Essays on Economics of Education

Viviana Rodriguez Andrade

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

© 2021

Viviana Rodriguez Andrade

All Rights Reserved

ABSTRACT

Essays on Economics of Education

Viviana Rodriguez Andrade

This dissertation consists of three studies on the economics of education and labor economics. The first essay seeks to deepen understanding of high school student engagement and effort response to changes in incentives. Changing the incentives students face is one lever for educators and policymakers to improve student learning in the presence of student disengagement. A statewide postsecondary admission policy which changes minimum college admissions standards for North Carolina high school students wishing to attend college in-state provides a setting to test how student effort responds to incentive structures. Regression discontinuity estimates show that students respond to the admission policy by increasing GPA and decreasing absences and suspensions. These effects suggest an increase in student engagement, however, the boost in GPA is driven by changes in course composition, with students substituting away from more demanding coursework. These unintended consequences of admission policies on student course-taking decisions can lead students to miss important learning opportunities in high school, possibly generating detrimental effects on student postsecondary success.

The second essay, coauthored with Hugh Macartney and Eric Nielsen, analyzes the effect of the Great Recession on racial employment inequality in the United States. It is well understood that adverse economic shocks affect workers non-uniformly. We explore a new channel through which unequal employment outcomes may emerge during a downturn: the extensive margin of establishment deaths. Intuitively, workers who are concentrated in less resilient establishments prior to an economic decline will be disproportionately affected by its onset. Using rich employment and establishment data, we show that black workers bore the brunt of the Great Recession in terms of within-industry employment changes arising from establishment deaths. This finding has important implications for the evolution of worker disparities during future downturns.

Finally, the third essay, coauthored with Clive Belfield and Brooks Bowden, examines the use

of benefit-cost analysis by the federal government on education regulations from 2006 to 2015. Benefit-cost analysis is an important part of regulatory decision-making, yet there are questions as to how often and how well it is performed. Here we examine 28 Regulatory Impact Assessments performed by the federal government on education regulations since 2006. We find many Regulatory Impact Assessments estimated costs, albeit using informal methods, but most failed to adequately report benefits. Also, most studies did not estimate net present value or clearly report methodological assumptions. In reviewing the relatively high quality studies we identified a number of discrepancies from best practice. Most importantly, few Regulatory Impact Assessments attempted a social benefit-cost analysis: Most examined ”administrative burdens” from compliance with legislation. This alternative focus on administrative burdens has significant implications for economic evaluation in practice.

Together, these essays advance what we know about higher education policy, labor market policy, and means of evaluating policies in both fields.

Table of Contents

List of Tables

List of Figures

Acknowledgements

The work contained in this dissertation would not have been possible without the help, support, and guidance of many people.

To start, I would like to thank my advisor, Alex Eble. Alex is an exceptional mentor who spares no effort in making sure his students succeed and accomplish their goals, whatever those may be. His guidance and encouragement through these years have been an invaluable input of my work.

I am forever indebted to Brooks Bowden, in whom I've found my greatest advocate, mentor, and friend. Under her wing, she has given me every opportunity to pursue my ideas, think critically about the rigor of my work, and understand the social importance of research in education. I hope to follow her as a scholar and a mentor.

The second chapter of this dissertation was born from a collaboration with Hugh Macartney and Eric Nielsen. Working with them has been a privilege. Even when I was a young PhD student from a different school, they both treated me as an equal from the start. That has meant to world to me.

I owe a special debt of gratitude to Sarah Cohodes and Judy Scott-Clayton. They provided incredible insights on the manuscript of the first chapter of this dissertation. Their input served to completely restructure the narrative of the paper and strengthen its importance and contribution. Furthermore, their guidance through important job market decisions, was timely and invaluable.

I would also like to thank faculty members at Teachers College and Duke University who provided important feedback that has shaped my research over the years. Among them, Henry Levin, Clive Belfield, Peter Bergman, Joydeep Roy, Jordan Matsudaira, Robert Garlick, and Peter Arcidiacono.

Throughout this journey, personal relationships have been a key source of support and strength. I will always look back with fond memories of innumerable hours with Takeshi Yanagiura and Tatiana Velasco as we cracked our brains solving problem sets, preparing for qualifiers, and supporting each other's research projects. Rahul Gupta, my dearest roommate, who I now consider to be family, has been an incredible source of love and support. To Adriana Fajardo and Camilo Acosta, my closet friends from Colombia, I count myself lucky to have such a strong friendship that has overcome the distance.

My parents, Edith and Gabriel, and my sister Paula, have been an unwavering source of support throughout my life. I do not believe that there are words to adequately express how indebted I am to them. They have given me every opportunity to grow and pursue my goals.

Finally, I am thankful for the years of support, encouragement, and unconditional love from my husband, Jonathan Moreno. Never did I imagine to be so lucky to find a partner who takes pride in my accomplishments and laments my disappointments as much as he does. I cannot imagine a life without him and our two puppies, Filomena and Ofelia.

Sincerest apologies to those I may have missed and thanks to all who have supported me throughout this journey. I hope we forge ahead together as I close the first chapter of my career and write the next.

Chapter 1

Student Effort Response to Shifts in University Admission Policies

1.1 Introduction

Incentive structures that shape student behavior in K-12 education, such as postsecondary admission policy, are abundant in education settings. Nonetheless, education economists and policymakers lack an understanding of the ways in which these incentive structures can shape student responses.

A change in admission requirements in one of the largest university systems in the U.S. provides a salient and high-stakes setting to examine how student effort responds to incentive structures. In an attempt to improve graduation rates at UNC campuses and prevent dropouts by only admitting college-ready students, the UNC Board of Governors set minimum GPA and SAT admission standards across the system in 2009, generating a discontinuity in the incentives for effort for high school students. Given that students take the SAT exam before completing their coursework, students right below and right above the threshold faced different returns to their effort in their remaining high school coursework.

Under a regression discontinuity design, I leverage rich administrative data for the universe of high school students in the state of North Carolina to compare the effort allocation of students right above and below the SAT threshold.[1] I find that students just above the SAT threshold have less 12th grade absences (-18.5%) and suspensions (-27.4%) than students right below. This suggests that students just above the threshold realize the possibility of a UNC college opportunity and increase effort in the form of attendance and good behavior. At the same time, students just above the SAT threshold obtain 12th grade GPAs that are 2.8% higher than students right below and are more likely to express intent to enroll ”in-state” after high school. However, higher student GPA can reflect either changes in effort allocation or substitution away from more demanding courses. In fact, I find that higher GPAs of students right above the SAT cutoff are driven by changes in course composition as students substitute away from less demanding courses.

Furthermore, I find that minority and less affluent students just above the cutoff have higher effort allocations via decreased absences and suspensions but engage to a lesser extent in gaming practices via course selection. This is consistent with the way the new admission standards affected admission practices across UNC campus selectivity. New admission cutoffs likely changed admission practices for the least selective campuses in the system which predominately serve minority students. These results are also consistent with Black and Hispanic students being less likely to have institutional knowledge and access to support systems such as academic counseling (Kirst and Venezia, 2004).

This paper contributes to three main strands of the literature on student effort and the incentives embedded in postsecondary education access policies. First, I show that students do change their effort in response to incentives. Student effort is a key contributor to academic success.[2] Yet,

[1] I follow previous literature and use student absences and suspensions as inverse proxies for effort and student GPA as an outcome that is a function of student effort (Hastings et al., 2012).

[2] Perhaps the strongest evidence of this relationship is provided by Stinebrickner and Stinebrickner (2008) who collect data on study time for first-semester college students at Berea College. The authors document strong positive causal effects of student effort on student GPA under an instrumental variable research design that exploits random assignment of roommates that are likely to exert low levels of effort. Measurement of student effort in K-12 settings is limited, mainly due to lack of data availability. However, the importance of effort in K-12 setting, as proxied by student absences and suspensions, still holds. Students who are absent and violate school rules are more likely to have poor grades and achievement and are less likely to graduate high school or enroll in college (Liu et al., 2019; Goodman, 2014; Christenson et al., 2012; Easton et al., 2017).

educators struggle to keep students engaged in schoolwork.[3] Incentive structures shape student behavior and can be important levers available to educators and policymakers to encourage students' investments in school. However, there is limited evidence of the ways in which students respond to incentives in K-12 settings. Generally, education accountability policies seek to shift incentives for effort of teachers and schools rather than students. Thus, there have not been many opportunities to evaluate the student incentive-effort connection outside of researcher designed pay-for-performance incentives schemes (Kremer et al., 2009; Fryer, 2011; Bettinger, 2010).[4]

Incentive structures embedded within policies for access to higher education provide an alternative, more realistic, setting to examine student effort. A recent and growing literature documents these high school student responses to incentives schemes embedded in postsecondary admission policies with a focus on percent plan programs (such as the Texas Top 10% program). For example, there is evidence of increased student effort and learning for Texas high school students as a response to the implementation of the Texas Top 10% program (Golightly, 2019; Cortes and Zhang, 2012). This paper instead evaluates salient and high-stakes shift in incentives for students at the *lower* end of the test score distribution who are more at risk of not obtaining postsecondary education.

Second, this paper expands on the above literature on student effort response to incentives by examining the channels through which students can adjust their behavior. In particular, I show that strategic course selection is an important dimension of student effort adjustments. These unintended consequences of admission policies can lead students to miss key learning opportunities in high school, which can have detrimental effects on future postsecondary outcomes and go against the policymaker's intent (Tincani et al., 2020).

While unintended consequences of incentives have been studied extensively in the literature of

[3]Barriers to students' optimal allocation of effort have been studied extensively in behavioral economics of education. These barriers include hyperbolic discounting (Bettinger and Slonim, 2007), present-bias behavior (Castillo et al., 2011; Wulfert et al., 2002; Harackiewicz et al., 2012), social norms and negative identities (Akerlof and Kranton, 2002; Eble and Hu, 2018), among others. For a review of the literature see Lavecchia et al. (2016).

[4]A notable exception is Fruehwirth (2013) who exploits a student accountability policy change in the state of North Carolina that required fifth grade students to achieve above a certain level to be automatically promoted to the next grade. Fruehwirth (2013) finds that this policy induced students in danger of scoring below the threshold to exert more effort in order to pass to the next grade.

education accountability programs, the focus has been on evaluating teacher and school response.[5] Unintended consequences of student response has been explored considerably less, but are observed in the evidence on percent plan programs which shows they might induce students to sort into lower-achieving high schools (Cullen et al., 2013), decrease student effort for seniors who already know their admission guarantee (Leeds et al., 2017), and decrease effort for students who overestimate the probability of being in the top of their class (Tincani et al., 2020).

Third, I show substantial heterogeneity of student response to incentives across demographic characteristics. While scarce in K-12 setting, the literature examining student response to incentives in postsecondary settings is abundant, with a focus on incentives generated by performance standards, merit aid requirements, and academic probation in college (Angrist et al., 2009; Lindo et al., 2010; Scott-Clayton, 2011; Casey et al., 2018; Cornwell et al., 2003).[6] This literature highlights how incentives have heterogeneous effects across student demographics and performance. In line with the findings of this literature, I find that minority and less affluent students increase effort via decreased absences and suspensions but engage to a lesser extent in gaming practices via course selection.

The remainder of the paper is organized as follows. The next section describes the Minimum Admission Requirement policy implemented by the UNC Board of Governors in 2009 and documents changes in admission practices across campus selectivity. Section 1.3 describes the data used in the analysis, and provides summary statistics. Section 1.4 establishes the research design and estimation procedure, as well as the method used for recovering bounds of regression discontinuity estimates which are presented as part of the validity exercises in Section 1.6. I present regression discontinuity results in Section 1.5 and conclude in Section 1.7.

[5] Unintended consequences of teacher and school accountability programs include: focusing on marginal students (Reback, 2008; Neal and Schanzenbach, 2010), teaching to the test (Jacob, 2005; Glewwe et al., 2010; Koretz, 2002), shifting away from untested subjects (Jacob, 2005; Dee and Jacob, 2011; Glewwe et al., 2010), manipulating student test pool (Cullen and Reback, 2006; Figlio, 2006), distorting effort across grades (Macartney, 2016), among other responses.

[6] This literature finds that incentives generated by postsecondary policy can induce students to increase effort but also induce students to strategic behavior such as strategic course-taking, course withdrawal, and effort allocation relative to the timing of eligibility criteria. These results are in line with the effects I find in this paper in a K-12 setting.

1.2 Institutional Background

University of North Carolina System is one of the largest university systems in the U.S., enrolling approximately 225,000 students in 16 university campuses across the state of North Carolina. Upwards of 70% of all Bachelor's degrees earned in the state of North Carolina are issued by the UNC system and around 85% of all UNC first-time freshmen undergraduate students are in-state students (UNC, 2015).

Beginning with the entering freshman class of 2009, in an attempt to improve graduation rates at all UNC campuses and prevent dropouts by only admitting college-ready students, the UNC Board of Governors set minimum admission standards for all campuses. These minimum admission standards consisted of a cumulative high school GPA of 2.5 and an SAT score of 800 or an ACT score of 17[7]. The SAT score requirement was based on the sum of the critical reading and mathematics subtests, a sum that has a possible range of 400-1600. For the reminder of this paper, SAT test scores presented will reflect this measure instead of the full SAT score which also includes the writing subtest.

The UNC Board of Governors implemented these requirements gradually starting with a cutoff of 2.0 GPA and a 700(15) SAT(ACT) score for the 2009 admission season. The cutoffs increased in 2011 to a GPA of 2.3 and an SAT(ACT) score of 750(16). Finally, the 2013 admission season cutoffs increased to the final target. Figure 1.1 shows how this transition was implemented over time. Note that the policy was first implemented in 2009 and affected UNC admission eligibility of students graduating high school in the Spring of 2009 and beginning college in the Fall of 2009. Therefore, the high school graduating cohort of 2009 presents the cleanest natural experiment to examine student response to incentives as, for these students, 12th grade performance was the only potential dimension of adjustment to this policy. Therefore, the focus of this paper will be on this

[7]Before 2009, UNC had no official minimum GPA, SAT, or ACT admission requirements. However, the system had required applicants to have a minimum high school course requirement since 2006. These requirements consisted of 4 credits of math, 4 credits of English, 3 credits of science, 2 credits of social studies, and 2 credits of language. After 2009, minimum GPA, SAT, and ACT requirements were implemented in addition to minimum course requirements. Any given campus could grant exceptions to these new requirements to up to 1% of their admitted pool.

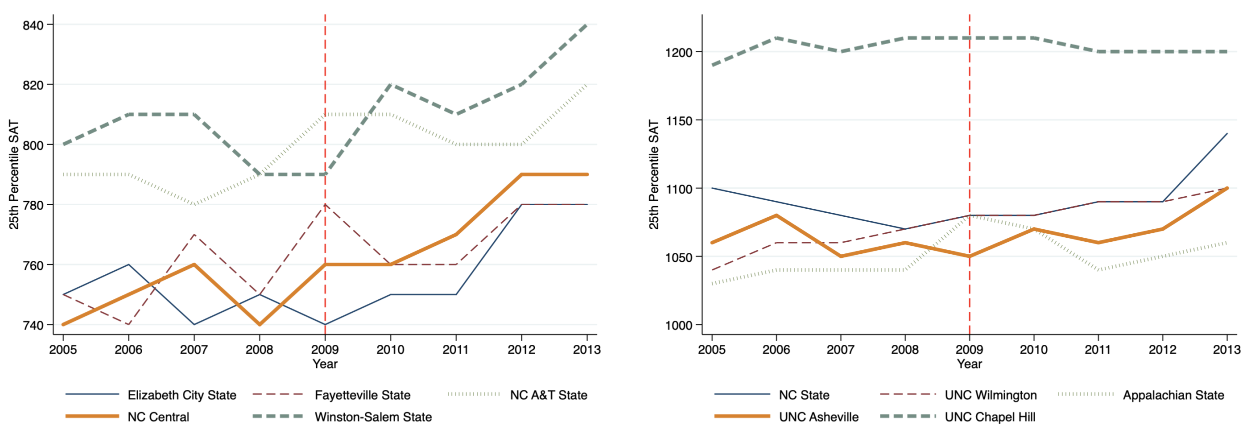
cohort of students.

Figure 1.1: UNC System Minimum Admission Requirements

A key feature of the UNC system is its substantial variation of selectivity across campuses. Given this variation, the minimum admission requirement policy might not have been binding for the most selective institutions within the system. I draw from IPEDS data to address this concern and explore campus-level trends in selectivity based on SAT test scores[8]. Admission officers at each institution report to IPEDS the 25th percentile SAT performance of their admitted undergraduate student pool for each academic year. Figure 1.2 presents the trend over time of 25th percentile SAT scores for the bottom and top five UNC campuses in terms of SAT scores.

Figure 1.2: 25th Percentile SAT of Admitted Students

(a) Bottom 5 Campuses (b) Top 5 Campuses



Notes: This figure shows the evolution of the 25th percentile SAT score for admitted students of UNC campuses over time. Panel (a) shows the evolution for the bottom five campuses. Panel (b) shows the evolution for the top five campuses. The dash red line represent the year in which the UNC Minimum Admission Requirements were implemented. Source: IPEDS

Panel (b) of Figure 1.2 shows how the new admission policy is unlikely to affect admission decisions for campuses that belong to the most selective campuses as only 25% of their admitted

[8]Similar trends are found when ACT 25th percentile scores are used instead of SAT scores (See Figure A.1).

pool scored SATs below 1040. However, Panel (a) of Figure 1.2 shows that the UNC admission policy is likely to change admission decisions for the least selective institutions as 25% of their admitted applicant pool score below 750 on the SAT. In fact, least selective campuses increase their 25th percentile SAT scores of their admitted pool beginning in 2009. Furthermore, these campuses predominately serve minority students, which is a possible mechanism that explains heterogeneity in treatment effects found in Section 1.5.3.

1.3 Data and Descriptive Statistics

I rely on rich administrative data from the North Carolina public schools. These data are stored and managed by the North Carolina Education Research Data Center (NCERDC). I am able to form an unbalanced panel of approximately 75,000 public high school students belonging to the graduation cohort of 2009. These data allow me to capture key information on student outcomes that proxy effort and engagement. These proxies include student GPA, credit-taking, absences and suspensions.

GPA and credit-taking patterns are recovered from administrative records of high school student transcripts. These transcripts record information on the courses taken by each student, the subject, the academic level (e.g. honors, AP, etc.) and the final mark obtained[9]. Information on absences are collected by the state of North Carolina for accountability purposes during each testing period (end of Fall and Spring terms). Finally, data on student suspensions come from administrative records on reportable incidents[10].

In addition to key behavior proxies, the NCERDC data also includes information of student SAT performance on their latest administration from 2009 to 2013. I also observe students' answers to the SAT questionnaire, which provide information on students' postsecondary plan at the time of their SAT administration.

[9]Although high school transcripts are available going back to the 2004-05 school year, coverage on transcript reporting increases over time until it becomes universal for the 2008-09 school year.

[10]Schools in North Carolina are required to report any incidents that occur within the school that lead to a student suspension, expulsion and/or transfer to an alternative learning program.

Table 1.1: Descriptive Statistics

	RD Sample			*Full Sample*		
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	SD	Obs	Mean	SD	Obs
Panel A. Student Demographics						
Female	0.61	0.49	3,745	0.52	0.50	75,656
Economically Disadvantaged	0.58	0.49	3,750	0.38	0.48	75,720
White	0.28	0.45	3,750	0.63	0.48	75,720
Black	0.61	0.49	3,750	0.26	0.44	75,720
Hispanic	0.05	0.21	3,750	0.05	0.23	75,720
Asian	0.03	0.16	3,750	0.02	0.15	75,720
Student with Disability	0.09	0.28	3,750	0.09	0.28	75,720
Academically Gifted	0.02	0.15	3,750	0.21	0.41	75,720
Took SAT	1.00	0.00	3,750	0.52	0.50	75,720
SAT Score	718.08	40.36	3,750	1004.80	190.87	39,222
Panel B. Pre-Treatment Effort and Engagement Outcomes						
GPA 9th Grade	2.63	0.61	3,392	3.02	0.87	67,313
GPA 10th Grade	2.53	0.61	3,441	3.00	0.92	68,629
GPA 11th Grade	2.56	0.63	3,555	3.07	1.00	71,061
Credits 9th Grade	8.28	1.65	3,393	8.07	1.53	67,347
Credits 10th Grade	8.25	1.59	3,441	8.06	1.52	68,649
Credits 11th Grade	8.17	1.57	3,556	7.97	1.52	71,081
Absences 9th Grade	4.91	4.75	3,399	5.48	5.60	67,269
Absences 10th Grade	5.64	5.53	3,500	6.39	6.37	69,541
Absences 11th Grade	6.51	6.34	3,645	7.48	7.77	72,103
Suspensions 9th Grade	0.23	0.84	3,453	0.18	0.78	68,261
Suspensions 10th Grade	0.18	0.71	3,531	0.17	0.71	69,979
Suspensions 11th Grade	0.18	0.72	3,662	0.17	0.74	72,623

Notes: This table shows descriptive statistics of the data. Panel A. provides summary statistics of student demographics. Panel B. provides summary statistics of student effort and engagement outcomes in pre-treatment years (pre 12th grade). Columns 1 through 3 show descriptive statistics of the sample used for the estimation of the main results shown in Table 1.2 using the optimal bandwidth estimated for GPA. Columns 4 through 6 show descriptive statistics of the all students belonging to the graduating cohort of 2009.

Table 1.1 provides an overview of the data for high school students belonging to the graduating cohort of 2009. Columns (1) through (3) present descriptive statistics for students who score close to the UNC cutoff on their SATs. These students make up the regression discontinuity sample, which are the data used for the main estimates of this paper. Columns (4) though (6) present descriptive statistics for the full graduating cohort of 2009.

Overall, students close to the UNC cutoff are more likely to be female, economically disadvantaged (EDS)[11], and more likely to be African American compared to the full cohort of North Carolina public school students. Furthermore, they are less likely to have a disability or be classified as academically gifted. By construction, all students in the RD sample take the SAT and score less than the average student who takes the SAT. Finally, in terms of pre-treatment effort measures, students belonging to the RD sample generally have lower measures of engagement and performance. They have lower GPAs, take slightly more coursework, are most absent, and are suspended more times than their counterparts throughout their high school years.

1.4 Research Design

This section details the empirical framework of this paper. I first describe the regression discontinuity design and estimation used for the main estimates which are presented in Section 1.5. I then describe the framework used to set identify regression discontinuity estimates in the presence of potential manipulation of the running variable following Gerard et al. (2020). Estimates of these bounds are presented in Section 1.6.

1.4.1 Regression Discontinuity Design

The sharp cutoffs set by UNC's new minimum admission requirements policy, I argue, generated changes in incentives for effort for students close to the cutoff. Students that score just below the SAT requirements have a zero probability of admission to any UNC campus. However, students who score just above, perceive a positive probability of admission. Therefore, the probability of admission to a UNC campus has a discontinuous jump at the policy cutoff. Students who perceive this jump in the probability of admission will have increased incentives to perform well during their remaining academic coursework. Several mechanisms could drive the change in student behavior at the cutoff. Students can choose to exert more effort in order to meet the UNC GPA admission requirement once they've fulfilled the SAT requirement. Students can also choose to exert more

[11] As proxied by student eligibility for free or reduced-price lunch.

effort in their remaining coursework to further increase probability of admission via increased GPAs. Finally, passing the first UNC requirement might increase motivation and effort regardless of student strategic behavior with respect to the admission policy.

I use student SAT test score data under a regression discontinuity framework to leverage quasi-random variation from students' SAT scores to compare effort of students who perform on either side of the threshold. Students just above and just below the SAT cutoff can be thought to be similar on several dimensions except for their UNC admission eligibility.

The regression discontinuity specification is given by:

$$y_i = \gamma + \beta D_i + f(SAT_i, D_i) + \epsilon_i \tag{1.1}$$

where i indexes students. The variable y_i represents the 12th grade outcome of interest, D_i is a treatment indicator variable that takes the value of one when students' SAT score meet the relevant UNC cutoff. SAT_i represents students' performance on their first SAT test and is the running variable of this specification. The estimated coefficient, $\hat{\beta}$, should be interpreted as the effect differential incentives embedded in the UNC admission policy on student engagement during their senior year.

Throughout this paper, estimation of Equation (1.1) is carried out via local linear regression (Gelman and Imbens, 2019), using optimal bandwidths and robust confidence intervals proposed by Calonico et al. (2014). Because optimal bandwidths are estimated separately for each student outcome, the number of observations in each estimation may vary. However, estimates are robust to bandwidth selection (see Appendix A.2). Finally, standard errors are clustered at the individual levels of the running variable to correct for misspecification due to discrete running variable (Lee and Card, 2008).

1.4.2 Regression Discontinuity Bounds under Manipulated Running Variable

A first order concern of the estimation presented in Equation (1.1) is the fact that I do not observe all student SAT administrations. Instead, I only observe the last one. Student SAT retaking can be an important source of endogeneity if students systematically retake the SAT in order to place themselves to the right of the cutoff to gain UNC admission eligibility. This selection into the right side of the cutoff can be an important threat to the internal validity of the regression discontinuity design presented in Equation (1.1)[12]. I address this concern by bounding the regression discontinuity point estimates following Gerard et al. (2020).

Gerard et al. (2020) develop a framework to estimate sharp bounds of regression discontinuity treatment effects under a potentially manipulated running variable. For the purposes of this paper, consider two types of students: manipulators ($M_i = 1$) and non-manipulators ($M_i = 0$). Manipulators have perfect control of the *side* of the cutoff they are in and manipulate in only one direction. Therefore, manipulators can perfectly retake the SAT until they are on the right side of the cutoff and they will never retake to be on the left side of the cutoff. Under this setting, the parameter of interest is the treatment on the treated for non-manipulators[13]:

$$\begin{aligned}\Gamma_0 &= E(Y_i(1) - Y_i(0)|X_i = c, M_i = 0) \\ &= E(Y_i|X_i = c^+, M_i = 0) - E(Y_i|X_i = c^-, M_i = 0)\end{aligned} \tag{1.2}$$

Note that, because manipulation is assumed to have one direction (students retake the SAT to be on the right side of the cutoff), we know that $E(Y_i|X_i = c^-, M_i = 0) = E(Y_i|X_i = c^-)$. However, we cannot identify the term $E(Y_i|X_i = c^+, M_i = 0)$ of Equation 1.2, because on the right side of the cutoff we cannot disentangle manipulators from non-manipulators. Therefore, we cannot point

[12]In the past, studies that use SAT scores as the running variable in regression discontinuity designs, as for example, Goodman et al. (2017), find no differences in estimated treatment effects when using students' first and last SAT administration scores.

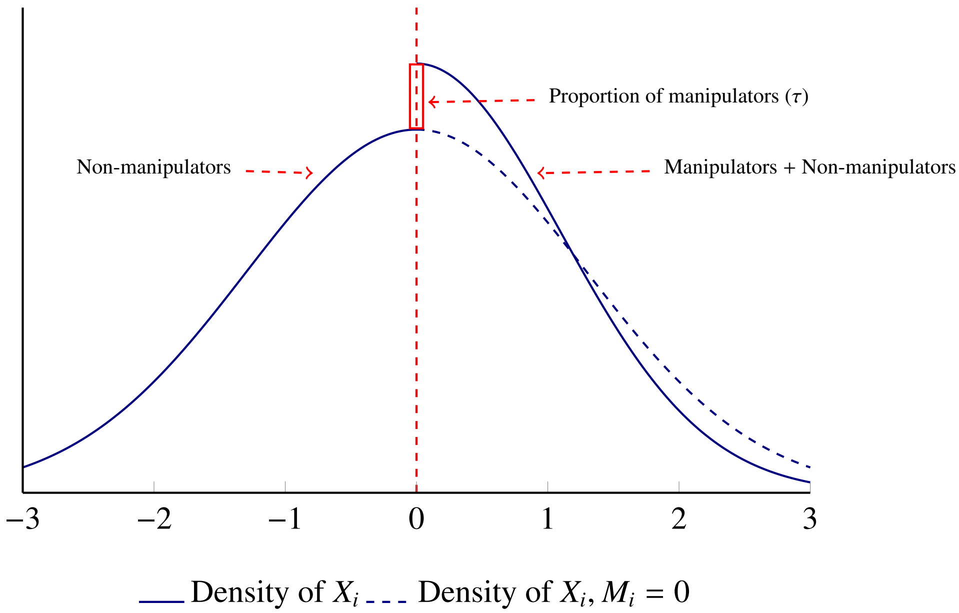
[13]The treatment effect for manipulators cannot be recovered as, by assumption, all manipulators on the right side of the cutoff and are treated (Gerard et al., 2020).

identify Γ_0. However, we can bound it using the estimated proportion of manipulators (τ) that is identified from the data:

$$\tau \equiv Pr(M_i = 1 | X_i = c^+) = \frac{f_x(c^+) - f_x(c^-)}{f_c(c^+)} \tag{1.3}$$

To understand why the proportion of manipulators (τ) is identified from the data, Figure 1.3 presents an illustration of the density of a manipulated running variable. All students to the left of the cutoff are non-manipulators by definition. However, students on the right side of the cutoff are both manipulators and non-manipulators. Since manipulators are able to control the side on the cutoff they are in, this generates a discontinuity of the density of the running variable at the cutoff. This discontinuity represents the proportion of students who manipulate the running variable.

Figure 1.3: Identification of Manipulators



Notes: This figure presents an illustration of how the proportion of manipulators can be recovered from the data. Source: Gerard et al. (2020).

Using the proportion of manipulators, lower bounds can be derived by trimming quantile $(1-\tau)$ of the outcome distribution, and therefore, excluding units with the highest outcomes from the estimation. Analogously, upper bounds can be derived by trimming quantile τ of distribution, and excluding the units with the lowest outcomes. Then, sharp lower and upper bounds on Γ_0 are given by:

$$\begin{aligned} \Gamma_0^L &= E(Y_i|X_i = c^+, Y_i \leq Q_{Y|X=c^+}(1-\tau)) - E(Y_i|X_i = c^-) \\ \Gamma_0^U &= E(Y_i|X_i = c^+, Y_i \geq Q_{Y|X=c^+}(\tau)) - E(Y_i|X_i = c^-) \end{aligned} \tag{1.4}$$

where $Q_{Y|X}$ is the conditional quantile function of the outcome variable. Estimation of both the density function and the conditional quantile are carried out via a local linear polynomial approximations. For more detailed explanation of estimation and inference see Gerard et al. (2020).

1.5 Results

This section presents regression discontinuity estimates of Equation (1.1) for the 2009 cohort. First, I present core results on student effort and engagement proxies in Section 1.5.1. Then I present mechanisms in Section 1.5.2, by exploring effects on student course-taking patterns. Finally, in Section 1.5.3, I present heterogeneity of the results across student demographics and aspirational postsecondary plans.

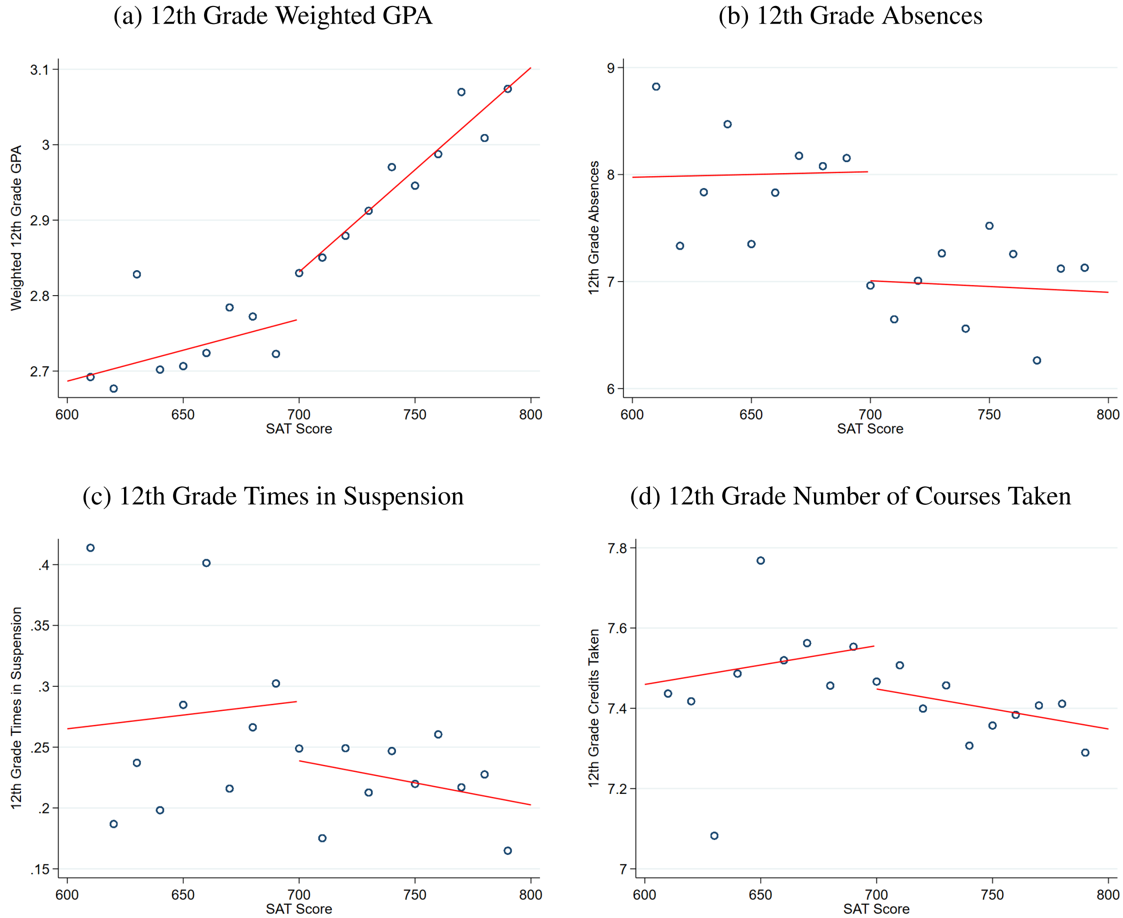
1.5.1 Core Results

In this section, I present results for estimating Equation (1.1) for the 2009 cohort of students. Panels (a) through (d) of Figure 1.4 plot average 12th grade GPA, absences, times in suspension, and credit-taking for each SAT score bin with a global linear fit on either side of the 700 cutoff[14]. Visual evidence provided by these plots suggest a possible student response on 12th grade GPA, absences, and times in suspension.

Table 1.2 presents regression discontinuity estimates of Equation (1.1) for GPA, absences, suspensions, and credit-taking in Columns (1) to (4), respectively. These estimates indicate that stu-

[14]The UNC minimum admission requirements sets a GPA threshold that is based on students' high school weighted GPA. Weighted GPA assigns additional grade points for Honors and AP/IB courses. I use weighted GPA instead of raw GPA as an outcome variable is to reflect this policy. However, Appendix A.3 shows that results are robust to using raw GPA as the outcome.

Figure 1.4: Global Linear Fit



Notes: This figure shows discontinuities in senior-year weighted GPA, absences, times in suspension, and course-taking (Panels (a), (b), (c), and (d). All panels are created by plotting the average outcomes across each SAT score bin and fitting the data using a linear regression without controls on either side of the cutoff.

dents who score just above the UNC SAT minimum admission threshold obtain GPAs that are, on average, 0.08 points higher compared to students who score just below. This represents an increase of 2.8% from the control group mean and an increase of 0.11 standard deviations. In addition, threshold crossing also reduces absences by 1.5 days (0.19 SD) and suspensions by 0.07 times (0.09 SD). Finally, Column (4) shows that there is no statistically significant effect on the number of courses students take in 12th grade. Overall, these results suggest an increase in student engagement as a result of the perceived changes in incentives brought by the policy.

Another dimension of student behavior that might shift as a result of meeting the UNC SAT cutoff is students' aspirations. Table 1.3 presents evidence of this. During the Spring of 12th

Table 1.2: Effects on Student Effort and Engagement Outcomes

	(1) GPA	(2) Absences	(3) Suspension	(4) Credits
RD Estimate	0.078**	-1.479***	-0.074*	-0.084
	(0.026)	(0.172)	(0.030)	(0.060)
Bandwidth	79.40	65.40	82.34	90.29
Observations	3,748	2,836	4,298	4,858
Control Mean	2.75	8.00	0.27	7.50
Control SD	0.68	7.98	0.78	1.92

Notes: This table shows regression discontinuity estimates of senior-year GPA, course-taking, absences and days in suspension. Optimal bandwidths are selected following Calonico et al. (2014). Local linear regressions are estimated on either side of the cutoff. Standard errors are clustered at the individual levels of the running variable to account for misspecification due to discrete running variable (Lee and Card, 2008). * p<0.1, ** p<0.05, *** p<0.01.

grade, all seniors in the state of North Carolina are surveyed and asked to report where they are bound for after high school graduation. From the responses to this survey, I can back out whether students report they are bound to a 4-year institution and whether the institution is a public in-state institution (UNC campus).

A concern with these data is that they are self-reported and can reflect intentions rather that actual transitions. As, I cannot actually observe where students enroll after high school graduation, I present these results with that caveat. While not statistically significant, estimates presented in Table 1.3 provides evidence that threshold crossing increases students' likelihood of reporting they are bound for a UNC campus by 0.08 percentage points (0.14 SD) and increases the likelihood of reporting they are bound for a 4-year institution by 0.05 percentage points (0.10 SD).

1.5.2 Mechanisms

I interpret decreased absences and suspensions shown in Table 1.2 as a reflection of increased engagement in school work due to the discontinuity in the probability of admission. Nevertheless, interpretation of increased student GPA is less straightforward. Students can increase their GPA by either exerting more effort in their coursework or by substituting away from more demanding

Table 1.3: Effects on Student Reported Intentions After High School

	(1) Bound UNC	(2) Bound 4-year
RD Estimate	0.075	0.045
	(0.039)	(0.032)
Observations	4806	4298
Control Mean	0.31	0.45

Notes: This table shows regression discontinuity estimates of senior-year reporting of their intentions after High School. Optimal bandwidths are selected following Calonico et al. (2014). Estimated bandwidths for UNC-going and 4-year going intentions 95.50 and 84.58, respectively. Local linear regressions are estimated on either side of the cutoff. Standard errors are clustered at the individual levels of the running variable to account for misspecification due to discrete running variable (Lee and Card, 2008). * p<0.1, ** p<0.05, *** p<0.01.

courses.

I explore this mechanism by splitting the effects on student overall GPA and course-taking into core and non-core subjects. I define core courses as courses in the subjects of English, Math, Science, and Social Studies. Non-core courses are all other course subjects. Presumable, course-taking in core subjects is more demanding. Panel A. of Table 1.4 presents regression discontinuity estimates for core and non-core GPA and course-taking in 12th grade.

Overall, effects of student GPA seem to be driven by increases in student GPA from non-core courses. Additionally, Columns (3) and (4) of Table 1.4 indicate that students are substituting away from core credit-taking and increasing course-taking in non-core subjects. Threshold crossing decreases core-course credit-taking by 0.225 (6.4%) credits while it increases non-core credit-taking by 0.221 credits (5.5%). I disaggregate these results further by presenting changes in credit-taking due to threshold crossing across core subjects in Panel B. of Table 1.4. Threshold crossing induces students to take less courses in math, social studies, and ELA. However, effects are larger and statistically significant for diminished credit-taking in social studies and ELA.

While Panels A and B of Table 1.4 provide evidence that students are substituting away from more demanding coursework as they cross the SAT cutoff, course subject categories are broad and might not be a precise measure of course difficulty. To address this concern, I recover school-

Table 1.4: Effects for Credit-Taking

	(1)	(2)	(3)	(4)
Panel A. Outcomes Across Core and Non-Core Subjects				
	Weighted GPA		*Credits*	
	Core	Non-Core	Core	Non-Core
RD Estimate	0.022	0.087*	-0.225***	0.221***
	(0.018)	(0.039)	(0.049)	(0.028)
Panel B. Credit-Taking Across Core Subjects				
	Math	Science	Social	ELA
RD Estimate	-0.014	-0.005	-0.084*	-0.109***
	(0.046)	(0.040)	(0.034)	(0.008)

Notes: This table shows regression discontinuity estimates of senior-year GPA, course-taking of core and non-core courses. I define core courses as courses in the subjects of English, Math, Science, and Social Studies. Non-core courses are all other course subjects. Optimal bandwidths are selected following Calonico et al. (2014). Estimated bandwidths for core GPA, non-core GPA, core credits, and non-core credits are 77.31, 100.7, 59.4, and 51.76, respectively. Estimated bandwidths for math, science, social studies, and ELA course-taking are 91.10, 80.85, 58.13, and 64.61, respectively. Local linear regressions are estimated on either side of the cutoff. Standard errors are clustered at the individual levels of the running variable to account for misspecification due to discrete running variable (Lee and Card, 2008). * p<0.1, ** p<0.05, *** p<0.01.

year-course-level grade averages. With these averages, I classify a course as an ”A course”, if on average, students obtain an A in that course. I do the analogous classification for ”B courses” and ”C courses”. I then recover the number of courses students take for each category. Table 1.5 presents regression discontinuity estimates on these variables. Even though none of the estimates are statistically significant, the direction of the coefficients follow the same underlying mechanism shown in Table 1.4. Students seem to be taking more courses where, on average, students obtain an A, and less courses where, on average, students obtain an C.

Overall Tables 1.4 and 1.5 suggest that students are substituting away from more demanding coursework as they cross the threshold and therefore, effects on GPA presented in Table 1.2 might not fully reflect increased effort but rather strategic course choice intended to boost 12th grade GPAs.

Table 1.5: Credit-Taking Effects Across Average A, B and C courses

	(1) A-Courses	(2) B-Courses	(3) C-Courses
RD Estimate	0.119	-0.047	-0.043
	(0.073)	(0.051)	(0.061)
Observations	3115	3142	2230

Notes: This table shows regression discontinuity estimates of senior-year course-taking across courses for which students on average obtain an A (Column (1)), a B (Column (2)), and a C (Column (3)). Optimal bandwidths are selected following Calonico et al. (2014). Estimated bandwidths for A, B, and C courses are 88.67, 90.29, 69.07, and 57.57, respectively. Local linear regressions are estimated on either side of the cutoff. Standard errors are clustered at the individual levels of the running variable to account for misspecification due to discrete running variable (Lee and Card, 2008). * p<0.1, ** p<0.05, *** p<0.01.

1.5.3 Heterogeneity

In this section I explore heterogeneity of the effects across different student populations. First, I explore heterogeneity effects across student demographics. Table 1.6 presents estimates of Equation (1.1) of GPA, absences, and times in suspension across student race, SES, and gender.

Columns (2) and (3) present regression discontinuity estimates for white and non-white students, respectively. Differences across these two groups are statistically significant for GPA, suspensions, core and non-core credit-taking. Thus, these results suggest that minority students respond to the incentives embedded by the UNC admission policy by obtaining higher GPA's while at the same time engaging in less strategic course-taking compared to the response of their counterparts. Columns (4) and (5) presents results for students who are economically disadvantaged and students who are not, respectively. Statistically significant differences across these two groups include absences, suspensions, core and non-core course-taking. These results suggest that low income students that make the SAT cutoff have higher effort and engage in less strategic course-taking than their counterparts. However, this change in behavior does not translate into higher GPAs.

Finally Columns (6) and (7) present estimate for female and male students, respectively. Statistically significant differences across these two groups include GPA, absences, suspensions, non-

Table 1.6: Effects Across Student Demographics

	(1) Full	(2) White	(3) Non-White	(4) EDS	(5) Non-EDS	(6) Female	(7) Male
Panel A. Effort and Engagement Outcomes							
GPA	0.078**	0.045	0.086***	0.071	0.055	0.043	0.175***
	(0.026)	(0.077)	(0.021)	(0.037)	(0.126)	(0.032)	(0.052)
Absences	-1.479***	-0.915	-1.523***	-1.802***	-0.706	-1.935***	-0.247
	(0.172)	(1.145)	(0.427)	(0.435)	(0.703)	(0.495)	(0.432)
Suspensions	-0.074*	0.023	-0.095*	-0.152*	0.090	-0.080***	-0.072
	(0.030)	(0.045)	(0.039)	(0.049)	(0.068)	(0.026)	(0.064)
Panel B. Student Credit-Taking							
Core	-0.225***	-0.449***	-0.119	-0.102	-0.358**	-0.220***	-0.193
	(0.049)	(0.103)	(0.073)	(0.081)	(0.130)	(0.028)	(0.130)
Non-Core	0.221***	0.566***	0.122***	0.156	0.201	0.095*	0.374***
	(0.028)	(0.061)	(0.034)	(0.103)	(0.119)	(0.043)	(0.051)

Notes: This table shows regression discontinuity estimates of senior-year GPA across student demographics. Optimal bandwidths are selected following Calonico et al. (2014). Local linear regressions are estimated on either side of the cutoff. Standard errors are clustered at the individual levels of the running variable to account for misspecification due to discrete running variable (Lee and Card, 2008). * p<0.1, ** p<0.05, *** p<0.01.

core course-taking. These results suggest that male students engage in more gaming practices via course selection and obtain higher GPAs than female students without increased effort.

As discussed in Section 1.2, the new admission requirements were likely to affect admission decisions of the least selective campuses within the UNC system which predominately serve minority students. As such, the heterogeneity results found in Table 1.6 are in line with minority and less affluent students responding more strongly to the policy because the foresee changes in admission practices of the UNC campuses they were intending to attend. Furthermore, these results are also consistent with Black and Hispanic students being less likely to have institutional knowledge and access to support systems such as academic counseling (Kirst and Venezia, 2004).

Table 1.7: Effects Across Student Postsecondary Plans

	(1) Full	(2) UNC	(3) Non-UNC
Panel A. Effort and Engagement Outcomes			
GPA	0.078**	0.158	-0.025
	(0.026)	(0.085)	(0.063)
Absences	-1.479***	-1.308	-1.681
	(0.172)	(0.797)	(1.049)
Suspensions	-0.074*	-0.102*	-0.041
	(0.030)	(0.046)	(0.037)
Panel B. Student Credit-Taking			
Core	-0.225***	-0.144*	-0.266***
	(0.049)	(0.073)	(0.079)
Non-Core	0.221***	0.470***	-0.109
	(0.028)	(0.079)	(0.122)

Notes: This table shows regression discontinuity estimates of senior-year GPA, absences and times in suspension for students all students in Column (1), for students who want to attend a UNC campus in Column (2), and for students who do not want to attend a UNC campus in Column (3). Optimal bandwidths are selected following Calonico et al. (2014). Local linear regressions are estimated on either side of the cutoff. Standard errors are clustered at the individual levels of the running variable to account for misspecification due to discrete running variable (Lee and Card, 2008). * p<0.1, ** p<0.05, *** p<0.01.

Second, I explore heterogeneity across students' postsecondary plans. Presumably, students' effort allocations are unaffected by the UNC admission policy if they have no intention to attend a UNC campus. I draw from student survey response from the SAT Questionnaire that is given to students prior to their SAT administration. This survey asks students about their postsecondary plans. Students report if they want to attend a public or private university, whether they want to attend a 2-year or 4-year institution, and whether they want to attend an institution in their home state. Answers to these questions allow me to back out whether students want to attend a UNC campus or not.

Table 1.7 provides regression discontinuity estimates for GPA, absences, and times in suspension for students who want to attend a UNC campus and for students who do not. Differences across these two groups are all statistically significant for all student outcomes except for absences. Overall, effects seem to be driven by students who want to attend a UNC campus as threshold crossing generates a greater increase in GPA and a greater decrease in suspensions than their counterparts. These results are in line with the the idea that the UNC admission policy change shifted students incentives for effort for those students who did want to attend a UNC campus. For students who did not, there are no perceived changes in incentives for their effort in school work.

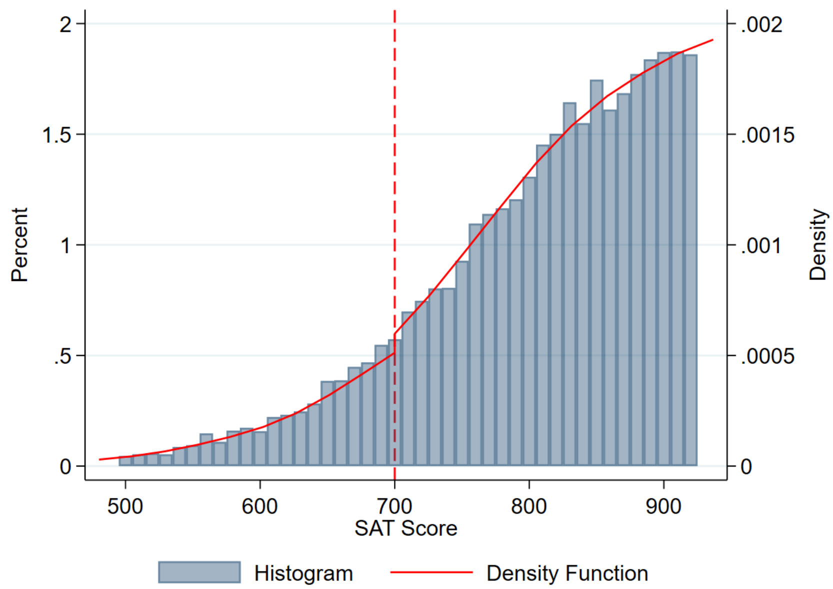
1.6 Internal Validity Checks

My results show that the discontinuity in the probability of admission to a UNC campus, generated by the 2009 admission policy shift, led students to the right of the cutoff to have higher effort levels, as proxied by absences and suspensions, but also led students to substitute away from more demanding coursework in order to boost their GPAs. The ability of my research design to produce causal estimates of these effects rests on the assumption that validity holds. An important threat to the validity of this design is student SAT retaking. Because I am only able to observe students' latest SAT administration, if students retake the SAT until they meet the UNC cutoff, then my results might be reflecting selection into SAT retaking rather than effects on student effort. In this section, I provide four types of evidence based on observables to address this validity concern.

First, I explore whether there is evidence of manipulation of the running variable which could be driven by SAT retaking. I implement a local polynomial density test following Cattaneo et al. (2019). This test estimates a local polynomial density on either side of the cutoff and tests for the null hypothesis that the limit of both functions when they approach the cutoff from either side are equal. Figure 1.5 presents the fit of the local polynomial density functions and the histogram of the SAT test scores. With a p-value of 0.376, I am not able to reject the null hypothesis that the density functions are equal at the cutoff. Therefore, this test provides evidence that there is no systematic manipulation of the running variable.

The second type of evidence examines the relevance of the policy cutoff. The idea behind this exercise is the following: if students are responding to the UNC admission policy, I should not find

Figure 1.5: Local Polynomial Density Test



Notes: This figure shows the histogram of the SAT test scores in blue bars and the local polynomial density estimation on either side of the cutoff in red solid lines estimated following Cattaneo et al. (2019).

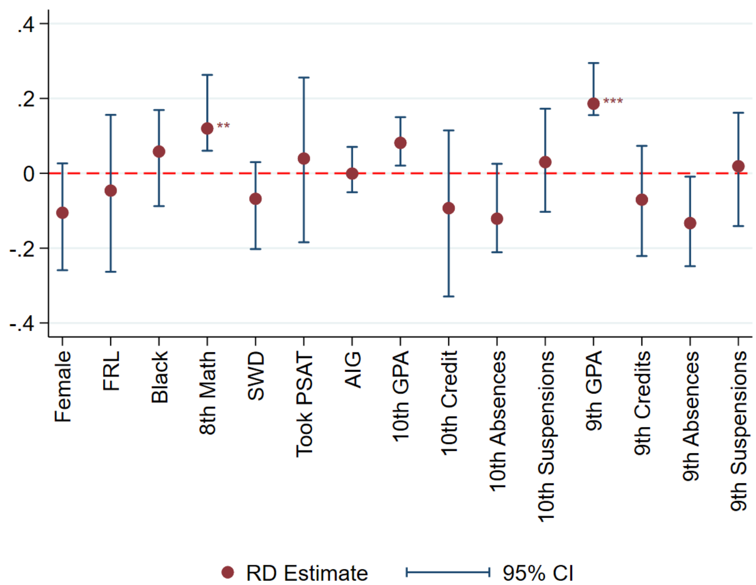
Table 1.8: Regression Discontinuity Bounds of Effects on Student Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	670	680	690	700	710	720	730
GPA	0.057	-0.030	-0.025	0.078**	0.053	0.026	0.015
	(0.033)	(0.039)	(0.029)	(0.026)	(0.035)	(0.019)	(0.015)
Absences	0.665*	-0.147	-0.717	-1.479***	-0.485	0.628*	0.644
	(0.266)	(0.332)	(0.446)	(0.172)	(0.514)	(0.263)	(0.359)
Suspensions	-0.124	-0.015	-0.019	-0.074*	-0.048	0.067*	0.033
	(0.064)	(0.073)	(0.045)	(0.030)	(0.032)	(0.031)	(0.038)

Notes: This table shows regression discontinuity estimates of senior-year GPA, absences and times in suspension for different SAT cutoffs. Optimal bandwidths are selected following Calonico et al. (2014). Local linear regressions are estimated on either side of the cutoff. Standard errors are clustered at the individual levels of the running variable to account for misspecification due to discrete running variable (Lee and Card, 2008). * p<0.1, ** p<0.05, *** p<0.01.

discontinuity of student outcomes on values of the running variable other than the relevant cutoff of 700. Table 1.8 presents evidence that this is the case. Discontinuity of student outcomes are only found in the relevant UNC admission policy cutoff.

Figure 1.6: Discontinuity in Student Demographics and Pre-Treatment Covariates



Notes: This figure shows regression discontinuity estimates of Equation (1.1) using standardized pre-treatment student covariates and outcomes as the dependent variable. Robust 95% confidence intervals are estimated following Calonico et al. (2014). I apply the Bonferroni correction to adjust significance for multiple comparisons. * p<0.1, ** p<0.05, *** p<0.01.

The third type of evidence for validity examines covariate smoothness across the cutoff. Figure 1.6 regression discontinuity estimates of Equation (1.1) using standardized pre-treatment student covariates and outcomes as the outcome variable. Scatter plots of these covariates, analogous to the ones shown in Figure 1.4 can be found in the Appendix A.4. Across most estimates, I find no evidence of a discontinuity except for two covariates: 8th grade math test scores and 9th grade GPA. Discontinuities in these variables could be driven by SAT retaking if higher achieving students systematically retake the SAT in order to be on the right side of the cutoff.

Despite evidence of no systematic manipulation of the running variable given by the local polynomial density test implemented in Figure 1.5, discontinuities of 8th grade math test scores and 9th grade GPA found in Figure 1.6 could indicate systematic manipulation of the running variable if the density test is underpowered to detect jumps in the density functions. In order to address this concern, I present a fourth type of evidence for validity that recovers bounds for the regression discontinuity estimates found in Table 1.2 under the framework presented in Section 1.4.2. These bounds are presented in Table 1.9. Overall, the estimated bounds of the regression discontinuity

Table 1.9: Regression Discontinuity Bounds of Effects on Student Outcomes

	(1) GPA	(2) Course-Taking	(3) Absences	(4) Suspension
RD Estimate	0.078**	-0.084	-1.479***	-0.074*
	(0.026)	(0.060)	(0.172)	(0.030)
Bounds	[0.069, 0.094]	[-0.135, -0.053]	[-2.351, -1.190]	[-0.108, -0.074]
Density Jump	0.007	0.006	0.037	0.007

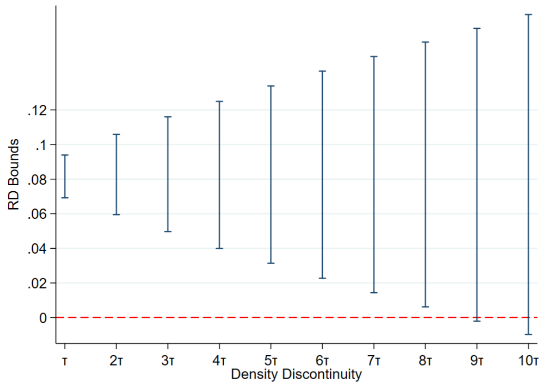
Notes: This table shows regression discontinuity estimates of senior-year GPA, absences and times in suspension for students all students in Column (1), for students who want to attend a UNC campus in Column (2), and for students who do not want to attend a UNC campus in Column (3). Optimal bandwidths are selected following Calonico et al. (2014). Local linear regressions are estimated on either side of the cutoff. Standard errors are clustered at the individual levels of the running variable to account for misspecification due to discrete running variable (Lee and Card, 2008). * p<0.1, ** p<0.05, *** p<0.01.

results are fairly tight. The reason why the bounds are fairly tight is due to the fact that the actual jump in density at the cutoff is estimated to be small.

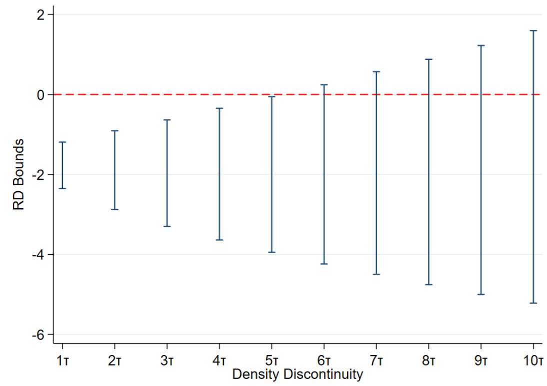
As explained in Section 1.4.2, regression discontinuity estimates are bounded using the estimated proportion of manipulators, τ, to trim upper and lower tails of the outcome distribution. This proportion is identified by the discontinuity of the density function at the cutoff via Equation 1.3. In order to assess how sensitive the bounds are to the estimated density discontinuity, τ, Figure 1.7 presents how regression discontinuity bounds change as τ changes. As τ increases, the bounds becomes larger. However, for most outcomes, the bounds are robust to the estimated proportion of manipulators. The estimated bounds for student 12th grade GPA and student 12th grade absences start to include zero after τ is set to be 9 times and 6 times the actual estimated value, respectively. Furthermore, the estimated bounds for student 12th grade times in suspension never include zero even if the estimated proportion of manipulators, τ, is set to be 10 times the actual estimated value. Thus, it is unlikely that manipulation of the running variable via SAT retaking is such that it completely drives away the results presented in Section 1.5.

Figure 1.7: Sensitivity of Bounds to Estimated Density Discontinuities

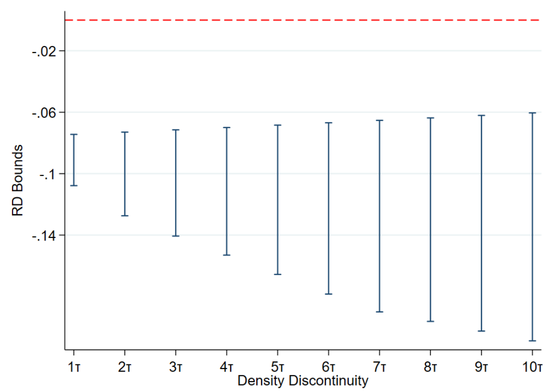
(a) 12th Grade Weighted GPA



(b) 12th Grade Absences



(c) 12th Grade Times in Suspension



Notes: This figure shows discontinuities in senior-year weighted GPA, absences and times in suspension (Panels (a), (b), and (c). All panels are created by plotting the average outcomes across each SAT score bin and fitting the data using a linear regression without controls on either side of the cutoff.

1.7 Conclusion

Changing the incentives students face is one lever for educators and policymakers to improve student learning in the presence of student disengagement. However, education economists and policymakers lack an understanding of the ways in which these incentive structures can shape student responses. Generally, education accountability policies seek to shift incentives for effort of teachers and schools rather than students. Thus, there have not been many opportunities to evaluate the student incentive-effort connection outside of researcher designed pay-for-performance incentives schemes.

Incentive structures embedded within policies for access to higher education provide an alternative, more realistic, setting to examine student effort. In this paper, I leverage exogenous variation from a change in admission requirements in one of the largest university systems in the U.S. that provided a salient and high-stakes setting to examine how student effort responds when incentive structures shift.

Using rich administrative data from the North Carolina public school system, I evaluate whether students' investments in schoolwork shifted as a results of this exogenous shift in incentives for effort. I find that students respond by increasing GPA and decreasing absences and suspensions. While these effects suggest an increase in student engagement as a result of the admission policy, further exploration of course-taking suggests that student increases in GPA are driven by changes in course composition as students substitute away from more demanding coursework. These unintended consequences of admission policies on student course-taking decisions can lead students to miss important learning opportunities in high school, possibly generating detrimental effects on student postsecondary success. Finally, I find important heterogeneity of student response across demographics. minority and less affluent students just above the cutoff have higher effort allocations via decreased absences and suspensions but engage to a lesser extent in gaming practices via course selection.

Chapter 2

Unequal Worker Exposure to Establishment Deaths*

(with Hugh Macartney† and Eric Nielsen‡)

Against a backdrop of rising inequality, the existence of labor market disparities between workers is a key public concern. Economists have studied such disparities extensively, focusing on trends in factors including differential productivity by worker type, the relative supply of workers, unionization, and skill-biased technical change.[1] Considerable attention has also been devoted to understanding the labor market effects of economic shocks, with a view toward worker inequality in several instances.[2] However, existing research has yet to determine whether and how any given

*The views and opinions expressed in this paper are solely those of the authors and do not reflect those of the Board of Governors or the Federal Reserve System.

†Department of Economics, Duke University

‡Federal Reserve Board

[1] Altonji and Blank (1999) and Katz and Autor (1999) offer excellent reviews of the relevant literature. Seminal early work includes Bound and Freeman (1992), Katz and Murphy (1992), Berman et al. (1994), Autor et al. (1998) and Card and DiNardo (2002), among others. Some notable recent additions to the literature are Autor et al. (2008) and Bayer and Charles (2018).

[2] A large literature examines how employment and earnings respond to changes in trade, government spending and defense spending (see Blau et al. 2000; Aizer 2010; Autor et al. 2013; Nakamura and Steinsson 2014; Bertrand et al. 2015; Pierce and Schott 2016, 2017; Goldsmith-Pinkham et al. 2018). Though these analyses place little emphasis on how labor outcomes might differ across worker types, several other works consider the effects of economic shocks through the prism of worker inequality. These include Bound and Holzer (2000), which applies the methodology proposed by Bartik (1991) and developed more fully by Blanchard and Katz (1992) to analyze decadal patterns in

shock-inequality connection is mediated by preexisting worker-firm matching.[3]

In this article, we explore a new and potentially important channel through which adverse economic conditions may affect inequality across workers: the extensive margin of establishment deaths. If workers are matched to establishments non-randomly by type (e.g., by race), then the type that is concentrated within establishments least able to survive the downturn will experience a disproportionate share of the employment decline.[4] Focusing on employment changes from establishment deaths (as opposed to changes from non-closing establishments) is key for understanding the connection between worker inequality and pre-downturn matching, given that, by definition, death-based changes cannot be affected by post-downturn employer favoritism toward one worker group over another. We use the Great Recession, a downturn that was abrupt, deep, and widespread, to uncover evidence of such sorting across establishments by demographic group. Under our approach, the degree to which workers are differentially distributed according to unobservable establishment resilience is revealed by the change in worker employment inequality explained by establishment deaths.

The key methodological contribution of our paper is to recover the component of the change in employment inequality that is explained by establishment deaths. In our context, this change is defined as the difference across worker types (e.g., white vs. black) in the aggregate percent change in employment over the Great Recession.[5] As direct and nationally comprehensive measures of worker-specific employment changes arising from establishment deaths are unavailable, we propose a method for statistically decomposing employment changes by worker type into four

inequality, as well as Couch and Fairlie (2010), Hoynes (1999), Hoynes et al. (2012) and Guvenen et al. (2014), which investigate whether recessions have differential effects by worker type.

[3]A growing literature studies the contribution of worker-firm matching to increasing earnings inequality over time (see Abowd et al. 1999; Card et al. 2013; Barth et al. 2016; Bonhomme et al. 2019; Song et al. 2019), but does not consider how the relationship (or its employment analogue) is affected by large negative shocks.

[4]Research analyzing the effect of demand shocks on firm/establishment closures includes trade-oriented papers, such as Yeaple (2005) and Egger and Kreickemeier (2009), which are not concerned with inequality per se and do not address differences by race or gender. Syverson (2011) reviews a literature documenting substantial productivity differences across firms within narrowly-defined industries.

[5]This measure differs somewhat from the metric commonly used in the literature: the change in a particular worker type's share of aggregate employment or earnings. Our variant is particularly conducive to analyzing demand shocks, as it reveals how each worker type is separately affected. In the case of the standard metric, it is not apparent whether an increase in the share of total employment for a particular type is due to growth for that type or to a contraction in total employment.

mutually exclusive and exhaustive components: those due to establishment deaths, births, contractions, and expansions. The approach exploits variation in each of these establishment-level causes while imposing restrictions so that the overall predicted employment changes by type and by cause match those observed in the data.[6]

With the decomposition in hand, we adapt the well-established procedure from the literature for isolating across- and within-industry variation,[7] separating the overall employment change attributed to establishment deaths into across and within components. These terms intuitively depend on the type-specific employment changes and employment shares at the industry level; the across component is a function of the average employment changes across types and the differences in shares by type, while the within component is a function of the differences in employment changes by type and the average employment shares across types. While the extent to which establishment deaths can explain across-industry patterns is interesting, focusing on the within-industry changes is particularly informative. Doing so rules out the deaths effects from being driven by a correlation between industry-specific preferences and industry vulnerability to demand shocks.

We apply our empirical framework to employment data from the Census Quarterly Workforce Indicators (QWI), broken down by worker type (e.g., race), geography, and industry (at the NAICS four-digit level). These publicly available data cover the vast majority of private sector employment for virtually all counties in the United States for the years surrounding the Great Recession. Our near-universal coverage is important – in order to identify the aggregate effect on worker inequality, it is necessary to trace out how *every* industry and *every* county is impacted by the Great Recession, rather than a subset. We supplement the QWI with Statistics of U.S. Businesses (SUSB) data from the Census on the number of establishment deaths, births, contractions, and expansions by county and four-digit NAICS.[8] We then link this establishment information to employment shifts

[6]While our decomposition method accounts for all categories of establishment-level cause, our establishment death results are especially prominent, given that it is the only component for which the employer cannot exercise discretion in firing.

[7]See Freeman (1975), Freeman (1980), Katz and Murphy (1992), Berman et al. (1994), Autor et al. (1998), Dunne et al. (1996), and Bernard and Jensen (1997).

[8]These data are not part of the standard SUSB distribution, but are available without restriction from Census for a

by industry, county, and worker type.

Applying our decomposition method reveals substantial changes in employment inequality by race over the Great Recession. In keeping with prior analyses, while all worker types lost employment, black workers were disproportionately affected by the downturn, with a decline that was about 25% larger than the loss for white workers and approximately five-sixths larger than Hispanic workers. Black workers fared even worse with respect to employment losses stemming from establishment deaths, losing at twice the rate of white and Hispanic workers.[9]

The decomposition into across- and within-industry components reveals that the more pronounced employment losses for black workers were driven entirely by within-industry declines, with white and Hispanic workers incurring greater losses as a result of their unfavorable preexisting distribution across industries. More importantly, the within-deaths estimates indicate that black workers were disadvantaged within industry precisely because they were concentrated in less resilient establishments prior to the recession. While not our primary focus, we also find that women fared worse than men within industry, but the statistically insignificant within-deaths point estimate is merely suggestive that women were disproportionately employed at weaker establishments pre-recession.

Our work also relates to a prior literature that examines the effects of individual establishment closures, arguing that these can be seen as natural experiments that allow researchers to evaluate the wage losses that result from job separations.[10] Our findings suggest that minority workers are more likely to be employed at firms that are less resilient to negative shocks and more likely to close. This has important implications for the interpretation of the estimates from this previous literature, as it suggests that using individual establishment closures may produce highly local treatment effects that do not necessarily generalize to other populations.

The remainder of the paper is organized as follows: The next section sets out our framework for

nominal fee.

[9]Importantly, local establishment deaths are only modestly correlated with local employment declines – there is meaningful variation in employment that is orthogonal to deaths. Our finding that deaths can explain a substantial fraction of the change in inequality between various worker types is not tautological.

[10]See Couch and Placzek (2010) for a thorough review of this literature.

exploring changes in inequality, detailing the procedure for determining the extent to which the gap in the employment growth rates between worker types can be attributed to each establishment-level cause (deaths, births, contractions, and expansions), as well as decomposing each gap into across- and within-industry components. Section 2.2 describes the data used in our analysis, and provides several stylized facts about the Great Recession. This lends context to our main empirical results, which we present in Section 2.3. Section 2.4 then justifies our approach by formalizing the sources and direction of potential bias, undertaking a sensitivity analysis, and assessing goodness-of-fit using an out-of-sample exercise. Section 2.5 concludes.

2.1 Empirical Framework

This section details our empirical framework, introducing relevant notation when needed. We first describe our novel statistical procedure for allocating employment losses to establishment categories (deaths, births, expansions, and contractions). We then set out our framework for decomposing changes in employment gaps over time into components that arise from across- and within-industry variation. Finally, we discuss how we conduct inference in our setting via the bootstrap.

2.1.1 Decomposition of Employment Changes by Establishment-Level Cause

We begin with the key methodological contribution of the paper: a method for using jurisdiction-industry-level variation in establishment categories to generate predicted employment changes due to each of these causes separately by worker type. Our headline estimates are then constructed by feeding the predicted changes due to establishment deaths into the across-within industry decomposition developed in the subsection that follows.

Let ΔE_{ij}^{τ} denote the change in the total employment of workers of type τ in industry i and jurisdiction j between periods $t = 0$ and $t = 1$ (corresponding to 2007 and 2009, respectively, in our context). Similarly, denote by θ_{ij}^{τ} the percent change in employment, and let θ^{τ} denote its aggregated counterpart (across i and j). The goal is to estimate the component of ΔE^{τ}_{ij} attributable to establishment deaths.

Our approach depends on worker type-agnostic establishment category counts predicting respective worker type-specific employment changes to a first-order approximation – we justify this assumption in Section 2.4. Defining d_{ij}, b_{ij}, c_{ij}, and ex_{ij} to represent the number of establishment deaths, births, contractions, and expansions, respectively, we recover the predicted changes in employment by estimating the following equation for each worker type and industry:

$$\Delta E^{\tau}_{ij} = \beta^{\tau}_{d,i} d_{ij} + \beta^{\tau}_{b,i} b_{ij} + \beta^{\tau}_{c,i} c_{ij} + \beta^{\tau}_{ex,i} ex_{ij} + \epsilon^{\tau}_{ij} \,. \tag{2.1}$$

Notably, our prediction equation omits an intercept, reflecting the fact that deaths, births, expansions, and contractions are the only channels through which employment can change.

We could simply estimate equation (2.1) using OLS, weighting by industry-jurisdiction employment in period 0. However, our goal is to construct the best estimates we can for employment changes due to each possible establishment change. To this end, we leverage additional information to constrain our estimation of equation (2.1) and improve the quality of our predictions.

In particular, we make use of the fact that, at the industry level, we know the total change in employment due to each type of establishment change, as well as the total change in employment (irrespective of cause) for each worker type.[11] For example, consider employment changes in industry i due to establishment deaths. Given estimates $\hat{\beta}^{\tau}_{d,i}$ and $\hat{\beta}^{\tau'}_{d,i}$, our predicted industry-level employment changes due to deaths are given by $\widehat{\Delta E^{\tau}_{i}}|_d = \hat{\beta}^{\tau}_{d,i} \sum_j d_{ij}$ and $\widehat{\Delta E^{\tau'}_{i}}|_d = \hat{\beta}^{\tau'}_{d,i} \sum_j d_{ij}$. While we do not observe the analogues of these changes in our data (which is the reason for carrying out this estimation procedure in the first place), we do observe the total industry-level change in employment due to establishment deaths, $\Delta E_i|d$. Therefore, it is natural to make the following restriction involving $\hat{\beta}^{\tau}_{d,i}$ and $\hat{\beta}^{\tau'}_{d,i}$:

$$\Delta E_i|_d = \widehat{\Delta E^{\tau}_{i}}|_d + \widehat{\Delta E^{\tau'}_{i}}|_d \quad \implies \quad \frac{\Delta E_i|_d}{\sum_j d_{ij}} = \hat{\beta}^{\tau}_{d,i} + \hat{\beta}^{\tau'}_{d,i} \,. \tag{2.2}$$

[11] See Section 2.2 for details.

The restriction in equation (2.2) ensures that the total predicted loss across τ and τ' due to an establishment death equals the average employment loss per establishment death observed in the data. Analogous restrictions apply for births, contractions, and expansions.

Similarly, any set of estimates for the parameters in equation (2.1) yield predicted overall changes in employment separately for τ and τ' workers which can be constrained to equal the observed total employment losses by type:

$$\begin{aligned}
\Delta E_i^\tau &= \widehat{\Delta E_i^\tau}|_d + \widehat{\Delta E_i^\tau}|_b + \widehat{\Delta E_i^\tau}|_c + \widehat{\Delta E_i^\tau}|_{ex} \\
&= \hat{\beta}_{d,i}^\tau \sum_j d_{ij} + \hat{\beta}_{b,i}^\tau \sum_j b_{ij} + \hat{\beta}_{c,i}^\tau \sum_j c_{ij} + \hat{\beta}_{ex,i}^\tau \sum_j ex_{ij} \, . \qquad (2.3)
\end{aligned}$$

In total, we can impose up to six restrictions for each (τ, τ') comparison (four restrictions as per equation (2.2) and two restrictions as per equation (2.3)). However, we have found that the optimization performs better (and is more stable across bootstraps) when the number of constraints is reduced by one. Therefore, our estimation does not impose equation (2.2) for expansions, as the expansion predictions are not of direct interest to us. Nevertheless, our estimated parameters yield implied aggregate employment changes by firm expansions that are quite close to the observed values.

It is worth emphasizing that because we cannot directly link employment in the QWI to particular establishments, the possibility remains that some of the employment losses we attribute to establishment deaths are actually attributable to a different cause. Nonetheless, the close correspondence between the predicted employment changes and the observed employment changes, using an out-of-sample goodness-of-fit assessment (see Section 2.4 for details), coupled with the fact that we simultaneously condition on all establishment variables, provides strong support for our attribution. It is difficult to conceive of a driver of employment changes that is distinct from, but correlated with, deaths and which would not be picked up by variation in births, contractions, or expansions. Furthermore, Section 2.4 also shows that our results would be directionally unchanged even in the face of significant bias in our estimates for the effects of establishment deaths on worker type-specific employment.

2.1.2 Decomposition of Employment Gaps by Industry

We now turn to the task of decomposing the across- and within-industry components of employment changes for different worker types during the Great Recession. Our measure of the overall change in the employment gap between type-τ and type-τ' workers is given by $[\theta^{\tau} - \theta^{\tau'}]$. To determine how much of this difference is due to *within-industry* and *across-industry* variation, we define *within-industry* variation as arising from within-industry differences in type-specific growth rates: that is, $\theta_i^{\tau} \neq \theta_i^{\tau'}$ for some i. *Across-industry* variation can then be recovered from the difference between overall and within-industry adjustments.

As detailed in Online Appendix B.5, the within-industry component is given by

$$[\theta^{\tau} - \theta^{\tau'}]_W = \sum_i (\theta_i^{\tau} - \theta_i^{\tau'}) \left[\frac{(E_{i0}^{\tau}/E_0^{\tau}) + (E_{i0}^{\tau'}/E_0^{\tau'})}{2} \right]. \tag{2.4}$$

Intuitively, it is the across-industry sum of the difference in the employment losses by type, weighted by the average share of employment in industry i. Similarly, the across-industry component is given by

$$[\theta^{\tau} - \theta^{\tau'}]_A = \sum_i \left(\frac{\theta_i^{\tau} + \theta_i^{\tau'}}{2} \right) \left(\frac{E_{i0}^{\tau}}{E_0^{\tau}} - \frac{E_{i0}^{\tau'}}{E_0^{\tau'}} \right). \tag{2.5}$$

It is the across-industry sum of the difference in industry-i employment shares by type, weighted by the (unweighted) average employment loss in i.

Note that the decomposition $[\theta^{\tau} - \theta^{\tau'}] = [\theta^{\tau} - \theta^{\tau'}]_W + [\theta^{\tau} - \theta^{\tau'}]_A$ can be carried out using either the observed employment changes by worker type (as in the explication above) or the changes predicted from establishment deaths (or any other establishment cause).

2.1.3 Interpreting Significance Under the Bootstrap

We construct confidence intervals and p-values for all of our estimates using 5,000 bootstrap iterations, sampling at the jurisdiction (county) level. To our knowledge, we are the first paper to conduct formal inference on across/within industry decompositions. While the construction and interpretation of the bootstrapped confidence intervals is straightforward, interpreting the p-values is more complicated. The problem is two-fold.

First, we do not know the sampling distribution of our statistics under the relevant null hypotheses, in part because many of these nulls are very likely not true. We address this difficulty by following the recommended approach of shifting the bootstrapped distributions so that they are centered around their respective nulls.

Second, shifting the bootstrapped distributions in this way assumes that they differ only in their locations under different null and alternative hypotheses. We are unaware of any *a priori* reason why this should be true, and the constraints we impose in our estimation (equations (2.2) and (2.3)) render simultaneous, independent shifts of constraint-linked bootstrapped distributions problematic. Therefore, while we report both p-values and confidence intervals for the results presented in Section 2.3, we view the confidence intervals as more reliable measures of the sampling variability of our estimates.

2.2 Data and Stylized Facts

2.2.1 Measures of Employment and Establishment Deaths

Our measure of employment comes from the Quarterly Workforce Indicators (QWI; U.S. Census Bureau 2019a), which is produced by the U.S. Census Bureau. The QWI provides local labor market information (including employment) by quarter-year, jurisdiction (county), industry (four-digit NAICS), and worker demographic (race, ethnicity and gender). These publicly available data are aggregated from the matched employer-employee micro-level Longitudinal Employer-Household Dynamics (LEHD) dataset, which is constructed using administrative records from state unemployment insurance fillings, social security data, federal tax records and other Census data. For our period of interest (2007-2009), the QWI data cover 95% of US private sector jobs and all but one state.[12]

We define each worker type from the QWI worker demographic characteristics. For race/ethnicity comparisons, the categories "white" and "black" consist of white non-Hispanic and black non-Hispanic workers, respectively, while "Hispanic" consists of Hispanic workers of any race. For gender, the process is straightforward: workers are either male or female. Using these definitions, we calculate the employment change during the recession for each worker type. While our employment measures do not condition on full- or part-time status, we draw upon the American Community Survey (ACS; Ruggles et al. 2019) to provide supplemental evidence that changes in the share of part-time workers do not drive our results (see Appendix B.2.1).

The first column of Panels A and B in Table 2.1 reports descriptive statistics for the QWI data. The sample contains 3,128 jurisdictions and 312 four-digit industries. It also covers 122 million jobs, with total employment reported separately by worker type, industry and jurisdiction.

To complement our measure of employment, we exploit information about the number of establishment deaths, births, contractions, and expansions for each county and four-digit NAICS industry code, using dynamic annual data from the Statistics of U.S. Businesses (SUSB; U.S. Census Bureau 2019b).[13] These data are constructed from the Business Information Tracking Series (BITS), which longitudinally tracks each establishment in the United States across successive Business Register records.[14] Establishment deaths (births) are defined in the SUSB data as the number of establishments that have positive (zero) employment in the first quarter of the initial year and zero (positive) employment in the first quarter of the subsequent year. Establishment contractions (expansions) are defined as the number of establishments that have larger (smaller) employment in the first quarter of the initial year than in the first quarter of the subsequent year.

[12]The QWI data did not report information for Massachusetts until 2010. Only 31 states participate in the LEHD program and, with the exception of a very small subset (8 as of this writing), restricted LEHD data must be obtained on a state-by-state basis (with most researchers obtaining only between 14 and 17 states).

[13]SUSB data are extracted from the Census Business Register which collects data on all known single and multi-establishment firms. These data come from several sources including the Economic Census, the Annual Survey of Manufacturers, the Current Business Surveys, and the administrative records of the Internal Revenue Service, the Social Security Administration, and the Bureau of Labor Statistics.

[14]Establishments that have undergone no ownership or organizational changes are matched across years according to their Census identifier. BITS is also able to match those that do change using Employer Identification Numbers, business names and addresses, and industry codes. Doing so guards against over-counting deaths or births.

Table 2.1: Descriptive Statistics

	QWI Sample	SUSB Subsample
Panel A: Industry-Jurisdiction Counts		
Industries	312	289
Jurisdictions	3128	3115
Industry-Jurisdictions	454,542	395,680
Panel B: Employment		
Total	122,068,351	113,480,392
White	82,291,944	77,058,720
Black	14,466,006	13,171,431
Hispanic	15,859,996	14,389,629
Male	61,627,520	57,172,056
Female	60,440,832	56,308,336
Average by Industry	391,245	392,666
	(792,667)	(788,487)
Average by Jurisdiction	39,024	36,430
	(149,751)	(141,356)
Panel C: Establishment Counts		
Initial Total		6,555,543
Deaths		1,495,878
Births		1,287,049
Contractions		3,813,287
Expansions		3,153,324

Notes: The QWI sample contains labor market outcomes for the universe of industries and counties across all states (except for MA) and the time period 2007-2009. The merged QWI-SUSB data used for the analysis is a subsample of the QWI sample, since the SUSB establishment data contain a slightly smaller subset of industries and counties. The SUSB initial establishment count pertains to 2007, while the changes (deaths, births, contractions and expansions) are from 2007 to 2009. Panels A and B present descriptive statistics for 2007.

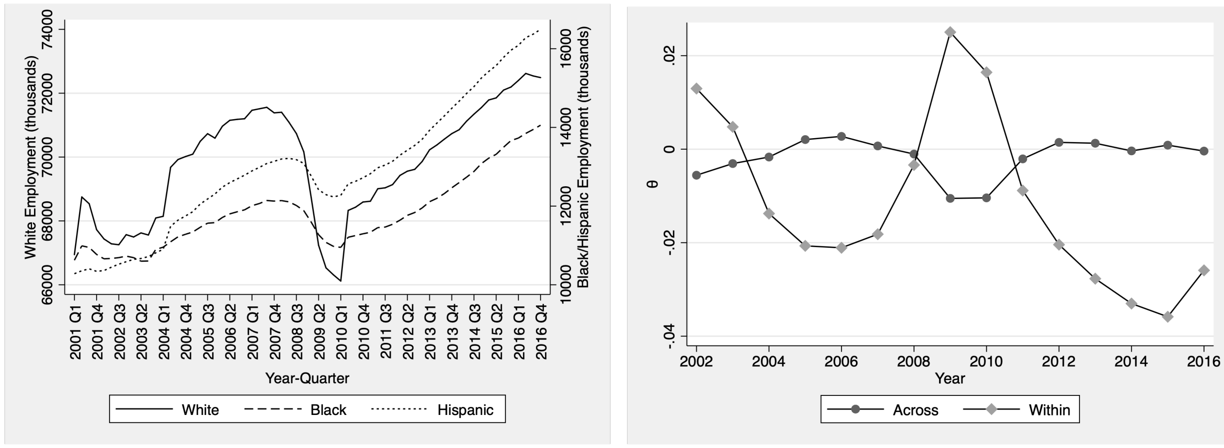
For our primary analysis, we merge the QWI data for 2007-08 and 2008-09 with SUSB establishment data for the corresponding years, forming the "SUSB Subsample" detailed in the second column of Table 2.1. Comparing the two columns, the subsample used for the analysis includes over 99.6% of jurisdictions and 92.6% of industries in the QWI data, accounting for 93% of total employment in the United States. The employment shares of each worker type are similar across samples. Panel C of Table 2.1 reveals that there are about 6.5 million establishments in our sample, with a higher rate of establishment deaths (contractions) than births (expansions) during our period of interest, as one would expect during a deep recession.

As one might assume during a period of low aggregate demand, employment changes and establishment deaths are correlated at the county level. However, the correlation is relatively small (-0.26), indicating that there are substantial employment changes at the county level that are orthogonal to establishment deaths. Births, contractions, and expansions are also correlated somewhat with county-level employment changes, with correlations of 0.07, -0.27, and 0.45, respectively.

2.2.2 Stylized Facts about the Great Recession

Before discussing our main results decomposing employment changes during the Great Recession into across- and within-industry components, as well as components stemming from establishment deaths, we present a number of motivating facts about the period of interest.

The Great Recession substantially impacted establishment deaths and births. The SUSB data reveal that establishment deaths began to increase in 2006, peaking in 2008 and 2009 during the height of the recession. Births, by contrast, rose in 2006 and 2007 before falling substantially during the recession. These establishment changes generated intuitive and significant employment changes, with the employment losses from deaths rising and the gains from births falling during the Great Recession, relative to their prior trends (see Online Appendix Figure B.5).



(a) Aggregate by Race

(b) Across/Within, White vs. Black

Figure 2.1: Aggregate Employment Trends and Decomposition by Worker Race

Notes: Panel (a) of this figure shows the quarterly evolution of seasonally adjusted employment by race (white, black, and Hispanic) from 2001 to 2016 inclusive. Panel (b) shows the evolution of the across and within components of the observed difference in the percent change in employment between black and white workers from 2001 to 2016 inclusive. We seasonally adjust employment trends using seasonal adjustment software by the US Census Bureau, entitled “X-13-ARIMA-SEATS.” Employment trends for the full population of workers (irrespective of race) are presented in Online Appendix Figure B.4.

Panel (a) of Figure 2.1 presents aggregate employment trends using the QWI data for white, black, and Hispanic workers from 2001Q1 to 2016Q4. These seasonally-adjusted employment trends show that while white workers experienced a sizable decrease in employment, their recovery following the Great Recession was faster than the recovery experienced by black workers. These trends also show that Hispanic workers fared the best of the three groups, in terms of having both a comparatively shallow decline during the recession and a very robust recovery.

While all worker types and most industries experienced declines during the Great Recession,[15] these losses were not even. Some industries were much more heavily affected than others, and variation in worker composition by industry, implies that the “across” inequality channel is likely to be important in many cases.[16] Moreover, in some industries, the Great Recession differentially affected worker types, suggesting that the “within” channel may frequently be salient as well.[17]

We now use the decomposition framework set out in Section 2.1 to understand how across- and

[15] Several industries actually experienced employment *increases* during the Great Recession (e.g., health, management services).

[16] As noted, worker types concentrated in industries particularly affected by the Great Recession will fare worse than those that are not.

[17] See figures in Online Appendix B.6.

within-industry employment inequality evolved between 2001 and 2016. In particular, we want to know whether the Great Recession had a pronounced effect on the across and within components, by assessing the extent to which there was a trend break during that time. Doing so isolates the effect of the demand-side recessionary shock from other long-run changes (e.g., prior trends and lagged effects of earlier shocks), which we assume continue to operate during the downturn. This provides suggestive evidence for the formal analysis that follows in Section 2.3.

Panel (b) of Figure 2.1 plots the year-over-year across and within components for the comparison between white and black workers. Changes in the across component generally contribute relatively little to the larger overall employment decline for black workers during the recession. Rather, this larger relative decline is mainly driven by the within component, which spiked from negative to positive during the recession. Black workers experienced much greater employment losses than white workers within industries during the downturn and had much greater employment growth at the industry level during the recovery. Interestingly, the across component does dip down slightly during the recession – the differential distribution of white and black workers across industries and the differential employment declines experienced by different industries on net partially counteracted the within-industry forces. For periods outside of the recession, the within component is mostly negative, consistent with faster employment growth experienced by black workers during the non-recession years. The patterns are similar for the white-Hispanic comparison (available upon request).

2.3 Main Results

In this section, we present our headline decomposition results. The previous section showed that the Great Recession generated substantial changes in employment inequality, with the within-industry component harming black workers and the across-industry component helping them. We now report formal estimates of these effects and then assess the degree to which they can be explained by establishment deaths. We carry out this analysis by implementing the various decompositions set out in Section 2.1.

Table 2.2: Employment Change Decompositions by Race

Panel A: Overall Employment Changes		
	Total	Deaths
θ^w	-0.058***	-0.084***
	[-0.061,-0.055]	[-0.090,-0.075]
	(0.000)	(0.000)
θ^b	-0.072***	-0.165***
	[-0.077,-0.067]	[-0.202,-0.133]
	(0.000)	(0.000)
θ^h	-0.039***	-0.081***
	[-0.047,-0.031]	[-0.120,-0.052]
	(0.000)	(0.001)

Panel B: Across-Within Comparisons (White vs. Black)		
	Total	Deaths
$[\theta^w - \theta^b]_A$	-0.011***	-0.011
	[-0.013,-0.009]	[-0.028,-0.003]
	(0.000)	(0.179)
$[\theta^w - \theta^b]_W$	0.025***	0.091***
	[0.021,0.028]	[0.062,0.139]
	(0.000)	(0.000)

Panel C: Across-Within Comparisons (White vs. Hispanic)		
	Total	Deaths
$[\theta^w - \theta^h]_A$	0.013***	-0.014***
	[0.010,0.016]	[-0.021,-0.006]
	(0.000)	(0.001)
$[\theta^w - \theta^h]_W$	-0.032***	0.010
	[-0.039,-0.027]	[-0.023,0.055]
	(0.000)	(0.579)

Notes: Panel A presents employment changes during the Great Recession (2007-2009) by worker race, both in total (agnostic to the establishment-level cause) and for those arising from establishment deaths. White, black and Hispanic workers are denoted by w, b and h, respectively. Panel B and C presents estimates of equations (2.4) and (2.5) for the white-black and white-Hispanic comparison, respectively, both in total and for establishment deaths. 95% confidence intervals and significance are calculated using 5,000 bootstrap iterations. Confidence intervals are reported in square brackets and p-values are reported in parentheses. *** denotes significance at the 1% level.

We first consider the total employment changes ("Total") by worker race, as well as the corresponding changes predicted from establishment deaths ("Deaths"). Panel A of Table 2.2 presents these results, which show how different workers fared over the Great Recession without adjusting for the distribution of employment across industries. Black workers lost modestly more total employment during the recession than white workers (7.2% versus 5.8%). Hispanic workers fared somewhat better, with total employment losses (3.9%) that were about two-thirds those of white workers. However, the deaths-based component paints a different picture: white and Hispanic workers lost about the same employment from establishment deaths (8.4% and 8.1%, respectively), but black employment declined by nearly twice as much in percentage terms (16.5%). The larger employment decline from establishment deaths for black workers suggests that they were particularly concentrated in less resilient establishments at the onset of the recession.

Having established the overall effects of the Great Recession, we now turn to the role that industries played (and perhaps more importantly, the role they did not play) in explaining these patterns. In particular, workers are not distributed evenly across industries, so that some of the differential employment losses reported in Panel A of Table 2.2 may be due to industry-level heterogeneity in the severity of the recession rather than differences in how different worker types are matched to firms within industry. The cleanest test of the hypothesis that disadvantaged workers tend to be concentrated in less resilient firms therefore is to compare the within-industry components of the total and predicted-from-deaths employment changes.

The bottom two panels of Table 2.2 carry out this comparison by reporting the across- and within-industry components for the total employment changes, as well as those predicted from establishment deaths, for white versus black workers (Panel B) and white versus Hispanic workers (Panel C). All of the within- and across-industry components for total employment are statistically and economically significant, and some of the deaths counterparts are significant as well.

Focusing on the total employment effects in the first column, it should be clear that the overall patterns presented in Panel A of Table 2.2 are potentially misleading as to the effects of the Great Recession on employment inequality. To see why, consider the white vs. black comparison in Panel B. The slightly larger total employment decline for black workers stems entirely from a much greater decline within-industry, with the distribution of black workers across industries actually serving to protect them somewhat from the downturn, relative to white workers. Accounting for that variation, the within-total effect becomes substantially more positive (2.5 versus 1.4 percentage points overall). The opposite story pertains to the white-Hispanic comparison in Panel C: white workers fared better than Hispanic workers across industries, but substantially worse within industries, with total white employment falling by 3.2 percentage points more than the decline for Hispanic workers.

For the white versus black comparisons, the across and within total employment effects agree with their deaths-based counterparts in sign if not in magnitude. Across industry, the total and deaths-based estimates are identical at -1.1 percentage points, indicating a slight relative advantage for black workers, although the deaths estimate is not statistically significant. Within-industry, both estimates suggest that black workers lost significantly more employment than white workers. However, the total effect, at 2.5 percentage points, is substantially smaller than the deaths-based effect, at 9.1 percentage points.

In terms of total and death-based effects agreeing in sign, the white-Hispanic comparison contrasts with the white-black results. Establishment closures reduced employment more for white workers across industries, with a 1.4 percentage point decline in white employment relative to Hispanic employment. However, within industries, Hispanic workers fared slightly worse, with a one percentage point decline in employment relative to white workers (though we cannot statistically reject this effect being zero).

The central takeaway from Table 2.2 is that black and Hispanic workers were disproportionately concentrated in less resilient establishments within industries at the onset of the recession, though only the estimate for the white-black comparison is statistically distinguishable from zero. Whether due to skill differences or discrimination (the determination of which is beyond the scope of this article), the differential within-industry concentrations of workers implied by these results has important consequences for inequality during future downturns.

While our primary focus is on race, the decomposition of employment changes by gender is also worth discussing. As Online Appendix Table B.1 shows, while both men and women experienced substantial employment losses between 2007 and 2009, the total decline for women was roughly half the decline for men (3.5% versus 7.1%). For both men and women, the employment losses predicted by deaths are greater than the observed losses and are highly significant. Decomposing these differences into across and within components reveals that women were advantaged relative to men by their distribution across industries, while losing more within industry. This pattern persists for the components predicted from establishment deaths – within-industry, the point estimate suggests that women were concentrated in relatively less robust establishments, leading to a relative employment loss versus men of 2.3 percentage points. While we are unable to statistically reject that this within-industry deaths effect is zero based on the p-value, we are able to do so using the confidence interval (see discussion in Section 2.1.3).

2.4 Justifying Our Approach

Recall from Section 2.1.1 that our approach relies on establishment category (deaths, births, contractions, and expansions) counts correctly predicting worker-type specific employment changes associated with that category. If they do, then the coefficients in estimating equation (2.1) should be unbiased. Assessing the extent to which this is the case requires setting out the econometrics of our approach in greater detail. As our primary interest centers on the death-based estimates in Table 2.2, we focus on that component of the decomposition.

Suppose the true data generating process for the death-based type-specific employment change is $\Delta E^{\tau}_{d,i,j} = -d_{ij} \cdot \bar{E}^{\tau}_{d,i,j}$, where $\bar{E}^{\tau}_{d,i,j}$ depends on the average size of establishments destined to close, and the proportion of type-τ workers in such establishments, at the $i - j$ level. Let $\Delta E^{\tau}_{d,i,j} = \widehat{\Delta E}^{\tau}_{i,j}|_d + \widetilde{\Delta E}^{\tau}_{d,i,j}$, where $\widehat{\Delta E}^{\tau}_{i,j}|_d = \hat{\beta}^{\tau}_{d,i} d_{ij}$, with analogous expressions and decompositions for the other establishment categories. Thus, we can be explicit about the error term in equation (2.1): $\epsilon^{\tau}_{ij} = \widetilde{\Delta E}^{\tau}_{d,i,j} + \widetilde{\Delta E}^{\tau}_{b,i,j} + \widetilde{\Delta E}^{\tau}_{c,i,j} + \widetilde{\Delta E}^{\tau}_{ex,i,j}$.

Defining $\eta^{\tau}_{\setminus d,i,j} \equiv \epsilon_{ij} - \widetilde{\Delta E}^{\tau}_{d,i,j}$, $\hat{\beta}^{\tau}_{d,i}$ (and thus $\widehat{\Delta E}^{\tau}_{i,j}|_{d}$) could be biased if d_{ij} is correlated with the unobserved determinants of births, contractions or expansions; that is, $cov(d_{ij}, \eta^{\tau}_{\setminus d,i,j}) \neq 0$. Based on Table 2.2, the problematic case for our analysis would be downward bias for black workers ($cov(d_{ij}, \eta^{b}_{\setminus d,i,j}) < 0$) and upward bias for white workers ($cov(d_{ij}, \eta^{w}_{\setminus d,i,j}) > 0$). This could only occur if the non-death establishments in jurisdictions with a higher number of establishment deaths systematically hire (fire) black workers at a lower (higher) rate than their white counterparts, conditional on the number of establishment births, contractions, and expansions. We view this as unlikely, particularly given the strong out-of-sample performance of our method, discussed below. Moreover, if firm deaths really do disproportionately affect black employment, as implied by our estimates, the larger supply of newly-unemployed black workers should push further against relatively low black hiring in deaths-intensive jurisdictions.

However, even if the direction of bias is problematic for our primary conclusion – that within-industry black workers lost greater employment than white workers from establishment deaths over the Great Recession – the magnitude of the bias would have to be fairly large to explain the effects we find. In particular, relative to the estimated coefficient, the true coefficient for white workers and black workers would need to be 60 percent larger and smaller, respectively, for the true within-deaths effect associated with Panel B of Table 2.2 to be zero (see Online Appendix Figure B.10).

Beyond this sensitivity analysis, a feasible way to assess the accuracy of our procedure for statistically allocating employment changes to establishment categories is to carry out a goodness-of-fit exercise. This entails determining how well predicted out-of-sample total employment changes by worker type, changes in total employment inequality between worker types, and across/within decompositions of these inequality changes accord with the observed analogues. We do so by randomly splitting the full sample of counties into two equally-sized groups, with the randomization stratified by county employment. We then estimate equation (2.1) on the first group of sampled counties to recover estimated coefficients for each type ($\hat{\beta}^{\tau}_{d,i}, \hat{\beta}^{\tau}_{b,i}, \hat{\beta}^{\tau}_{c,i}, \hat{\beta}^{\tau}_{ex,i}$). Finally, we predict the employment changes for the other group of counties (the hold-out group) using these estimated parameters. We repeat this procedure 500 times.

Table 2.3 reports the results of this exercise for race, comparing white workers to black workers. On average, the predicted values of θ^{τ}, $\theta^{\tau'}$, $\theta^{\tau}-\theta^{\tau'}$, $[\theta^{\tau}-\theta^{\tau'}]_A$, and $[\theta^{\tau}-\theta^{\tau'}]_W$ for the hold-out samples are very close to their observed values. Moreover, for all measures, the variance of the difference between the predicted and observed value is uniformly small, suggesting that our approach rarely makes large errors. Online Appendix Tables B.2 and B.3 repeat the analysis for the white-Hispanic and male-female comparisons, respectively. In each case, the predicted values are very close to the observed values, with relatively modest variances. The consistently accurate out-of-sample predictions of our method lends credence to our main estimates.

Table 2.3: Out-Of-Sample Fit by Race (White vs. Black)

	(1) Mean Observed θ	(2) Mean Predicted θ	(3) Mean Difference	(4) SD of Difference
θ^w	-0.058	-0.058	-0.000	0.003
θ^b	-0.072	-0.074	0.002	0.008
$\theta^w - \theta^b$	0.014	0.016	-0.002	0.007
$[\theta^w - \theta^b]_A$	-0.011	-0.012	0.001	0.002
$[\theta^w - \theta^b]_W$	0.025	0.028	-0.003	0.008

Notes: This table presents statistics related to the out-of-sample prediction outlined in the main text. All statistics are based on 500 randomly-drawn estimation and holdout samples, stratified on county-level employment. Columns (1) through (3) present averages of estimates obtained for each random draw. Column (4) presents the standard deviation of Column (3).

2.5 Conclusion

In this paper, we considered the extent to which the evolution of worker inequality during an economic downturn is dictated by preexisting worker-firm matching. We did so by developing an approach that makes use of publicly available data, exploiting variation in employment by worker type and establishment deaths across jurisdictions. Focusing on employment changes from establishment deaths (as opposed to changes from non-closing establishments) is key for understanding the connection between worker inequality and pre-downturn matching, given that, by definition, death-based changes cannot be affected by post-downturn employer favoritism toward one worker group over another. Applying our method to analyze employment losses during the Great Recession, we found that, within industry, black workers were disproportionately matched with less resilient establishments prior to the downturn, which resulted in larger subsequent employment losses than experienced by their white or Hispanic counterparts. Our paper lays the foundation toward a deeper understanding of the role that economic downturns play in generating inequality.

Chapter 3

Evaluating Regulatory Impact Assessments in Education Policy

(with Clive Belfield[†] and Brooks Bowden[‡])

3.1 Introduction

Today's era is considered "the golden age of evidence-based policy" due to record growth and utilization of rigorous research to address social problems (Haskins, 2016). This movement is intended to develop higher standards and to efficiently allocate resources toward building and replicating programs and practices that work. A critical component inherent in this is examining efficiency through economic evaluations.

Over prior decades, a series of executive orders have been issued to establish a regulatory planning and review process to make federal policy more efficient, that is, in favor of policies where the benefits justify the costs (since Executive Order No. 12,291, 1981). Generally, government utilizes regulation to intervene when market failure exists to ensure that the economy continues to progress

[†]Department of Economics, Queens College, City University of New York
[‡]Education Policy Division, Graduate School of Education, University of Pennsylvania

according to the values established by society (Nas, 2016). This Executive Order built upon that framework by requiring economic evaluation of the costs and benefits of any significant federal regulation having an effect on the economy of US$100 million or more, an adverse material effect, or an inconsistency with another agency or action (Executive Order No. 12,866, 1993). President Obama reaffirmed this order in 2011 to ensure that the regulatory system protects "public health, welfare, safety, and our environment while promoting economic growth, innovation, competitiveness, and job creation" (Executive Order No. 13,563, 2011). The most recent Executive Order on Regulatory Reform (January 26, 2017) proposes a "two-for-one" regulatory exchange process: Each new regulation can only be adopted if two existing regulations are removed (Peacock, 2016).

Generally, the federal role in education is to guide discussion of the policy agenda, to provide supplemental support for educational programming at the K-12 level, and to provide policies and support for postsecondary education institutions and students. In order to continue to build our use of evidence in policy-making within education and to ensure that our policies are efficient, it is important that we examine the economic evaluations conducted of federal education regulations to continue to improve the quality and relevance of this evidence base.

The primary method for economic evaluation is benefit-cost analysis (BCA; see Boardman et al., 2017; Institute of Medicine, The National Research Council, 2014; Karoly, 2012; Levin et al., 2017). As described in detail by Vining and Weimer (2010), this method can be adapted to evaluate social policy interventions, including those in education. Straightforwardly, policy makers and education professionals should decide in favor of policies that have greater benefits than costs. As a tool that guides policy, the quality of BCA becomes crucial: Costs and benefits should be estimated with precision to the extent possible; if elements that are not easily quantifiable are included in the analysis, transparent statements of their assumptions should be made available; and benefits should be mapped into costs in a clear and explicit way. Hence, good quality analysis should yield an economic metric that can guide investment of public (and private) funds. Of course, BCA is only one tool that is available to the policy maker to aid in decision-making; ultimately, the decision maker must make a reasoned determination, recognizing society's preferences and goals as well as existing practices.

Fundamentally, BCA is justified insofar as it improves the quality of decision-making (Posner, 2000; Revesz and Livermore, 2008). If the results of BCAs help to improve policy, these analyses should be performed. In order to improve decision-making, however, BCAs need to be performed to a high methodological standard (see Farrow and Zerbe, 2013).

In this article, we critically evaluate prior BCA of education regulations at the federal level. We begin by documenting the many practical challenges in performing BCA and reviewing the quality of BCAs in other policy fields. We then evaluate the quality of education policy BCAs as performed by the federal government. First, we collate findings from a checklist appraisal of each BCA. Next, we review a subset of (relatively) high-quality BCAs. This appraisal and review illustrates the many ways in which federal attempts at benefit-cost evaluation differ from social BCA. We conclude with discussion of the implications for policy when BCAs fall short of accepted methodological standards.

3.2 BCA From Policy to Theory to Practice

BCA plays an important role in determining the efficiency of federal policies. Executive branch agencies promulgate regulations to implement laws enacted by the Congress. Under Executive Order No. 12,866 (1993), each agency must prepare a unified regulatory agenda containing the regulatory plans of the most important significant regulatory actions that it expects to issue in that fiscal year. These plans are then forwarded it to the Office of Information and Regulatory Affairs (OIRA).

Together with the relevant agency, OIRA plays an integral role in reviewing these regulatory plans and ensuring they are not in conflict with other policies.[1] Once OIRA notifies the agency

[1] For example, plans by one agency are usually revised by other agencies to identify possible conflicts and comments; these conflicts must be notified to the administrator of Office of Information and Regulatory Affairs (OIRA) who in turn must notify the affected agencies and other relevant parties. Additionally, any other planned regulatory action that the administrator of OIRA believes to be inconsistent with the President's priorities or may be in conflict with another policy must also be notified. The administrator of OIRA must also chair a regulatory working group to serve as a forum to assist agencies in identifying and analyzing important regulatory issues. Finally, the administrator

that the review is completed and that there are no further considerations, the regulatory action then becomes a final rule and is published in the Federal Registry. Major final rules must go through Congress to be approved. One important component of this review is the application of BCA; this form of analysis helps policy makers come to a ”reasoned determination” as to whether to implement policy.

The basic theory of BCA is well-established. There are several excellent textbook treatments of BCA, its principles, and theoretical foundations (Adler et al., 2006; Boardman et al., 2017; Farrow and Zerbe, 2013). Recently, however, it is the practical application of BCA that has received more scrutiny (for an overview, see Belfield (2015); for how to read BCAs, see Dudley et al. (2017)). This scrutiny has highlighted two main concerns - the lack of BCAs and the quality of the BCAs that are performed.

A number of studies have drawn attention to the limited application of BCA. In a recent review, Ellig (2016) concludes that ”regulatory agencies often adopt regulations without knowing whether a given regulation will really solve a significant problem, whether a more effective alternative solution exists, or whether a more targeted solution could achieve the same result at lower cost.” So, there are not enough BCAs to determine whether the most efficient policy has been implemented and, even when attempted, BCAs are often incomplete. Checklist studies have counted the ways in which BCA is not fully performed. Hahn and Dudley (2007) reviewed 78 regulations and found that only 65% considered costs, 22% considered benefits, and 12% considered benefits minus costs. This work is summarized in (Table 1). This lack of BCA is not just a concern across federal departments. Even as other agencies conform to different standards, their economic evaluations have also been found to be incomplete. For example, in a review of practices at the state level, Schwartz (2010) found BCAs to be very infrequently performed by state legislatures. Also, there has been extensive inquiry into the inadequacy and inconsistency of regulatory impact assessments (RIAs) across the European Union and the United Kingdom (Dunlop and Radaelli, 2016; Fritsch et al., 2017).

of OIRA must meet quarterly with representatives of state, local, and tribal governments to identify both existing and proposed regulations that may uniquely or significantly affect those governmental entities.

The second concern is that the BCAs that are performed - even the incomplete ones - are not of high quality. On the one hand, there are many challenges in estimating costs. These include gaining access to proprietary data, identifying resources across agencies, miss-specifying business-as-usual, pre-implementation or noncompliance behaviors, technological change, and program infidelity (see Harrington et al., 2000).

On the other hand, there are also many challenges in estimating benefits. Uncertainty in the estimation of benefits can come from several sources. First, it is unclear how, and the extent to which, institutions comply with regulations. The compliance strategy institutions adopt dictate both the cost structure and the benefits generated by the regulatory action. Second, once institutions have adopted the new regulation, its causal impact on the target population (such as students and institutions) is usually hard to identify. Some policies are not easily amenable to BCA (e.g., if they mostly involve transfers or have strong equity implications that cannot be precisely modeled or necessitate normative approaches). Third, even if the causal estimates are identified, monetization of the benefits (and costs) is subject to the derivation of shadow prices for nonmarket goods (Ellig and McLaughlin, 2012), which are sensitive to the assumptions used to recover them. Sunstein (2014b) proposes a number of ways to respond; nevertheless, these challenges must usually be overcome in each BCA.

Moreover, the lack of BCAs makes it harder to perform high-quality BCAs. With few studies, there are even fewer methodological inquiries (e.g., for sensitivity testing or shadow pricing). Also, although there are guidance manuals (e.g., Office of Management and Budget Circular A-4), there are little precedence to help evaluators harmonize with, or compare against, their analysis.

Here, we examine the extent to which these concerns - particularly regarding the quality of BCA - are valid for economic evaluation of education policies and programs. There are two main reasons why it is important to evaluate the practice of BCA in education policy at the federal level.

One reason is that the current quantity of BCA in education is insufficient. Although there are a few high-quality BCAs within education, these focus almost exclusively on the returns to one reform - preschool (García et al., 2017; Heckman et al., 2010; Nores et al., 2005). Moreover, these BCAs are of small-scale programs that do not enroll many children and have budgets of less than US$10 million; and, typically, they are retrospective in that the preschool programs have already been implemented. By contrast, federal regulations affect multiple cohorts of students across the United States with economic effects of billions of dollars; typically, these regulations require a benefit-cost appraisal that is prospective and so might influence whether and how the policy is implemented. Thus, even within a policy realm where there are too few BCAs generally, education policy receives relatively little attention.

A second reason is that there are specific challenges in undertaking BCAs of educational policies. One challenge is that many educational policies involve large-scale transfers. Posner (2003) has discussed this aspect at length: Transfers are not amenable to BCA (they almost certainly fail a benefit-cost test because transfers are not benefits); and evaluators have been reluctant to apply cost-effectiveness analysis as an alternative (despite its inclusion in Executive Order No. 12,866, 1993). A second challenge arises because education policies have features that make them, according to Sunstein (2014b), ”Hard Cases”: They have significant distributional and equity consequences and/ or their benefits are heavily loaded on specific individuals (students). Also, as Vining and Weimer (2010) note, social BCAs require significant sensitivity testing; this leaves open the possibility that an efficiency ruling cannot be clearly determined. Finally, the field of educational research does not place a strong emphasis on BCA. Indeed, in its most recent guidance document on evidence to strengthen education investments, the US Department of Education (2016) does not consider the application of cost analysis or BCA. Thus, even as BCAs are challenging across many policy domains (e.g., environmental and health policy), there are a number of specific challenges for BCAs in education policy.

3.3 Method for Evaluating BCAs

Our evaluation focuses on federal education policies for which a BCA is expected.[2] Our evidence is taken from OIRA reports on all 28 education regulations over the most recent decade from 2006 to 2015 that were deemed economically significant and so for which some form of RIA was performed. A regulatory action is defined to be economically significant if the regulation is expected to have an annual effect on the economy of US$100+ million; may create a serious inconsistency or interfere with an action taken or planned by another agency; alters the budgetary impact of entitlements, grants, user fees, or loan programs or the rights and obligations of recipients; or raises novel legal or policy issues. A summary list of the 28 RIAs is given in Table C.3 of Appendix C.

Under Executive Order No. 12,866, 1993, the issuing agency must provide an assessment of the potential costs and benefits of all economically significant regulatory actions. The assessment of these potential costs and benefits must be, to the extent feasible, quantified. In addition, these costs and benefits should be compared to potentially effective and reasonably feasible alternatives of the planned regulation. Thus, each RIA should include some attempt at BCA.

We use two approaches to evaluating these education RIAs. The first is a checklist method of all 28 RIAs. Evaluators have criticized checklists and scorecards on several grounds: the use of unclear terms for scoring; ad hoc weighting of scores across items; and the vagueness of using the same scorecard for accountability, communications, and process improvement purposes (Fritsch and Kamkhaji, 2016; Radaelli and Fritsch, 2012). However, this method has been used successfully by Hahn and Dudley (2007) and Shapiro and Morrall (2016). Also, our checklist is adapted to the specifics of educational interventions and looks at the basic requirements for a satisfactory BCA. Thus, it provides only a rudimentary understanding of why RIAs take the form they do and whether that is actually the most appropriate form of evaluation. Therefore, our second approach to evaluating these RIAs is by direct review. We examine in detail the seven most complete BCAs to assess their validity as an economic evaluation of the proposed policy.

[2] For many regulations and policies, benefit-cost analyses are not required. The reviewed studies are on regulations and policies where there is a requirement.

3.3.1 Checklist Approach

There is great variation in the way assessment of benefits and costs is done across regulatory actions. Therefore, we try to capture in our data the depth of analysis each planned regulation uses in their BCA. The information is classified and gathered as follows.

Costs. The cost analysis performed in each regulatory report is coded into four levels. These levels represent progressively more extensive cost analyses. First, cost analyses are classified as ”stated” if there is any reference in the report to resource use other than budget allocations and if there is some description of how the resources required correspond to the policy being implemented. Second, cost analyses are classified as ”ingredients” if there is a list of resource ingredients needed or description of them with detail. Third, cost analyses are coded as ”agency”if there is any discussion of costs incurred by other agencies. Finally, cost analyses are coded as ”dollar value calculated” if there is a dollar value assigned to cost. This last code is the most extensive type of cost analysis (although we note that regulations can and do calculate costs without considering ingredients or costs to other agencies).

Benefits. Similarly, benefit analyses are also classified into four levels that reflect the depth of the analysis. Analogous to the costs classification, benefits are classified as stated if there is any reference in the report to monetized benefits of the proposed policy. Next, benefit analyses are coded as ”described” if there is a description of how the benefits are estimated. Third, benefit analyses are coded agency if there is any discussion of benefits from multiple perspectives. Finally, benefit analyses are classified as ”calculated” if there is a dollar value assigned to the benefits.

BCA. In order to capture how benefits are weighed against costs, we attempted to assign two different measures. The first coding is ”net present value” (NPV) and refers to analyses where an NPV (benefit minus cost) dollar value is reported. The second coding is ”B/C” and is for analyses which plausibly assert that the benefits exceeded the costs (but do not provide a numerical estimate).

Additional information. In addition to the information collected above, one other coding was derived from the OIRA documentation. This coding captured the methodological transparency of the analysis. Studies were coded according to whether the report clarifies assumptions used in terms of interest rates, inflation, discount rates, and time frame. Information about these assumptions is necessary for researchers to adjudicate on the quality of each BCA and to compare results from these BCAs with other analyses.

For each of the 28 RIAs, the information was coded separately by two persons. There were very few discrepancies; these were reconciled in discussions.

3.3.2 Text Review

Given the results from our checklist, we reviewed the RIAs to determine what efforts were made to perform an economic evaluation congruent with Executive Order No. 12,866, 1993. For the text review, we selected the seven highest quality RIAs for detailed interpretation and assessment. (These RIAs are shaded in Table C.3 of Appendix C). The review allowed us to assess in detail what type of analysis was performed, how the analysis was structured, and the most salient modeling assumptions and shadow price valuations. These seven specimens are used to assess the quality and features of BCAs-conditional on BCA actually having been performed.

3.4 Evaluating Educational BCAs: Summary Findings

Summary findings on the quality of educational BCAs at the federal level are derived from the checklist analysis. These checklist results are given in Table 3.1. Most of the RIAs perform some form of cost analysis. Almost all (93%) make a statement about social costs, that is, costs beyond simple budgetary allocations. One quarter (25%) make an attempt to estimate costs by distinguishing the quantity of inputs and their prices. Almost one half (54%) investigate cost implications for different government agencies or levels. In addition, approximately three quarters (71%) of the RIAs is a dollar value of costs reported.

The RIA benefit analyses are significantly weaker. Only four fifths (82%) state the benefits that

Table 3.1: Checklist Evaluation of 28 Federal RIAs

Checklist Criteria	Percentage Performing Activity
Cost	
Statement of costs beyond budgetary statement (stated)	93
Estimated costs using ingredients method (ingredients)	25
Estimated costs separately by agency (agency)	54
Dollar value of costs reported (calculated)	71
Benefits	
Statement of monetized benefits (stated)	82
Description of benefits estimation (described)	29
Estimated benefits separately by agency (agency)	57
Dollar value of benefits reported (calculated)	4
Net Present Value (NPV; benefits minus costs)	
Dollar value calculated (NPV)	0
Stated as positive (B>C)	100
Methodology	
Assumptions described (year, inflation, discount rate)	36

Notes: RIAs from 2006 to 2016 are listed in Appendix Table C.3. RIA=regulatory impact assessment.

are expected from the legislation and only one third (29%) describe these benefits in some detail. Half (57%) of the RIAs do look at different benefits across government. Notably, only one study (4%) calculated a dollar value of benefits.

Together, these frequencies mean that no RIA calculated a dollar value for the NPV or was able to assert how benefits might exceed costs. Also, we find that only one third (36%) of RIAs adequately documented key methodological assumptions.

For the seven highest quality BCAs, the checklist results are given in Table 3.2. For these studies, almost all performed a full cost analysis: They estimated resource use using a version of the ingredients method, separated out costs by agency, and reported a dollar value for costs. With respect to benefits, most of the studies stated the benefits, described them, and apportioned them across agency; however, none calculated the dollar value of benefits. Therefore, even across these high-quality studies, no NPV estimate was reported. Also, less than half of these studies clearly reported the assumptions used.

Table 3.2: Checklist Evaluation of Seven High-Quality Federal Regulatory Impact Assessment

Checklist Criteria	Percentage Performing Activity
Cost	
Statement of costs beyond budgetary statement (stated)	100
Estimated costs using ingredients method (ingredients)	86
Estimated costs separately by agency (agency)	100
Dollar value of costs reported (calculated)	86
Benefits	
Statement of monetized benefits (stated)	86
Description of benefits estimation (described)	57
Estimated benefits separately by agency (agency)	86
Dollar value of benefits reported (calculated)	0
Net Present Value (NPV; benefits minus costs)	
Dollar value calculated (NPV)	0
Stated as positive (B>C)	100
Methodology	
Assumptions described (year, inflation, discount rate)	43
Other	
Average budget allocation ($billion)	US$3.532
Median year	2010

As an additional exercise, we reported the checklist results for groups of regulations by year and budget allocation. The results are reported in Tables C.1 and C.2 of Appendix C. These results show that the quality of BCAs is not improving over time, although there is some indication that higher cost regulations receive a more intensive application of BCA.

Overall, these findings show that most educational BCAs performed by the federal government are incomplete. These findings are similar to reviews of economic evaluations in other sectors (e.g., from Hahn and Tetlock, 2008; Ellig, 2016; Shapiro and Morrall, 2016). However, we emphasize two points. First, although we contend there are significant informational and policy gains from performing complete BCAs, complete analyses may not be essential for justifying policy decisions. As noted above, policy decisions may be obvious after a basic review of costs and benefits; or policy decisions may be justifiable based on a ”reasoned determination,” taking account of the fact that ”some costs and benefits are hard to quantify” (Executive Order No. 12,866, 1993). Second, in no way should we conclude that these policies have a negative NPV, rather the magnitude of the NPV - and possibly its sign - is uncertain.

3.5 Assessing the Quality of Educational BCAs

We base our assessment of the quality of educational RIAs on a thorough review of seven of the highest quality BCAs. These RIAs included sufficient information for us to assess the quality of BCA when a reasonable attempt at analysis has been made. We find that the practice of BCA differs considerably from, and in some important ways does not meet, the methodologies and protocols as prescribed in textbooks (Boardman et al., 2017). These differences extend across estimation of both costs and benefits as well as across how a BCA is structured. However, the most important distinction is that these RIAs do not perform social BCA as it is commonly understood: Their actual analyses are far more constricted and narrow.

3.5.1 Estimating Costs

The first concern is that some costs are not accounted for in federal BCAs. Notably, some RIAs only considered the resources required to perform new tasks and did not consider the resources required to be in a position to perform these new tasks. Crucially, in order to perform new tasks, workers must be trained and new managerial and organizational procedures need to be developed. The resources required for training, management, and organization were not taken into account. For example, regulations to limit the eligibility length for direct subsidized student loans have cost consequences for colleges. The RIA only includes the ”reporting and financial aid counseling activity” to conform to this rule; changes colleges must make to program offerings (as student enrollment patterns change) are not included.

Second, although most regulations did account for the time burden on professional staff, they used very low estimates of time costs. That is, they typically assigned a value in the range of US$25 - US$30 per hour. This is a low value for a professional occupation: Bureau of Labor Statistics (2015) data show estimates of education administrators time in postsecondary education at US$49 per hour. It is also an underestimate in that there was no information on whether employer costs of compensation or overheads were included. It is therefore likely to be an underestimate of the opportunity cost of professional staff, that is, their value in alternative roles. Also, studies rarely reported any resource use by the professional staff such as office space or computer use.[3]

Review of the RIAs generated additional concerns. An implicit assumption was that all inputs are variable and neither fixed costs nor economies of scale were salient. Hence, it would be possible to either reduce inputs and save the full expenditure (input times input price) or increase inputs without facing an upward-sloping supply of inputs (increasing input prices). In practice, it would seem plausible to assume that there are some fixed costs and that marginal costs would exceed average costs. A final concern is that, for all regulations, in few of the RIAs were the sourcing and methods for calculating costs clear. In particular, few RIAs specify whether the accounting, survey, or full ingredients method was used even as results will likely differ depending on which method was used (Levin et al., 2017, Chapter 4).

3.5.2 Estimating Benefits

As shown in the checklist analysis, few RIAs reported benefits in sufficient detail. Even in the highest quality BCAs, benefits usually were only stated and were not explored further.

As well as precluding calculation of NPV, this omission is especially important for educational BCAs. For educational BCAs, distributional issues are salient; in some cases, they are central or integral to the policy. (We recognize that distributional issues are important in many BCAs: On these issues for environmental BCAs, see Robinson et al., 2016). Regulatory actions often require resource commitments from - and convey benefits to - various groups such as students, state ed-

[3]An important exception is RIN 1840-AD01 (High School Equivalency Program and College Assistance Migrant Program, The Federal TRIO Programs, and Gaining Early Awareness and Readiness for Undergraduate Program, 2010), where the issuing agency specifically accounts for an overhead at 50% of the salary and computer time and printing costs.

ucation agencies, local education agencies, or the federal government. It is important to establish whether each group receives benefits that outweigh the costs and that each group is treated equitably. These distributional considerations were rarely reported. For example, Regulation Identifier Number (RIN) 1840-AD01, a regulation of the College Assistance Migrant Program, states that

> there is no need to discuss the changes to the regulations . . . because the changes to regulations for these programs were minor. The most significant changes . . . address who can be considered an immediate family member of a migrant individual in order to be eligible for program services.(High School Equivalency Program and College Assistance Migrant Program, The Federal TRIO Programs, and Gaining Early Awareness and Readiness for Undergraduate Program, 2010)

The fact that agencies do not address equity issues is not a feature of the benefit analysis only. Differential burden of costs across populations is not considered either in any of the regulatory actions reviewed.[4]

3.5.3 Analyzing Costs and Benefits Together

We identify three areas where the joint analysis of costs and benefits should be improved so as to conform more closely to recommended standards and thresholds (Zerbe et al., 2013).

Few BCAs considered alternative policy options. Under Executive Order No. 12,866, 1993, costs and benefits of the proposed regulation are to be compared to costs and benefits of alternative, feasible regulations. Only a few of the RIAs included any consideration of alternatives and those RIAs typically referred to an alternative of "no implementation" (rather than the next best alternative).

Also, no sensitivity analyses of the estimated costs, benefits, or NPVs are reported.[5]. Sensitivity analysis is especially important for prospective BCAs where decision makers need to know

[4]For example, RIN 1840-AD13 (William D. Ford Federal Direct Loan Program, 2013), a regulatory action that proposes changes to the William D. Ford Federal Direct Loan Program, considers costs to student that become ineligible for the program's loans due to the proposed changes. However, there are no equity issues addressed as to which students might carry this burden.

[5]For example, RIN 1840-AC94 (High School Equivalency Program and College Assistance Migrant Program, The Federal TRIO Programs, and Gaining Early Awareness and Readiness for Undergraduate Program, 2010; William D. Ford Federal Direct Loan Program, 2013) that proposes to amend the Federal Perkins Loan Program, Federal Family Education Loan Program, and William D. Ford Federal Direct Loan Program makes projections of costs with clear

the possible downside risks. This is particularly interesting for some regulatory actions where projections of costs over time are made.

Finally, the RIAs typically fail to apply the proportionality principle. That is, they fail to allocate the most attention to the most important factors in their BCA.[6] Application of the proportionality principle is especially important, given the limited resources federal staff has to perform BCAs.

3.5.4 Administrative Burden BCA

On review, we find that the educational RIAs performed by the federal government are a particular type of BCA. Based on how the RIAs interpret costs, these BCAs do not calculate costs of implementing a policy. Instead, they calculated costs of documenting compliance with regulations (in accordance with the Paperwork Reduction Act of 1995). Instead of evaluating a policy as a social investment and applying social BCA, the RIAs were actually evaluating how to implement a policy or a change in policy. We refer to these as ”administrative burden” BCAs. They differ in three distinct ways from social BCA.

First, these administrative burden BCAs use a very particular approach to calculate costs. As one example, we refer to RIN 1840-AD02 (Program Integrity Issues, 2010) that relates to institutional eligibility under the Higher Education Act of 1965. The RIA does not attempt to calculate the costs and benefits of the new number of eligible institutions. Instead, the RIA is directed toward estimating the administrative burden of changing eligibility status. The administrative burden increase for this regulation is estimated at US$126 million over 5 years. This burden is calculated as a product of the hours worked and the hourly wage rate, but it is not derived from a formal costing exercise using the ingredients method. Instead, the costs are estimated using what we refer to as a ”conveyor belt” approach. For each student, the regulation necessitates a change in status or

assumptions of program and job participation, working hours, time frame, and discounting used. However, none of these assumptions are varied to see how they might affect the final estimate.

[6]In RIN 1840-AD02 (Program Integrity Issues, 2010), for example, there are 15 new regulatory changes. Each one receives approximately equal attention in the costing exercise. However, the regulatory changes vary dramatically in their cost implications: One change requires only 628 hr of new administrative time; another change requires 2,080,800 hr of regulatory time. Clearly, the latter change should be subject to much greater analytical scrutiny than the former.

eligibility; this change is estimated to take x minutes of personnel time. Therefore, if the college has 100 students, then the cost is 100x multiplied by the wage rate of personnel (if 10,000 students, the cost is 10,000x times the wage rate).

This conveyor belt approach is incomplete and potentially misleading. First, it does not account for other costs such as overheads, management, or facilities. Instead, the conveyor belt approach focuses on marginal costs of performing an administrative task; it assumes administrative changes can be placed on top of existing reporting structures and do not require additional capital, managerial personnel, or computer systems. Second, it does not account for the scale of operations: It assumes that there are no fixed costs to changing administrative burdens and no resources are required for training personnel to meet new administrative requirements. Instead, the full ingredients approach requires identification of the inputs and their prices to implement a regulatory change at a college of a prescribed scale, that is, to estimate a cost function with fixed and variable costs.

Second, the administrative burden approach significantly restricts what can appropriately be considered as a benefit. With an administrative burden approach, the benefits are from compliance with a regulation and it is difficult to describe and calculate these benefits. It may be tempting to consider regulations as unnecessarily onerous and so having zero benefits: Any reduction in compliance costs should therefore have a positive NPV. Nevertheless, it is still important to specify the benefits of reducing administrative burdens. For example, the stated benefits of RIN 1840-AD02 (Program Integrity Issues, 2010) are that these administrative changes will lead to more accurate determination of status as a college student eligible for federal support and ”greater transparency for borrowers” (p. 66971). Neither of these benefits is quantified; no attempt is made even to bound the value of these benefits so that they might be compared to the costs. (There are also other vague benefits, including ”increased clarity about incentive compensation for employees at institutions of higher education.”) Administrative regulations typically require colleges to provide information; an economic evaluation would consider whether the administrative burden is justified in terms of the information obtained. The value of this information depends on how much the new information changes the expected value of a given policy; this value is often difficult to estimate (see the discussion in Boardman et al., 2017, Chapter 10).

The distinction between a social BCA and an administrative burden BCA is most clearly illustrated with RIN 1840-AD15 (Program Integrity: Gainful Employment, 2014), a regulation to change eligibility for Pell grants. This regulatory change would cost US$126 million to implement, but it would yield savings of US$4.3 billion in Pell grants not taken over the subsequent decade. Yet these ”savings” are only benefits if we adopt an administrative burden perspective (in some BCAs, they would be considered as transfers). From a social perspective, the reduction in Pell grants is not a benefit unless all those Pell grants produced zero human capital for the recipients (and even under that very unlikely scenario, policy makers might still be concerned about the regressive effect of reducing Pell grants).

The third and final observation about administrative burden BCAs relates to their correspondence to new regulatory directives. As noted in the Introduction section, the January 2017 Executive Order proposes a ”two-for-one” regulatory exchange process. (Before this Executive Order, the U.S. approach to reviewing existing regulations was a more formal and structured ”look-back” approach (Sunstein, 2014a).) This order accords directly with BCAs that are essentially about administrative burdens: A new regulation with an administrative burden of $X million can only be implemented if two existing regulations with administrative burdens of $2X million are rescinded. (Similar approaches have been adopted in other countries, e.g., Canada, the United Kingdom, and Australia.) Given the lack of consideration of benefits in the ”two-for-one” regulatory exchange process, the desirability of regulations is yet harder to define, when considering either social or administrative scopes. Leaving aside the desirability, efficiency, and practicality of this type of Executive Order, it prioritizes administrative burden BCA.

3.6 Summary and Conclusion

Methods and theories for performing BCA are clearly described in textbooks (Boardman et al., 2017; Zerbe et al., 2013). And, the need for BCA is clear: As summarized by Dudley et al., 2017, p.201, ”Regulatory impact analysis can be an invaluable method for transparently evaluating contentious policy choices before they are put in effect.” However, the practice of BCA is important.

Our findings for regulatory impact assessments of education policies at the federal level demonstrate that - in the infrequent instances in which it is undertaken - practice falls short of methodological standards in a number of ways. Cost estimates appear to be underestimated and lack transparency with respect to method and assumptions. Benefit estimates are very infrequent and analyses often lack sensitivity testing, proportionality, and a reasonable counterfactual. Overall, these findings on costs, benefits, and economic metrics echo those in other reviews of regulatory review standards (e.g. Hahn and Tetlock, 2008; Shapiro and Morrall, 2016; Shapiro and Morrall III, 2012). Most notably, these RIAs are not attempts at social BCA but instead are evaluations of administrative burdens; this focus unavoidably affects the structure of economic evaluation and impairs our ability to use BCA as a tool in policy-making.

We share the concern of Gordon (2016) that BCA should guide decision-making, rather than being an instrument of justification of an already determined policy. In addition, in order for BCA to provide said guidance, it must be of high quality if it is to be informative in the decision-making process. There are a number of remedies beyond simply exhorting analysts to perform more rigorous analyses. Certainly, more time and funding should be provided, so that analysts can undertake more rigorous study. A full application of the ingredients method and shadow pricing of benefits requires a similar level of research resources as an impact evaluation. More training for analysts may be desirable, as well as a greater emphasis on harmonizing studies for comparative purposes.

In addition, we suggest the following recommendations that could help strengthen the BCAs we have reviewed. These recommendations correspond to the general tips listed by Dudley et al. (2017).

3.6.1 Benefits

Several RIAs include a description of possible benefits to students, but none attempt to estimate the number of students receiving these benefits or the dollar value of these benefits. We encourage the government agencies to assign a number, along with a measure of its precision or a confidence interval. This estimate of the number of beneficiaries of a new regulation could be varied in the sensitivity analysis to verify the importance of the assumption on the overall results.

Performing a rigorous estimation of benefits may not be plausible for government agencies, given limited time and other resource constraints. Therefore, we suggest that government agencies refer to external rigorous benefit calculations, such as those provided by the Washington State Institute for Public Policy. The match between prior estimates and the RIA under consideration need not be exact as this can be varied in sensitivity analyses.

3.6.2 Costs

Most RIAs only refer to costs incurred by the funding agency or by other government agencies. This is what we refer to as administrative burden BCA. Only a few RIAs consider costs to students and education institutions. We encourage government agencies to consider multiple perspectives and all associated costs of regulations.

3.6.3 Benefit-Cost

A very important step missing for all RIAs analyzed is that none of them match the costs with the benefits. This may be in part due to the lack of quantified benefits within the RIAs reviewed. We strongly suggest that whenever benefits are quantifiable, government agencies should include in the BCAs a comparison of the benefits to the costs in a simple subtraction (B-C) or a ratio (B:C). Ultimately, the objective of quantifying and valuing costs and benefits is to provide guidance in the decision-making process. Weighing costs against benefits, whether it is a subtraction or a ratio, is the most important and useful result of a BCA to policy makers, and it should be estimated whenever possible.

At a more basic level, RIAs need to be more explicit in three respects. One is the alternative policy options that are under consideration: Explaining why specific options were chosen is necessary but often these choices are presumed. A second is that the theory of action is often unclear: Education policies are complex and may bring about change in many different ways. (These 2 items match with the primary tips Dudley et al. (2017) suggest to readers of RIAs.) Finally, the third basic clarification relates to which type of analysis - social BCA or administrative burden BCA - is being undertaken. The choice between them has fundamental consequences for economic evaluation. In particular, we believe social BCAs should be prioritized over administrative BCAs. Currently, the latter are being undertaken, perhaps giving the impression that they are equivalent to the former. They are not. If administrative BCAs are an important analysis needed for government agencies, these should be done as a part of the broader social BCA.

Notwithstanding these deficiencies, we believe these economic evaluations are helpful in substantiating education policy decisions. Even when RIAs fail to follow best practices or face methodological challenges, they may still be an improvement over the alternative of no information. More importantly, we emphasize that a lack of evidence does not imply that these regulations are inefficient. Hundreds of studies have established the economic value of education in terms of higher earnings, as well as in terms of private and social benefits, for those with more education (Autor, 2014; Barrow and Malamud, 2015; Belfield and Levin, 2007). Therefore, we expect these regulations are efficient in the sense that the benefits of the regulations exceed the costs. Nevertheless, it is still important to substantiate this expectation with high-quality BCAs.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Adler, M. D., Posner, E. A., Posner, E., et al. (2006). *New foundations of cost-benefit analysis*. Harvard University Press.

Aizer, A. (2010). The Gender Wage Gap and Domestic Violence. *American Economic Review*, 100, 4:1847–1859.

Akerlof, G. A. and Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 40(4):1167–1201.

Altonji, J. G. and Blank, R. M. (1999). Race and gender in the labor market. *Handbook of Labor Economics*, 3:3143–3259.

Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–63.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the ”other 99 percent”. *Science*, 344(6186):843–851.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, 103, 6:2121–2168.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in us wage inequality: Revising the revisionists. *Review of Economics and Statistics*, 90(2):300–323.

Autor, D. H., Katz, L. F., and Krueger, A. B. (1998). Computing inequality: have computers changed the labor market? *Quarterly Journal of Economics*, 113(4):1169–1213.

Barrow, L. and Malamud, O. (2015). Is college a worthwhile investment? *Annu. Rev. Econ.*, 7(1):519–555.

Barth, E., Bryson, A., Davis, J. C., and Freeman, R. (2016). It's where you work: Increases in the dispersion of earnings across establishments and individuals in the united states. *Journal of Labor Economics*, 34(S2):S67–S97.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* Upjohn Institute for Employment Research.

Bayer, P. and Charles, K. K. (2018). Divergent paths: A new perspective on earnings differences between black and white men since 1940. *Quarterly Journal of Economics*, 133(3):1459–1501.

Belfield, C. (2015). Cost-benefit analysis and public value. In Crosby, B. C., Bryson, J. M., and Bloomberg, L., editors, *Valuing public value*. Georgetown University Press.

Belfield, C. R. and Levin, H. M. (2007). *The price we pay: Economic and social consequences of inadequate education*. Brookings Institution Press.

Berman, E., Bound, J., and Griliches, Z. (1994). Changes in the demand for skilled labor within us manufacturing: evidence from the annual survey of manufactures. *Quarterly Journal of Economics*, 109(2):367–397.

Bernard, A. B. and Jensen, J. B. (1997). Exporters, skill upgrading, and the wage gap. *Journal of International Economics*, 42(1-2):3–31.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender Identity and Relative Income within Households. *Quarterly Journal of Economics*, 130, 2:571–614.

Bettinger, E. and Slonim, R. (2007). Patience among children. *Journal of Public Economics*, 91(1-2):343–363.

Bettinger, E. P. (2010). Paying to learns: The effect of financial incentives on elementary school test scores. Working Paper 16333, National Bureau of Economic Research.

Blanchard, O. J. and Katz, L. F. (1992). Regional Evolutions. *Brookings Papers on Economic Activity*, 1:1–75.

Blau, F. D., Kahn, L. M., and Waldfogel, J. (2000). Understanding Young Women's Marriage Decisions: The Role of Labor and Marriage Market Conditions. Working Paper 7510, National

Bureau of Economic Research.

Boardman, A. E., Greenberg, D. H., Vining, A. R., and Weimer, D. L. (2017). *Cost-benefit analysis: concepts and practice*. Upper Saddle River, NJ: Prentice Hall.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739.

Bound, J. and Freeman, R. B. (1992). What went wrong? the erosion of relative earnings and employment among young black men in the 1980s. *Quarterly Journal of Economics*, 107(1):201–232.

Bound, J. and Holzer, H. J. (2000). Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s. *Journal of Labor Economics*, 18, 1:20–54.

Bureau of Labor Statistics (2015). Occupational outlook handbook (2014 - 2015 ed.). Technical report, U.S. Department of Labor. Retrieved from https://www.bls.gov/ooh/home.htm.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 6(82):2295–2326.

Card, D. and DiNardo, J. E. (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics*, 20(4):733–783.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics*, 128(3):967–1015.

Casey, M. D., Cline, J., Ost, B., and Qureshi, J. A. (2018). Academic probation, student performance, and strategic course-taking. *Economic Inquiry*, 56(3):1646–1677.

Castillo, M., Ferraro, P. J., Jordan, J. L., and Petrie, R. (2011). The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics*, 95(11-12):1377–1385.

Cattaneo, M. D., Jansson, M., and Ma, X. (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 0(0):1–7.

Christenson, S. L., Reschly, A. L., and Wylie, C. (2012). *Handbook of Research on Student Engagement*. Springer Science & Business Media.

Cornwell, C., Lee, K. H., and Mustard, D. B. (2003). The effects of merit-based financial aid on course enrollment, withdrawal and completion in college. Technical report, IZA Institute of Labor Economics.

Cortes, K. E. and Zhang, L. (2012). The incentive effects of the top 10% plan. Technical report.

Couch, K. A. and Fairlie, R. (2010). Last hired, first fired? black-white unemployment and the business cycle. *Demography*, 47(1):227–247.

Couch, K. A. and Placzek, D. W. (2010). Earnings losses of displaced workers revisited. *American Economic Review*, 100(1):572–89.

Cullen, J. B., Long, M. C., and Reback, R. (2013). Jockeying for position: Strategic high school choice under texas' top ten percent plan. *Journal of Public Economics*, 97:32–48.

Cullen, J. B. and Reback, R. (2006). Tinkering toward accolades: School gaming under a performance accountability system. *Advances in Applied Microeconomics*, 14(1):1–34.

Dee, T. S. and Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3):418–446.

Dudley, S., Belzer, R., Blomquist, G., Brennan, T., Carrigan, C., Cordes, J., Cox, L. A., Fraas, A., Graham, J., Gray, G., et al. (2017). Consumer's guide to regulatory impact analysis: Ten tips for being an informed policymaker. *Journal of Benefit-Cost Analysis*, 8(2):187–204.

Dunlop, C. A. and Radaelli, C. M. (2016). The politics and economics of regulatory impact assessment. In Dunlop, C. A. and Radaelli, C. M., editors, *Handbook of regulatory impact assessment*. Edward Elgar Publishing.

Dunne, T., Haltiwanger, J., and Troske, K. R. (1996). Technology and jobs: Secular changes and cyclical dynamics. Technical report, National Bureau of Economic Research.

Easton, J. Q., Johnson, E., and Sartain, L. (2017). The predictive power of ninth-grade GPA. *Chicago, IL: University of Chicago Consortium on School Research*, pages 2018–10.

Eble, A. and Hu, F. (2018). Stereotypes, role models, and the formation of beliefs.

Egger, H. and Kreickemeier, U. (2009). Firm heterogeneity and the labor market effects of trade liberalization. *International Economic Review*, 50(1):187–216.

Ellig, J. (2016). Evaluating the quality and use of regulatory impact analysis: The Mercatus Center's regulatory report card, 2008-2013. Retrieved from https://ssrn.com/abstract=2821088.

Ellig, J. and McLaughlin, P. A. (2012). The quality and use of regulatory analysis in 2008. *Risk Analysis: An International Journal*, 32(5):855–880.

Executive Order No. 12,291. (1981). 3 C.F.R.

Executive Order No. 12866. (1993). 3 C.F.R.

Executive Order No. 13563. (2011). 3 C.F.R.

Farrow, S. and Zerbe, R. O. (2013). *Principles and standards for benefit-cost analysis*. Edward Elgar Publishing.

Figlio, D. N. (2006). Testing, crime and punishment. *Journal of Public Economics*, 90(4-5):837–851.

Freeman, R. B. (1975). Supply and salary adjustments to the changing science manpower market: Physics, 1948-1973. *American Economic Review*, 65(1):27–39.

Freeman, R. B. (1980). An empirical analysis of the fixed coefficient" manpower requirements" model, 1960-1970. *Journal of Human Resources*, 15(2):176–199.

Fritsch, O. and Kamkhaji, J. C. (2016). Implementing in the laboratory: Scorecards for appraising regulatory impact assessment. In Dunlop, C. A. and Radaelli, C. M., editors, *Handbook of Regulatory Impact Assessment*. Edward Elgar Publishing.

Fritsch, O., Kamkhaji, J. C., and Radaelli, C. M. (2017). Explaining the content of impact assessment in the united kingdom: Learning across time, sectors, and departments. *Regulation & Governance*, 11(4):325–342.

Fruehwirth, J. C. (2013). Identifying peer achievement spillovers: Implications for desegregation and the achievement gap. *Quantitative Economics*, 4(1):85–124.

Fryer, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics*, 126(4):1755–1798.

García, J. L., Heckman, J. J., Leaf, D. E., and Prados, M. J. (2017). Quantifying the life-cycle benefits of a prototypical early childhood program. Working Paper 23479, National Bureau of

Economic Research.

Gelman, A. and Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*.

Gerard, F., Rokkanen, M., and Rothe, C. (2020). Bounds on treatment effects in regression discontinuity designs with a manipulated running variable. *Quantitative Economics, forthcoming*.

Glewwe, P., Ilias, N., and Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 2(3):205–27.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018). Bartik Instruments: When, What, Why and How. Working Paper 24408, National Bureau of Economic Research.

Golightly, E. (2019). Does college access increase high school effort? evaluating the impact of the texas top 10% rule. Technical report.

Goodman, J. (2014). Flaking out: Student absences and snow days as disruptions of instructional time. Working Paper 25018, National Bureau of Economic Research.

Goodman, J., Hurwitz, M., and Smith, J. (2017). Access to 4-year public colleges and degree completion. *Journal of Labor Economics*, 35(3):829–867.

Gordon, N. (2016). Protecting and promoting the use of evidence in the regulatory process. Technical report, Brookings. Retrieved from https://www.brookings.edu/research/protecting-and-promoting-the-use-of-evidence-in-the-regulatory-process.

Guvenen, F., Ozkan, S., and Song, J. (2014). The nature of countercyclical income risk. *Journal of Political Economy*, 122(3):621–660.

Hahn, R. W. and Dudley, P. M. (2007). How well does the us government do benefit-cost analysis? *Review of Environmental Economics and Policy*, 1(2):192–211.

Hahn, R. W. and Tetlock, P. C. (2008). Has economic analysis improved regulatory decisions? *Journal of Economic Perspectives*, 22(1):67–84.

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., and Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science*, 23(8):899–906.

Harrington, W., Morgenstern, R. D., and Nelson, P. (2000). On the accuracy of regulatory cost estimates. *Journal of Economic Perspectives*, 19(2):297–322.

Haskins, R. (2016). Under the radar: Getting social policy done in a divided washington. Technical report, Brookings. Retrieved from https://www.brookings.edu/opinions/under-the-radar-getting-social-policy-done-in-a-divided-washington.

Hastings, J. S., Neilson, C. A., and Zimmerman, S. D. (2012). The effect of school choice on intrinsic motivation and academic outcomes. Working Paper 18324, National Bureau of Economic Research.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010). The rate of return to the highscope perry preschool program. *Journal of Public Economics*, 94(1-2):114–128.

High School Equivalency Program and College Assistance Migrant Program, The Federal TRIO Programs, and Gaining Early Awareness and Readiness for Undergraduate Program (2010). 75 Fed. Reg. RIN 1840-AD01.

Hoynes, H. (1999). The employment, earnings, and income of less skilled workers over the business cycle. Technical report, National bureau of economic research.

Hoynes, H., Miller, D. L., and Schaller, J. (2012). Who suffers during recessions? *Journal of Economic perspectives*, 26(3):27–48.

Institute of Medicine, The National Research Council (2014). *Considerations in Applying Benefit-Cost Analysis to Preventive Interventions for Children, Youth, and Families: Workshop Summary*. National Academies Press.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the chicago public schools. *Journal of Public Economics*, 89(5-6):761–796.

Karoly, L. A. (2012). Toward standardization of benefit-cost analysis of early childhood interventions. *Journal of Benefit-Cost Analysis*, 3(1):1–45.

Katz, L. F. and Autor, D. H. (1999). Changes in the wage structure and earnings inequality. *Handbook of Labor Economics*, 3:1463–1555.

Katz, L. F. and Murphy, K. M. (1992). Changes in relative wages, 1963–1987: supply and demand factors. *Quarterly Journal of Economics*, 107(1):35–78.

Kirst, M. and Venezia, A. (2004). *From high school to college: Improving opportunities for success*. San Francisco: Jossey-Bass.

Koretz, D. M. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources*, 37(4):752–777.

Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to learn. *Review of Economics and Statistics*, 91(3):437–456.

Lavecchia, A. M., Liu, H., and Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. In *Handbook of the Economics of Education*, volume 5, pages 1–74. Elsevier B.V., 1 edition.

Lee, D. S. and Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 2(142):665–674.

Leeds, D. M., McFarlin, I., and Daugherty, L. (2017). Does student effort respond to incentives? evidence from a guaranteed college admissions program. *Research in Higher Education*, 58(3):231–243.

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., and Shand, R. (2017). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis*. Thousand Oaks, CA: Sage.

Lindo, J. M., Sanders, N. J., and Oreopoulos, P. (2010). Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics*, 2(2):95–117.

Liu, J., Lee, M., and Gershenson, S. (2019). The short- and long-run impacts of secondary school absences. Technical report, IZA Institute of Labor Economics.

Macartney, H. (2016). The dynamic effects of educational accountability. *Journal of Labor Economics*, 34(1):1–28.

Nakamura, E. and Steinsson, J. (2014). Fiscal stimulus in a monetary union: Evidence from US regions. *American Economic Review*, 104(3):753–92.

Nas, T. F. (2016). *Cost-benefit analysis: Theory and application*. Thousand Oaks, CA: Sage.

Neal, D. and Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92(2):263–283.

Nores, M., Belfield, C. R., Barnett, W. S., and Schweinhart, L. (2005). Updating the economic impacts of the high/scope Perry preschool program. *Educational Evaluation and Policy Analysis*, 27(3):245–261.

Peacock, M. (2016). Implementing a two-for-one regulatory requirement in the U.S. Technical report, Washington, DC: George Washington.

Pierce, J. R. and Schott, P. K. (2016). The Surprisingly Swift Decline of US Manufacturing Employment. *American Economic Review*, 106, 7:1632–1662.

Pierce, J. R. and Schott, P. K. (2017). Investment Responses to Trade Liberalization: Evidence from the U.S. Industries and Plants. Working Paper 24071, National Bureau of Economic Research.

Posner, E. A. (2003). Transfer regulations and cost-effectiveness analysis. *Duke Law Journal*, 53:1067.

Posner, R. A. (2000). Cost-benefit analysis: Definition, justification, and comment on conference papers. *The Journal of Legal Studies*, 29(S2):1153–1177.

Program Integrity: Gainful Employment (2014). 79 Fed. Reg. RIN 1840-AD15.

Program Integrity Issues (2010). 75 Fed. Reg. RIN 1840-AD02.

Radaelli, C. and Fritsch, O. (2012). Measuring regulatory performance: evaluating regulatory management tools and programmes.

Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92(5-6):1394–1415.

Revesz, R. L. and Livermore, M. A. (2008). *Retaking rationality: How cost-benefit analysis can better protect the environment and our health*. Oxford University Press.

Robinson, L. A., Hammitt, J. K., and J Zeckhauser, R. (2016). Attention to distribution in us regulatory analyses. *Review of Environmental Economics and Policy*, 10(2):308–328.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M. (2019). Ipums usa: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2019.

https://doi.org/10.18128/D010.V10.0 (accessed June 5, 2019).

Schwartz, J. A. (2010). *52 Experiments with regulatory review: The political and economic inputs into state rulemakings*. Institute for Policy Integrity, New York University School of Law.

Scott-Clayton, J. (2011). On money and motivation a quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, 46(3):614–646.

Shapiro, S. and Morrall, J. (2016). Does haste make waste? how long does it take to do a good regulatory impact analysis? *Administration & Society*, 48(3):367–389.

Shapiro, S. and Morrall III, J. F. (2012). The triumph of regulatory politics: Benefit–cost analysis and political salience. *Regulation & Governance*, 6(2):189–206.

Song, J., Price, D. J., Guvenen, F., Bloom, N., and von Wachter, T. (2019). Firming Up Inequality. *Quarterly Journal of Economics*, 134(1):1–50.

Stinebrickner, R. and Stinebrickner, T. R. (2008). The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis & Policy*, 8(1).

Sunstein, C. R. (2014a). The regulatory lookback. *Boston University Law Review*, 94:579–602.

Sunstein, C. R. (2014b). *Valuing life: Humanizing the regulatory state*. University of Chicago Press.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature*, 49(2):326–65.

Tincani, M. M., Kosse, F., and Miglino, E. (2020). Student beliefs and the (perverse) incentives of preferential college admissions. Technical report.

UNC (2015). Statistical abstract of higher education in north carolina 2014-15. Technical report, The University of North Carolina Chapel Hill, North Carolina.

U.S. Census Bureau (2019a). Quarterly workforce indicators (qwi) [dataset]. Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program [distributor]. Data retrieved from Quarterly Workforce Indicators, `https://lehd.ces.census.gov/data/#qwi(accessedJune5,2019)`.

U.S. Census Bureau (2019b). Statistics of u.s. businesses (susb) [dataset]. Washington, DC: U.S.

Census Bureau. Data available through custom request via https://www.census.gov/programs-surveys/susb/data/custom-tabulations.html (received June 13, 2019).

US Department of Education (2016). Non-regulatory guidance: Using evidence to strengthen education investments.

Vining, A. and Weimer, D. L. (2010). An assessment of important issues concerning the application of benefit-cost analysis to social policy. *Journal of Benefit-Cost Analysis*, 1(1):1–40.

William D. Ford Federal Direct Loan Program (2013). 78 Fed. Reg. RIN 1840-AD13. (to be codified at 34 C.F.R. pt. 685).

Wulfert, E., Block, J. A., Santa Ana, E., Rodriguez, M. L., and Colsman, M. (2002). Delay of gratification: Impulsive choices and problem behaviors in early and late adolescence. *Journal of Personality*, 70(4):533–552.

Yeaple, S. R. (2005). A simple model of firm heterogeneity, international trade, and wages. *Journal of International Economics*, 65(1):1–20.

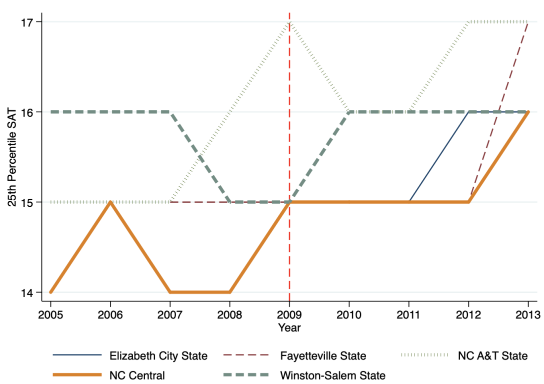
Zerbe, R. O., Davis, T. B., Garland, N., and Scott, T. (2013). Conclusion: Principles and standards for benefit-cost analysis. In Farrow, S. and Zerbe, R. O., editors, *Principles and standards for benefit-cost analysis*, pages 364–445. Edward Elgar Publishing.

Appendix A: Chapter 1

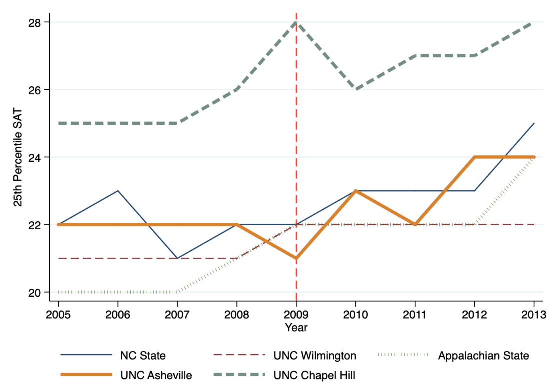
A.1 ACT Admission Trends for UNC Campuses

Figure A.1: 25th Percentile ACT Among Admitted Students

(a) Bottom 5 Campuses



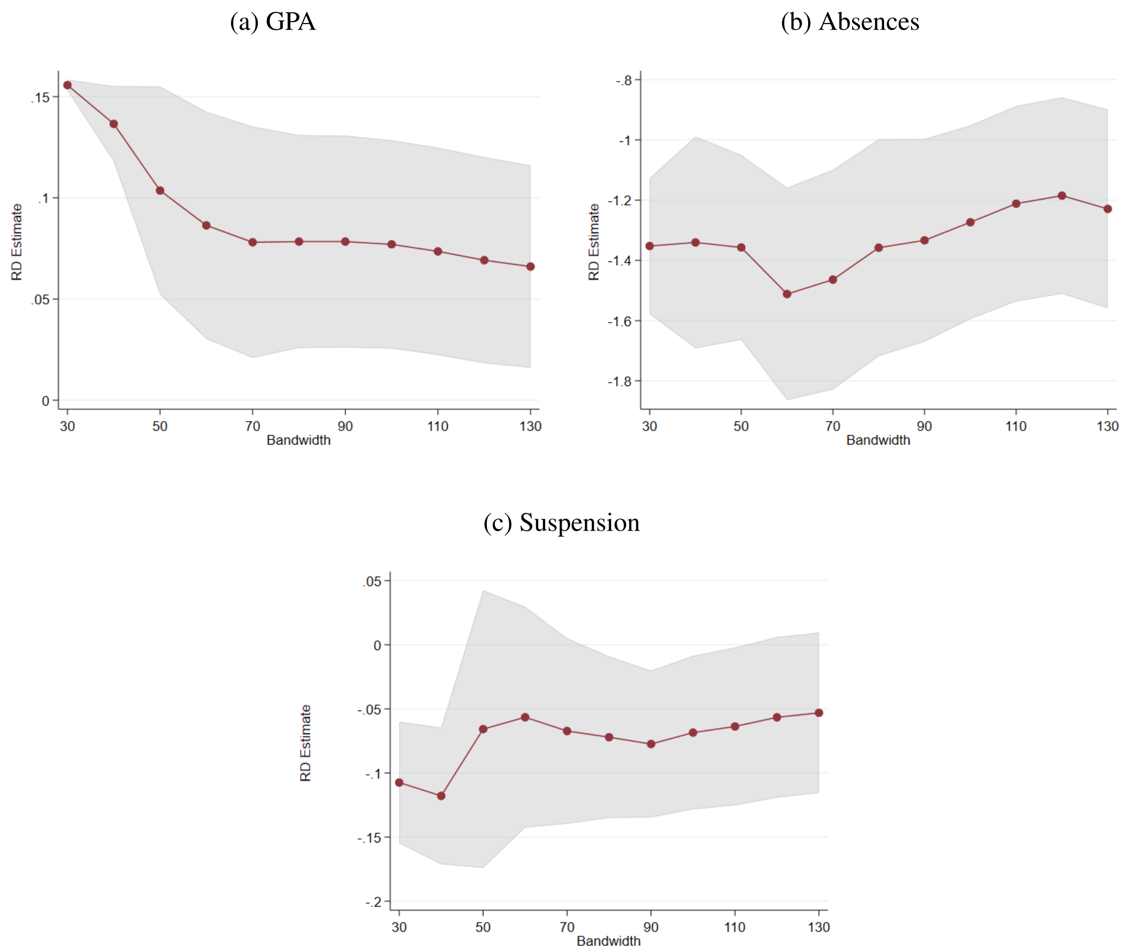
(b) Top 5 Campuses



Notes: This figure shows the evolution of the average 25th percentile ACT score for admitted students of UNC campuses over time. Panel (a) shows the evolution for the bottom five campuses. Panel (b) shows the evolution for the top five campuses. The dash red line represent the year in which the UNC Minimum Admission Requirements were implemented. Source: IPEDS.

A.2 Robustness of Regression Discontinuity Estimates to Bandwidth Selection

Figure A.2: Regression Discontinuity Estimates with Different Bandwidths

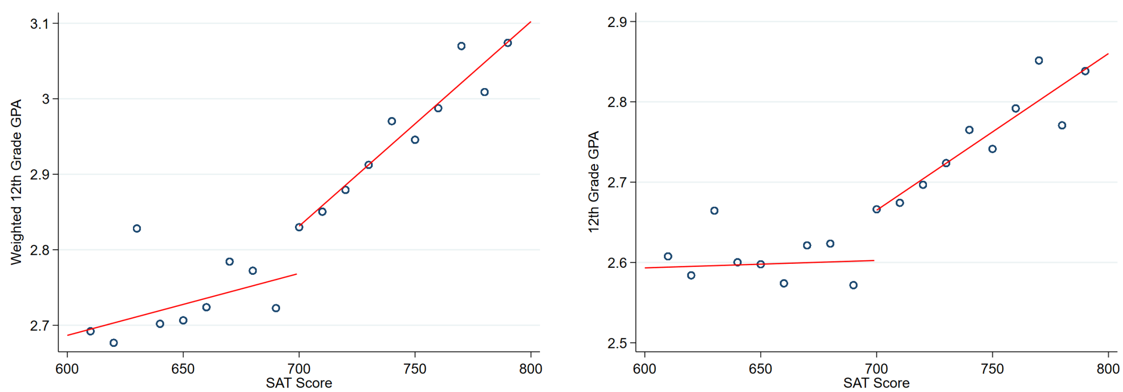


Notes: This figure shows regression discontinuity estimates for student senior-year outcomes as the bandwidth used for estimation increases from 30 to 120. Panel (a) presents estimates for student GPA, panel (b) presents estimates for student absences, and panel (c) presents estimates for student suspensions.

A.3 Results for Raw GPA v. Weighted GPA

Figure A.3: Global Linear Fit

(a) 12th Grade Weighted GPA (b) 12th Grade GPA

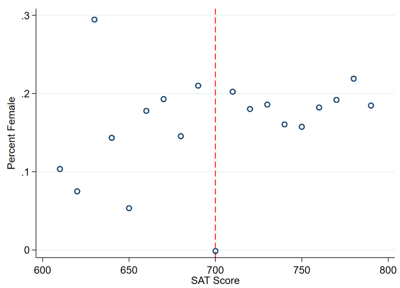


Notes: This figure shows discontinuities in senior-year weighted and unweighted GPA. All panels are created by plotting the average outcomes across each SAT score bin and fitting the data using a linear regression without controls on either side of the cutoff.

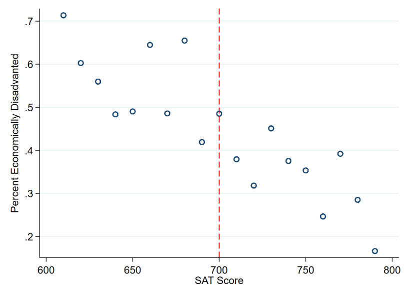
A.4 Covariate Smoothness Scatter Plots

Figure A.4: Smoothness of Student Pre-Treatment Covariates

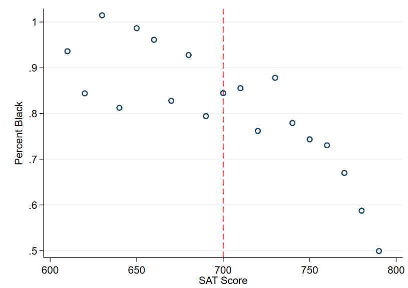
(a) Gender - Female



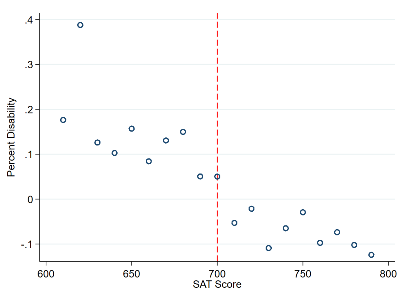
(b) EDS



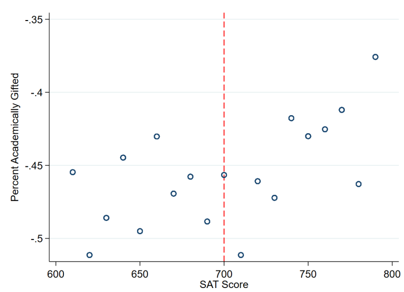
(c) Race - African American



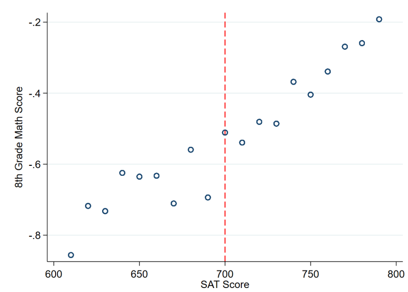
(d) SWD



(e) AIG

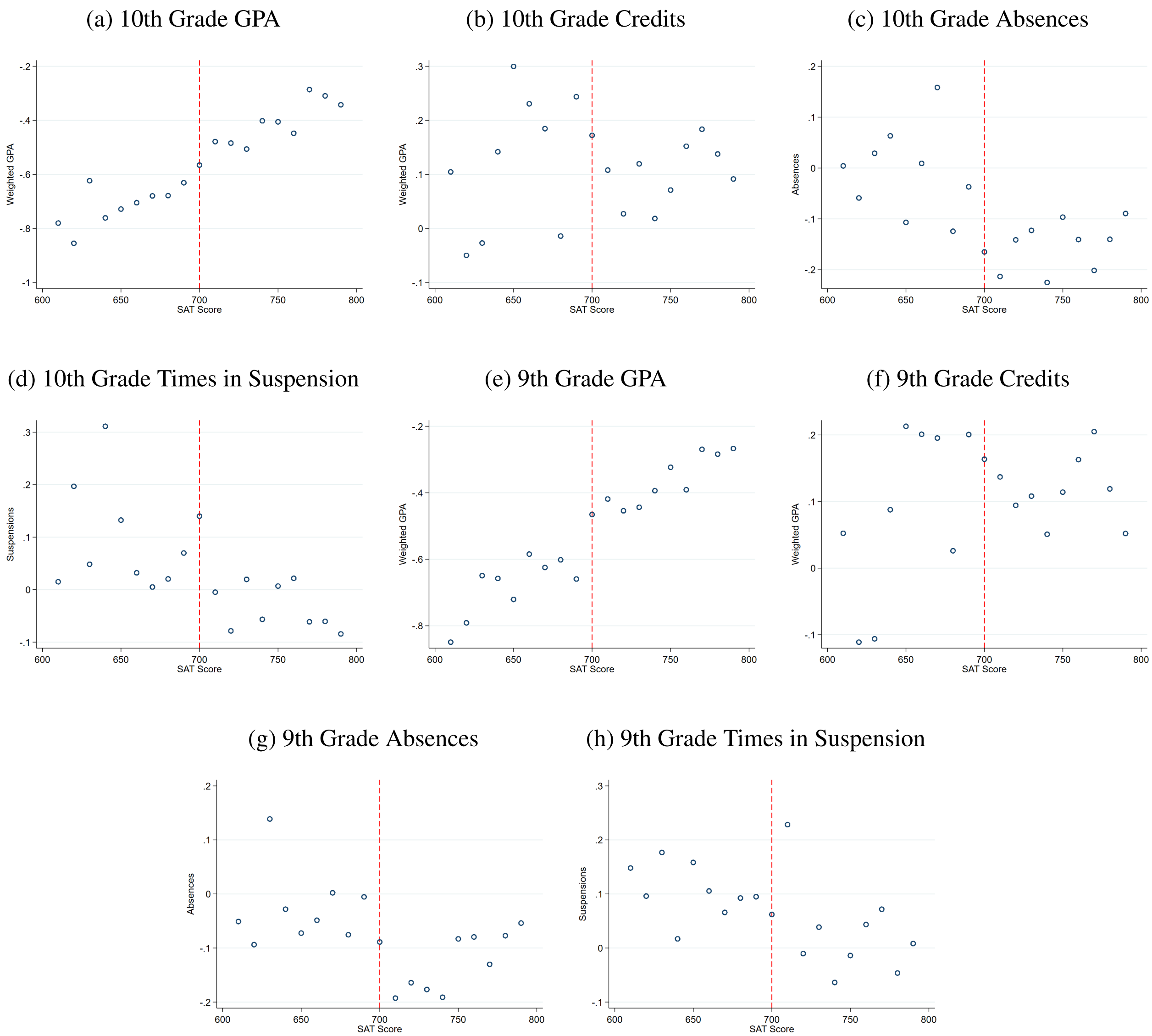


(f) 8th Grade Math



Notes: This figure shows discontinuities in student pre-treatment covariates. All panels are created by plotting the average outcomes across each SAT score bin and fitting the data using a linear regression without controls on either side of the cutoff.

Figure A.5: Smoothness of Student Pre-Treatment Outcomes



Notes: This figure shows discontinuities in student pre-treatment outcomes. All panels are created by plotting the average outcomes across each SAT score bin and fitting the data using a linear regression without controls on either side of the cutoff.

Appendix B: Chapter 2

B.1 Additional Detail about Merging QWI with SUSB

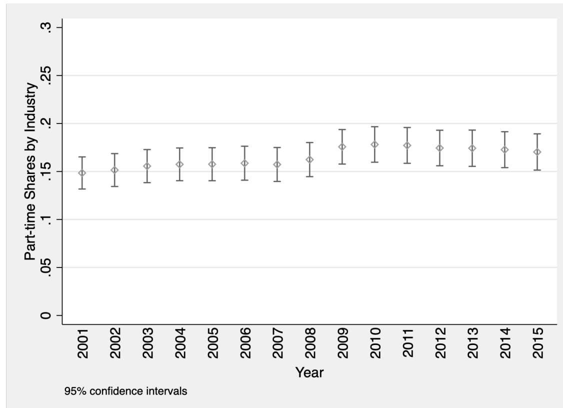
We merge QWI and SUSB data at the county-industry level from 2007 to 2009. The 2007-08 SUSB dataset uses the 2002 NAICS classification while the 2008-09 dataset uses the 2007 classification. As the QWI uses the 2012 classification, we convert all industry categories to the 2012 definition using equivalences published by the Census, before merging the data sources. While the SUSB data contain six-digit industry codes, we use the more aggregated four-digit measure to match the aggregation level of our QWI employment data.

Additionally, SUSB excludes some NAICS codes, including crop and animal production (NAICS 111,112), rail transportation (NAICS 482), postal service (NAICS 491), pension, health, welfare, and vacation funds (NAICS 525110, 525120, 525190), trusts, estates, and agency accounts (NAICS 525920), private households (NAICS 814), and public administration (NAICS 92).

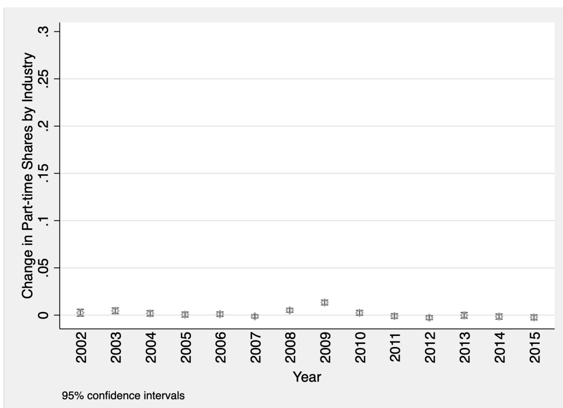
B.2 Additional Stylized Facts

B.2.1 Part-Time Employment

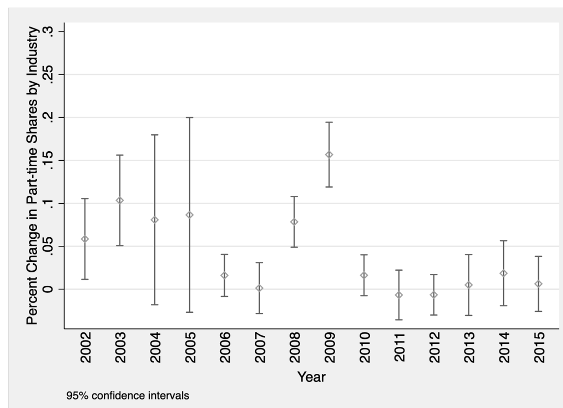
The QWI data does not distinguish between full-time and part-time employment. This complicates the interpretation of our results because the Great Recession may cause shifts on both the employed/not-employed margin and the full-time/part-time margin. Our method and data will not detect industry responses which shift workers between part-time and full-time roles. Although we cannot address this concern directly, we can nonetheless present evidence that the share of part time work is reasonably stable (1) across industries and (2) within industries over time.



(a) Across-Industry Levels



(b) Within-Industry Changes



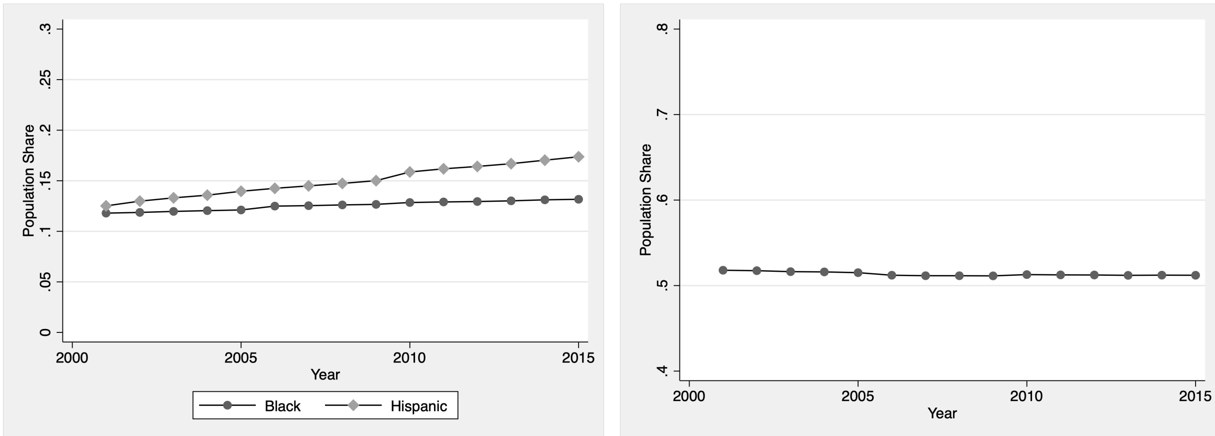
(c) Within-Industry Percent Changes

Figure B.1: Part-Time Employment Shares Across and Within Industries (2001-2015)

Notes: This figure shows the evolution of part-time employment shares across industries from 2001 to 2015 inclusive. In particular, panels (a), (b) and (c) show yearly variation in the level, change and percent change in part-time employment, respectively.

We use the ACS to estimate the share of part-time and full-time workers by four-digit industry code. Panel (a) of Figure B.1 plots the mean and 95% confidence interval for the share of part time workers by industry for each ACS year (2001 to 2015). The means are around 0.15-0.18, which matches aggregate estimates from the BLS quite well.[1] Moreover, the variance in this share across industries is fairly small each year, with the 95% confidence intervals generally spanning only about 0.025 percentage points. The part-time share increased modestly from 0.16 to 0.18 during the recession, suggesting across-industry stability.

[1]Compare BLS series LNS11000000 (total labor force) to LNS12600000 (part time labor force).



(a) Black and Hispanic

(b) Female

Figure B.2: Population Share by Type (2001-2015)

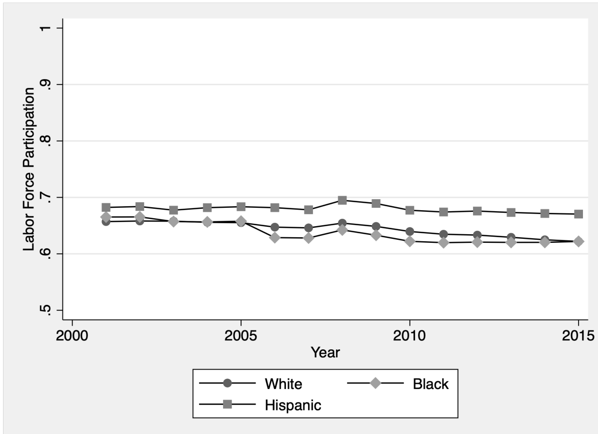
Notes: This figure shows the evolution of population share by gender and race from 2001 to 2015 inclusive. In particular, panel (a) and (b) depict the trends for female, Black, and Hispanic people, respectively.

With respect to within-industry stability, Panels (b) and (c) of Figure B.1 show that the part-time shares did not change very much within industry over our sample period. In 2008 and 2009, the part time share increased for most industries, while the typical change aside from these two years is around 0. However, even in 2009, the typical industry only saw its part-time share increase by about 0.015 percentage points (off of a base of about 0.16 percentage points). Taken together, these figures suggest that the distinction between part- and full-time is not an important source of within-industry variation.

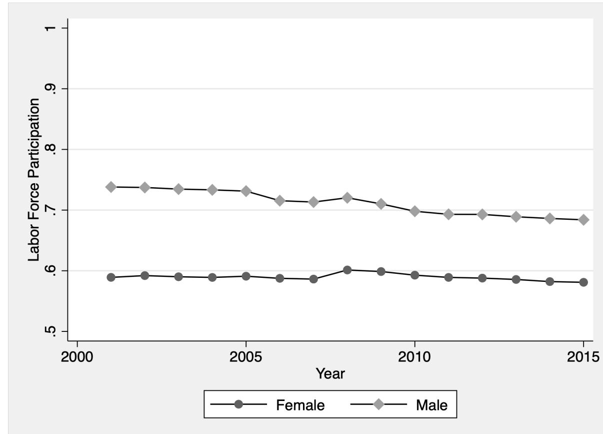
B.2.2 Population Share and Labor Force Participation by Type

To provide additional context for our results, we document trends in the population share and labor force participation by worker type surrounding the Great Recession. Using supplemental ACS data, Figure B.2 plots the population share for black and Hispanic people, and females over time. As one would expect, the share of females in the population is unchanged over time. The share of black and Hispanic people is rising during the 2000s, but the trend does not change during the onset of the recession.

We also use the ACS data to plot the labor force participation by type in Figure B.3. There is a



(a) White, Black and Hispanic



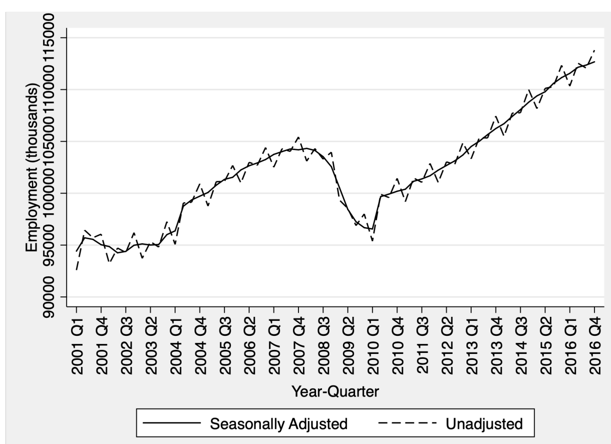
(b) Male and Female

Figure B.3: Labor Force Participation by Type (2001-2015)

Notes: This figure shows the evolution of labor force participation by gender and race from 2001 to 2015 inclusive. In particular, panels (a) and (b) depict the trend for gender (male and female) and race (white, black and Hispanic), respectively.

temporary uptick in the participation of all worker types during the recession, against a generally declining trend. However, given that it is very small relative to the level of labor force participation for each worker type, we do not expect our results in Section 2.3 to be driven by such patterns.

B.2.3 Aggregate Employment 2001-2016

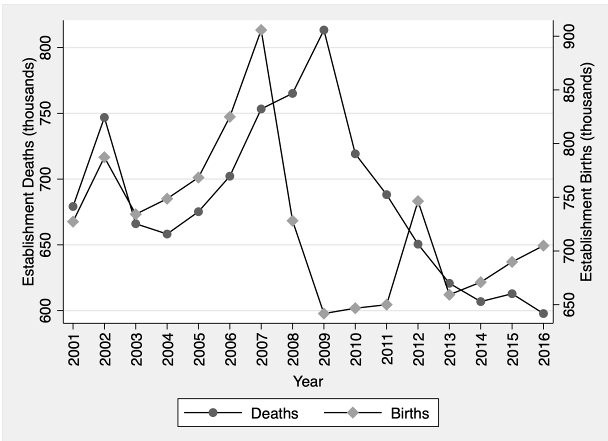


(a) Aggregate

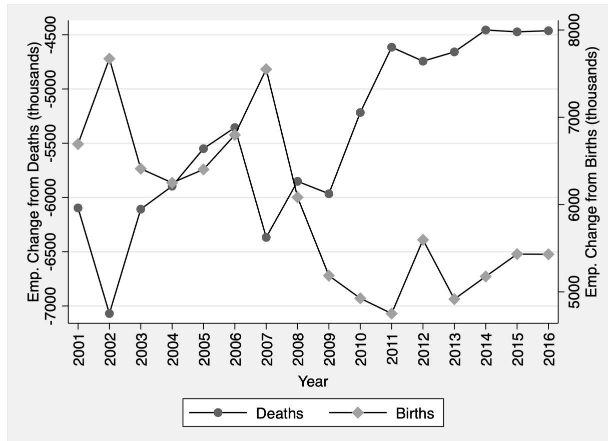
Figure B.4: Aggregate Employment Trends, Overall and by Worker Race (2001-2016) – QWI Data

Notes: This figure shows the quarterly evolution of aggregate employment from 2001 to 2016 inclusive with and without seasonality adjustments.

B.2.4 Establishment Deaths and Births in the SUSB



(a) Establishment Deaths and Births



(b) Employment Loss due to Deaths and Births

Figure B.5: Trends in Establishment Deaths and Births (2001-2016) – SUSB Data

Notes: Panel (a) shows the evolution of establishment deaths and births from 2001 to 2016, while Panel (b) shows the trend in employment loss (gain) due to deaths (births) over time.

B.3 Results by Gender

Table B.1: Employment Change Decompositions by Gender

Panel A: Overall Employment Changes

	Total	Deaths
θ^m	-0.071***	-0.089***
	[-0.075,-0.066]	[-0.095,-0.078]
	(0.000)	(0.000)
θ^f	-0.035***	-0.099***
	[-0.038,-0.032]	[-0.112,-0.092]
	(0.000)	(0.000)

Panel B: Across-Within Comparisons (Male vs. Female)

	Total	Deaths
$[\theta^m - \theta^f]_A$	-0.045***	-0.013*
	[-0.048,-0.042]	[-0.029,-0.005]
	(0.000)	(0.062)
$[\theta^m - \theta^f]_W$	0.009***	0.023
	[0.008,0.010]	[0.007,0.060]
	(0.000)	(0.133)

Notes: Panel A presents employment changes during the Great Recession (2007-2009) by worker gender, both in total (agnostic to the establishment-level cause) and for those arising from establishment deaths. Male and female workers are denoted by m and f, respectively. Panel B presents estimates of equations (2.4) and (2.5) for the male-female comparison, both in total and for establishment deaths. 95% confidence intervals and significance are calculated using 5,000 bootstrap iterations. Confidence intervals are reported in square brackets and p-values are reported in parentheses. *** and * denotes significance at the 1% and 10% level, respectively.

B.4 Spatial Variation

Figure B.6: Percent Employment Change 2007-2009 by County

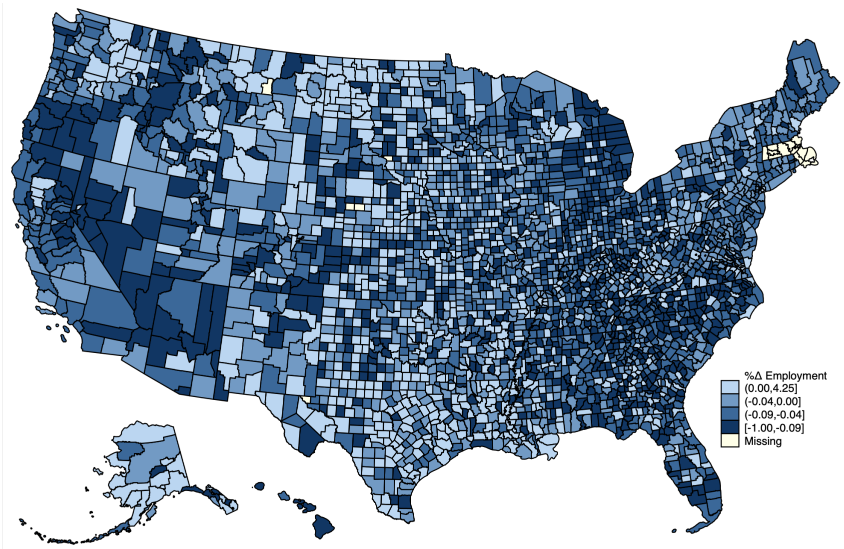
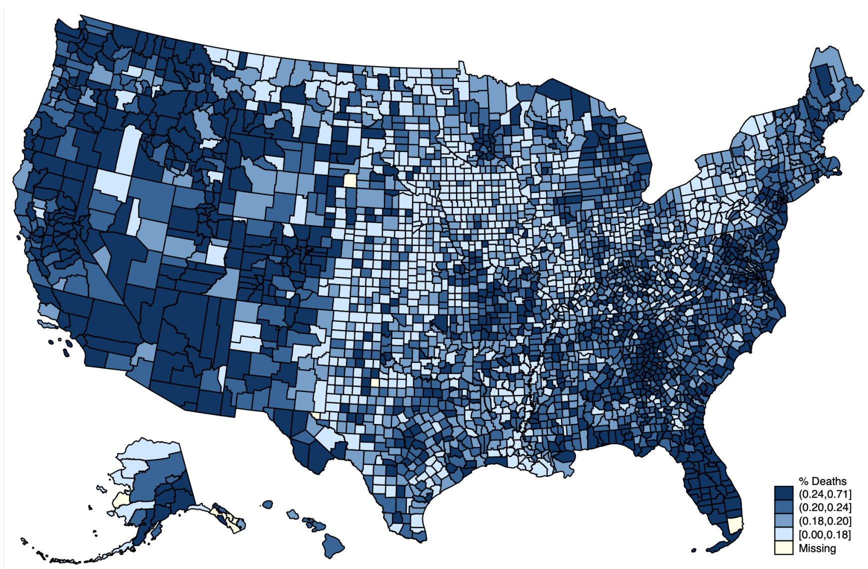


Figure B.7: Percent of Establishment Deaths from 2007-2009 by County



B.5 Decomposition Derivations

$$
\begin{aligned}
\theta^{\tau} - \theta^{\tau'} &\equiv \frac{\Delta E^{\tau}}{E_0^{\tau}} - \frac{\Delta E^{\tau'}}{E_0^{\tau'}} = \frac{\sum_i \Delta E_i^{\tau}}{\sum_i E_{i0}^{\tau}} - \frac{\sum_i \Delta E_i^{\tau'}}{\sum_i E_{i0}^{\tau'}} \\
&= \frac{\sum_i \theta_i^{\tau} E_{i0}^{\tau}}{\sum_i E_{i0}^{\tau}} - \frac{\sum_i \theta_i^{\tau'} E_{i0}^{\tau'}}{\sum_i E_{i0}^{\tau'}} \\
&= \frac{\sum_i \pi_i^{\tau} \theta_i^{\tau} E_{i0}}{\sum_i \pi_i^{\tau} E_{i0}} - \frac{\sum_i \pi_i^{\tau'} \theta_i^{\tau'} E_{i0}}{\sum_i \pi_i^{\tau'} E_{i0}} \\
&= \sum_i (\tilde{\pi}_i^{\tau} \theta_i^{\tau} - \tilde{\pi}_i^{\tau'} \theta_i^{\tau'}) E_{i0} \,,
\end{aligned}
\tag{B.1}
$$

To separate this overall adjustment into within (W) and across (A) components, we expand the expression by subtracting and adding the average share $\tilde{\pi}_i \equiv \frac{\tilde{\pi}_i^{\tau} + \tilde{\pi}_i^{\tau'}}{2}$ on the right-hand side in the following way:

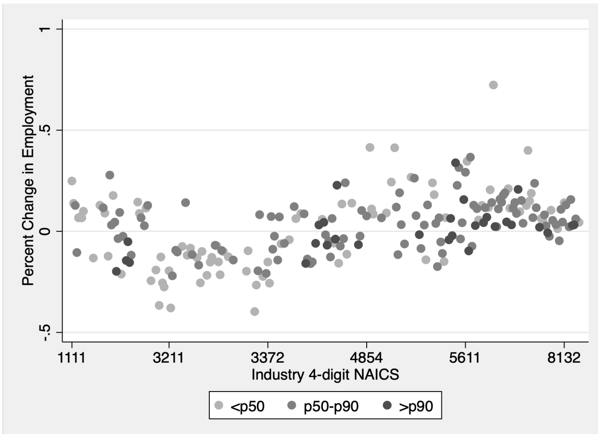
$$
\theta^{\tau} - \theta^{\tau'} = \sum_i [\underbrace{\theta_i^{\tau}(\tilde{\pi}_i^{\tau} - \tilde{\pi}_i) - \theta_i^{\tau'}(\tilde{\pi}_i^{\tau'} - \tilde{\pi}_i)}_{\text{A}} + \underbrace{\tilde{\pi}_i(\theta_i^{\tau} - \theta_i^{\tau'})}_{\text{W}}] E_{i0} \,.
\tag{B.2}
$$

Then, we simplify each component as follows:

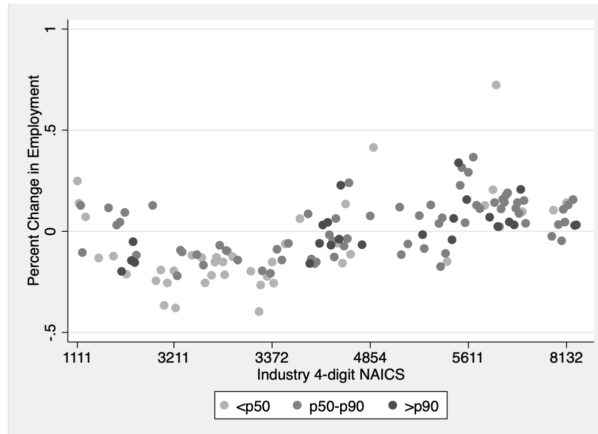
$$
\begin{aligned}
\theta_W^{\tau} - \theta_W^{\tau'} &\equiv \sum_i \tilde{\pi}_i (\theta_i^{\tau} - \theta_i^{\tau'}) E_{i0} \\
&= \sum_i \left(\frac{\tilde{\pi}_i^{\tau} + \tilde{\pi}_i^{\tau}}{2} \right) (\theta_i^{\tau} - \theta_i^{\tau'}) E_{i0} \\
&= \sum_i (\theta_i^{\tau} - \theta_i^{\tau'}) \left(\frac{(E_{i0}^{\tau}/E_0^{\tau}) + (E_{i0}^{\tau'}/E_0^{\tau'})}{2} \right) .
\end{aligned}
\tag{B.3}
$$

$$
\begin{aligned}
\theta_A^{\tau} - \theta_A^{\tau'} &\equiv \sum_i [\theta_i^{\tau}(\tilde{\pi}_i^{\tau} - \tilde{\pi}_i) - \theta_i^{\tau'}(\tilde{\pi}_i^{\tau'} - \tilde{\pi}_i)] E_{i0} \\
&= \sum_i \left[\theta_i^{\tau} \left(\frac{\tilde{\pi}_i^{\tau} - \tilde{\pi}_i^{\tau'}}{2} \right) - \theta_i^{\tau'} \left(\frac{\tilde{\pi}_i^{\tau'} - \tilde{\pi}_i^{\tau}}{2} \right) \right] E_{i0} \\
&= \sum_i \left(\frac{\theta_i^{\tau} + \theta_i^{\tau'}}{2} \right) (\tilde{\pi}_i^{\tau} - \tilde{\pi}_i^{\tau'}) E_{i0} \\
&= \sum_i \left(\frac{\theta_i^{\tau} + \theta_i^{\tau'}}{2} \right) \left(\frac{E_{i0}^{\tau}}{E_0^{\tau}} - \frac{E_{i0}^{\tau'}}{E_0^{\tau'}} \right) .
\end{aligned}
\tag{B.4}
$$

B.6 Heterogeneous Effects by Industry



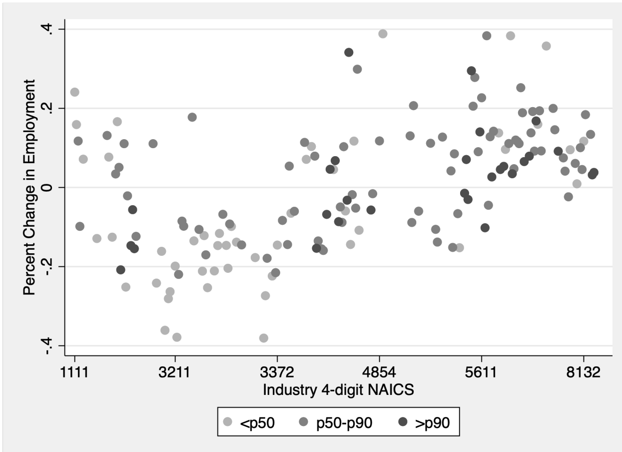
(a) No Correction



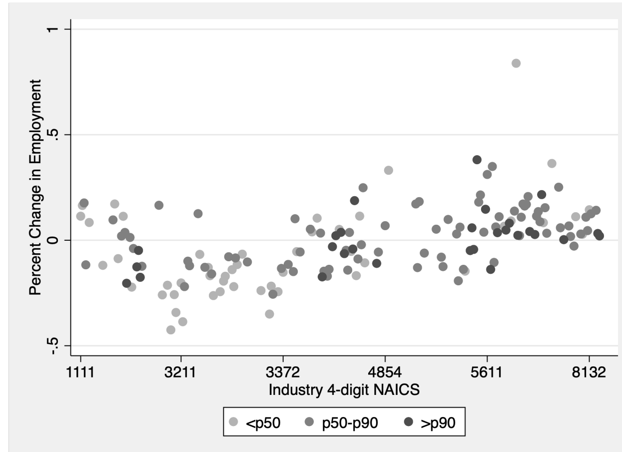
(b) Holm Bonferroni

Figure B.8: Percent Change in Employment by Industry (2007-2009)

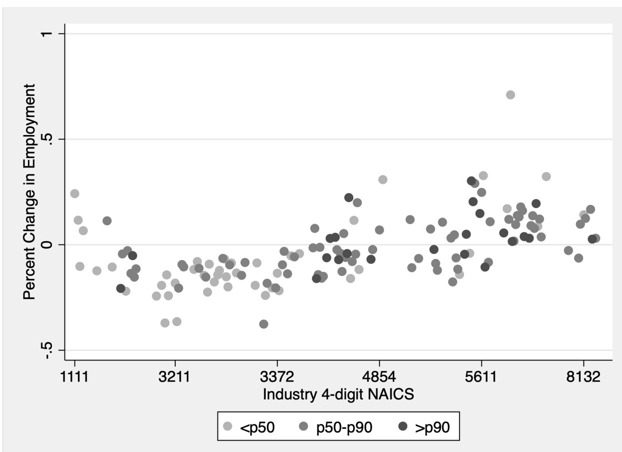
Notes: This figure shows the industry-level average percent change of employment from 2007 to 2009 using the full sample. Light gray points indicate industries below the median industry size (as measured by total employment in 2007), gray points indicate industries between the median and the 90th percentile of size, and dark gray points indicate the largest industries (above the 90th percentile of size). Panel (a) plots the average percent change in employment by industry that is statistically different than zero at the 10% level, while panel (b) applies the Holm-Bonferroni correction to adjust significance for multiple comparisons.



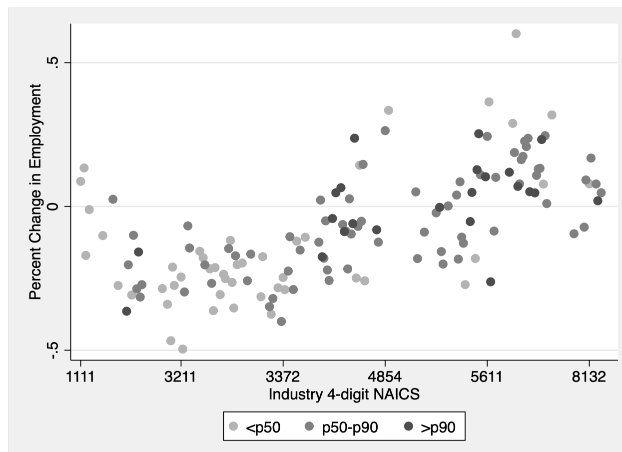
(a) Male



(b) Female



(c) White



(d) Black

Figure B.9: Percent Change in Employment by Industry and Gender (2007-2009)

Notes: This figure shows the percent change of employment from 2007 to 2009 inclusive for different worker types, using the QWI sample and the Holm-Bonferroni correction. Light gray points indicate industries below the median industry size (as measured by total employment in 2007), gray points indicate industries between the median and the 90th percentile of size, and dark gray points indicate the largest industries (above the 90th percentile of size). Panels (a) and (b) show these changes by gender (male and female), while panels (c) and (d) show them by race (white and black).

B.7 Sensitivity Analysis and Goodness of Fit

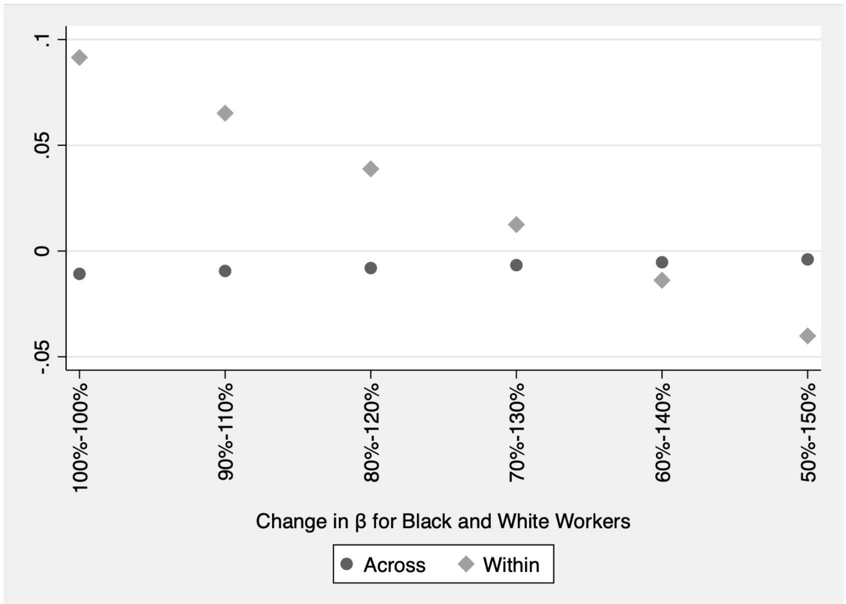


Figure B.10: Sensitivity of Estimated Within and Across Components to Changes in β^b and β^w

Notes: This figure shows the estimates of across and within components due to deaths when estimated coefficients of black employment change are reduced stepwise by 10% and estimated coefficients of white employment change are increased stepwise by 10%.

Table B.2: Out-Of-Sample Fit by Race (White vs. Hispanic)

	(1) Mean Observed θ	(2) Mean Predicted θ	(3) Mean Difference	(4) SD of Difference
θ^w	-0.058	-0.058	-0.000	0.003
θ^h	-0.039	-0.038	-0.001	0.009
$\theta^w - \theta^h$	-0.019	-0.020	0.000	0.010
$[\theta^w - \theta^h]_A$	0.013	0.013	0.000	0.002
$[\theta^w - \theta^h]_W$	-0.032	-0.032	0.000	0.010

Notes: This table presents statistics related to the out-of-sample prediction outlined in Section 2.4. All statistics are based on 500 randomly-drawn estimation and holdout samples, stratified on county-level employment. Columns (1) through (3) present averages of estimates obtained for each random draw. Column (4) presents the standard deviation of Column (3).

Table B.3: Out-Of-Sample Fit by Gender

	(1) Mean Observed θ	(2) Mean Predicted θ	(3) Mean Difference	(4) SD of Difference
θ^m	-0.071	-0.070	-0.001	0.003
θ^f	-0.035	-0.035	-0.000	0.003
$\theta^m - \theta^f$	-0.036	-0.035	-0.001	0.002
$[\theta^m - \theta^f]_A$	-0.045	-0.044	-0.000	0.003
$[\theta^m - \theta^f]_W$	0.009	0.009	-0.000	0.002

Notes: This table presents statistics related to the out-of-sample prediction outlined in Section 2.4. All statistics are based on 500 randomly-drawn estimation and holdout samples, stratified on county-level employment. Columns (1) through (3) present averages of estimates obtained for each random draw. Column (4) presents the standard deviation of Column (3).

Appendix C: Chapter 3

Table C.1: Checklist Evaluation of 28 Federal Regulatory Impact Assessments by Years

	Percentage Performing Activity	
Checklist Criteria	2006-2010	2011-2015
Cost		
Statement of costs beyond budgetary statement (stated)	100	85
Estimated costs using ingredients method (ingredients)	33	15
Estimated costs separately by agency (agency)	60	46
Dollar value of costs reported (calculated)	80	62
Benefits		
Statement of monetized benefits (stated)	87	77
Description of benefits estimation (described)	47	8
Estimated benefits separately by agency (agency)	67	46
Dollar value of benefits reported (calculated)	7	0
Net Present Value (NPV; benefits minus costs)		
Dollar value calculated (NPV)	0	0
Stated as positive (B>C)	100	100
Methodology		
Assumptions described (year, inflation, discount rate)	33	38

Table C.2: Checklist Evaluation of 28 Federal Regulatory Impact Assessments by Budgetary Allocation

Checklist Criteria	Percentage Performing Activity	
	Lower Budgets	Higher Budgets
Cost		
Statement of costs beyond budgetary statement (stated)	86	100
Estimated costs using ingredients method (ingredients)	14	36
Estimated costs separately by agency (agency)	43	64
Dollar value of costs reported (calculated)	71	71
Benefits		
Statement of monetized benefits (stated)	86	79
Description of benefits estimation (described)	14	43
Estimated benefits separately by agency (agency)	50	64
Dollar value of benefits reported (calculated)	0	7
Net Present Value (NPV; benefits minus costs)		
Dollar value calculated (NPV)	0	0
Stated as positive (B>C)	100	100
Methodology		
Assumptions described (year, inflation, discount rate)	21	50

Table C.3: Twenty-Eight Federal Regulatory Impact Assessments

Year	Order	Dept	Regulation	Description
2009	1810-AB04	OESE	State Fiscal Stabilization Fund Program–Notice of Proposed Requirements, Definitions, and Approval Criteria	Formula grants to States to help stabilize State and local budgets in order to minimize and avoid reductions in education and other essential services, in exchange for a State's commitment to advance essential education reform in key areas.
2009	1810-AB06	OESE	School Improvement Grants–Requirements under American Recovery and Reinvestment Act of 2009; Title I of the Elementary and Secondary Education Act of 1965	Funds for school improvement, School Improvement Grants are used to improve student achievement in Title I schools identified for improvement, corrective action, or restructuring so as to enable those schools to make adequate yearly progress (AYP) and exit improvement status.
2009	1810-AB07	OESE	Race to the Top Fund–Notice of Proposed Priorities, Requirements, Definitions, and Selection Criteria	Encourages and rewards States that are creating the conditions for education innovation and reform; achieving significant improvement in student outcomes, closing achievement gaps, improving high school graduation rates, and ensuring student preparation for success in college and careers; and implementing ambitious plans in four core education reform areas.
2010	1810-AB08	OESE	Teacher Incentive Fund–Priorities, Requirements, Definitions, and Selection Criteria	Supports projects that develop and implement PBCSs for teachers, principals, and other personnel in order to increase educator effectiveness and student achievement, measured in significant part by student growth in high-need schools.

2012	1810-AB12	OESE	Teacher Incentive Fund	Support projects that develop/implement PBCSs for teachers, principals, and other personnel to increase educator effectiveness and student achievement, measured in significant part by student growth, in high-need schools.
2012	1810-AB15	OESE	Race to the Top–Early Learning Challenge Phase 2	Improves the quality of early learning and development and close the achievement gap for children with high needs.
2013	1810-AB17	OESE	Race to the Top–District	Supports bold, locally directed improvements in learning and teaching that will directly improve student achievement and educator effectiveness.
2013	1810-AB18	OESE	Race to the Top–Early Learning Challenge	Improves the quality of early learning and development and close the educational gaps for Children with High Needs.
2006	1840-AC86	OPE	Student Assistance General Provisions and Federal Student Aid Programs–Academic Competitiveness and National Science and Mathematics Access To Retain Talent Grant Programs	These interim final regulations implement certain provisions of the Higher Education Reconciliation Act of 2005.
2006	1840-AC87	OPE	Institutional Eligibility Under the Higher Education Act of 1965, as Amended; Student Assistance General Provisions and Federal Student Aid Programs*	Reflect the provisions of the HERA that affect students, borrowers, postsecondary educational institutions, lenders, and other program participants in the Federal student aid programs authorized under Title IV of the HEA.
2007	1840-AC89	OPE	Federal Perkins Loan Program, Federal Family Education Loan Program, and William D. Ford Federal Direct Loan Program	Amending these regulations to strengthen and improve the administration of the loan programs authorized under Title IV of Higher Education Act of 1965.

2008	1840-AC93	OPE	Title IV of the Higher Education Act of 1965, as Amended (TEACH Grant Program)	Non-need-based grant program for students who are enrolled in an eligible program and who agree to teach in a high-need field, at a low-income elementary or secondary school for at least four years within eight years of completing the program for which the TEACH Grant was awarded.
2008	1840-AC94	OPE	Federal Perkins Loan, Federal Family Education Loan (FFEL), and William D. Ford Federal Direct Loan (DL) Programs*	The Secretary amends the Federal Perkins Loan (Perkins Loan) Program, Federal Family Education Loan (FFEL) Program, and William D. Ford Federal Direct Loan (Direct Loan) Program regulations to implement provisions of the College Cost Reduction and Access Act of 2008 (CCRAA)
2009	1840-AC96	OPE	Student Assistance General Provisions; TEACH Grant, Federal Pell Grant, and Academic Competitiveness Grant, and National Science and Mathematics Access To Retain Talent Grant Programs	Amends the regulations for the Academic Competitiveness Grant (ACG) and National Science and Mathematics Access to Retain Talent Grant (National SMART Grant) Programs. These interim final regulations are needed to implement provisions of the Higher Education Act of 1965 (HEA), as amended by the Ensuring Continued Access to Student Loans Act of 2008 (ECASLA) and the Higher Education Opportunity Act of 2008 (HEOA).
2009	1840-AC99	OPE	General and Non-Loan Programmatic Issues	Amends the regulations for Institutional Eligibility Under HEA 1965, the Student Assistance General Provisions, the Federal Work-Study (FWS) Programs, the Teacher Education Assistance for College and Higher Education (TEACH) Grant Program, the Federal Pell Grant Program, and the Leveraging Educational Assistance Partnership Program (LEAP).

2010	1840-AD01	OPE	Federal TRIO Programs, Gaining Early Awareness and Readiness for Undergraduate Program, and High School Equivalency and College Assistance Migrant Programs	Helps migrant and seasonal farmworkers and their immediate family members obtain a GED; CAMP assists students to complete their first academic year of college and continue in postsecondary education.
2010	1840-AD02	OPE	Institutional Eligibility Under the Higher Education Act of 1965; Student Assistance General Provisions	Improving integrity in programs authorized under title IV of HEA 1965 by amending the regulations for Institutional Eligibility, the Secretary's Recognition of Accrediting Agencies, the Secretary's Recognition Procedures for State Agencies, the Student Assistance General Provisions, FFEL Program, the TEACH Grant Program, the Federal Pell Grant Program, and the AGC and National Smart Grant Programs.
2010	1840-AD04	OPE	Program Integrity: Gainful Employment	Amend the Student Assistance General Provisions to establish measures for determining whether certain postsecondary educational programs lead to gainful employment in recognized occupations, and the conditions under which these educational programs remain eligible for the student financial assistance programs authorized under title IV of HEA 1965.
2012	1840-AD05	OPE	Federal Perkins Loan Program, Federal Family Education Loan Program, and William D. Ford Federal Direct Loan Program	Income-Contingent Repayment (ICR) plans in the Direct Loan program, incorporate recent statutory changes to the IBR plan in the Direct Loan and FFEL programs, and streamline and add clarity to the TPD discharge process for borrowers in loan programs under title IV of HEA 1965.

2011	1840-AD06	OPE	Program Integrity: Gainful Employment-Measures	Amend the Student Assistance General Provisions to establish measures for determining whether certain postsecondary educational programs lead to gainful employment in recognized occupations, and the conditions under which these educational programs remain eligible for the student financial assistance programs authorized under title IV of HEA 1965.
2013	1840-AD11	OPE	Federal Pell Grant Program	Amends four sections of the Federal Pell Grant Program regulations to make them consistent with recent changes in the law that prohibit a student from receiving two consecutive Pell Grants in a single award year.
2013	1840-AD12	OPE	Transitioning from the FFEL Program to the Direct Loan Program and Loan Rehabilitation Under the FFEL, Direct Loan, and Perkins Loan Programs	Amends the FFEL and Direct Loan program regulations to reflect changes made to HEA 1965, by the SAFRA Act included in the Health Care and Education Reconciliation Act of 2010; incorporate statutory changes to interest rates and other recent statutory changes in the Direct Loan Program regulations; update, strengthen, and clarify various areas of the Student Assistance General Provisions, Perkins Loan, FFEL, and Direct Loan program regulations; and provide for greater consistency in the regulations governing the title IV, HEA student loan programs.
2013	1840-AD13	OPE	William D. Ford Federal Direct Loan Program	Amendments to ensure a borrower may not receive Direct Subsidized Loans for more than 150 percent of the published length of the educational program in which the borrower is enrolled.

2014	1840-AD15	OPE	Gainful Employment	Establish measures for determining whether certain postsecondary educational programs prepare students for gainful employment in a recognized occupation, and the conditions under which these educational programs remain eligible under the Federal Student Aid programs authorized under title IV of the HEA.
2015	1840-AD18	OPE	REPAYE	The REPAYE plan is modeled on the existing Pay As You Earn repayment plan, and will be available to all Direct Loan student borrowers regardless of when the borrower took out the loans.
2010	1855-AA06	OII	Investing in Innovation–Priorities, Requirements, Definitions, and Selection Criteria	Supports LEAs and nonprofit organizations to provide competitive grants to applicants with a record of improving student achievement and attainment.
2013	1855-AA09	OII	Investing in Innovation	Establish priorities, requirements, definitions, and selection criteria that will enable effective grant making, resulting in the selection of high-quality applicants who propose to implement activities that are most likely to have a significant national impact on educational reform and improvement.
2015	1855-AA12	OII	Charter Schools Grants to SEAs	Priorities, requirements, definitions, and selection criteria for CSP Grants to SEAs.
2011	1894-AA01	OII	Race to the Top Fund Phase 3	Requirements for Phase 3 of the Race to the Top program. In this phase the Department intends to make awards to States that were finalists but did not receive funding under the Race to the Top Fund Phase 2 competition held in fiscal year 2010.

Notes: All orders ruled ”consistent with change” except * denotes ”consistent without change”. OESE=Office of Elementary and Secondary Education; OPE= Offie of Postsecondary Education; OII=Office of Innovation and Improvement; SEA=state education agency.