves of the NLSY79, restricted to single-establishment firms as described in Section 2.4.2.2.

McCrary (2008) proposes a test of whether the density of the assignment variable

is continuous, but we can see immediately from Figure 2.2 that the test would reject the hypothesis of no discontinuity in firm employment at any of the 15-, 101-, 201-, or 501-employee cutoffs. Figure 2.2 shows bunching, or "heaping," in firm employment at 100, 200, and 500 employees, just below regions of higher liability under the Civil Rights Act of 1991. We also see heaps at 15 employees and at other multiples of 5 or 10 employees, especially from respondents who work for larger firms, which raises the question: Do some firms respond to antidiscrimination law in choosing how many employees to hire, or do NLSY respondents report the size of their employer with error, in particular, by rounding? Or both?

The choice for business owners of how many workers to hire depends on a variety of factors, including demand for their good or service, their production technology, worker productivity and, of course, the supply of labour. One might think that liability for employment discrimination under federal law would be unlikely to dominate any of these considerations. However, evidence from France suggests that firms respond strongly to labour and other regulations that take effect once they employ 50 workers. Garicano, Lelarge, and Van Reenen (2016) examine the effect of these regulations on the distribution of firm employment, which bunches sharply at 49 employees in France, and they use this result to identify a model of size-dependent regulations and to estimate the efficiency cost of regulation. They estimate that the regulations are equivalent to a 1 percent increase in labour costs. Gourio and Roys (2014) model the French regulations as a sunk cost upon reaching the regulatory threshold and a labour tax. Their structural model suggests that the French regulations operate more as a sunk cost than as a per-period labour tax, and the efficiency cost of regulation may be small if the number of firms can adjust.

The France example does not exactly translate to the case of federal antidiscrimination law in the United States, since the number and cost of French regulations taking effect at 50 employees likely exceeds the burden of Title VII coverage or higher liability

under the Civil Rights Act of 1991. Still, it suggests that firms might respond to liability by hiring fewer employees, and this essay therefore uses some of the methods from Garicano, Lelarge, and Van Reenen (2016) and Gourio and Roys (2014) to analyse the distribution of firm employment in the United States.

The structure of the limits on Title VII claims allows us to infer something about the relative importance of firm behaviour in response to liability for employment discrimination versus measurement error by NLSY respondents. From Figure 2.2, we see that there are heaps of firms reported to have 15, 100, 200, and 500 employees (as well as several other heaps). While bunching at 100, 200, and 500 employees could represent precise manipulation of the assignment variable in response to treatment (that is, to avoid greater liability applied to firms with 101, 201, and 501 workers), bunching at 15 employees, if due to firm behaviour, would imply firms sort *into* treatment (that is, increased penalties for unequal hiring, firing, and compensation). This is strong evidence that heaping in the survey data results from rounding by respondents, not firm behaviour in response to cutoffs in Title VII coverage or liability under the Civil Rights Act of 1991. That there are other heaps at multiples of 5 and 10 employees, and that these heaps become more common as firm sizes become larger, substantiates this claim. I should point out, however, that there is a small uptick in the number of firms with 14 employees relative to 13 employees, as shown in Figure 2.2.

To understand whether firms adjust their employment in response to antidiscrimination law, we need to move beyond data from surveys of individuals and examine administrative data on firms. Figure 2.3 shows the distribution of firm sizes in the United States using data from two different administrative sources. The top-left panel of Figure 2.3 shows the distribution of firms with fewer than 50 employees using the U.S. Census Bureau's Synthetic Longitudinal Business Database (SynLBD).[12] The SynLBD

[12]Without access to a secure Census computing site in the United States, I can only obtain a synthetic version of the longitudinal business microdata, which is generated by fitting models to the original data set and simulating replacement values to mimic its properties.

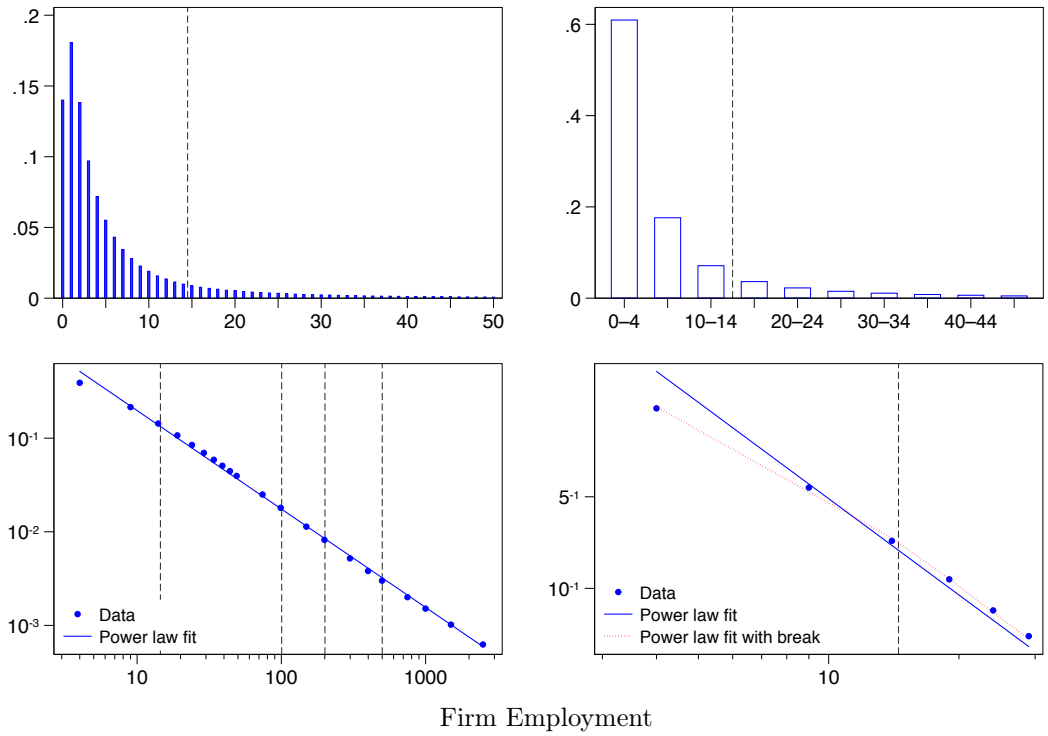
sample is limited to single-establishment firms for comparison with the subsample of the NLSY79. Looking at the figure, we can see that the distribution of firm employment declines smoothly through the cutoff in Title VII coverage at 15 employees. The distribution is smooth through higher cutoffs in legal penalties under the Civil Rights Act of 1991, too. This is evidence that firms do not manipulate their employment in response to antidiscrimination law.

The top-right panel of Figure 2.3 shows the distribution of firms with fewer than 50 employees using the Census Bureau's Statistics of U.S. Businesses (SUSB). As you can see, these data are coarse, not granular like the Longitudinal Business Database. Firms are grouped into employment bins with bin width equal to 5. Although it would appear that the proportion of firms with 10-14 employees does not "jump" above the density implied by the top-right panel of Figure 2.3, the bottom two panels offer a closer look.

Gabaix (2009) shows that a large number of empirical regularities in economics and finance follow a power law. For example, the distribution of firm sizes is well described by Zipf's law (Axtell 2001; Fujiwara et al. 2004; Gabaix & Landier 2008), whereby $\Pr(\text{size} > S) = kS^{-\zeta}$, which corresponds to the density $f(S) = k\zeta S^{-(\zeta+1)}$. Zipf's law says that $\zeta = 1$. A power law fitted to the SUSB data is shown in the bottom two panels of Figure 2.3. Note that these panels show the countercumulative distribution of firms, that is, $\Pr(\text{size} > S)$, using a logarithmic scale. I estimate $\widehat{\zeta} = 1.052$, which is consistent with Axtell (2001) ($\widehat{\zeta} = 1.059$) and Zipf's law ($\zeta = 1$). We can then ask the question: does the proportion of firms with 10–14 employees deviate from the power law regularity? Visual inspection of the figure suggests that there is not an unusual amount of firms sized 10–14 employees. Note that if there were bunching just below the 15-employee cutoff, we would expect the countercumulative observation for 10–14 employees to fall below the power law trend. It falls marginally above.

This does not account for the effect of very small firms on the power law estim-

Figure 2.3: U.S. Firm Size Distribution, 1994–2000 & 2006



Source: U.S. Census Bureau

Top-left: Distribution of single-establishment firms in the Synthetic Longitudinal Business Database (SynLBD), 1994–2000. Top-right: Statistics of U.S. Businesses (SUSB), 2006. Bottom: Countercumulative distribution of firms and power law estimation using SUSB data. Dashed lines indicate cutoffs in Title VII coverage and liability following the Civil Rights Act of 1991.

ation, however. Axtell (2001) points out that especially small and large firms do not follow the power law as closely. Re-estimating the power law using the SUSB data but excluding the three smallest size categories yields an observation for 10–14 employees slightly below the power law trend. However, a statistical test of whether the observation for 10–14 employees "jumps" off (below) the power law trend comfortably fails to reject the hypothesis of no deviation from the trend, lending support to the hypothesis that firms do not adjust their employment in response to liability for employment discrimination.

The bottom-right panel of Figure 2.3 focuses on the 15-employee threshold and adds another power law specification. Similar to Garicano, Lelarge, and Van Reenen (2016) and Gourio and Roys (2014), I estimate a "broken" power law, allowing for

different slopes and intercepts before and after the 15-employee cutoff. The resulting break in the fitted countercumulative distribution is the opposite of what we would expect if there was bunching in firm employment at 14 employees or fewer, after which Title VII applies.

Altogether, this evidence suggests that firms do not change their employment to opt out of Title VII coverage or higher liability under the Civil Rights Act of 1991. This is good news from the standpoint of identification in the regression discontinuity design, although, if firms changed their behaviour in response to these laws, that would be an interesting finding. In this discussion of firm sizes, it is worth noting again that, viewed as data on firms, the NLSY79 is subject to size-biased sampling. Since it surveys individuals rather than establishments or firms, it is more likely to include workers from large firms. This does not threaten the regression discontinuity design because real wages are estimated conditional on firm size (Hahn, Todd & van der Klaauw 1999).

Measurement error and heaping-induced bias. If we believe that firms do not manipulate their employment in response liability for employment discrimination, we are left to interpret the heaping in Figure 2.2 as measurement error on the part of NLSY respondents. This is pretty obvious from the heaps at multiples of 5 or 10 employees. From the figure, it appears that rounding is especially common when reporting the number of employees at large firms.

If individuals report the number of employees at the place where they work with error, leading to heaping in the assignment variable in the regression discontinuity design, what are the consequences for identification? One worry is that treatment may be mismeasured for some observations. For example, if there are 14 employees where an individual works, and they round this number to 15, from the viewpoint of the econometrician, the employer is subject to Title VII coverage. Similarly, someone whose employer has 105 employees might round and say their employer has 100 employees. While punitive and compensatory damages in a discrimination lawsuit could equal

$100,000 for this employer, in this study, they would be subject to only $50,000 in damages. The latter example points out that rounding could plausibly be greater than to the nearest multiple of 5 employees.

To account for the possibility of mismeasured treatment, this essay includes a set of estimates that drop observations with firm employment "at" the cutoffs, in particular, respondents who report that their employer has 15, 100, 200 or 500 employees. As just explained, this may not eliminate mismeasurement in treatment status if rounding is more severe, but estimation at the higher cutoffs is already difficult due to limited data. Dropping observations at the cutoff, sometimes called "donut RD," is common in applied regression discontinuity designs. Another set of estimates keeps some data—chosen randomly—at the cutoff in order to match the implied (smooth) distribution of firm sizes.[13] Treatment effects are estimated by averaging results from 100 such simulations. Not surprisingly, these estimates fall in between those from the specification that includes all observations at the cutoff and the specification that drops them.

Dong (2015) and Barreca, Lindo, and Waddell (2016) show that regression discontinuity estimates may be biased in the presence of non-random heaping (e.g., rounding) in the running variable, even when treatment is not mismeasured. Dong (2015) offers a bias correction when the distribution of the measurement error is known. Barreca, Lindo, and Waddell (2016) use Monte Carlo simulations to demonstrate that regression discontinuity estimates are biased in a number of scenarios involving heaping, including when (i) mean outcomes differ for heaped and non-heaped types, (ii) slopes of the relationship between the outcome and assignment variables differ, (iii) treatment effects differ, or some combination of (i)–(iii). Regression discontinuity estimates can be biased regardless of whether heaps are close to the treatment threshold or far away (but within the bandwidth). Thus, "donut RD," that is, dropping heaped data near the treatment

[13]This is a way of addressing the issue with heaping, but one might still be concerned about including observations for which treatment could be mismeasured due to rounding in the assignment variable, especially since including these observations does not shrink the standard errors enough to make the results significant.

threshold only, will not automatically produce unbiased estimates. The direction of the bias depends on differences between heaped and non-heaped types, the location of heaps around the cutoff, and the chosen bandwidth (Barreca, Lindo & Waddell 2016).

To begin to consider the role of heaping in these estimates, I distinguish heaped and non-heaped data in the plots of binned average outcomes by assignment variable (except for higher cutoffs from the Civil Rights Act of 1991, around which nearly all observations are heaped). This allows for visual inspection of how the relationship between firm employment and real hourly wages differs for data at heap points. To obtain consistent estimates of treatment effects, Barreca, Lindo, and Waddell (2016) recommend stratifying the data. Dropping heaped data yields unbiased estimates of the treatment effect for non-heaped types, while dropping non-heaped data yields unbiased estimates of the average treatment effect across the two types, weighted by the share of heaped and non-heaped data observed at data heaps. (This requires that there be enough heaps within a reasonable bandwidth of the treatment threshold, however.) To recover an unbiased estimate of the unconditional average treatment effect, we can compute a population-weighted average of the two estimates (Barreca, Lindo & Waddell 2016). The next section includes results using this method.

State laws. Another threat to the identification strategy, and perhaps the most serious one, is the existence of state antidiscrimination laws with cutoffs that coincide exactly with cutoffs under federal law. As described in Section 2.2, several states have antidiscrimination statutes that apply to firms with 15 or more employees, such as Illinois, Florida, and Texas. In Texas, there are cutoffs in legal liability at the state level at each of the federal cutoffs established by the Civil Rights Act of 1991. The existence of state antidiscrimination law is not in itself a problem. It is only a problem when these laws apply discontinuously at the same thresholds used to estimate the effect of federal law. In all other cases, estimates of the effect of liability under federal antidiscrimination law are conditional on existing state laws.

There are at least two possible ways forward. The first is to simply conclude that the method used here likely overestimates the effect of federal antidiscrimination law, since it could pick up the effect of state laws, especially in large states such as Texas and Florida. If we found that federal antidiscrimination law has little to no effect on black-white relative wages, then the existence of state antidiscrimination law with identical cutoffs would strengthen this conclusion.

A second way forward is to seek different data. (The version of the NLSY79 that I have access to does not include state identifiers.) With access to a data set with information on race, wages, exact firm size, and U.S. state, the analysis could be limited to states that do not have cutoffs in antidiscrimination law that coincide with federal cutoffs, or one could estimate the effect of state cutoffs themselves when they differ from federal ones. In my experience, the most demanding of the data requirements is the need for an accurate, granular measure of firm employment, as many data sets have information on firm employment in bins. A data set that satisfies these requirements, including state information, is the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) microdata. This is matched employer-employee data, with firm employment derived from administrative data sources. The LEHD microdata is available at secure Census Research Data Centers in the United States, upon approval by the U.S. Census Bureau.

2.4.3 Results

This essay considers separately the effect of Title VII coverage, which takes effect at 15 employees, and Title VII liability, which increases from $50,000 to $100,000, $200,000, and $300,000 for firms with 101, 201, and 501 employees, respectively, under the Civil Rights Act of 1991 (see Figure 2.1 above). Table 2.A3 and Table 2.A4 in the Appendix show the results of the first stage from the two-stage estimation procedure described in the preceding section. In these tables, I estimate wage regressions separately

for men and women, and the results align with the literature in terms of racial wage gaps and the effect of schooling and pre-market skill. As in Neal and Johnson (1996), differences in hourly wages between black and white individuals narrow substantially (or even reverse, in the case of women) once you control for a measure of pre-market skill, such as AFQT score. However, they re-emerge once you account for education, as in Lang and Manove (2011). (Conditional on AFQT score, black individuals in the NLSY79 have higher levels of education.) I use specification (4) from Table 2.A3 and Table 2.A4 to form residuals for later analysis.

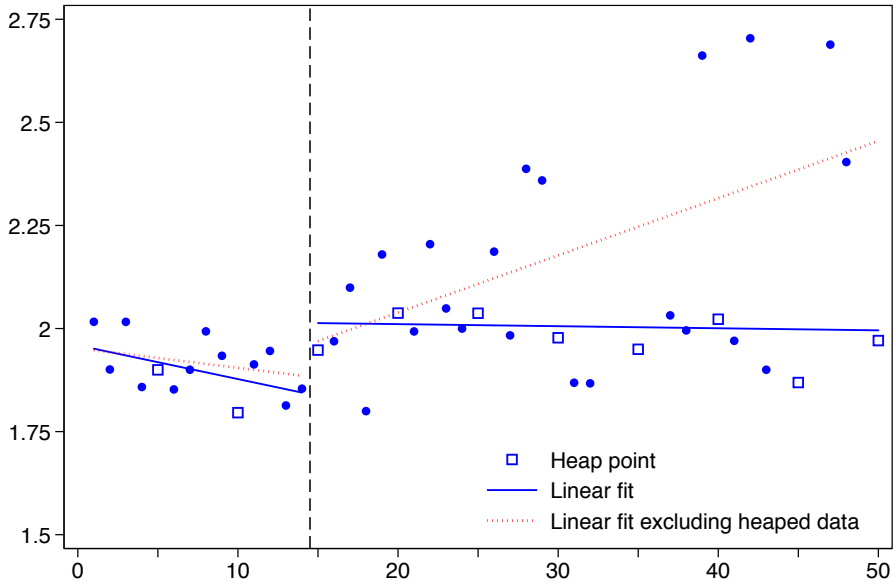
2.4.3.1 Title VII Coverage

Figure 2.4 summarises the identification strategy and baseline results. The figure plots log real hourly wages for white and black men as a function of firm employment, with linear trends fitted separately below and above the cutoff for Title VII coverage. The plot points are binned averages with bin width equal to one, but the fitted trend lines use the underlying data. We can see from the figure that hourly wages for black men jump at the cutoff, whereas hourly wages for white men do not. White wages vary continuously through the cutoff, although there is a "kink" point in their slope. I identify observations that fall at heap points to understand how heaping in the assignment variable—firm employment—may affect these estimates. The top panel shows that dropping the heaped data decreases the regression discontinuity estimate of the effect of Title VII coverage on wages for black men (compare dotted lines with solid lines at the dashed cutoff). Heaping does not appear to affect the estimates for white men. Visually, what I estimate is the discontinuity for black men (top panel) minus the discontinuity for white men (bottom panel). This is the black-white difference in treatment effect at the first cutoff, $\tau_{b1} - \tau_{w1}$, described earlier.

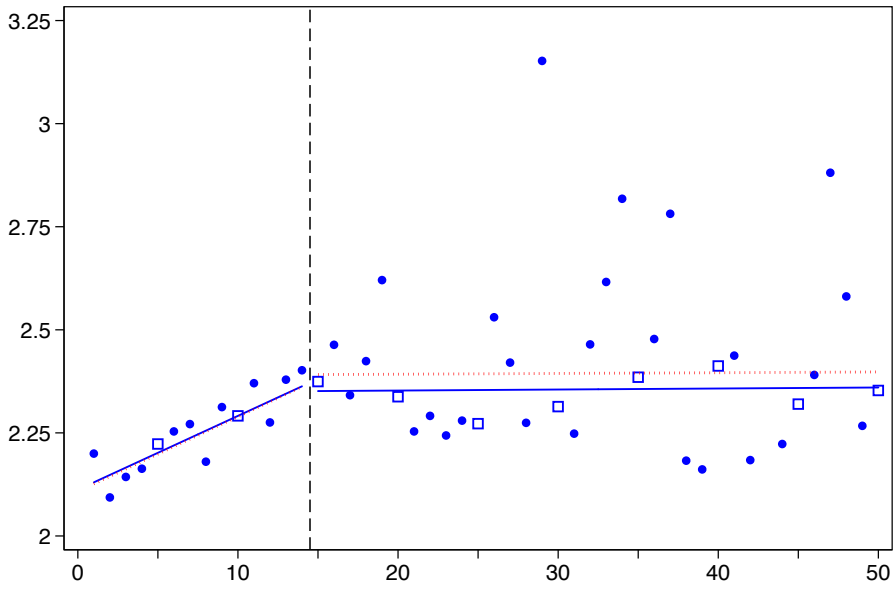
Regression discontinuity estimates of the effect of Title VII coverage on black-white relative wages can be found in Table 2.1. The table shows estimates from local

Figure 2.4: Wages and Title VII Coverage: Detail (Men)

(a) Black Men



(b) White Men



Firm Employment

Source: NLSY79, 1994–2010

Note: Binned averages of log real hourly wages with bin width equal to 1. Linear fit uses underlying data. Sample restricted as described in Section 2.4.2.2. Dashed line indicates cutoff in Title VII coverage.

linear regression using a rectangular (uniform) kernel.[14] The baseline specification uses log real hourly wages as the dependent variable, and it does not condition on any covariates. We can see from the table that the estimates are sensitive to the choice of bandwidth, with smaller bandwidths producing the most volatile estimates. A baseline estimate, using the central bandwidth of $h = 7$, is that Title VII coverage increases wages for black men relative to white men by 0.174 log points, or 19 percent, at the cutoff in coverage. However, we cannot reject the hypothesis that the effect is zero. The estimate increases to 0.245 log points after halving the bandwidth and decreases to 0.131 after doubling the bandwidth.

The other estimates in Table 2.1 condition on a set of covariates or residualise wages as described in the previous section. A specification that conditions on covariates but does not account for mismeasured treatment or heaping in the assignment variable puts the effect of Title VII coverage on black-white relative wages for men at 0.115 log points, or 12.3 percent, somewhat lower than the baseline specification. Again, the estimates are rather sensitive to the choice of bandwidth, and we cannot reject the hypothesis of zero effect. Using residualised wages as the dependent variable rather than including covariates produces similar results.

To account for rounding in employer size, which leads to heaping in the assignment variable and possibly mismeasured treatment, I compute the alternative estimates (a)–(e) in Table 2.1. Each of these estimates conditions on covariates. The first randomly replaces observations at the 15-employee heap to match the implied (smooth) distribution of firms. I repeat this 100 times and calculate the average treatment effect, which equals or exceeds the previous estimate, depending on the bandwidth. To compute standard errors for this specification, I average t-statistics. Although this approach

[14] A rectangular kernel means that observations within the bandwidth are weighted equally regardless of how close they are to the point of interest, whereas a triangular or Epanechnikov kernel would weight observations closer to the point of interest more heavily. See Section 2.4.2 for reasons to use a rectangular kernel. However, I should note that NLSY sample weights are used in this analysis, so individual observations are not weighted equally in the end.

Table 2.1: RD Estimates of the Effect of Title VII Coverage on Log Real Hourly Wages for Black Men Relative to White Men: Detail

Black-white difference in treatment effect at cutoff c_k estimated using bandwidth h			
$c_1 = 15$, $h = 7$	$h/2$	h	$2h$
Baseline	0.245	0.174	0.131
	(0.217)	(0.136)	(0.103)
Including covariates	0.238	0.115	0.127
	(0.212)	(0.128)	(0.096)
(a) Random resampling	0.274	0.115	0.138
	(0.242)	(0.147)	(0.106)
	[525]	[1,573]	[4,045]
(b) "Donut RD"	0.649	0.127	0.146
	(0.323)	(0.159)	(0.115)
	[467]	[1,515]	[3,987]
(c) Non-heaped observations		0.037	0.063
		(0.160)	(0.140)
		[954]	[2,860]
(d) Heaped observations			0.321
			(0.215)
			[1,127]
(e) Weighted average of (c) and (d)			0.135
			(0.088)
			[3,987]
Residuals	0.243	0.121	0.119
	(0.211)	(0.126)	(0.096)
Observations	684	1,732	4,204
Individuals	462	948	1,698

Source: NLSY79, 1994–2010

Note: Estimates of $\tau_{b1} - \tau_{w1}$ for men using local linear regression and a rectangular (uniform) kernel. Heteroscedasticity-robust standard errors clustered by individual reported in parentheses. Number of observations reported in brackets where different from column total. Sample restricted as described in Section 2.4.2.2. Estimates that include covariates or that use residuals as the dependent variable condition on the covariates in Table 2.A3.

lessens the heap near the cutoff, it does not account for the possibility that treatment is mismeasured for individuals who report that the number of persons at the place where they work is 15. The next specification ("donut RD") drops all observations at the 15-employee heap. This increases the estimates further, especially for the smallest bandwidth. Note that standard errors also increase. The estimated effect of Title VII coverage on the hourly wages of black men relative to white men grows to 0.127 log points, or 13.5 percent, using the central bandwidth, $h = 7$.

Following Barreca, Lindo, and Waddell (2016), I estimate the effect of Title VII separately for data at heap points and data not at heap points. The latter yields a consistent estimate of the treatment effect for non-heaped types, and the former yields a consistent estimate of the average treatment effect across the two types, weighted by the share of non-heaped data observed at data heaps. For each of these, I continue to exclude data at the cutoff ("donut RD"). Dropping the heaped data, the estimates of the effect of Title VII for men decrease considerably, to 0.037 and 0.063 log points using bandwidths \$h\$ and $2h$, respectively. (Note that I do not include the estimate using bandwidth $h/2$ in Table 2.1 because it coincides exactly with the "donut RD" estimate in the previous row.) This decrease in the estimated effect is consistent with Figure 2.4. To estimate the effect of Title VII coverage using data only at heap points, I must use the largest bandwidth, $2h$. Consistent with the heap points in Figure 2.4, this estimate is larger, at 0.321 log points for men. To recover an estimate of the unconditional average treatment effect across heaped and non-heaped data, I can compute a weighted average of the two estimates using population sample weights. This equals 0.135 log points, or 14.5 percent, for estimates using a bandwidth equal to $2h$. The standard error is computed using the delta method.

Restricting our attention to the right two columns of Table 2.1 and to specifications that include covariates or residualise wages, we can see that estimates of the effect of Title VII coverage on hourly wages for black men relative to white men range

from 0.115 to 0.146 log points, or 12.2–15.7 percent. Note that the optimal bandwidth procedures described in the previous section suggest using the larger bandwidth (rightmost column). An estimate such as (e), at 14.5 percent, accounts for the possibility of mismeasured treatment and for heaping in the assignment variable as a result of rounding in firm employment. Estimates in this range amount to sizeable (but not unreasonable) gains in black wages relative to white wages that result from Title VII coverage. However, the estimates do not differ significantly from zero.

A note about the interpretation of these estimates: If the raw black-white wage gap for men was –0.331 log points, or 28.2 percent, and we estimated that Title VII coverage increases black wages relative to white wages by 0.135 log points, or 14.5 percent, it is wrong to conclude that Title VII coverage roughly halves the black-white wage gap for men, for two reasons. First, the regression discontinuity design estimates the effect of treatment *at the cutoff*. This is a key feature, or weakness, of the design. While the design has strong internal validity, the resulting estimates cannot simply be applied to the overall black-white wage gap for men. Second, the estimate of the change in black wages relative to white wages of 14.5 percent is indeed a percent change in relative wages, not a percentage point change in the black-white wage gap. That is, if the ratio of black wages to white wages for men is 0.718 at the cutoff, I would estimate that Title VII increases this ratio by 14.5 percent to 0.822, a 10.4 percentage point decrease in the black-white wage gap (from 28.2 percent to 17.8 percent).

The bottom of Table 2.1 lists the number of observations and the number of distinct individuals for each bandwidth. (I pool data from the 1994–2010 waves of the NLSY79.) There are, on average, 1.5 to 2.5 observations per individual, depending on the bandwidth, and the distribution of observations per individual is more skewed within smaller bandwidths. This implies "between" variation drives the estimates. Where the number of observations differs from the column total in Table 2.1, e.g., "donut RD" or heaped and non-heaped specifications, I include it in brackets.

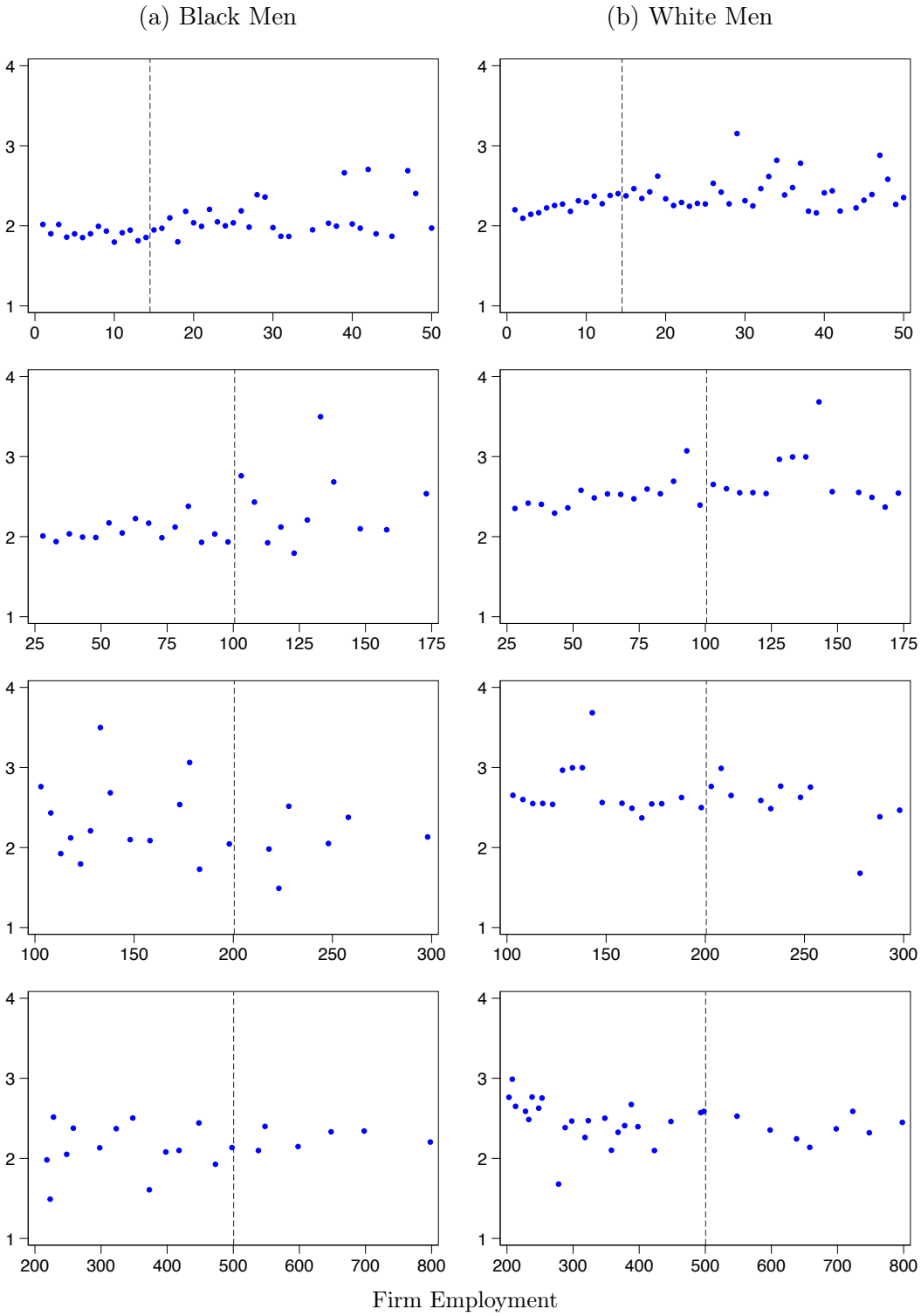
Figure 2.A1 and Table 2.A1 in the Appendix show the analogous results for women. In this results section, I focus on men because Title VII applies to black men but not white men, whereas both black women and white women are covered to some degree under the law. (The law provides legal remedies for gender discrimination in addition to race discrimination.) Obviously, black women have more avenues for litigation under Title VII because of their race, but there is not a clean difference like there is for men. (Also, there are other laws enforced by the EEOC that address discrimination based on gender, such as the Equal Pay Act and the Pregnancy Discrimination Act.) Since both black women and white women can bring charges of discrimination under Title VII, it is perhaps not surprising that the estimates in Table 2.A1 are smaller than the corresponding estimates for men. Often they are close to zero, except when the bandwidth is small or when we drop heaped observations (which is clear from Figure 2.A1). We cannot reject the hypothesis that the effect of Title VII coverage on hourly wages for black women relative to white women is zero.

2.4.3.2 Title VII Liability

Figure 2.5 plots log real hourly wages for white and black men as a function of firm employment around all cutoffs in Title VII liability following the Civil Rights Act of 1991. (The top panel repeats the plots in Figure 2.4.) The plot points are binned averages with bin width equal to 1 for the top panel and bin width equal to 5 for the bottom three panels. I do not distinguish observations at heap points, since these constitute nearly all of the data around the higher cutoffs (see Figure 2.2 above). From looking at Figure 2.5, there are not any obvious discontinuities in black or white wages at the higher cutoffs in Title VII liability. What is clear is that hourly wages are more volatile around the higher cutoffs, especially for black men.

Table 2.2 shows estimates of the effect of increased Title VII liability at these higher cutoffs for different choices of bandwidth. The central estimates for the cutoff

Figure 2.5: Wages and Title VII Liability: All Cutoffs (Men)



Source: NLSY79, 1994–2010

Note: Binned averages of log real hourly wages with bin width equal to 1 (top panel) and bin width equal to 5 (bottom three panels). Sample restricted as described in Section 2.4.2.2. Dashed lines indicate cutoffs in Title VII liability following the Civil Rights Act of 1991.

Table 2.2: RD Estimates of the Effect of Title VII Liability on Log Real Hourly Wages for Black Men Relative to White Men: Upper Cutoffs

Black-white difference in treatment effect at cutoff c_k estimated using bandwidth h			
$c_2 = 101$, $h = 43$	$h/2$	h	$2h$
Baseline	0.526	0.376	-0.011
	(0.361)	(0.284)	(0.167)
Including covariates	0.109	0.119	-0.026
	(0.327)	(0.229)	(0.129)
(b) "Donut RD"	0.978	0.371	0.025
	(0.443)	(0.269)	(0.142)
	[251]	[713]	[2,712]
Residuals	-0.026	0.080	-0.053
	(0.265)	(0.206)	(0.121)
Observations	469	931	3,147
Individuals	339	577	1,437
$c_3 = 201$, $h = 50$			
Baseline	1.090	0.187	-0.229
	(2.141)	(0.776)	(0.277)
Including covariates	-0.057	-0.241	-0.114
	(2.325)	(0.511)	(0.201)
Residuals	0.064	0.011	0.004
	(2.273)	(0.548)	(0.209)
Observations	186	286	711
Individuals	152	219	462
$c_4 = 501$, $h = 150$			
Baseline		0.029	0.120
		(0.248)	(0.286)
Including covariates		-0.171	0.044
		(0.301)	(0.231)
Residuals		-0.279	0.027
		(0.309)	(0.221)
Observations		238	537
Individuals		169	339

Source: NLSY79, 1994–2010

Note: Estimates of $\tau_{bk} - \tau_{wk}$, $k = 2, 3, 4$, for men using local linear regression and a rectangular (uniform) kernel. Heteroscedasticity-robust standard errors clustered by individual reported in parentheses. Number of observations reported in brackets where different from column total. Sample restricted as described in Section 2.4.2.2. Estimates that include covariates or that use residuals as the dependent variable condition on the covariates in Table 2.A3.

$c_2 = 101$ are quite close to the results for overall Title VII coverage at cutoff $c_1 = 15$ from Table 2.1, once you include covariates or use residualised wages as the dependent variable. Doing so makes sense at these higher levels since the bandwidths are so much larger. The central estimates of the effect of increasing the limit on punitive and compensatory damages to $100,000 for firms with 101 employees on wages for black men relative to white men are 0.119 and 0.080 log points, once you include covariates or residualise wages, respectively. However, these estimates are sensitive to the choice of bandwidth and to excluding observations with firm size equal to 100 employees, as shown by the "donut RD" row in Table 2.2. Still, most of the estimates for the 101-employee cutoff are positive or close to zero.

The 201- and 501-employee cutoffs, on the other hand, produce some estimates that do not have the predicted sign. The standard errors for these estimates are large, and we cannot reject the hypothesis of no effect. Empty cells in Table 2.2 indicate that there were not enough observations within the given bandwidth for estimation. Returning to Figure 2.5, it is perhaps not surprising that we find no effect (or some noisy effect) around these upper cutoffs, which is consistent with the figure. Overall, there is less evidence that higher liability under Title VII increases wages for black men relative to white men than there is that Title VII coverage, which begins at 15 employees, has that effect. Figure 2.A2 and Table 2.A2 in the Appendix show estimates of the effect of Title VII liability for women, and they tell the same story. Figure 2.A2 shows no discernible effect of the higher cutoffs on the wages of black women relative to white women.

2.5 Conclusion

What are the wage effects of liability for employment discrimination? In an equilibrium search and matching model with discrimination in the labour market, a

law that prohibits discrimination in hiring, firing, and pay has an ambiguous effect on wages. The net effect depends on the relative size of firing and (non-)hiring costs introduced by the law and the strength of the equal pay provision. This is roughly the same as the standard competitive model.

A number of empirical studies have attempted to estimate the effect of anti-discrimination laws in the United States. Doing so is difficult because these laws typically have wide coverage. For this reason, much of the work on this topic uses "before and after" comparisons. This essay aims to contribute to this literature by proposing a new empirical strategy. A sharp regression discontinuity design could exploit cutoffs in Title VII coverage and Title VII liability following the Civil Rights Act of 1991. Title VII coverage applies to firms with 15 or more employees, and firms with 101, 201, and 501 or more employees face higher damages in discrimination lawsuits.

This project will aim to use confidential matched employer-employee data from the U.S. Census Bureau. In the meantime, this essay uses survey data to estimate the effects of coverage and liability nonparametrically using local linear regression. If the assumptions for identification hold, it obtains consistent estimates of the treatment effect of the law. Identification requires that workers do not precisely sort into covered firms and that firms do not precisely manipulate their employment to avoid coverage or larger penalties under the law. Administrative data on the distribution of firms in the United States suggest that the second assumption holds, despite the existence of heaps in the data near cutoffs in coverage and liability. Identification also requires that state antidiscrimination laws or other laws that could affect relative wages do not apply discontinuously at the thresholds in Title VII coverage or liability under the Civil Rights Act of 1991. Several states do have fair employment laws with cutoffs that coincide with federal antidiscrimination law, which may bias the estimates upward.

A preliminary estimate is that Title VII coverage increases black relative wages by 0.115 to 0.146 log points, or 12 to 16 percent, after including covariates or resid-

ualising wages, but these estimates are sensitive to shrinking the bandwidth, and they do not differ significantly from zero. This essay provides several alternative specifications to check whether the estimates are robust to measurement error in the assignment variable. These include dropping observations near the cutoff ("donut RD") and estimating the treatment effect separately for heaped and non-heaped data and computing a population-weighted average. An estimate that does both of these suggests that Title VII coverage increases the wages of black men relative to white men by 0.135 log points, or 14.5 percent, at the cutoff. The magnitude of the estimate for the effect of coverage on relative wages for black women is smaller, likely because white women have some protections under Title VII. There is not strong evidence that greater Title VII liability, that is, higher damages allowable in discrimination lawsuits, affects black relative wages. However, data around these upper cutoffs are especially noisy.

Overall, these estimates suggest that Title VII coverage may increase wages for black men relative to white men, but higher damages awardable in discrimination lawsuits do not have an effect. This would make sense if overall coverage under federal antidiscrimination law, which allows for charges of discrimination, EEOC investigations, and possibly litigation, affects the behaviour of prejudiced firms, but an increase in damages awardable in court from $50,000 to $100,000 (which may translate into an even smaller increase in the expected cost of discrimination), for example, does not. With a more accurate measure of firm employment using matched employer-employee data and a larger sample of workers, more precise estimates could be obtained, and something more definitive about the effect of the damages caps could be said. Moreover, data with geographic identifiers would allow a researcher to account for interactions with state antidiscrimination laws. Further research will aim to use of confidential data held by the U.S. Census Bureau for this purpose.

2.6 References

Acemoglu, Daron, and Joshua D. Angrist. 2001. "Consequences of Employment Protection? The Case of the Americans with Disabilities Act." *Journal of Political Economy* 109 (5): 915–57.

Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review* 30 (2): 175–87.

Akerlof, George. 1985. "Discriminatory, Status-Based Wages Among Tradition-Oriented, Stochastically Trading Coconut Producers." *Journal of Political Economy* 93 (2): 265–76.

Albrecht, James, and Boyan Jovanovic. 1986. "The Efficiency of Search Under Competition and Monopsony." *Journal of Political Economy* 94 (6): 1246–57.

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics*, vol. 3, edited by Orley Ashenfelter and David Card, 3143–259. Amsterdam: North-Holland.

Altonji, Joseph G., and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116, no. 1 (February): 313–50.

Arrow, Kenneth J. 1972. "Some Mathematical Models of Race Discrimination in the Labor Market." In *Racial Discrimination in Economic Life*, edited by Anthony H. Pascal, 187–204. Lexington, Massachusetts: Lexington Books.

———. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton, New Jersey: Princeton University Press.

Axtell, Robert. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science* 293 (5536): 1818–20.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell. 2016. "Heaping-Induced Bias in Regression-Discontinuity Designs." *Economic Inquiry* 54, no. 1 (January): 268–93.

Becker, Gary S. 1957. *The Economics of Discrimination.* Chicago: University of Chicago Press.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94, no. 4 (September): 991–1013.

Black, Dan A. 1995. "Discrimination in an Equilibrium Search Model." *Journal of Labor Economics* 13, no. 2 (April): 309–34.

Blau, Francine D., and Lawrence M. Kahn. 2000. "Gender Differences in Pay." *Journal of Economic Perspectives* 14 (4): 75–99.

Borjas, George J., and Stephen G. Bronars. 1989. "Consumer Discrimination and Self-Employment." *Journal of Political Economy* 97 (3): 581–605.

Bowlus, Audra J., and Zvi Eckstein. 2002. "Discrimination and Skill Differences in an Equilibrium Search Model." *International Economic Review* 43, no. 4 (November): 1309–45.

Butler, Richard, and James J. Heckman. 1997. "The Government's Impact on the Labor Market Status of Black Americans: A Critical Review." In *Equal Rights and Industrial Relations*, edited by Leonard J. Hausman, Orley Ashenfelter, Bayard Rustin, Richard F. Schubert, and Donald Slaiman, 235–81. Madison, Wisconsin: Industrial Relations Research Association.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82, no. 6 (November): 2295–326.

Card, David, and Alan B. Krueger. 1992. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Quarterly Journal of Economics* 107, no. 1 (February): 151–200.

Carneiro, Pedro, James J. Heckman, and Dimitriy V. Masterov. 2005. "Labor Market Discrimination and Racial Differences in Premarket Factors." *Journal of Law and Economics* 48, no. 1 (April): 1–39.

Carrington, William J., Kristin McCue, and Brooks Pierce. 2000. "Using Establishment Size to Measure the Impact of Title VII and Affirmative Action." *Journal of Human Resources* 35 (3): 503–23.

Charles, Kerwin Kofi, and Jonathan Guryan. 2008. "Prejudice and Wages: An Empirical Assessment of Becker's *The Economics of Discrimination*." *Journal of Political Economy* 116 (5): 773–809.

———. 2011. "Studying Discrimination: Fundamental Challenges and Recent Progress." *Annual Review of Economics* 3:479–511.

Chay, Kenneth Y. 1998. "The Impact of Federal Civil Rights Policy on Black Economic Progress: Evidence from the Equal Employment Opportunity Act of 1972." *Industrial and Labor Relations Review* 51, no. 4 (July): 608–32.

Chiswick, Barry R. 1973. "Racial Discrimination in the Labor Market: A Test of Alternative Hypotheses." *Journal of Political Economy* 81 (6): 1330–52.

Corcoran, Mary, and Greg J. Duncan. 1979. "Work History, Labor Force Attachment,

and Earnings Differences between the Races and Sexes." *Journal of Human Resources* 14 (1): 3–20.

Darity, William A., and Patrick L. Mason. 1998. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." *Journal of Economic Perspectives* 12 (2): 63–90.

Dong, Yingying. 2015. "Regression Discontinuity Applications with Rounding Errors in the Running Variable." *Journal of Applied Econometrics* 30:422–46.

Donohue, John J. 2007. "Antidiscrimination Law." In *Handbook of Law and Economics*, vol. 2, edited by A. Mitchell Polinsky and Steven Shavell, 1387–472. Amsterdam: North-Holland.

Donohue, John J., and James J. Heckman. 1991. "Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks." *Journal of Economic Literature* 29, no. 4 (December): 1603–43.

Eckstein, Zvi, and Kenneth I. Wolpin. 1999. "Estimating the Effect of Racial Discrimination on First Job Wage Offers." *Review of Economics and Statistics* 81, no. 3 (August): 384–92.

Epstein, Richard A. 1992. *Forbidden Grounds: The Case Against Employment Discrimination Laws*. Cambridge, Massachusetts: Harvard University Press.

Equal Employment Opportunity Commission (EEOC). 2018. Litigation Statistics, FY1997 to FY2017. Accessed 30 September 2018. http://www.eeoc.gov/eeoc/statistics.

Fan, Jianqing, and Irene Gijbels. 1996. *Local Polynomial Modelling and Its Implications*. London: Chapman and Hall.

Freeman, Richard B. 1973. "Changes in the Labor Market for Black Americans, 1948–72." *Brookings Papers on Economic Activity* 1973 (1): 67–120.

Fryer, Roland G., Jr., and Matthew O. Jackson. 2008. "A Categorical Model of Cognition and Biased Decision Making." *The B.E. Journal of Theoretical Economics* 8 (1): 1–42.

Fryer, Roland G., Jr., Devah Pager, and Jörg L. Spenkuch. 2013. "Racial Disparities in Job Finding and Offered Wages." *Journal of Law and Economics* 56, no. 3 (August): 633–89.

Fujiwara, Yoshi, Corrado Di Guilmi, Hideaki Aoyama, Mauro Gallegati, and Wataru Souma. 2004. "Do Pareto–Zipf and Gibrat Laws Hold True? An Analysis with European Firms." *Physica A* 335:197–216.

Gabaix, Xavier. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics* 1:255–93.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123, no. 1 (February): 49–99.

Garicano, Luis, Claire Lelarge, and John Van Reenen. 2016. "Firm Size Distortions and the Productivity Distribution: Evidence from France." *American Economic Review* 106, no. 11 (November): 3439–79.

Goldberg, Matthew. 1982. "Discrimination, Nepotism, and Long-Run Wage Differentials." *Quarterly Journal of Economics* 97, no. 2 (May): 307–19.

Goldin, Claudia, and Cecilia Rouse. 2001. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review* 90, no. 4 (September): 715–41.

Gourio, François, and Nicolas Roys. 2014. "Size-Dependent Regulations, Firm Size Distribution, and Reallocation." *Quantitative Economics* 5, no. 2 (July): 377–416.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 1999. "Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design." NBER Working Paper No. 7131.

———. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69, no. 1 (January): 201–09.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79, no. 3 (July): 933–59.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142, no. 2 (February): 615–35.

Johnson, George E., and Frank P. Stafford. 1998. "Alternative Approaches to Occupational Exclusion." In *Women's Work and Wages*, edited by Inga Persson and Christina Jonung. London: Routledge.

Lang, Kevin, and Jee-Yeon K. Lehmann. 2012. "Racial Discrimination in the Labor Market: Theory and Empirics." *Journal of Economic Literature* 50, no. 4 (December): 959–1006.

Lang, Kevin, and Michael Manove. 2011. "Education and Labor Market Discrimination." *American Economic Review* 101, no. 4 (June): 1467–96.

Lang, Kevin, Michael Manove, and William T. Dickens. 2005. "Racial Discrimination in Labor Markets with Posted Wage Offers." *American Economic Review* 95, no. 4

(September): 1327–40.

Lazear, Edward P. 1990. "Job Security Provisions and Employment." *Quarterly Journal of Economics* 105, no. 3 (August): 699–726.

Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics* 142, no. 2 (February): 675–97.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48, no. 2 (June): 281–355.

Ludwig, Jens, and Douglas L. Miller. 2005. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." NBER Working Paper No. 11702.

McAdams, Richard H. 1995. "Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination." *Harvard Law Review* 108, no. 5 (March): 1003–84.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142, no. 2 (February): 698–714.

Neal, Derek A. 2004. "The Measured Black-White Wage Gap Among Women Is Too Small." *Journal of Political Economy* 112, no. S1 (February): S1–28.

Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104, no. 5 (October): 869–95.

O'Neill, June. 1990. "The Role of Human Capital in Earnings Differences Between Black and White Men." *Journal of Economic Perspectives* 4 (4): 25–45.

Oettinger, Gerald S. 1996. "Statistical Discrimination and the Early Career Evolution of the Black-White Wage Gap." *Journal of Labor Economics* 14, no. 1 (January): 52–78.

Oyer, Paul, and Scott Schaefer. 2000. "Layoffs and Litigation." *RAND Journal of Economics* 31 (2): 345–58.

———. 2002. "Sorting, Quotas, and the Civil Rights Act of 1991: Who Hires When It's Hard to Fire?" *Journal of Law and Economics* 45, no. 1 (April): 41–68.

Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74, no. 5 (October): 777–99.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62, no. 4 (September): 659–61.

Reimers, Cordelia. 1983. "Labor Market Discrimination Against Hispanic and Black Men." *Review of Economics and Statistics* 65, no. 4 (November): 570–79.

Riach, Peter A., and Judith Rich. 2006. "An Experimental Investigation of Sexual Discrimination in Hiring in the English Labor Market." *The B.E. Journal of Economic Analysis & Policy* 5 (2): 1–22.

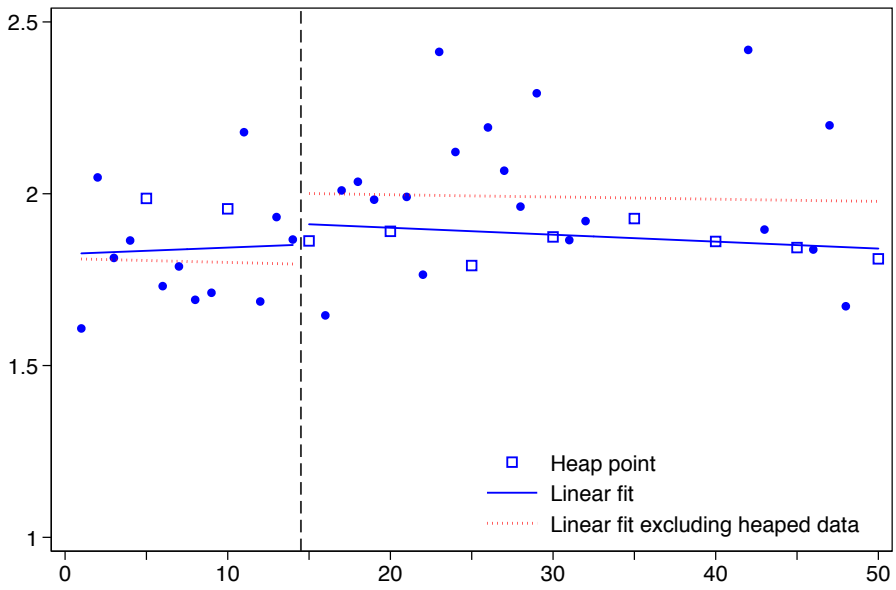
Smith, James P., and Finis R. Welch. 1989. "Black Economic Progress After Myrdal." *Journal of Economic Literature* 27, no. 2 (June): 519–64.

Thistlethwaite, Donald L., and Donald T. Campbell. 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology* 51 (6): 309–17.

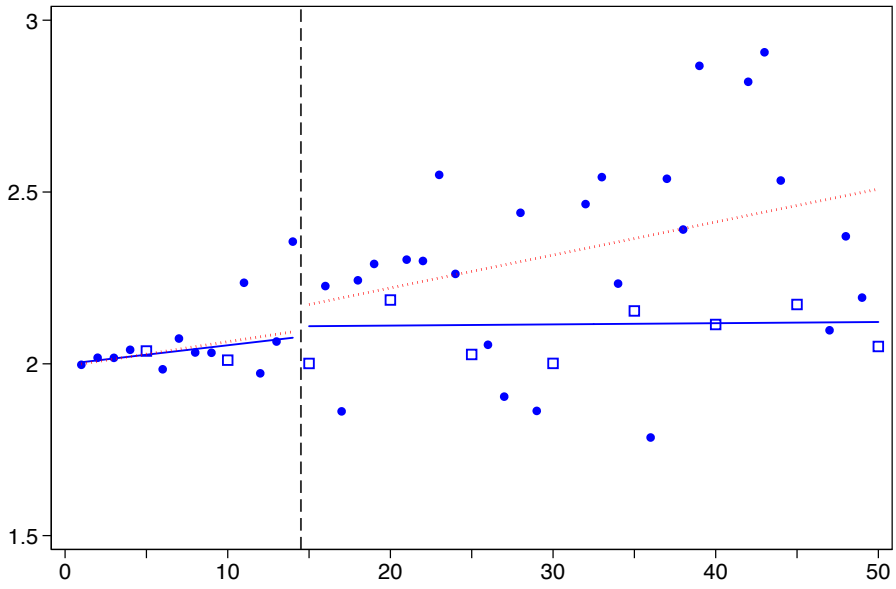
2.7 Appendix

Figure 2.A1: Wages and Title VII Coverage: Detail (Women)

(a) Black Women



(b) White Women

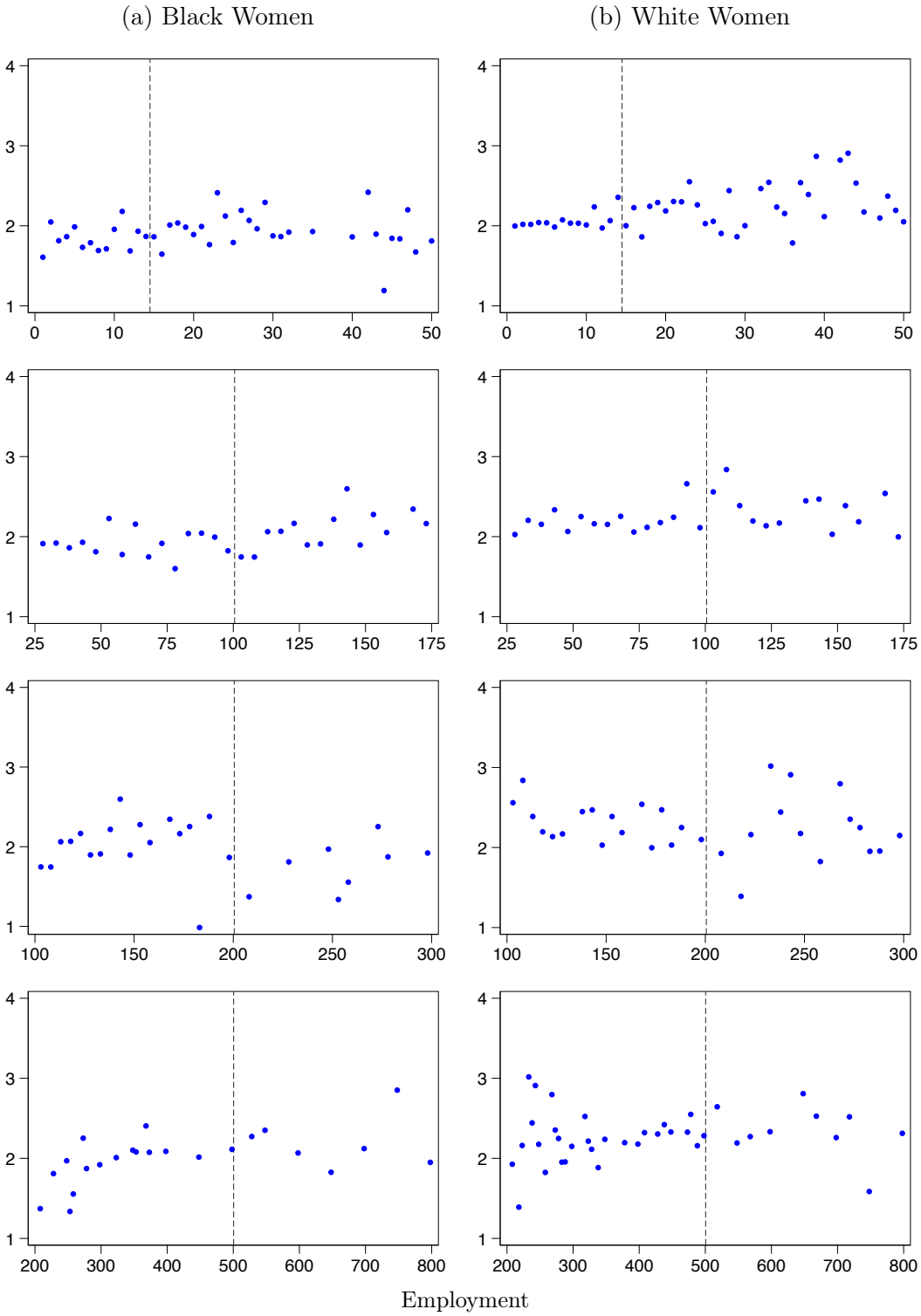


Employment

Source: NLSY79, 1994–2010

Note: Binned averages of log real hourly wages with bin width equal to 1. Linear fit uses underlying data. Sample restricted as described in Section 2.4.2.2. Dashed line indicates cutoff in Title VII coverage.

Figure 2.A2: Wages and Title VII Liability: All Cutoffs (Women)



Source: NLSY79, 1994–2010

Note: Binned averages of log real hourly wages with bin width equal to 1 (top panel) and bin width equal to 5 (bottom three panels). Sample restricted as described in Section 2.4.2.2. Dashed lines indicate cutoffs in Title VII liability following the Civil Rights Act of 1991.

Table 2.A1: RD Estimates of the Effect of Title VII Coverage on Log Real Hourly Wages for Black Women Relative to White Women: Detail

Difference in treatment effect at cutoff c_k estimated using bandwidth h			
$c_1 = 15$, $h = 7$	$h/2$	h	$2h$
Baseline	0.340	0.054	0.024
	(0.297)	(0.164)	(0.122)
Including covariates	0.311	-0.062	-0.001
	(0.249)	(0.141)	(0.105)
(a) Random resampling	0.250	-0.046	0.012
	(0.287)	(0.178)	(0.115)
	[409]	[1,305]	[3,125]
(b) "Donut RD"	0.183	0.019	0.024
	(0.360)	(0.198)	(0.129)
	[367]	[1,263]	[3,083]
(c) Non-heaped observations		0.067	0.080
		(0.197)	(0.144)
		[809]	[2,210]
(d) Heaped observations			-0.196
			(0.279)
			[873]
(e) Weighted average of (c) and (d)			0.003
			(0.095)
			[3,083]
Residuals	0.292	-0.041	0.006
	(0.237)	(0.138)	(0.104)
Observations	570	1,466	3,286
Individuals	385	859	1,456

Source: NLSY79, 1994–2010

Note: Estimates of $\tau_{b1} - \tau_{w1}$ for women using local linear regression and a rectangular (uniform) kernel. Heteroscedasticity-robust standard errors clustered by individual reported in parentheses. Number of observations reported in brackets where different from column total. Sample restricted as described in Section 2.4.2.2. Estimates that include covariates or that use residuals as the dependent variable condition on the covariates in Table 2.A4.

Table 2.A2: RD Estimates of the Effect of Title VII Liability on Log Real Hourly Wages for Black Women Relative to White Women: Upper Cutoffs

Black-white difference in treatment effect at cutoff c_k estimated using bandwidth h			
$c_2 = 101$, $h = 43$	$h/2$	h	$2h$
Baseline	-1.209	-0.482	-0.091
	(0.398)	(0.292)	(0.163)
Including covariates	-0.993	-0.298	-0.039
	(0.359)	(0.252)	(0.140)
(b) "Donut RD"	-1.041	-0.331	-0.008
	(0.584)	(0.313)	(0.160)
	[211]	[582]	[2,235]
Residuals	-1.075	-0.318	-0.049
	(0.403)	(0.262)	(0.143)
Observations	454	825	2,681
Individuals	304	523	1,357
$c_3 = 201$, $h = 50$			
Baseline		-0.353	-0.202
		(0.267)	(0.210)
Including covariates		-0.255	-0.329
		(0.283)	(0.189)
Residuals		-0.433	-0.428
		(0.258)	(0.195)
Observations		341	767
Individuals		269	496
$c_4 = 501$, $h = 150$			
Baseline		0.408	-0.077
		(0.297)	(0.259)
Including covariates		0.408	-0.046
		(0.479)	(0.238)
Residuals		0.369	-0.045
		(0.588)	(0.257)
Observations		268	616
Individuals		203	403

Source: NLSY79, 1994–2010

Note: Estimates of $\tau_{bk} - \tau_{wk}$, $k = 2, 3, 4$, for women using local linear regression and a rectangular (uniform) kernel. Heteroscedasticity-robust standard errors clustered by individual reported in parentheses. Number of observations reported in brackets where different from column total. Sample restricted as described in Section 2.4.2.2. Estimates that include covariates or that use residuals as the dependent variable condition on the covariates in Table 2.A4.

Table 2.A3: Determinants of Log Real Hourly Wages: Men (First Stage)

	Log Real Hourly Wage			
	(1)	(2)	(3)	(4)
Black	-0.386	-0.296	-0.122	-0.177
	(0.019)	(0.016)	(0.019)	(0.018)
Hispanic	-0.197	-0.076	-0.011	-0.012
	(0.023)	(0.020)	(0.021)	(0.020)
Age	0.053	0.049	0.048	0.047
	(0.015)	(0.014)	(0.014)	(0.014)
$Age^2/100$	-0.045	-0.042	-0.042	-0.041
	(0.019)	(0.018)	(0.018)	(0.018)
Urban	0.136	0.045	0.062	0.035
	(0.017)	(0.016)	(0.016)	(0.015)
Schooling (years)		0.101		0.070
		(0.004)		(0.004)
AFQT			0.255	0.138
			(0.009)	(0.011)
$AFQT^2$			0.017	-0.005
			(0.009)	(0.009)
Constant	0.977	-0.198	1.101	0.243
	(0.288)	(0.282)	(0.275)	(0.280)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	30,341	30,341	30,341	30,341
Individuals	4,445	4,445	4,445	4,445

Source: NLSY79, 1994–2010

Note: Heteroscedasticity-robust standard errors clustered by individual in parentheses. Sample restricted to individuals with real wages between $2 and $100 per hour. Regressions include indicator variables for missing values on each covariate.

Table 2.A4: Determinants of Log Real Hourly Wages: Women (First Stage)

	Log Real Hourly Wage			
	(1)	(2)	(3)	(4)
Black	-0.190	-0.134	0.056	-0.018
	(0.018)	(0.015)	(0.017)	(0.017)
Hispanic	-0.097	0.014	0.101	0.092
	(0.022)	(0.018)	(0.020)	(0.018)
Age	0.014	0.013	0.026	0.020
	(0.015)	(0.014)	(0.014)	(0.014)
$Age^2/100$	-0.017	-0.014	-0.025	-0.019
	(0.018)	(0.017)	(0.018)	(0.017)
Urban	0.140	0.076	0.096	0.070
	(0.017)	(0.014)	(0.015)	(0.014)
Schooling (years)		0.102		0.078
		(0.003)		(0.004)
AFQT			0.243	0.123
			(0.009)	(0.010)
$AFQT^2$			-0.013	-0.023
			(0.010)	(0.009)
Constant	1.698	0.346	1.391	0.520
	(0.301)	(0.286)	(0.285)	(0.283)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	28,719	28,719	28,719	28,719
Individuals	4,386	4,386	4,386	4,386

Source: NLSY79, 1994–2010

Note: Heteroscedasticity-robust standard errors clustered by individual in parentheses. Sample restricted to individuals with real wages between $2 and $100 per hour. Regressions include indicator variables for missing values on each covariate.

Chapter 3

Police Discrimination and Body Cameras: Evidence from London

3.1 Introduction

There is growing concern over police discrimination. In the United States, for example, police misconduct targeting African Americans has been on national display, from high-profile police killings of Michael Brown and Eric Gardner to documented racial bias in New York's "stop-and-frisk" policy (Gelman, Fagan & Kiss 2007). In response, the Black Lives Matter movement has started a national conversation about racial profiling and police brutality. However, activists and policymakers are still searching for solutions. Body-worn video cameras have been proposed as a means to reduce bias in policing, but the effects of the cameras are not well-understood. Existing research on body cameras has focused on general patterns of police use of force and citizen complaints (Ariel, Farrar & Sutherland 2015; Ariel et al. 2016; Jennings, Lynch & Fridell 2015; Ready & Young 2015; Yokum, Ravishankar & Coppock 2019), not discrimination.

This essay investigates discrimination in police searches and evaluates the effects of body cameras, using data from the London Metropolitan Police. In London, police officers have the power to stop and search individuals for weapons, drugs, and items to be used in crimes, among other reasons. However, officers must have "reasonable grounds for suspicion," which include an objective basis for suspicion that would lead

a reasonable person to reach the same conclusion. Physical appearance, such as ethnic appearance, gender, or age, can never support reasonable grounds, meaning an officer may not rely on "[g]eneralisations or stereotypical images that certain groups or categories of people are more likely to be involved in criminal activity" (UK Home Office 2014a).

Using administrative data on over 400,000 searches between April 2013 and April 2015, I find that, in the year prior to a body camera trial, the London police disproportionately stopped ethnic minorities and men. Black individuals made up approximately 16% of the London population but 33% of searched persons, making them 2.5 times more likely to be searched than whites. This disparity was greater for drug searches, for which black Londoners were 2.9 times more likely to be searched than whites. South Asian (predominantly Indian and Pakistani) individuals were 1.4 times more likely to be searched for drugs than whites, but approximately equally as likely to be searched overall. Men made up 94% of searched persons but only half of the overall population, making them 15.8 times more likely to be searched.

A simple comparison of search rates does not take into account the possibility that the distribution of characteristics that give rise to reasonable suspicion might vary by group. For example, if men are more likely to carry illegal weapons, the "facts, information and/or intelligence which are relevant to the likelihood that the object in question will be found"—that is, the objective basis needed for reasonable grounds for suspicion (UK Home Office 2014a)—will lead to gender disproportionality. (This is similar to, but distinct from, a claim of statistical discrimination *based on* gender.) To distinguish the possibility that police officers engaged in (unlawful) preference-based discrimination, rather than (unlawful) statistical discrimination based on ethnic appearance, gender, or age, or (lawful) statistical discrimination based on observable characteristics that correlate with ethnic appearance, gender, or age, this essay examines search success, or "hit," rates across different groups. A further possibility is that police officers en-

gage in inaccurate statistical discrimination, that is, statistical discrimination based on incorrect beliefs (Bohren et al. 2019). If police officers aim to maximise the rate at which they discover illegal behaviour, and if they do not engage in preference-based or inaccurate statistical discrimination, search success rates should be equal across groups (under certain assumptions described in Section 3.3).

Considering all searches in the year prior to a body camera trial, search success rates were on average 0.7 percentage points (p.p.) lower for black persons and 2.0 p.p. lower for South Asians than for whites, consistent with preference-based or inaccurate statistical discrimination against these ethnic minorities. However, after including a full set of control variables, the black-white difference falls to zero. That said, conditional on further action by the police, searches of black persons were 2.5 p.p. more likely to result in arrest and 2.7 p.p. less likely to result in a warning. The overall disparities between ethnic groups appear to be driven by drug searches (the most frequent kind), for which search success rates were 2.4 p.p. and 4.6 p.p. lower for black and South Asian persons, respectively, as well as searches of young persons. In general, young persons (particularly those under 20 years old) experienced lower search success rates, and the age differences were most striking for South Asians and for drug searches. With respect to gender, the search success rate was on average 6–7 p.p. lower for men than for women, and this pattern was fairly stable across different search grounds and age groups.

Using variation in the timing and location of body camera use resulting from a one-year camera trial by the London police, this essay estimates the effect of body cameras on differences in search rates and search success rates between ethnic and gender groups. The cameras slightly increased the relative frequency of searches of black persons, slightly decreased the relative frequency of searches of South Asians, and had no effect on the relative frequency of searches of men. However, controlling for borough-specific time trends eliminates these ethnicity effects, calling into question the validity of the empirical design. A triple difference-in-differences approach suggests

that, consistent with the changes in search frequency, the cameras slightly decreased the relative search success rate for black persons and slightly increased the relative success rate for South Asian persons. (And these findings are made stronger by the inclusion of borough- and group-specific time trends.) That said, the estimated effects on search success rates are somewhat small. Finally, there was an economically large and statistically significant (positive) effect on the search success rate for men relative to women, suggesting the cameras may have reduced gender discrimination. Therefore, when it comes to reducing discrimination, the usefulness of body cameras is mixed. If the primary goal is to reduce ethnic disparities, body cameras are not a cure-all.

The rest of this essay proceeds as follows. Section 3.2 provides background information about stop and search practices in London, including the legal grounds necessary for search, protections against discrimination, and reporting requirements. Section 3.3 describes the conceptual framework that guides the analysis, and Section 3.4 describes the data used throughout. Section 3.5 analyses searches and search outcomes, with an eye toward distinguishing different types of discrimination. Section 3.6 evaluates the effects of body cameras, using variation in camera use due to a one-year trial. Section 3.7 concludes.

3.2 Background

The London Metropolitan Police Service (MPS or "Met") has the power to stop and search individuals. The Police and Criminal Evidence Act (PACE) of 1984, which governs this practice, provides that police officers may conduct searches to enforce Section 1 of PACE (stolen goods and offensive weapons), Section 23 of the Misuse of Drugs Act of 1971 (drugs), Section 47 of the Firearms Act of 1968 (firearms), Section 60 of the Criminal Justice and Public Order Act of 1994 (anticipated violence), and, in rare circumstances, the Terrorism Act of 2000. Almost all searches require "reasonable

grounds for suspicion"[1] (UK Home Office 2014a). Physical appearance, such as ethnic appearance, gender, or age, can never support reasonable grounds. In particular, police officers may not rely on "[g]eneralisations or stereotypical images that certain groups or categories of people are more likely to be involved in criminal activity." In addition, the Equality Act of 2010 stipulates that officers "have a duty to have due regard to the need to eliminate unlawful discrimination, harassment and victimisation."[2]

Despite legal protections against discrimination in its application, "stop and search" is the subject of ongoing criticism of police across the United Kingdom—and the London police, in particular (London Assembly 2014; Oyeniran 2017). In 2010, the UK Equality and Human Rights Commission (EHRC) concluded that "the evidence indicates that [stop and search] may be being used in a discriminatory and unlawful manner" (EHRC 2010). The primary criticism of the practice is that it disproportionately affects ethnic minorities (Oyeniran 2017). In 2014, then–Home Secretary Theresa May remarked: "Nobody wins when stop and search is misapplied. It is a waste of police time. It is unfair, especially to young black men. It is bad for public confidence in the police" (UK Home Office 2014b). In many ways, the criticisms of stop and search in London echo those of New York City's controversial "stop-and-frisk" program, namely, that the practice entails racial profiling (Gelman, Fagan & Kiss 2007) and does not reduce crime (Rosenfeld & Fornango 2014).

By law, all stops and searches in London must be documented, electronically or on paper, on the spot. If doing so is impracticable, the officer must make a record

[1]"Reasonable grounds for suspicion" is a two-part test. First, a police officer must form "a genuine suspicion in their own mind that they will find the object for which the search power being exercised allows them to search." Second, the officer must have "an objective basis for that suspicion based on facts, information and/or intelligence which are relevant to the likelihood that the object in question will be found, so that a reasonable person would be entitled to reach the same conclusion" (UK Home Office 2014a). Section 60 searches do not require reasonable grounds, but these searches are relatively rare. (They constituted less than 1 percent of all searches between 2013 and 2015.)

[2]The Equality Act also requires police officers to have due regard for the need "to advance equality of opportunity between people who share a 'relevant protected characteristic' and people who do not share it, and to take steps to foster good relations between those persons," where "relevant protected characteristic" includes age, disability, gender reassignment, pregnancy and maternity, race, religion or belief, sex and sexual orientation (UK Home Office 2014a).

as soon as possible after the search. In addition, the officer must offer the searched person a copy of the search record, furnishing it immediately if they accept. Appendix Figure 3.A1 displays the search record (Form 5090) used by the London police for this purpose. The record must always include the date, time, and place of the person or vehicle search; the legal grounds for search; the self-defined ethnicity of the person searched ("SDE code"); and, if different, their ethnic appearance ("EA code"). These records are held by the London police as electronic data. Searched persons receive a white copy of the form in Figure 3.A1 that includes information about their rights and responsibilities, as well as instructions for filing a complaint.

Although the police are legally required to document searches, some have expressed concerns about under-reporting (London Assembly 2014). According to the MPS, about 5 percent of searches go unrecorded (*Id.*) If true, this relatively low rate is probably a result of the widely known requirement to provide a copy of the search record, which police accountability groups highlight in their educational materials (Odogwu 2012). Nonetheless, under-reporting is an important issue, one that this essay returns to in Section 3.5.

3.3 Conceptual Framework

Economic models of discrimination began with the seminal work of Becker (1957), who proposed a model of labour market discrimination based on discriminatory preferences. Phelps (1972) and Arrow (1973) later developed a theory of statistical discrimination involving inferences based on group averages. To distinguish the possibility that police officers engage in (unlawful) preference-based discrimination, rather than (unlawful) statistical discrimination based on ethnic appearance, gender, or age, or (lawful) statistical discrimination based on observable characteristics that correlate with ethnic appearance, gender, or age, this essay uses a Becker-type outcome test that

examines search success, or "hit," rates across different groups.[3] A search is defined to be successful if it results in an arrest, warning, or other action (a "hit"). The intuition behind this approach is that, if the search success rate is higher for group A than for group B, that is evidence that the threshold for suspicion is lower for group B. That is, the police have searched too many members of group B, and they could improve the rate at which they discover illegal behaviour by searching more members of group A. Note that this could be true whether group B members make up 5% or 95% of all searches.

A complication, which much of the literature ignores, is that the equality of average search success rates does not imply, and is not implied by, the equality of marginal search success rates (Anwar & Fang 2006). This is known as the "infra-marginality problem." A comparison of marginal search success rates is better because it more directly addresses the question of whether officers use a different threshold of suspicion for different groups. One way to justify using average, rather than marginal, search success rates is to assume that individual decisions to carry illegal objects are endogenously determined by group-specific search rates, such that average and marginal search success rates are equal (Knowles, Persico & Todd 2001). However, this theoretical result requires fairly restrictive assumptions. Researchers have developed tests to overcome the so-called infra-marginality problem (e.g., Anwar & Fang 2006; Antonovics & Knight 2009; Arnold, Dobbie & Yang 2018), but these tend to require demographic information about searching officers or quasi-random assignment of decision-makers. The London police have not been able to provide officer demographics at the search level, although they may provide demographic information about officers at the borough-month level. The following theoretical model suggests a method to overcome the infra-marginality problem using aggregated officer demographic information, should it become available.

[3] A further possibility is that police officers engage in inaccurate statistical discrimination, that is, statistical discrimination based on incorrect beliefs (Bohren et al. 2019). Distinguishing this possibility would require eliciting officers' preferences or observing them over time (Bohren, Imas & Rosenberg 2019).

To begin, consider the model proposed by Antonovics and Knight (2009). Suppose, for simplicity, that potential search subjects fall into one of two groups, denoted $g \in \{M, W\}$. For example, M could be "men" and W "women," M "minority" and W "white," and so on. Each subject has characteristic c, which could be informative to the police. The subject receives a pay-off of zero (or normalised to zero) if he does not carry an illegal object. If he carries an illegal object, he receives a benefit $v(c, g)$ if he is not searched, but he incurs a cost Z, distributed according to $F(z)$, no matter what, plus a cost $j(c, g)$ if he is searched. Let $\gamma^j(c, g)$ be the probability that officers from group j search subjects from group g, and let ρ be the proportion of officers with $j = M$. Then the subject has the following expected pay-off from carrying an illegal object:

$$-\gamma(c, g)j(c, g) + [1 - \gamma(c, g)]v(c, g) - Z,$$

where $\gamma(c, g) = \rho\gamma^M(c, g) + (1 - \rho)\gamma^W(c, g)$.

Suppose the police receive a pay-off of one (or normalised to one) if a search is successful. Police officers incur a group-specific cost, $t_g^j \in (0, 1)$, where j is the group of the officer and g is the group of the subject, as well as an idiosyncratic cost U, distributed according to $H(u)$. Let $\pi(c, g)$ be the probability that a subject carries an illegal object. Then the expected pay-off to an officer from a search is:

$$\pi(c, g) - t_g^j - U.$$

If officers maximise the expected pay-off of search, and subjects maximise the expected pay-off from carrying an illegal object, the following conditions must hold in equilibrium:

$$\gamma^{M*}(c, g) = H(\pi^*(c, g) - t_g^M),$$

$$\gamma^{W*}(c, g) = H(\pi^*(c, g) - t_g^W), \text{ and}$$

$$\pi^*(c,g) = F(-\gamma^*(c,g)j(c,g) + [1-\gamma^*(c,g)]v(c,g)),$$

where $\gamma^*(c,g) = \rho\gamma^{M*}(c,g) + (1-\rho)\gamma^{W*}(c,g)$.

Antonovics and Knight demonstrate that if average search costs are equal across officer groups, i.e., $t_W^W + t_M^W = t_M^M + t_W^M$, and at least one group prefers to search subjects of the other group, then search rates will differ across officer groups in equilibrium, i.e., $\gamma^{W*}(c,g) \neq \gamma^{M*}(c,g)$ for a given g. In contrast, without a discriminatory preference, search rates may differ across subject groups, i.e., $\gamma^*(c,W) \neq \gamma^*(c,M)$ (indicating statistical discrimination based on c or g or both), but not officer groups, i.e., $\gamma^{W*}(c,g) = \gamma^{M*}(c,g) = \gamma^*(c,g)$. This forms the basis for their empirical test using officer-subject mismatch.

Antonovics and Knight focus on differential search rates. Consider the following extension. First, observe that the equilibrium search success rate, $S^*(c,g)$, is decreasing in the equilibrium search rate, $\gamma^*(c,g)$. (Antonovics and Knight focus their empirical analysis on search rates, but the data set used in this essay (described in Section 3.4) contains only searches, not stops that never resulted in a search.) Second, note that if there is no preference-based discrimination, such that search rates do not differ across officer groups, then the subject group–specific search rate, $\gamma^*(c,g)$, and the subject group–specific search success rate, $S^*(c,g)$, do not depend on the fraction of M officers, ρ, even if there is statistical discrimination. On the other hand, in the presence of preference-based discrimination, for example, if $\gamma^{W*}(c,g) > \gamma^{M*}(c,g)$ for fixed g, an increase in ρ would correspond with a decrease in the average search rate, $\gamma^*(c,g) = \rho\gamma^{M*} + (1-\rho)\gamma^{W*}$, and therefore an increase in the average search success rate. This forms the basis for an empirical test using average search success rates and *aggregated* officer demographic information, should this information become available. In the absence of preference-based discrimination, the coefficient on the interaction between a group g indicator and the fraction of M officers (ρ) in a regression on search success rates should equal zero. A non-zero coefficient, then, would be indicative of

preference-based discrimination.

3.4 Data

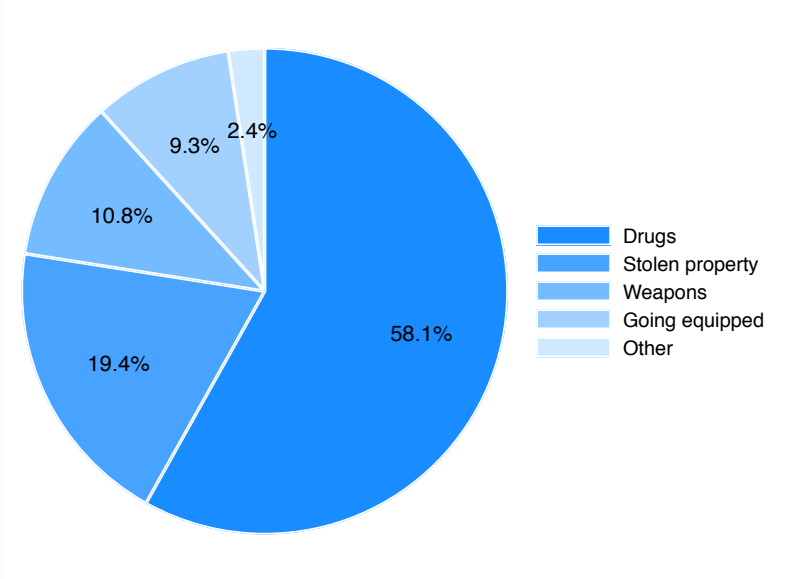
This essay uses administrative data provided by the London Metropolitan Police. The data set includes all searches recorded between April 2013 and April 2015, amounting to 472,533 observations. Each observation is an individual search. This data set is the product of the reporting requirements described in Section 3.2. Therefore, each search record contains the exact date and time of the search; search type (person, vehicle, or person and vehicle); borough location; gender, age, self-defined ethnicity, and ethnic appearance of the searched person; search grounds; and search outcome.

As Appendix Table 3.A1 shows, 79% of searches are of persons only, while 21% are of persons and vehicles, and less than 1% are of vehicles only. That is, most searches are of pedestrians, and a vehicle rarely gets searched without a person getting searched, too. Table 3.A2 displays the distribution of searches across London's thirty-two boroughs. The table shows that searches vary significantly by location. Compare, for example, Westminster, which is the home of Parliament (8.0% of all searches), and Southwark, which has one of the highest crime rates (7.1% of all searches), with Merton, the home of Wimbledon (1.3% of all searches).

Figure 3.1 displays the distribution of legal grounds for search. The largest category by far is drugs, which makes up more than half of all searches (58%), followed by stolen property (19%), weapons including firearms (11%), and "going equipped" (9%). "Going equipped" means possession of an article for use in connection with a burglary or theft. All other search grounds, including anticipated violence and terrorism, make up less than 3% of searches. Appendix Table 3.A3 lists search grounds and their frequency in more detail. The most striking feature of these data is the prevalence of drug searches, of which there were about 265,000 over the two-year period between April

2013 and April 2015.

Figure 3.1: Search Grounds



Source: London Metropolitan Police Service; author calculations.
Note: "Going equipped" means possession of an article for use in connection with any burglary or theft. "Other" includes the other search grounds listed in Table 3.A3.

Appendix Tables 3.A4 and 3.A5 list the categories for ethnic appearance and self-defined ethnicity, respectively. A couple of things are worth noting. First, ethnic appearance—that is, the ethnicity of the searched person as perceived by the police officer—will be the primary variable used in this analysis. The reason is that discrimination will be defined in terms of decisions made by police officers, based in part on their subjective beliefs. Second, as Tables 3.A4 and 3.A5 show, "Asian" in the United Kingdom means Indian, Pakistani, or Bangladeshi, while East Asian (for example, Chinese or Japanese) individuals form a much smaller group, frequently categorised together with all other ethnicities. For simplicity, and to reflect the major ethnic groups in London, the primary groups of analysis in what follows will be termed white, black, South Asian, and other. In the latest census, these groups constituted 60%, 16%, 18%, and 6% of the London population, respectively (UK Office for National Statistics 2011).

Finally, Table 3.1 displays the frequency of search outcomes. A search can result in: no action; an arrest; a warning, such as a cannabis warning or a verbal warning;

or (less frequently) some other action, such as an alcohol confiscation or dispersing of groups. Most searches result in no action (70%). Fewer result in an arrest (17%), warning (10%), or other action (3%). Appendix Table 3.A6 displays the different search outcomes in more detail. Prior to 2015, "advised" was a nuance for "no further action." However, in 2014, the MPS removed "advised" to simplify recording, leaving "no further action" for everything except formal criminal justice outcomes. The data show that, around the time of this change, "advised" disappeared and "no further action" increased by the same amount. Therefore, this essay combines the two categories into "no action," meaning no formal criminal justice outcome, such as an arrest or warning.

Table 3.1: Search Outcomes

	Freq.	Percent	Cum.
No action	317,252	69.61	69.61
Arrest	78,572	17.24	86.85
Warning	45,247	9.93	96.78
Other action	14,684	3.22	100.00
Total	455,755	100.00	

Source: London Metropolitan Police Service; author calculations.
Note: "Other action" includes the other search outcomes listed in Table 3.A6, besides "advised," which the MPS combined with "no further action" in 2014.

After removing observations with missing values or uncoded fields, the resulting data set comprises 455,755 searches between April 2013 and April 2015, 298,782 of which were completed before the body camera trial (discussed in Section 3.6), which began in May 2014. Appendix Table 3.A7 displays summary statistics for the main variables used in the following analysis.

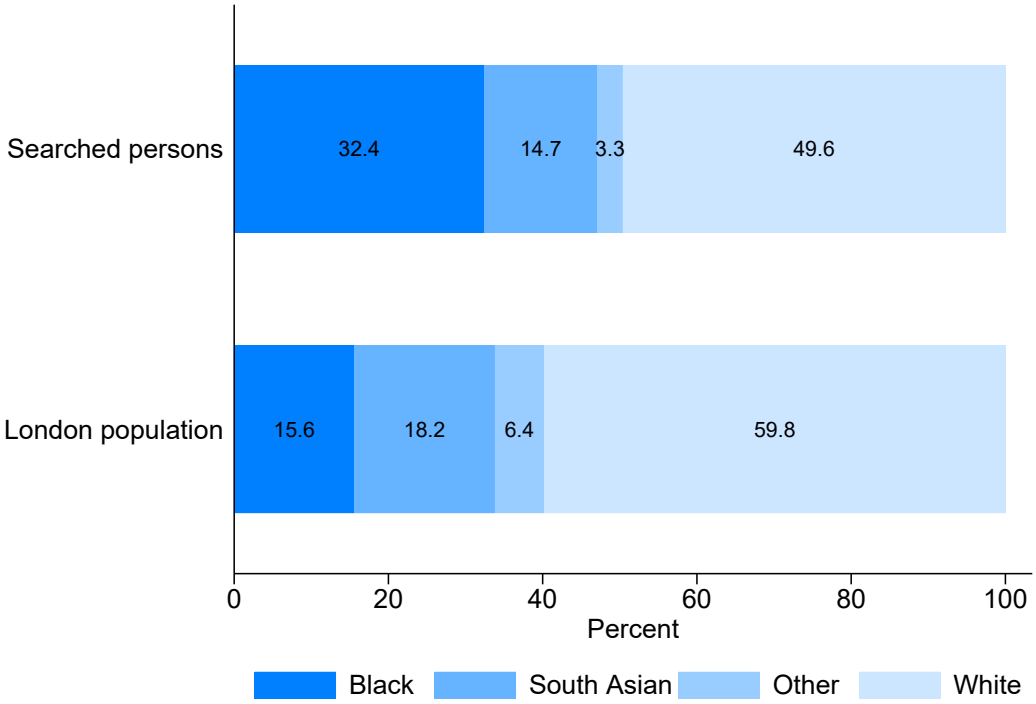
3.5 Results: Searches and Search Outcomes

To investigate discrimination in police searches, this essay begins by examining disproportionality in search rates—that is, the extent to which police officers search

members of certain groups relative to the groups' fraction of the population. The following analysis considers only searches completed before the start of the London body camera trial. Figure 3.2 displays the ethnic makeup of searched persons versus the overall London population. As the figure shows, black individuals make up only 16% of the London population but 32% of searched persons over this time period. In other words, searches disproportionately affected black Londoners. In contrast, the South Asian and white shares of searched persons (15% and 50%, respectively) were lower than the corresponding population shares (18% and 60%). These percentages can be translated into likelihood ratios using Bayes's formula. Using this formula, we see that black Londoners were 2.5 times more likely to be searched than whites. South Asians were about equally likely to be searched as whites. Figure 3.3 displays the ethnic makeup of searched persons by search grounds, with the overall London population for comparison. Here, we can see that, for all search grounds, black Londoners were stopped at rates that exceeded their population share. This disparity was greatest for weapon and drug searches. Drug searches, which were by far the most common type of search (see Figure 3.1), is also the search category for which South Asians represented their largest share (and the only category exceeding their population share). Again using Bayes's formula, black Londoners were 2.9 times and South Asians 1.4 times more likely to be searched for drugs than whites.

Search disparities in terms of gender were even greater. Figure 3.4 displays the gender makeup of searched persons versus the London population. Approximately 94% of searched persons were men, despite men making up only half of the population. This means that men were 15.8 times more likely than women to be searched. As Appendix Figure 3.A2 shows, this disparity was relatively constant across search grounds, although it was greatest for weapons and "going equipped" (i.e., possession of an article for use in a burglary) searches. It was smallest for stolen property searches, but, even for those, men exceeded 90% of searches.

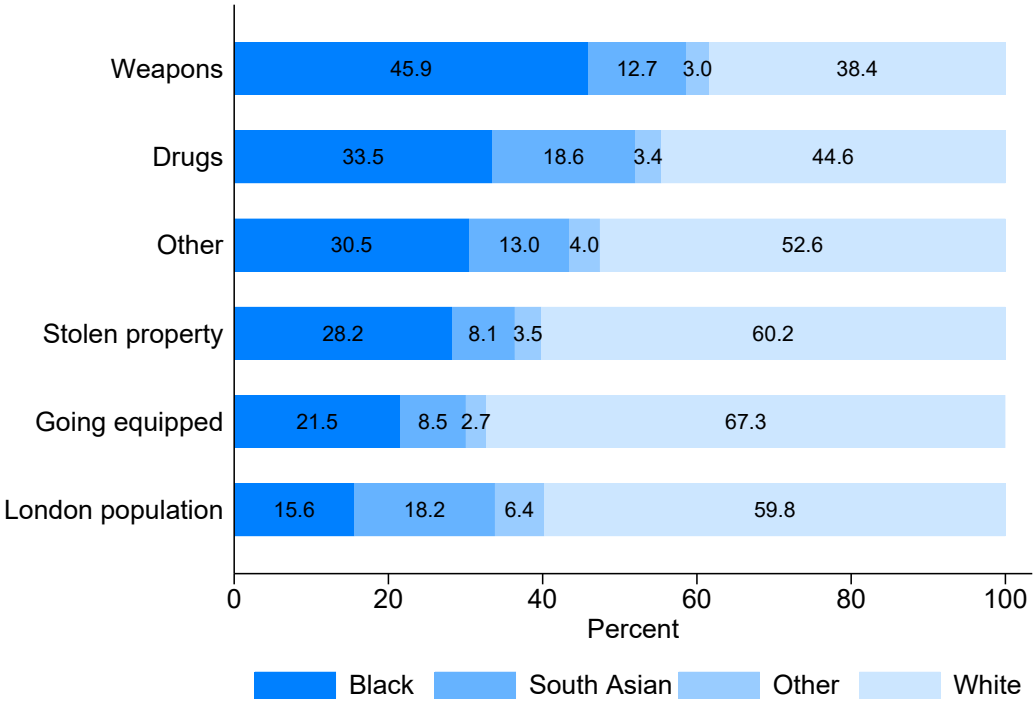
Figure 3.2: Ethnic Makeup of Searched Persons versus London Population



Source: London Metropolitan Police Service; London 2011 census; author calculations.
Note: All searches completed prior to the body camera trial, which began in May 2014. "Other" includes East Asian and Middle Eastern individuals.

The foregoing analysis has several limitations. First, as mentioned in Section 3.2, the police may under-report searches. According to the London MPS, about 5% of searches go unrecorded (London Assembly 2014), which should serve as a lower bound. Whether this means searches were more or less disproportionate is difficult to know. On the one hand, if a police officer searches a minority-ethnic person and nothing turns up, the officer might not want to document the encounter out of fear of being accused of ethnic profiling. On the other hand, knowing that ethnic profiling is a source of concern among the public, the officer might be keen to document the encounter out of fear that failure to do so will be noticed by the searched individual and become part of a complaint against the police. In addition, the knowledge or willingness to ask for a copy of the search record might vary by group. Unfortunately, it is impossible to tell with the data on hand. A second limitation is that there may be meaningful differences

Figure 3.3: Ethnic Makeup of Searched Persons by Search Grounds

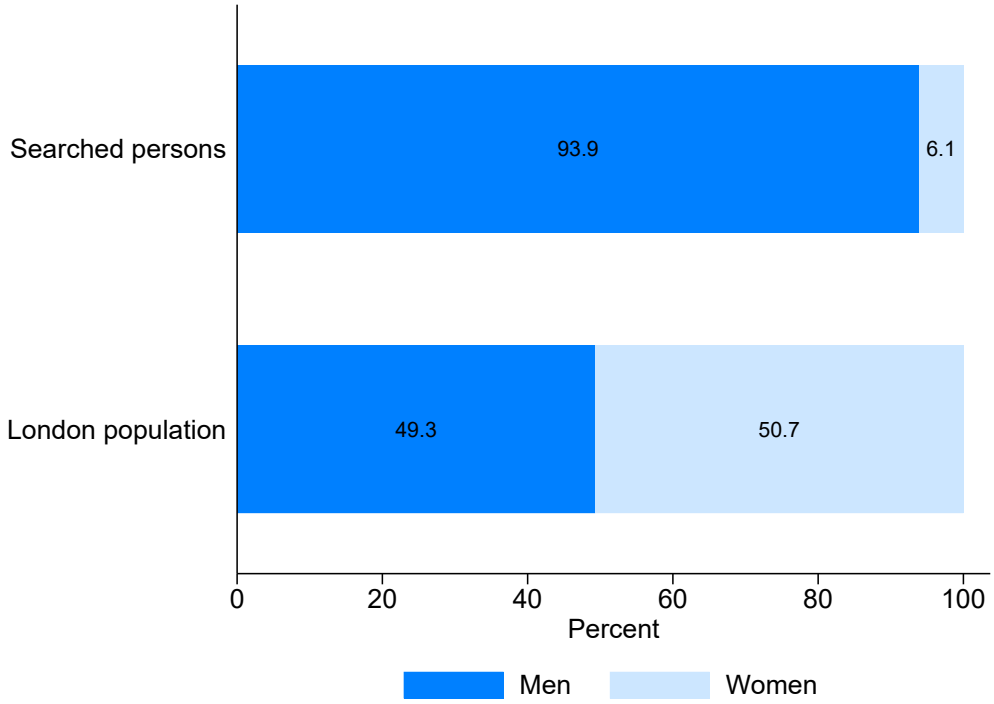


Source: London Metropolitan Police Service; London 2011 census; author calculations. Note: All searches completed prior to the body camera trial, which began in May 2014. London population shown for comparison. "Other" search grounds include the grounds listed in Table 3.A3. "Other" includes East Asian and Middle Eastern individuals.

in the resident population, as measured by the census, and the "street" or "available" population that is subject to search (Waddington et al. 2004).

But, more importantly (and as described in Section 3.3), the distribution of characteristics that give rise to reasonable suspicion might vary by group. For example, if men are more likely to carry illegal weapons, the "facts, information and/or intelligence which are relevant to the likelihood that the object in question will be found"—that is, the objective basis needed for reasonable grounds for suspicion—likely leads to more searches of men. Note that this claim is distinct from, but related to, a claim of statistical discrimination *based on* gender. If men are more likely to carry illegal weapons and a police officer uses this information to search more men, this would constitute statistical discrimination based on gender. However, doing so would be illegal under UK law,

Figure 3.4: Gender Makeup of Searched Persons versus London Population



Source: London Metropolitan Police Service; London 2011 census; author calculations.
Note: All searches completed prior to the body camera trial, which began in May 2014.

which specifies that physical appearance, such as ethnic appearance, gender, or age, can never support reasonable grounds for suspicion, and that police officers may not rely on "[g]eneralisations or stereotypical images that certain groups or categories of people are more likely to be involved in criminal activity" (UK Home Office 2014a). Still, the distribution of *other* characteristics giving rise to suspicion might vary by group. For example, if men carry illegal weapons at higher rates than women, then whatever observable characteristics (besides gender) that correlate with weapon possession also correlate with gender. In that case, police officers would end up searching more men than women, even if they did not discriminate against men based on gender stereotypes (statistical discrimination) or animus (preference-based discrimination).

To distinguish (unlawful) preference-based discrimination from (unlawful) statistical discrimination based on ethnicity or gender or (lawful) statistical discrimination

based on observable characteristics that correlate with ethnicity or gender,[4] I estimate the following linear probability model using ordinary least squares (OLS) regression:

$$Hit_{ibt} = \alpha + \beta_1 Ethnicity_{ibt} + \beta_2 Gender_{ibt} + \beta_3 Age_{ibt} + \gamma X_{ibt} + \delta_b + \eta_t + \varepsilon_{ibt}, \quad (3.1)$$

where Hit_{ibt} is a binary variable for search success;[5] $Ethnicity_{ibt}$ and $Gender_{ibt}$ are factor variables for ethnic appearance and gender; Age_{ibt} is age in years; X_{ibt} is a vector of search-level covariates, including age squared, search type (person, vehicle, or person and vehicle), and search grounds;[6] δ_b is a borough fixed effect for borough b; η_t is a month-by-year fixed effect; and ε_{ibt} is an error term. The results of this regression are shown in Table 3.2. Columns 1 and 2 display the results excluding search-level controls (including age) and fixed effects. That is, these columns show nonparametric estimates of average search success rates for different ethnic and gender groups. Column 3 adds search-level controls, including age, and column 4 adds borough and month-by-year fixed effects. Heteroscedasticity-robust standard errors are reported in parentheses.

Columns 1 and 2 of Table 3.2 show that, prior to the camera trial, the search success rates for black and South Asian persons were 0.7 p.p. and 2.0 p.p. lower than for white persons, and the search success rate for men was 7.3 p.p. lower than for women. These results are displayed graphically in Figures 3.5 and 3.6. Figure 3.5 shows that search success rates were 29.7%, 28.9%, and 27.6% for white, black, and South

[4]Statistical discrimination may be further subdivided into accurate statistical discrimination (based on correct beliefs) and inaccurate statistical discrimination (based on incorrect beliefs) (Bohren et al. 2019). Differences in search success rates may be due to preference-based or inaccurate statistical discrimination (or both). In the present context, the two possibilities cannot be distinguished. However, the absence of a difference in search success rates is evidence against both possibilities, unless the two forms of discrimination exactly offset one another. That is, equal search success rates, when coupled with disproportionality, are evidence of accurate statistical discrimination.

[5]A search is defined to be successful if it results in an arrest, warning, or other action. A search is unsuccessful if it results in no action.

[6]It should be noted that search type and search grounds may be endogenous. For example, if an officer searches a person suspected of carrying illegal drugs and finds a weapon, he might record the search as a weapons search. Put differently, we might think of search grounds as a possible outcome variable or proxy control affected by the variable of interest. Fortunately, the results that follow are robust to including only search-level covariates that can be thought of as "fixed" at the time of search, such as age, age squared, borough, and month.

Table 3.2: Search Success Rates

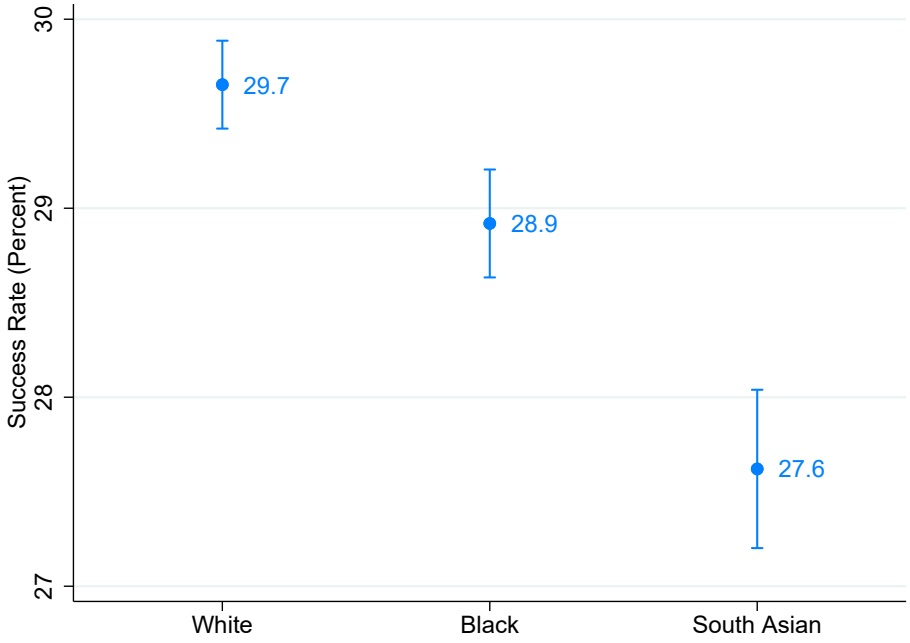
	(1)	(2)	(3)	(4)
Black	-0.007		0.001	-0.000
	(0.002)		(0.002)	(0.002)
South Asian	-0.020		-0.014	-0.019
	(0.002)		(0.002)	(0.003)
Male		-0.073	-0.060	-0.060
		(0.004)	(0.004)	(0.004)
Age			0.015	0.015
			(0.000)	(0.000)
Search-level controls	No	No	Yes	Yes
Borough fixed effects	No	No	No	Yes
Month-by-year fixed effects	No	No	No	Yes
Observations	298,782	298,782	298,782	298,782

Note: This table displays results from ordinary least squares regressions with a binary search success (or "hit") variable as their estimand. White women are the base category. A search is defined to be successful if it results in an arrest, warning, or other action. A search is unsuccessful if it results in no action. "Search-level controls" include age squared, search type (person, vehicle, or person and vehicle), and search grounds. Heteroscedasticity-robust standard errors are in parentheses. All searches completed prior to the body camera trial, which began in May 2014.

Asian persons, respectively. Figure 3.6 displays the average search success rates for women (36.1%) and men (28.8%). Columns 3 and 4 of Table 3.2 show that the results for South Asians and men persist when using search-level controls and borough and month-by-year fixed effects, although the differences narrow somewhat—to 1.9 p.p. for South Asians and 6.0 p.p. for men. However, the black-white difference disappears after including these controls. The results are robust to alternative specifications, such as probit and logit.

Table 3.2 shows that there was a significant (positive) relationship between age and search success. That is, the search success rate was increasing in age. This finding is explored nonparametrically in Figure 3.7. The figure displays search success rates for different age groups with roughly equal observations. Young persons, particularly

Figure 3.5: Search Success Rates by Ethnic Appearance

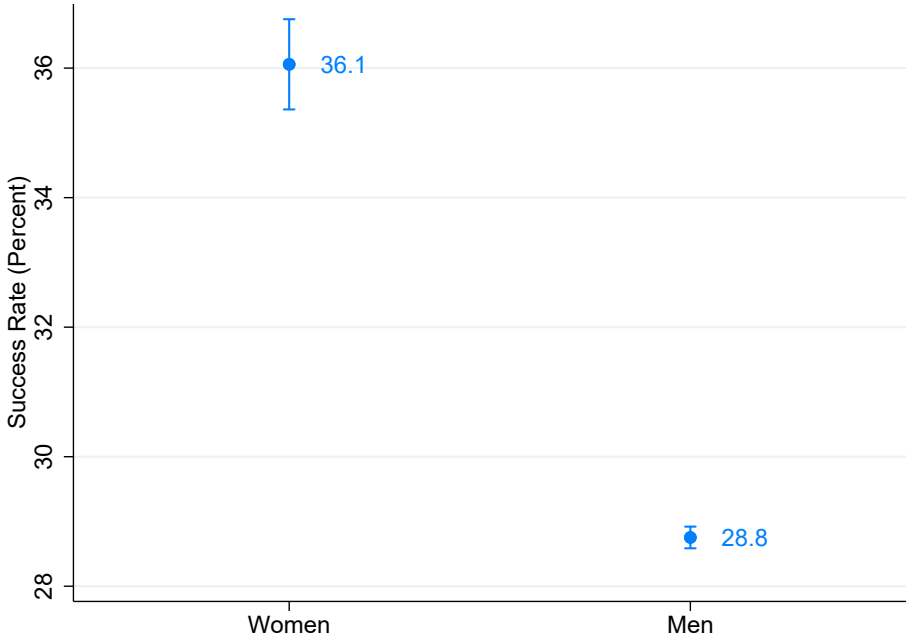


Source: London Metropolitan Police Service; author calculations.
Note: Error bars represent 95% confidence intervals. All searches completed prior to the body camera trial, which began in May 2014.

teenagers, had the lowest search success rates—less than 25 percent. The search success rate increases linearly with age until about age 30 years old, when it levels off around 35 percent. This is evidence of preference-based or inaccurate statistical discrimination against younger individuals.

Appendix Figures 3.A3–3.A5 explore heterogeneity in these results. For example, Figure 3.A3 displays search success rates by ethnic appearance and age. The figure can be interpreted in a couple of ways. First, every ethnic group exhibited upward-sloping search success rates with respect to age, but the slope was steepest for South Asians. Second, the lower average search success rates for black and South Asian persons appear to be driven by searches of young persons, particularly those between ages 19 and 29 years old. Figure 3.A4 shows that search success rates were consistently lower for men than for women, regardless of age. (And search success rates increased with age for both genders.) Finally, Figure 3.A5 shows that search success rates were lower for men

Figure 3.6: Search Success Rates by Gender



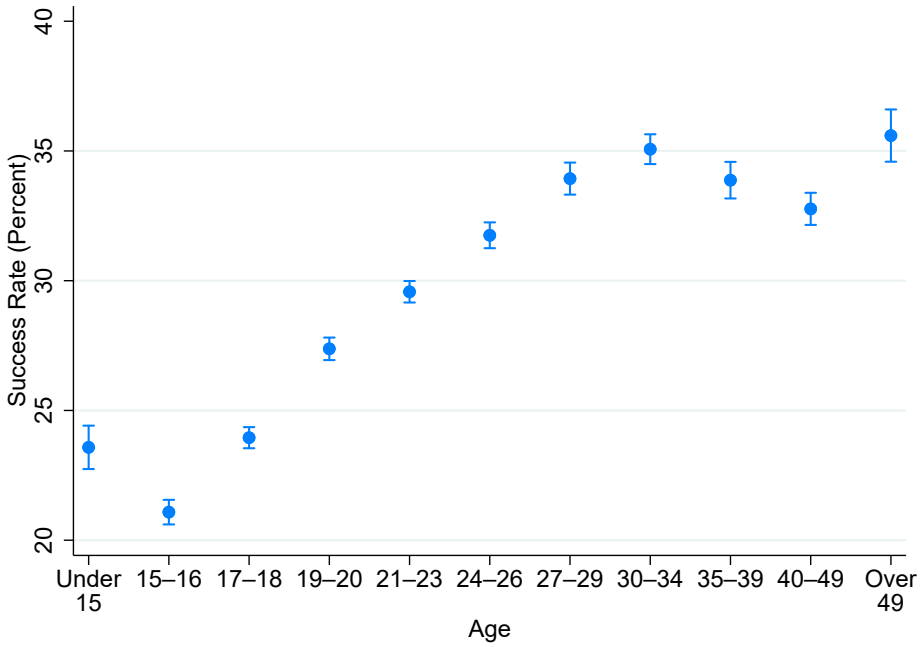
Source: London Metropolitan Police Service; author calculations.
Note: Error bars represent 95% confidence intervals. All searches completed prior to the body camera trial, which began in May 2014.

than for women, regardless of ethnic appearance, but that ethnic differences in search success rates were driven by men.

So far, this analysis has considered overall search success rates—that is, the rates at which searches result in criminal justice outcomes, rather than no action by police. However, we can further investigate specific outcomes. Appendix Table 3.A8 shows search success (or "hit") rates overall and by specific outcome.[7] Men were consistently less likely to experience a criminal justice outcome, whether an arrest, warning, or some other action. South Asian persons were also less likely to be arrested or receive a warning, conditional on search. However, whereas search success rates were equal for black and white persons (after conditioning on covariates), black persons were 2.5 p.p.

[7]Whether a specific outcome, such as arrest, should be the primary outcome of analysis is a good question. Because there is not detailed information about what the search found (e.g., a quantity of drugs), this essay takes the approach that is least likely to be confounded by differential arrest rates conditional on search findings. The main outcome of interest could be rephrased in terms of its converse: what is the rate at which a search results in no further action?

Figure 3.7: Search Success Rates by Age



Source: London Metropolitan Police Service; author calculations.
Note: Error bars represent 95% confidence intervals. All searches completed prior to the body camera trial, which began in May 2014.

more likely to be arrested and 2.7 p.p. less likely to receive a warning, conditional on search (and conditional on further action by the police). This finding is striking. Of course, if the characteristics of the offence differed systematically by ethnic group, the differences in outcomes may not be due to different treatment. However, in that case, it may be odd that the overall search success rates were equal.

Finally, we can examine search success rates separately by search grounds. Appendix Table 3.A9 displays these results, using a full set of controls. As the table shows, for the largest category of search—drug searches—search success rates were 2.4 p.p. and 4.6 p.p. lower for black and South Asian Londoners, respectively. These magnitudes and the relative frequency of drug searches appear to drive the overall disparity in search success rates for black and South Asian persons. The success rates for these groups were also lower for weapons searches, but they were higher for stolen property searches (and going equipped searches, for black persons), moderating the overall disparity. Search

success rates were consistently and significantly (between 6 and 16.7 p.p.) lower for men than for women, with the exception of drug searches, for which they were 0.8 p.p. higher. The coefficient on age is twice as high or more for drugs and weapons searches (0.02) than for searches for stolen property (0.01) or going equipped (0.003).

Overall, these results are consistent with statistical discrimination against black Londoners (and South Asians for drug searches) and men, as these groups were searched more frequently than their population share. Furthermore, there is evidence of preference-based or inaccurate statistical discrimination against black and South Asian persons, as they had lower search success rates, on average, than whites. The black-white difference disappears once we account for a full set of control variables, but conditional on search and on further action by the police, black persons were significantly more likely than whites to be arrested, and they were less likely to receive a warning. Overall differences in search success rates by ethnic group appear to be driven by drug searches (the most frequent kind), for which black and South Asian persons had significantly lower search success rates, and by searches of young persons. In general, young persons (particularly those under 20 years old) had significantly lower search success rates, regardless of ethnic appearance. Age disparities also appear to be driven by drug searches. Finally, there is considerable evidence of preference-based or inaccurate statistical discrimination against men, as their average search success rate was 6–7 p.p. lower than women, and the direction of this difference was fairly stable across different search grounds.

3.6 Results: Body Cameras

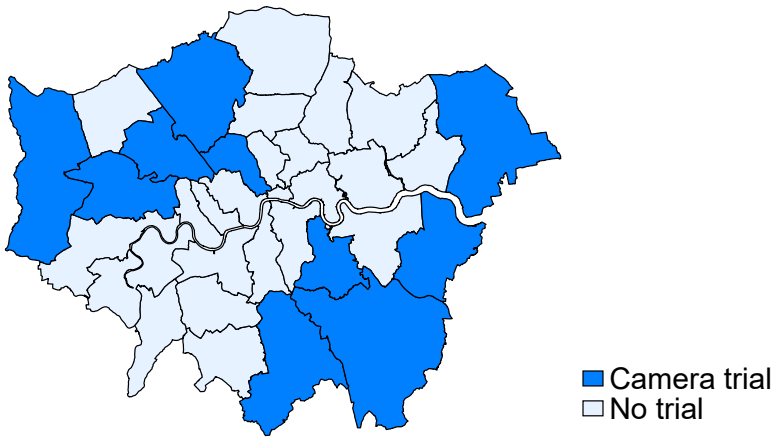
If police discriminate against, for example, young minority-ethnic men, what can be done to stop it? This section evaluates the potential of body-worn video cameras. Body cameras are frequently invoked as a way to reduce police bias (e.g., Oyeniran 2017), but the effect of the cameras on discrimination are not well-understood. Between

May 2014 and April 2015, the London police conducted a randomised trial of body cameras in ten boroughs. Using variation in the timing and location of camera use, this essay estimates the effect of cameras on differences in search rates and search success rates between ethnic and gender groups. It finds that the cameras slightly increased the relatively frequency of searches of black persons and decreased the relative frequency of searches of South Asians, although this finding may reflect pre-existing borough-level trends. Consistent with relatively more searches of black persons and relatively fewer searches of South Asians, the black search success rate decreased and the South Asian success rate increased. Although there was not a detectable effect of camera use on gender differences in terms of search frequency, there was an economically large and statistically significant (positive) effect on the search success rate for men relative to women. Overall, these findings suggest that body cameras may have lessened, to some degree, the disparities described in the previous section.

Between May 2014 and April 2015, the London police conducted a randomised trial of body-worn video cameras. This trial took place in ten London boroughs: Barnet, Brent, Bexley, Bromley, Camden, Croydon, Ealing, Havering, Hillingdon, and Lewisham. These boroughs are displayed in Figure 3.8. In each borough, five Emergency Response Teams (ERTs) were assigned to either treatment or control conditions. Treated ERTs were required to wear body cameras, while control teams were not. The London MPS collected information from both sets of teams to evaluate whether the cameras affected complaints against police; the frequency of stops and searches; or arrests (Grossmith et al. 2015).

As a result of the body camera trial, the London MPS made several findings. First, the cameras reduced the frequency of complaints against the police (Grossmith et al. 2015). Whether this was due to changed officer behaviour or changed citizen behaviour (or both) is not certain, but when surveyed, officers reported that the cameras made them feel protected against complaints. Some officers told anecdotes of camera

Figure 3.8: Body Camera Trial Boroughs



Source: London Metropolitan Police Service.
Note: The 10 trial boroughs are: Barnet, Brent, Bexley, Bromley, Camden, Croydon, Ealing, Havering, Hillingdon, and Lewisham.

use as a way to pre-empt potential complaints. Second, the cameras did not affect the overall frequency of stops and searches (Grossmith et al. 2015). However, arrests following searches were less likely among officers with cameras.

The London MPS's internal trial was consistent with other studies of body cameras, which tend to focus on general patterns of police behaviour and citizen complaints (Ariel, Farrar & Sutherland 2015; Jennings, Lynch & Fridell 2015; Ready & Young 2015; Yokum, Ravishankar & Coppock 2019). In what follows, this essay uses variation in the timing and location of camera use to investigate the effects of the cameras on demographic patterns of stop and search.

At this time, the London police are unable to provide officer-level data on the body-worn camera trial. Therefore, this essay is not able to directly compare treatment and control groups. However, because camera use varied across London boroughs and across time, I propose a difference-in-differences approach to overcome this limitation (Card & Krueger 1994). The ten trial boroughs were selected for the camera trial in part

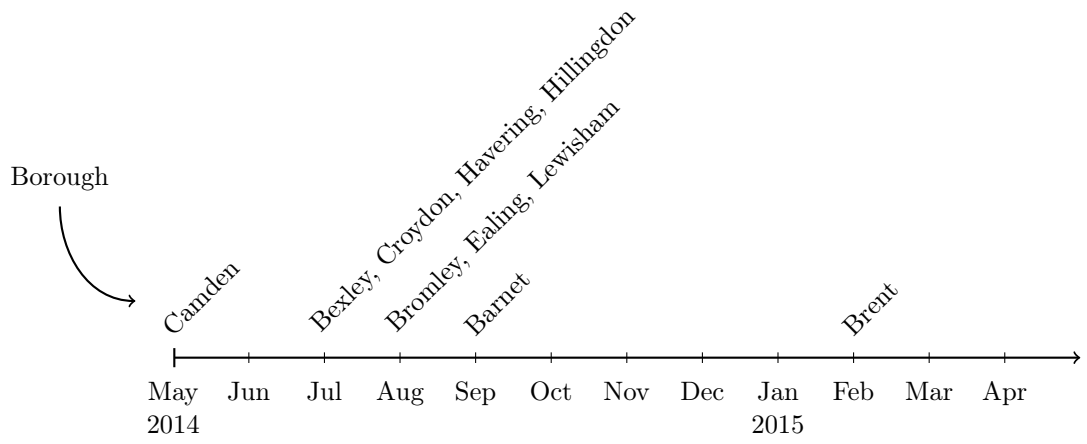
based on their characteristics, including their crime rate, stop and search rate, and prior experience with cameras (Grossmith et al. 2015). Thus, a simple comparison between trial boroughs and non-trial boroughs would suffer from selection bias. However, a peculiar feature of the London body camera trial is that the cameras were not introduced simultaneously. Because of limited resources for training and technical support, the cameras were introduced in a staggered fashion over the course of a year. Figure 3.9 displays a timeline of the camera roll-out.[8] We can use variation in the timing and location of camera use to estimate its effect on race and gender differences in search rates and search success rates. To begin, I estimate the following linear equation using OLS regression:

$$y_{ibt} = \delta_b + \eta_t + \beta Treat_{ibt} + \gamma X_{ibt} + \varepsilon_{ibt}, \tag{3.2}$$

where y_{ibt} is the fraction of searches devoted to a particular group, such as men, black persons, or South Asian persons; $Treat_{ibt}$ is an indicator for whether a particular search was conducted in a trial borough *after* the introduction of cameras in that borough; and the other variables are defined as before (see Section 3.5). The coefficient of interest is β. Because of the inclusion of borough and month-by-year fixed effects (δ_b and η_t) and the definition of $Treat_{ibt}$, this is a difference-in-differences estimate. Table 3.3 shows the results, with standard errors robust to intra-borough correlation in parentheses. As the table indicates, the fraction of search subjects that were black increased by 0.8 p.p. (column 1), and the fraction that were South Asian decreased by 1 p.p. (column 3). However, only the second difference is statistically significant when using clustered standard errors. There does not appear to be an effect on the fraction of search subjects that were men (columns 5–6). Furthermore, the purported effects on searches by ethnic

[8]Boroughs are listed according to when 50 percent or more officers in a treatment team uploaded an operational video clip. To avoid the possibility that treatment is mismeasured due to camera use prior to the date listed in Figure 3.9, the analysis that follows drops months after May 2014 but prior to 50 percent operation. The results are robust to including these months (which, in some cases, makes the results even stronger).

Figure 3.9: Timing of Introduction of Body Cameras



Source: London Metropolitan Police Service.

group disappear with the inclusion of borough-specific time trends (columns 2 and 4), a robustness check suggested by Besley and Burgess (2004) and Wolfers (2006). This suggests that the parallel trends assumption necessary for a valid difference-in-differences design may not hold.

Table 3.3: Effect of Body Cameras on Demographic Shares of Searches

	Black		South Asian		Men	
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	0.008	-0.006	-0.010	0.006	0.002	0.001
	(0.009)	(0.014)	(0.004)	(0.011)	(0.004)	(0.006)
Borough-specific time trends	No	Yes	No	Yes	No	Yes
Observations	441,598	441,598	441,598	441,598	441,598	441,598

Note: This table displays results from difference-in-differences (DD) regressions with demographic indicator variables listed in the column headings as estimands. All regressions use search-level controls, including age, age squared, search type (person, vehicle, or person and vehicle), and search grounds, as well as borough and month-by-year fixed effects. Standard errors robust to intra-borough correlation are in parentheses. A positive coefficient indicates that body cameras increased the relative frequency of search for that group.

To further investigate the effects of the cameras on differences in search success rates, I estimate the following linear probability model using OLS:

$$Hit_{ibt} = \delta_{bg} + \eta_{tg} + \beta_1 Treat_{ibt} + \beta_2 Group_{ibt} + \beta_3 Treat \times Group_{ibt} + \gamma X_{ibt} + \varepsilon_{ibt}, \quad (3.3)$$

where $Group_{ibt}$ is either a factor variable for ethnicity or an indicator variable for gender, and the other variables are defined as before, except that the borough and month-by-year fixed effects are estimated separately by group g, that is, by ethnic or gender group. This results in a triple difference-in-differences (DDD) estimate of the effect of the body cameras. That is, the comparison is between, for example, searches of black and white persons, before and after camera use, relative to the same double difference in boroughs that did not take part in the camera trial.

Table 3.4 shows the triple difference-in-differences results. Columns 1 and 2 indicate that the cameras did not have statistically significant effects on the search success rates for black or South Asian persons relative to whites, although the direction of the estimates is consistent with columns 1 and 3 of Table 3.3. However, in this case, the inclusion of borough- and group-specific time trends magnifies the estimated effects. As for gender, columns 3 and 4 suggest that the cameras had a significant effect on the gender difference in search success rates. This effect ranged from 3.9 p.p. to 8.3 p.p. (when including borough- and group-specific time trends), eliminating as much as half or all of the pretrial difference (see Table 3.2 and Figure 3.6).

3.7 Conclusion

Using data on over 400,000 stops and searches by the London Metropolitan Police Service, this essay shows that, prior to the introduction of body cameras, black Londoners were 2.5 times more likely to be searched than whites. This disparity was even greater for drug searches, which constituted approximately half of all searches, for which black Londoners were 2.9 times more likely to be search than whites. South

Table 3.4: Effect of Body Cameras on Search Success Rates

	(1)	(2)	(3)	(4)
Treat	-0.008	-0.002	-0.044	-0.084
	(0.012)	(0.019)	(0.023)	(0.032)
Treat × Black	-0.001	-0.019		
	(0.014)	(0.025)		
Treat × South Asian	0.008	0.022		
	(0.016)	(0.023)		
Treat × Male			0.039	0.083
			(0.022)	(0.029)
Borough- and group-specific time trends	No	Yes	No	Yes
Observations	441,598	441,598	441,598	441,598

Note: This table displays results from difference-in-difference-in-differences (DDD) regressions with a binary search success (or "hit") variable as their estimand. A search is defined to be successful if it results in an arrest, warning, or other action. All regressions use search-level controls, including age, age squared, search type (person, vehicle, or person and vehicle), and search grounds, as well as borough and month-by-year fixed effects, estimated separately by comparison group, i.e., by ethnic group for columns 1 and 2 and by gender for columns 3 and 4. Standard errors robust to intra-borough correlation are in parentheses. A positive interaction coefficient indicates that body cameras increased the relative search success rate for that group.

Asians were 1.4 times more likely to be searched for drugs than whites, but approximately equally as likely to be searched overall. The overall search success, or "hit," rates were 0.7 p.p. lower for black persons and 2.0 p.p. lower for South Asians, consistent with preference-based or inaccurate statistical discrimination against them. However, after including a full set of control variables, the black-white difference reduces to zero. Conditional on the search resulting in further action, black persons were 2.5 p.p. more likely than whites to be arrested and 2.7 p.p. less likely to receive a warning. Differences between ethnic groups appear to be driven by drug searches (the most frequent kind) and searches of young persons, particularly those between 19–29 years old. In general, young persons had significantly lower search success rates, and this was true regardless of ethnicity or gender. Men were 15.8 times more likely to be searched, and their average search success rate was 7.3 p.p. lower than women, which is evidence of gender-based discrimination in searches. Body cameras may have reduced these disparities, particularly the gender disparity, but this finding is complicated by pre-existing trends in boroughs that participated in a camera trial. In any event, the estimated effects of the cameras on ethnic differences are small. For activists and policymakers looking to address ethnic disparities in police searches, body cameras should not be viewed as a panacea.

3.8 References

Antonovics, Kate, and Brian G. Knight. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." *Review of Economics and Statistics* 91, no. 1 (February): 163–77.

Anwar, Shamena, and Hanming Fang. 2006. "An Alternative Test of Racial Bias in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96, no. 1 (March): 127–51.

Ariel, Barak, William A. Farrar, and Alex Sutherland. 2015. "The Effect of Police Body-Worn Cameras on Use of Force and Citizens' Complaints Against the Police: A Randomized Controlled Trial." *Journal of Quantitative Criminology* 31:509–35.

Ariel, Barak, Alex Sutherland, Darren Henstock, Josh Young, Paul Drover, Jayne Sykes, Simon Megicks, and Ryan Henderson. 2016. "Wearing Body Cameras Increases Assaults Against Officers and Does Not Reduce Police Use of Force: Results from a Global Multi-Site Experiment." *European Journal of Criminology* 13 (6): 744–55.

Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Quarterly Journal of Economics* 133, no. 4 (November): 1885–1932.

Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton, New Jersey: Princeton University Press.

Becker, Gary S. 1957. *The Economics of Discrimination.* Chicago: University of Chicago Press.

Besley, Timothy, and Robin Burgess. 2004. "Can Labor Regulation Hinder Economic Performance? Evidence from India." *Quarterly Journal of Economics* 19, no. 1 (February): 91–134.

Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G. Pope. 2019. "Inaccurate Statistical Discrimination." University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2019-86.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg. 2019. "The Dynamics of Discrimination: Theory and Evidence." *American Economic Review* 109, no. 10 (October): 3395–436.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84, no. 4 (September): 772–93.

Equality and Human Rights Commission (EHRC). 2010. *Stop and Think: A Critical Review of the Use of Stop and Search Powers in England and Wales.* https://bit.ly/2gKaT6i.

Gelman, Andrew, Jeffrey Fagan, and Alex Kiss. 2007. "An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias." *Journal of the American Statistical Association* 102, no. 479 (September): 813–23.

Grossmith, Lynne, Catherine Owens, Will Finn, David Mann, Tom Davies, and Laura Baika. 2015. "Police, Camera, Evidence: London's Cluster Randomised Controlled Trial of Body Worn Video." London: College of Policing and Mayor's Office of Policing and Crime (MOPAC). https://bit.ly/2OfdCKe.

Jennings, Wesley G., Mathew D. Lynch, and Lorie A. Fridell. 2015. "Evaluating the Impact of Police Officer Body-Worn Cameras (BWCs) on Response-to-Resistance and Serious External Complaints: Evidence from the Orlando Police Department (OPD) Experience Utilizing a Randomized Controlled Experiment." *Journal of Criminal Justice* 43:480–486.

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109 (1): 203–29.

London Assembly. 2014. *Stop and Search: An Investigation of the Met's New Approach.* https://bit.ly/2HnlvcR.

Odogwu, Ifeanyi. 2012. *Know Your Rights on Stop and Search.* StopWatch. https://bit.ly/2HjD2CG.

Oyeniran, Femi. 2017. "Stop and Search in the UK is Fundamentally Flawed, and Disproportionally Affects People of Colour." *Independent*, August 13, 2017. https://bit.ly/32JpI1f.

Persico, Nicola, and Petra Todd. 2006. "Generalising the Hit Rates Test for Racial Bias in Law Enforcement, with an Application to Vehicle Searches in Wichita." *The Economic Journal* 116 (November): F531–67.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62, no. 4 (September): 659–61.

Ready, Justin T., and Jacob T.N. Young. 2015. "The Impact of On-Officer Video Cameras on Police-Citizen Contacts: Findings from a Controlled Experiment in Mesa, AZ." *Journal of Experimental Criminology* 11:445–58.

Rosenfeld, Richard, and Robert Fornango. 2014. "The Impact of Police Stops on Precinct Robbery and Burglary Rates in New York City, 2003–2010." *Justice Quarterly* 31 (1): 96–122.

Waddington, P.A.J., Kevin Stenson, and David Don. 2004. "In Proportion: Race, and Police Stop and Search." *British Journal of Criminology* 44 (6): 889–914.

UK Home Office. 2014a. Police and Criminal Evidence Act 1984 (PACE) – Code A. https://bit.ly/2HMxD6n.

UK Home Office. 2014b. *Stop and Search: Theresa May Announces Reform of Police Stop and Search*. https://bit.ly/1sB9QFO.

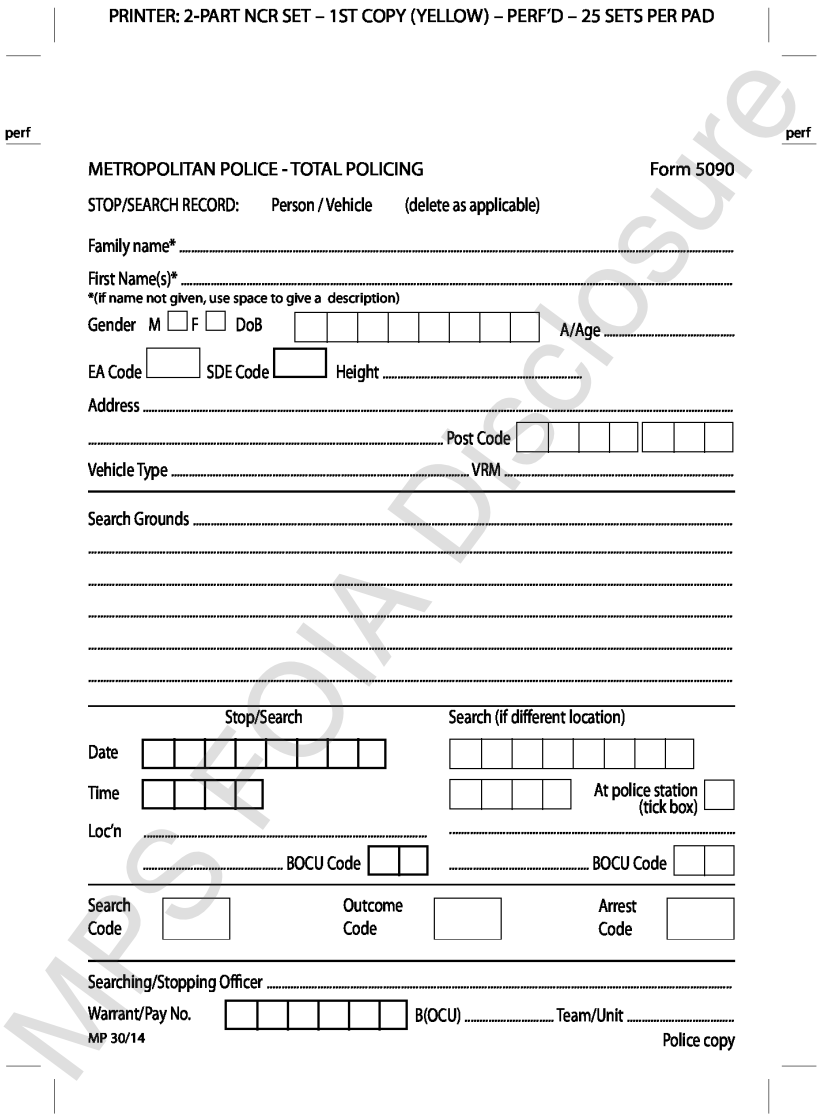
UK Office for National Statistics. 2011 Census. https://data.london.gov.uk/census/.

Wolfers, Justin. 2006. "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results." *American Economic Review* 96, no. 5 (December): 1802–20.

Yokum, David, Anita Ravishankar, and Alexander Coppock. 2019. "A Randomized Control Trial Evaluating the Effects of Police Body-Worn Cameras." *Proceedings of the National Academy of Sciences* 116, no. 21 (May) 10329–32.

3.9 Appendix

Figure 3.A1: Police Stop and Search Form



PRINTER: 2-PART NCR SET – 1ST COPY (YELLOW) – PERF'D – 25 SETS PER PAD

perf perf

METROPOLITAN POLICE - TOTAL POLICING Form 5090

STOP/SEARCH RECORD: Person / Vehicle (delete as applicable)

Family name* .....

First Name(s)* .....

*(if name not given, use space to give a description)

Gender M ☐ F ☐ DoB ☐☐☐☐☐☐☐☐ A/Age .....

EA Code ☐☐ SDE Code ☐☐ Height .....

Address .....

..... Post Code ☐☐☐☐☐☐☐

Vehicle Type ..... VRM .....

Search Grounds .....

.....

.....

.....

.....

.....

	Stop/Search	Search (if different location)
Date	☐☐☐☐☐☐☐☐	☐☐☐☐☐☐☐☐
Time	☐☐☐☐	☐☐☐☐ At police station (tick box) ☐
Loc'n	.....	.....
	..... BOCU Code ☐☐	..... BOCU Code ☐☐

Search Code ☐ Outcome Code ☐ Arrest Code ☐

Searching/Stopping Officer .....

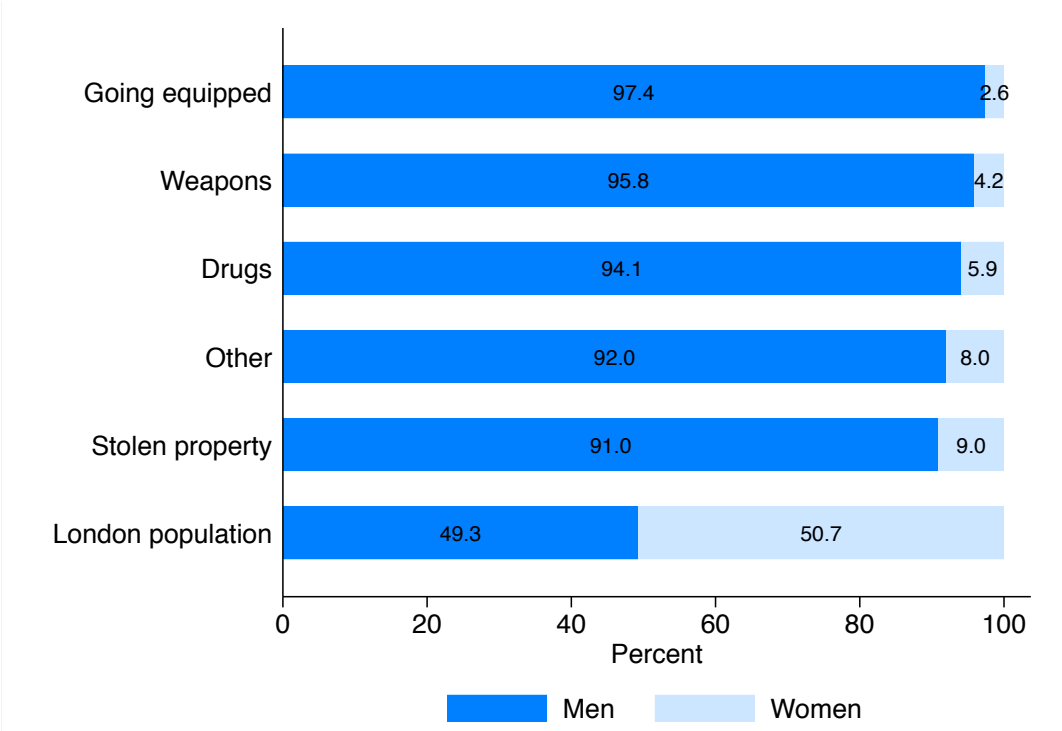
Warrant/Pay No. ☐☐☐☐☐☐ B(OCU) ..... Team/Unit .....

MP 30/14 Police copy

MPS FOIA Disclosure

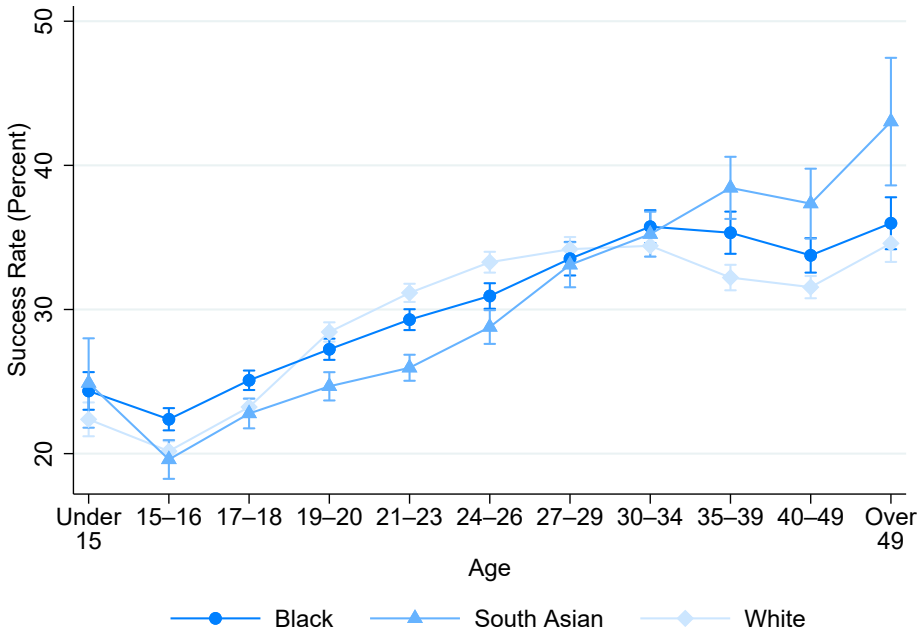
Source: London Metropolitan Police Service.

Figure 3.A2: Gender Makeup of Searches by Search Grounds



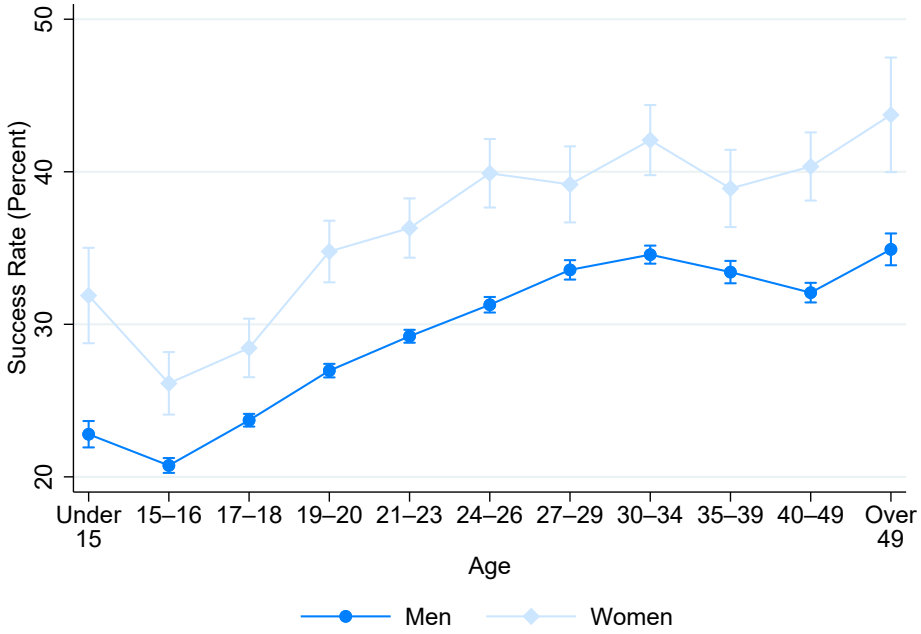
Source: London Metropolitan Police Service; London 2011 census; author calculations.
Note: All searches completed prior to the body camera trial, which began in May 2014. London population shown for comparison. "Other" search grounds include the grounds listed in Table 3.A3.

Figure 3.A3: Search Success Rates by Ethnic Appearance and Age



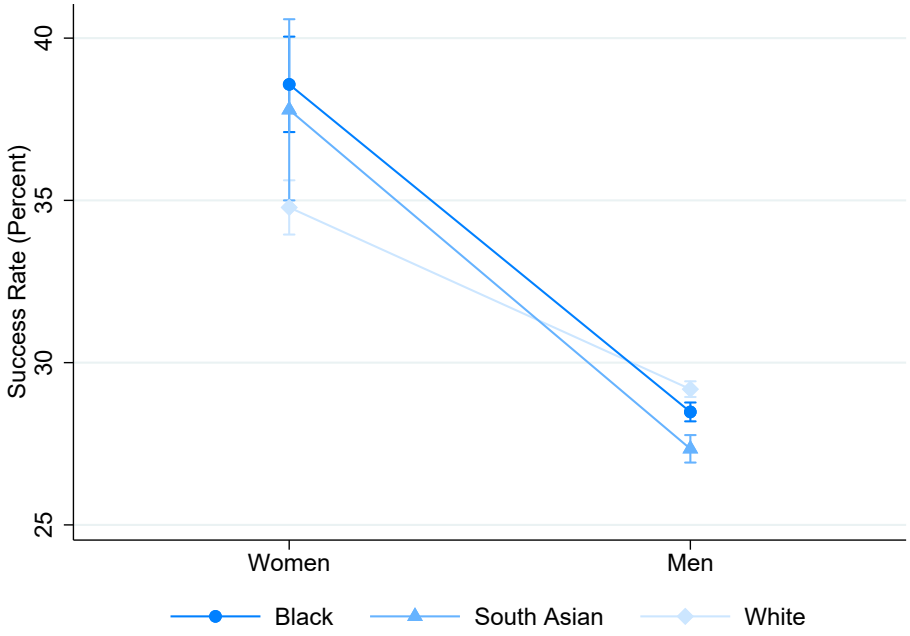
Source: London Metropolitan Police Service; author calculations.
Note: Error bars represent 95% confidence intervals. All searches completed prior to the body camera trial, which began in May 2014.

Figure 3.A4: Search Success Rates by Gender and Age



Source: London Metropolitan Police Service; author calculations.
Note: Error bars represent 95% confidence intervals. All searches completed prior to the body camera trial, which began in May 2014.

Figure 3.A5: Search Success Rates by Ethnic Appearance and Gender



Source: London Metropolitan Police Service; author calculations.
Note: Error bars represent 95% confidence intervals. All searches completed prior to the body camera trial, which began in May 2014.

Table 3.A1: Search Types

	Freq.	Percent	Cum.
PERSON AND VEHICLE	95,655	20.99	20.99
PERSON ONLY	359,715	78.93	99.92
VEHICLE ONLY	385	0.08	100.00
Total	455,755	100.00	

Source: London Metropolitan Police Service; author calculations.

Table 3.A2: Searches by London Borough

	Freq.	Percent	Cum.
BARKING AND DAGENHAM	9,083	1.99	1.99
BARNET	11,777	2.58	4.58
BEXLEY	7,802	1.71	6.29
BRENT	22,483	4.93	11.22
BROMLEY	19,255	4.22	15.45
CAMDEN	19,065	4.18	19.63
CROYDON	14,881	3.27	22.90
EALING	12,888	2.83	25.72
ENFIELD	14,119	3.10	28.82
GREENWICH	15,748	3.46	32.28
HACKNEY	11,834	2.60	34.87
HAMMERSMITH & FULHAM	20,607	4.52	39.39
HARINGEY	11,041	2.42	41.82
HARROW	9,962	2.19	44.00
HAVERING	7,336	1.61	45.61
HILLINGDON	8,173	1.79	47.41
HOUNSLOW	14,402	3.16	50.57
ISLINGTON	9,723	2.13	52.70
KENSINGTON AND CHELSEA	7,909	1.74	54.43
KINGSTON-UPON-THAMES	9,076	1.99	56.43
LAMBETH	24,592	5.40	61.82
LEWISHAM	12,279	2.69	64.52
MERTON	5,826	1.28	65.79
NEWHAM	18,261	4.01	69.80
REDBRIDGE	16,539	3.63	73.43
RICHMOND-UPON-THAMES	7,314	1.60	75.03
SOUTHWARK	32,114	7.05	82.08
SUTTON	5,042	1.11	83.19
TOWER HAMLETS	18,498	4.06	87.25
WALTHAM FOREST	12,374	2.72	89.96
WANDSWORTH	9,089	1.99	91.96
WESTMINSTER	36,663	8.04	100.00
Total	455,755	100.00	

Source: London Metropolitan Police Service; author calculations.

Table 3.A3: Search Grounds: Detail

	Freq.	Percent	Cum.
A - STOP & SEARCH FOR STOLEN PROPERTY (S1 PACE)	87,701	19.24	19.24
B - DRUGS (S23 MISUSE OF DRUGS ACT)	265,234	58.20	77.44
C - FIREARMS (S47 FIREARMS ACT)	4,574	1.00	78.44
D - OFFENSIVE WEAPONS (S1 PACE)	38,214	8.38	86.83
E - OFFENSIVE WEAPONS (S139 - CJA, SCHOOLS)	6,462	1.42	88.25
F - GOING EQUIPPED (S1 PACE)	42,713	9.37	97.62
G - OTHER POWER	5,443	1.19	98.81
H - TERRORISM 44(1)	1	0.00	98.81
J - TERRORISM 44(2)	2	0.00	98.81
K - ANTICIPATED VIOLENCE (S60 CJPO)	1,983	0.44	99.25
L- ARTICLES TO CAUSE CRIMINAL DAMAGE	2,635	0.58	99.83
N - S.43 TERORISM ACT 2000	785	0.17	100.00
Q - S.47A TERRORISM	3	0.00	100.00
R - TERRORISM 47A(3)	5	0.00	100.00
Total	455,755	100.00	

Source: London Metropolitan Police Service; author calculations.

Table 3.A4: Ethnic Appearance of Searched Persons

	Freq.	Percent	Cum.
WHITE - NORTH EUROPEAN	185,984	40.81	40.81
WHITE - SOUTH EUROPEAN	38,044	8.35	49.16
BLACK	151,043	33.14	82.30
ASIAN	65,798	14.44	96.73
CHINESE, JAPANESE, AND OTHER	3,073	0.67	97.41
MIDDLE EASTERN	11,813	2.59	100.00
Total	455,755	100.00	

Source: London Metropolitan Police Service; author calculations.

Table 3.A5: Self-Defined Ethnicity of Searched Persons

	Freq.	Percent	Cum.
A1 - INDIAN	10,637	2.33	2.33
A2 - PAKISTANI	10,191	2.24	4.57
A3 - BANGLADESHI	15,960	3.50	8.07
A9 - ANY OTHER ASIAN BACKGROUND	25,869	5.68	13.75
B1 - CARIBBEAN	39,217	8.60	22.35
B2 - AFRICAN	42,696	9.37	31.72
B9 - ANY OTHER BLACK BACKGROUND	43,527	9.55	41.27
M1 - WHITE AND BLACK CARIBBEAN	7,696	1.69	42.96
M2 - WHITE AND BLACK AFRICAN	1,861	0.41	43.37
M3 - WHITE AND ASIAN	1,203	0.26	43.63
M9 - ANY OTHER MIXED BACKGROUND	7,746	1.70	45.33
N1 - OFFICER REQUIRED ELSEWHERE	708	0.16	45.49
N2 - PUBLIC DISORDER	634	0.14	45.63
N3 - SUBJECT DOES NOT UNDERSTAND	5,042	1.11	46.73
N4 - SUBJECT DECLINES TO DEFINE	23,539	5.16	51.90
O1 - CHINESE	1,219	0.27	52.17
O9 - ANY OTHER ETHNIC GROUP	9,651	2.12	54.28
UNCODED	32	0.01	54.29
W1 - BRITISH	134,511	29.51	83.80
W2 - IRISH	6,610	1.45	85.25
W9 - ANY OTHER WHITE BACKGROUND	67,206	14.75	100.00
Total	455,755	100.00	

Source: London Metropolitan Police Service; author calculations.

Table 3.A6: Search Outcomes: Detail

	Freq.	Percent	Cum.
01 - NO FURTHER ACTION	229,151	50.28	50.28
02 - ADVISED	88,101	19.33	69.61
03 - VERBALLY WARNED	13,001	2.85	72.46
04 - ARRESTED	78,572	17.24	89.70
05 - OTHER	8,978	1.97	91.67
06 - DISPERSE GROUPS PART 4 ASB ACT	1,273	0.28	91.95
07 - DIRECTED TO LEAVE ALCOHOL LOCN.	790	0.17	92.13
08 - ALCOHOL CONFISCATION	788	0.17	92.30
09 - CANNABIS WARNING	32,246	7.08	99.37
11 - PENALTY NOTICE (PND/FPN)	1,722	0.38	99.75
12 - POSTAL CHARGE REQUISITION / SUMMONS	392	0.09	99.84
13 - COMMUNITY RESOLUTION	649	0.14	99.98
14 - CAUTION (SIMPLE OR CONDITIONAL)	92	0.02	100.00
Total	455,755	100.00	

Source: London Metropolitan Police Service; author calculations.

Table 3.A7: Summary Statistics

	Min.	Max.	Mean	Std. Dev.
Black	0	1	0.331	0.471
South Asian	0	1	0.144	0.351
Other race	0	1	0.0327	0.178
Male	0	1	0.938	0.241
Age	5	92	25.23	9.663
Stolen property	0	1	0.192	0.394
Drugs	0	1	0.582	0.493
Weapons	0	1	0.108	0.310
Going equipped	0	1	0.0937	0.291
Other search grounds	0	1	0.0238	0.152
Arrested	0	1	0.172	0.378
Warning	0	1	0.0993	0.299
Other action	0	1	0.0322	0.177
Hit	0	1	0.304	0.460
No action	0	1	0.696	0.460
Treat	0	1	0.0720	0.259
Observations	455,755			

Source: London Metropolitan Police Service; author calculations.

Table 3.A8: Search Success Rates: Detail

	Overall	Specific Outcome		
	(1)	(2) Arrest	(3) Warning	(4) Other action
Black	-0.000	0.025	-0.027	0.002
	(0.002)	(0.002)	(0.001)	(0.001)
South Asian	-0.019	-0.008	-0.013	0.002
	(0.003)	(0.002)	(0.002)	(0.001)
Male	-0.060	-0.021	-0.019	-0.019
	(0.004)	(0.003)	(0.002)	(0.002)
Age	0.015	0.010	0.005	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Search-level controls	Yes	Yes	Yes	Yes
Borough fixed effects	Yes	Yes	Yes	Yes
Month-by-year fixed effects	Yes	Yes	Yes	Yes
Observations	298,782	298,782	298,782	298,782

Note: This table displays results from ordinary least squares regressions with binary search outcome variables listed in the column headings as estimands. White women are the base category. A search is defined to be successful (a "hit") if it results in an arrest, warning, or other action. A search is unsuccessful if it results in no action. "Search-level controls" include age squared, search type (person, vehicle, or person and vehicle), and search grounds. All searches completed prior to the body camera trial, which began in May 2014.

Table 3.A9: Search Success Rates by Search Grounds

	All Searches	By Search Grounds				
	(1)	(2) Stolen property	(3) Drugs	(4) Weapons	(5) Going equipped	(6) Other
Black	-0.000	0.060	-0.024	-0.024	0.028	-0.040
	(0.002)	(0.004)	(0.003)	(0.006)	(0.006)	(0.013)
South Asian	-0.019	0.051	-0.046	-0.025	0.000	-0.047
	(0.003)	(0.007)	(0.003)	(0.009)	(0.009)	(0.017)
Male	-0.060	-0.167	0.008	-0.074	-0.070	-0.060
	(0.004)	(0.007)	(0.005)	(0.014)	(0.016)	(0.019)
Age	0.015	0.010	0.020	0.020	0.003	0.029
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Search-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Borough fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month-by-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	298,782	62,282	168,670	31,117	28,861	7,852

Note: This table displays results from ordinary least squares regressions with a binary search success (or "hit") variable as their estimand, estimated separately by search grounds. White women are the base category. A search is defined to be successful if it results in an arrest, warning, or other action. A search is unsuccessful if it results in no action. "Search-level controls" include age squared and search type (person, vehicle, or person and vehicle). Heteroscedasticity-robust standard errors are in parentheses. All searches completed prior to the body camera trial, which began in May 2014.

Conclusion

This thesis comprised three essays in empirical labour and public economics. All three essays used experimental or quasi-experimental methods to evaluate a government activity or program. Two of the three essays focused on discrimination.

My first essay, co-authored with Anne Brockmeyer, Marco Hernandez, and Stewart Kettle, was motivated by the finding that the majority of firms in developing countries are informal, yet even among firms registered with the government, tax filing rates are low. We argued that nonfiling of taxes among registered firms constitutes an important intermediate form of informality, which can be tackled cost-effectively. Using a randomised experiment in Costa Rica, we showed that credible enforcement emails increased tax filing and payment rates among previously nonfiling firms. Emails that highlight third-party reports of a firm's transactions further increased compliance. The effect persisted over two years, and treated firms became more likely to report transactions with other firms, facilitating future tax enforcement.

My second essay analysed the wage effects of liability for employment discrimination. In both a standard competitive model and a monopsony model of the labour market, the net theoretical effect of liability for employment protection is ambiguous. Therefore, whether liability for employment discrimination increases the wages of protected workers and, if so, by how much, is an important empirical question. I proposed an empirical strategy to estimate the effects of liability for employment discrimination, using firm size–based discontinuities in coverage of United States antidiscrimination law and in damages awardable in discrimination lawsuits. The essay included a pre-

liminary investigation using survey data, but future work will aim to use confidential employer-employee matched data held by the U.S. Census Bureau. I am working with the Census Bureau to gain access to the confidential data. This project is an avenue for future research.

Finally, my third essay used data from the London Metropolitan Police Service to investigate discrimination in police searches and to estimate the effects of body cameras on discrimination. I found that, in the year prior to a body camera trial, black individuals were 2.5 times more likely to be searched than white or South Asian (Indian or Pakistani) individuals, and men were 15.8 times more likely to be searched than women. These disparities cannot be completely explained by accurate statistical discrimination, as there were significant differences in search success rates between groups—differences consistent with preference-based or inaccurate statistical discrimination against young minority-ethnic men. Using variation in body camera use resulting from a one-year camera trial, I found suggestive but not conclusive evidence that cameras reduced these disparities somewhat, particularly the gender difference in search success rates. However, the estimated effects of body cameras on ethnic disparities were small. These findings would be bolstered by the analysis of additional data on the camera trial, and I am working with the London police to obtain such data for future research.