ESSAYS ON PUBLIC FINANCE AND DEVELOPMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ali Abbas

August 2021

ESSAYS ON PUBLIC FINANCE AND DEVELOPMENT

Ali Abbas, Ph.D.
Cornell University 2021

This dissertation studies fiscal policy, specifically the efficiency and electoral constraints on tax policy, and the measurement of public sector health to increase the effectiveness of government spending.

Chapter 1 studies high-income taxpayer responses through the tax base channel to changes in marginal income tax rates in the United States. Prior research that has used bunching methods to estimate the taxable income response of high earners has presented no evidence of taxpayer responses at the top kink in the regular, federal income tax schedule. I argue that at the federal level, a combination of the regular and Alternative Minimum Tax schedules identifies the actual tax-related incentives that apply to high earners. I use annual income tax codes and publicly available samples of Internal Revenue Service individual income tax return data from 1993-2011 to characterize the combined schedule for each taxpayer. I discover previously undetected bunching at the top kink in this schedule and use it to estimate the elasticity of taxable income with respect to the net-of-tax rate for high earners to be between 0.15 to 0.28. This implies an upper bound on the efficiency cost of income taxation of 45 cents to a dollar, and a lower bound on the optimal top marginal tax rate of 70 percent, suggesting an optimal rate that is higher than prevailing top rates. I also mitigate an emerging endogeneity concern with bunching estimators that use kink points fixed in taxable income. By using effective top kinks that vary across taxable income for each

taxpayer, I separate variation in marginal tax rates from variation in taxable income, making my bunching estimates more methodologically robust than earlier estimates.

Chapter 2 estimates the impact of tax reforms on citizens' voting behavior. We examine the effect of changing income tax burdens on voting behavior in presidential and House elections across the United States. To do so, we use a novel simulated instrumental variable approach in conjunction with survey, administrative, and voting data for the years 2010 to 2020 to isolate changes in tax burdens that arise purely due to variation in tax policy from changes caused by demographic shifts. We estimate that an increase in tax burdens by about half a standard deviation increases the vote share for the Republican party by one to six percentage points. This relationship is strongest, both statistically and in terms of magnitude, for presidential elections. For House elections, we find suggestive, but not definitive evidence that this relationship holds. Our analysis shows that contrary to popular belief, taxpayers continue to vote in their economic self-interest.

In Chapter 3, I develop a tool for measuring the multidimensional performance of the public sector in the spirit of multidimensional measures of poverty. The framework allows fiscally constrained policymakers to measure a sector's resource base, assess it over time, and optimize spending. The measure's decompositional properties provide for easy identification of the sources of deprivation along various dimensions and across subgroups, such as geographical areas and subsectors. In an application to the public education sector in Sindh province, Pakistan, I show that 27 percent of public schools are multidimensionally deprived and the weakest dimensions are physical infrastructure and facilities. Single-sex, rural schools, where instruction is in the native Sindhi language contribute the most to the overall measurement of sectoral weakness.

BIOGRAPHICAL SKETCH

Ali Abbas spent his early childhood and teenage years in Karachi, Pakistan where he attended the Saint Patrick's High School. While keen on pursuing the natural sciences, the socioeconomic and political turbulence of the early 2000s inspired Ali to pivot to the social sciences at the Lahore University of Management Sciences in Lahore, Pakistan where he completed his undergraduate degree in economics in 2010. Between 2010 and 2016, Ali worked at a diverse set of policy and research institutions including the Pakistan Policy Group, the Brookings Institution, the Center for Economic Research in Pakistan, and the World Bank. From 2012-2014, Ali attended the Master of Public Policy program at the University of Minnesota, Twin Cities as a Fulbright Scholar.

In 2016, Ali enrolled in the doctoral program at the Charles H. Dyson School of Applied Economics and Management at Cornell University. Ali is a member of the American Economic Association, the National Tax Association, the International Institute of Public Finance, and the Mahbub ul Haq Public Finance Cluster at the Lahore University of Management Sciences. Upon completion of his PhD, Ali will join the International Monetary Fund as part of the Economist Program from September 2021.

This dissertation is dedicated to my mother, Gulnaz.

ACKNOWLEDGMENTS

Completing this dissertation would not have been possible without the love, support, and encouragement of my family, friends, and colleagues. There are more people than I can name here who made it possible for me to complete my doctorate. Amongst those who stand out, are my parents, Ghulam Abbas and Gulnaz Abidi who instilled confidence in me through thick and thin; my sister, Fatima, for her strength and patience; and my loving wife and best friend, Maham, who sacrificed much to see me through it all. To them, I am eternally grateful.

I am privileged to have worked with a wonderful set of advisers. I thank Nancy for her valuable guidance, unfettered support, and her unique ability to pinpoint potential pitfalls that others could not see. I thank Mike, for his energy, positivity, and his deep interest in my work. And I am grateful to Ravi, for always encouraging me to take a step back and ask the right questions. The impact of their mentorship goes miles beyond this dissertation.

I would like to thank Linda Sanderson and Bridget Richardson for their encouragement, and their consistent and effective administrative support over the past five years. I am indebted to the faculty at the Charles H. Dyson School of Applied Economics and Management, and the Economics Department at Cornell University for preparing me academically to produce this work. To my peers, thank you for the support and the camaraderie which made Cornell and Ithaca feel like home. Finally, to Raza, I am grateful, for he showed me the way.

TABLE OF CONTENTS

1. Using the Alternative Minimum Tax to Estimate the Elasticity of Taxable Income for High Earners 12
1.1 Introduction 13
1.2 Prior Literature 22
1.3 Conceptual Framework for the Combined Income Tax Schedule 29
1.3.1 Tax Brackets, Marginal Tax Rates, and Taxable Income 31
1.3.2 Interaction of the Regular Income Tax and AMT Schedules 34
1.4 Empirical Methodology 40
1.4.1 Data 40
1.4.2 Solving for the Top Kink in the Combined Schedule 43
1.4.3 Estimation Method 46
1.4.4 Robustness to Endogeneity Concern 49
1.4.5 Parameter Selection and the Functional-Form Assumption 50
1.5 Results 53
1.5.1 Graphical Evidence 53
1.5.2 Elasticity Estimates 60
1.6 Sensitivity to Model Parameters 69
1.7 Efficiency Cost and the Optimal Top Marginal Tax Rate 72
1.8 Conclusion 77

2. Tax Reforms and Voting Behavior in the United States 80
2.1 Introduction 81
2.2 Prior Literature 84
2.3 Data 85
2.4 Empirical Methodology 89
2.5 Key Tax Reforms 95
2.5.1 American Taxpayer Relief Act of 2012 95
2.5.2 Tax Cuts and Jobs Act of 2017 96
2.6 Results 97
2.6.1 Presidential Elections 97
2.6.2 House Elections 105
2.7 Concluding Remarks 116

3. Multidimensional Measurement of Sectoral Performance: Evidence from Public Schools in Pakistan 118
3.1 Introduction 119
3.2 Methodology 122

3.2.1 Construction of the Multidimensional Sectoral Deprivation Index (MSDI) 122
3.2.2 Decomposability ... 125
3.2.3 Comparison to Alternative Approaches ... 128
3.3 Application: Public Education in Sindh, Pakistan ... 130
3.3.1 Overview of Pakistan's Education Sector Performance ... 130
3.3.2 Data ... 134
3.3.3 Choice of Dimensions: Public Education in Sindh ... 135
3.3.4 Choice of Weights and Cutoffs ... 141
3.4 Discussion of Results ... 145
3.4.1 Descriptive Statistics ... 145
3.4.2 Preliminary Results ... 148
3.5 Sensitivity Analysis ... 159
3.6 Conclusion ... 162
Bibliography ... 165
Appendix ... 172

LIST OF FIGURES

Figure 1: Historical Estimates of the ETI for High-Income Taxpayers 26
Figure 2: The Regular Income Tax and the AMT Schedules 37
Figure 3: The Combined Schedule 38
Figure 4: Distribution of the Intersection Kink Relative to Regular Taxable Income 39
Figure 5: Relationship between Deductions and the Intersecton Kink 44
Figure 6: Graphical Evidence of Bunching 55
Figure 7: Bunching Responses around the Top Kinks in the Combined and Regular Schedules 56
Figure 8: Bunching Responses of Wage Earners and the Self-Employed 58
Figure 9: Bunching Responses Related to Long-Term Capital Gains 60
Figure 10: Distribution of Observed Versus Counterfactual Taxable Incomes 61
Figure 11: Distribution of Observed versus Counterfactual Taxable Income (Disaggregated) 64
Figure 12: Elasticity Estimates for High Earners by Population Type 68
Figure 13: Sensitivity of Elasticity Estimates to Binwidth 70
Figure 14: A Growing Education Gap (Source: Pew Research Center | Jacob Turcotte/Staff 101
Figure 15: Scatter Plot of Republican Vote Share and Simulated Liability Per Capita 102
Figure 16: Relationship between College Share and Simulated Liability Per Capita 103
Figure 17: Residual Scatter Plot Republican Vote Share versus Simulated Liability Per Capita 103
Figure 18: Data Envelopment Analysis versus Stochastic Frontier Analysis (Source: Aparicio et al., 2014) 129
Figure 19: Sindh Province with its Districts 132
Figure 20: Division-wise Contribution to MSDI 150
Figure 21: District-wise Contribution to MSDI (by Division) 153
Figure 22: Other Subgroup-wise Contribution to MSDI 155
Figure 23: Contribution to MSDI per Indicator 157
Figure 24: Dimensional Contributions to MSDI (by Division) 158
Figure 25: Sensitivity Analysis 161

LIST OF TABLES

Table 1: Fraction of Taxpayers Facing the Combined AMT-Regular Tax Schedule 30
Table 2: Taxable Income Brackets and Marginal Tax Rates in Year 2000 32
Table 3: Summary of Elasticity Estimates for High Earners 67
Table 4: Sensitivity of Elasticity Estimates to Bandwidth 72
Table 5: Summary Statistics for Demographic Controls 89
Table 6: Presidential – Cross-Sectional (2012 & 2016) 98
Table 7: Presidential – FD OLS (2012-2016) 99
Table 8: Presidential – FD Reduced-Form (2012-2016) 99
Table 9: Presidential – FD First Stage (2012-2016) 100
Table 10: Presidential – FD IV (2012-2016) 101
Table 11: Presidential – FD Reduced-Form (2016-2020) 104
Table 12: Presidential – FD Reduced-Form (2012-2020) 105
Table 13: House – FD OLS (2012-2016) 106
Table 14: House – FD Reduced-Form (2012-2016) 107
Table 15: House – FD IV (2012-2016) 107
Table 16: House – FD Reduced-Form (2016-2020) 108
Table 17: House – FD Reduced-Form (2012-2014) 109
Table 18: House – FD Reduced-Form (2016-2018) 110
Table 19: House – FD First Stage (2012-2014) 111
Table 20: House – FD First Stage (2016-2018) 111
Table 21: House – FD IV (2012-2014) 112
Table 22: House – FD IV (2016-2018) 112
Table 23: House – FD OLS (2010-2018) 113
Table 24: House – FD Reduced-Form (2010-2018) 114
Table 25: House – FD First Stage (2010-2018) 115
Table 26: House – FD IV (2010-2018) 115
Table 27: Indicators and their impacts on key student outcomes 137
Table 28: Description of Dimensions, Indicators, and Cutoffs 140
Table 29: Description of Dimensions, Indicators, and Cutoffs 143
Table 30: Summary Statistics 146
Table 31: Division-wise Decomposition Statistics 149
Table 32: District-wise Decomposition Statistics 151
Table 33: Other Subgroup-wise Decomposition Statistics 155
Table 34: Dimensional Contribution to MSDI (by Division) 158

1. Using the Alternative Minimum Tax to Estimate the Elasticity of Taxable Income for High Earners*

Abstract

Personal Income tax revenue in the United States draws heavily from high-income taxpayers. How high earners respond to tax changes has repercussions for tax revenue, the efficiency costs of taxation and the optimal progressivity of the tax schedule. Prior research that uses bunching methods to estimate the taxable income response of taxpayers has presented no evidence of high-income bunching at the top kink in the regular, federal income tax schedule. I argue that the regular schedule does not identify the actual tax-related incentives that apply to high-income individuals. At the federal level, high earners are subject to a combination of the regular income tax and the Alternative Minimum Tax. I use annual tax codes and publicly available samples of Internal Revenue Service individual income tax return data from 1993-2011 to characterize the combined schedule for each taxpayer. I discover previously undetected bunching at the top kink in this schedule and use it to estimate the elasticity of taxable income with respect to the net-of-tax rate for high earners to be between 0.15 to 0.28. This estimate implies a lower bound on the optimal top marginal tax rate of 70 percent, suggesting an optimal rate that is higher than prevailing top rates. I also use this setting to make a unique methodological contribution: I show that the location of the top kink in the

* I am grateful to Michael Lovenheim, Nancy Chau, Ravi Kanbur, Ariel Ortiz-Bobea, Zhuan Pei, Douglas Miller, Nathan Seegert, Anil Kumar, and Nicole Bosch for their comments. This research has also benefited from comments received at seminars at the Economics Department and the Dyson School of Applied Economics and Management at Cornell University. A special thanks to participants at the National Tax Association Annual Conference 2019 and the International Institute of Public Finance Annual Congress 2020 for their valuable feedback. I am grateful to Daniel Feenberg at the NBER for his guidance on the use of the income tax microdata.

combined schedule for each taxpayer varies across the distribution of taxable income. This generates novel variation in marginal tax rates that is separable from variation in taxable income, allowing me to mitigate a key endogeneity concern associated with the use of bunching estimators on fixed kink points.

1.1 Introduction

Personal income taxation is a key source of revenue for financing public goods and redistributive schemes. However, non-lump sum personal income taxes alter the after-tax price of labor, incentivizing individuals to change their labor supply (Pencavel, 1986; Hausman, 1991; MaCurdy et al., 1991), shelter earned income from taxation by consuming more tax-deductible items such as healthcare and housing (Glaeser and Shapiro, 2003), or illegally under-declare income (IRS, 2016). These responses can generate deadweight loss in the economy and cause tax revenue loss if the size of the taxable economy shrinks, making the marginal tax rate structure a highly debated policy and political issue. These debates rest in large part on how high-income individuals respond to changes in the marginal tax rate in the top income bracket.

I focus on high income taxpayers because of three reasons. First, the top quintile (percentile) of income earners by households in the United States contribute approximately 88 percent (38 percent) of personal income tax revenue (Tax Policy Center, 2019), so taxable income responses in this group can have substantial revenue consequences.

Second, the response to tax changes in the right-tail of the income distribution can itself be higher relative to the rest of the distribution, given high earners' access to diverse financial strategies including income-shifting across tax bases, retiming of income realization, and the increased use of itemized deductions such as home mortgage and business expense allowances (Saez, Slemrod and Giertz, 2012). For example, part of the income of high earners such as executives could be in the form of stock options, which face lower marginal tax rates on the capital gains schedule as compared to the top marginal tax rate on the income tax schedule (Hanlon et al., 2005). Taxpayers can also retime capital gains realizations, as documented by Goolsbee (2000b), Parcell (1995) and Samartino and Weiner (1997). Increased bargaining power of these taxpayers such as top executives can also allow them to substitute taxable income with non-taxable fringe benefits at work, such as improved work facilities and better healthcare benefits (Piketty et al., 2014). Top earners also have access to sophisticated tax planning services, and self-employment income that is not reported by third parties creates space for tax evasion (Slemrod, 2007; Hurst et al., 2010).

Third, the magnitude of the ETI parameter for high earners is highly contested in the public finance literature. However, it is the hypothesized high responsiveness of top earners and the sensitivity of revenues to the high-income tax base that served as a factor in the Reagan tax cuts of 1981 that reduced the top marginal tax rate from 70 percent to 50 percent; and again, in 1986 when the top rate was decreased to 28 percent. Contested views on the responsiveness of the high-income tax base continue to pervade the policy and political discourse.

Measuring taxpayer responses along the labor, avoidance, and evasion margins separately is infeasible due to the inability to observe all the dimensions of behavior. Instead, Feldstein's (1999) canonical model shows that all such margins of taxpayer responses that affect taxable income and generate deadweight burden are captured by the elasticity of taxable income (ETI) with respect to the net-of-tax rate.[1] This makes the ETI a sufficient statistic for estimating efficiency costs of income taxation and conducting welfare analyses[2], assuming no transfer costs[3] of sheltering (Chetty, 2009) and no fiscal externalities[4] (Slemrod, 1998; Saez, 2004). This makes the ETI a core parameter in the public economics literature. Previous work on estimating the magnitude of high-income taxpayer responsiveness has generated mixed results. For example, Feldstein (1995) estimates the ETI for high earners to be as high as 1.7, while others studying bunching behavior around the top kink in the regular income tax schedule have found no response (Saez, 2010; Mortenson & Whitten, 2016).

In this paper, I employ a bunching estimator to study the responsiveness of top earners in the United States to changes in marginal tax rates by using the intersection of the Alternative Minimum Tax (AMT) and the regular income tax schedules. To date, no prior research has studied the combined schedule, especially for estimating the ETI for high earners.

Existing literature on estimating the ETI for high earners in the United States predominantly uses two approaches. The first approach uses taxable income responses

---

[1] The net-of-tax rate is the post-tax, take-home portion of the marginal dollar earned by a taxpayer.
[2] This is because in the canonical model, the marginal private value of sheltering an additional dollar of income and the marginal social value of earning an additional dollar of income are both pegged to the tax rate.
[3] Chetty (2009) shows that transfers costs to taxpayers of avoiding or evading taxes can be offset by a positive externality on other agents. For example, penalties paid to the government due to tax evasion are redistributed; and an increase in deductible charitable contributions generates positive externalities for other agents in the economy.
[4] Tax revenue losses due to income shifted from one stream can be partially offset by taxation in another stream.

related to tax reforms that change top marginal tax rates to estimate the ETI. However, rising inequality that differentially affects secular growth rates in different parts of the taxable income distribution presents a challenge, since it becomes difficult to disentangle the effect of secular income growth on taxable income from the effect of tax rates. Estimates vary significantly, in the range of 0 to 1.7. Initial estimates tended to be high (Lindsey, 1986; Feldstein, 1995). More recent studies that have attempted to isolate variation in taxable income from secular income growth have generated lower estimates. These issues and the relevant literature are discussed in more detail in Section 1.2.

The second approach uses bunching methods on cross-sectional data to avoid identification issues created by secular income growth. The use of this approach in the United States involves estimating the magnitude of bunching at kinks in the regular, federal income tax schedule. This bunching is presumably a result of taxpayers strategically locating on the side of the kink that offers the lower marginal tax rate. Excess bunching is then compared to the tax rate differential around the kink point to estimate the ETI. This approach has revealed no high-income responses around the top kink in the regular, federal income tax schedule (Saez, 2010; Mortenson and Whitten, 2016). On the other hand, the estimated ETI for low-income individuals in these studies is higher, in the range of 0.1 - 0.3, raising the question of why bunching estimators have failed to show evidence of economically significant elasticities for high earners who have more margins along which they can respond.

I argue that the federal, regular income tax schedule used by previous bunching studies does not identify the actual, tax-related incentives that apply to high-income individuals. At the federal level, high earners respond to a combination of the regular income tax and the AMT. The AMT is a concomitant income tax schedule with its own definition of taxable income and marginal tax rates. The purpose of the AMT is to ensure that high-income taxpayers do not take disproportionate advantage of deductions – which reduce taxable income – offered by the regular income tax schedule. The AMT disallows major deductions such as personal exemptions, the standard deduction, and important itemized deductions such as the state and local tax (SALT) deduction, and miscellaneous deductions used primarily by business owners.[5] By redefining taxable income, the AMT causes a larger part of earned income to be counted as taxable that is otherwise sheltered from taxation on the regular income tax schedule. However, the AMT provides a substantial fixed deduction that prevents low- to middle-income taxpayers from being affected by it.

Taxpayers separately calculate their income tax liabilities on the regular income tax and the AMT schedules and are liable for the higher of the two taxes. The effective schedule is, therefore, the upper envelope of the interaction of the two schedules. The intersection kink – where the two schedules cross – is the top kink in the combined schedule. I find that between 1993-2011, less than 0.5 percent of taxpayers with real 2007 adjusted gross income (AGI) of less than $100,000 were subject to the combined AMT-regular schedule. For real AGIs between $100,000 to $200,000, this rate rises to approximately 3 percent. Amongst taxpayers with real AGI above $300,000, more than 65 percent were subject to the combined schedule, implying a large proportion of high earners for whom the correct schedule to analyze is the combined, rather than just the regular income tax schedule.

[5] The AMT also partially disallows medical and dental deductions, accelerated depreciation, and deductions on home mortgage interest on non-primary property, among others.

Studying taxpayer behavior around the top kink in the regular income tax schedule in isolation can reveal low ETI estimates for two reasons. First, the top kink in the combined schedule does not systematically align with the top kink in the regular schedule. Studying bunching behavior only around the latter will introduce measurement error and bias estimates of the ETI downward. The top kink in the regular schedule does not affect taxpayers who are subject to the combined schedule. For these taxpayers, strategic decision-making occurs around the top kink in the combined schedule. Second, the difference in marginal tax rates on the two sides of the top kink in the regular schedule can be too small to elicit a substantial bunching response, even in the absence of the combined schedule. Larger tax rate differentials around kinks create stronger incentives for taxpayers to bunch on the side of the kink point that offers a lower tax rate (Chetty et al. 2011). Such differentials exist on the combined schedule. The marginal tax rate changes from 28 percent to the left of the top kink on the combined schedule to approximately 38 percent to its right, as compared to the approximately 36 to 39 percent (33 to 35 percent) change across the top kink in the regular income tax schedule between 1993-2002 (2003-2011).

Using annual federal income tax codes and publicly available Internal Revenue Service (IRS) income tax return data from 1993-2011, I construct the two piecewise linear functions associated with the regular income tax and the AMT schedules for each taxpayer in each tax year, adjusting for taxpayer-level deductions. Across years, the shape of the two tax functions is determined by legislative rules related to the size of income tax brackets and corresponding marginal tax rates. Within years and across individuals, the location of the intersection kink is determined by the amount of deductions allowed by the regular income tax relative to the AMT. Once constructed, I solve these two tax functions for each taxpayer in my sample to find the complete set of intersection kinks. The intersection kink for these taxpayers lies on average, at $430,200 for time period 1993-2002, and at $679,307 for time period 2003-2011.

Since the location of the intersection kink varies for each taxpayer, I recenter these kink points and overlay the observed distribution of taxable income to provide visual evidence of the aggregate bunching response of high earners. I estimate this excess mass as compared to an estimated counterfactual density[6] that is a fitted polynomial of the seventh order and use it in a standard bunching estimator to measure the ETI for high earners. I also test the robustness of my estimates by using weaker assumptions for the functional form of the counterfactual density. Earlier bunching studies that have estimated the ETI for high earners have made stronger functional-form assumptions. For example, the counterfactual density is assumed to be linear in Saez (2010) and a polynomial of order seven in Chetty et al. (2011). I use the method proposed in Bertanha et al. (2020) to estimate non-parametric bounds on the ETI, by using the area of the observed distribution as a constraint on the counterfactual density to restrict its range of slopes in the bunching region.

The location of the top kink in the combined schedule varies across the distribution of taxable income for each taxpayer, providing novel variation in marginal tax rates that is separable from variation in taxable income. This feature allows me to mitigate an important endogeneity concern associated with the use of bunching estimators on fixed kink points (Blomquist and Newey, 2017; Bertanha et al., 2020).

[6] The counterfactual density is the underlying distribution if there was no kink point and therefore, no differential taxpayer response to changing tax rates.

I estimate the average ETI for high earners to be 0.15. This estimate is bounded below at 0.12 and above at 0.17. The estimated ETI for high earners rises to 0.20 for taxpayers who are unaffected by the additional complexity of the capital gains schedule. I also generate the estimates by time-period. The Jobs and Growth Tax Relief Reconciliation Act (JGTRRA) of 2003 was followed by annual increases in the AMT fixed deduction amount, pushing the intersection kink to higher income levels where taxpayers are plausibly more responsive to changes in marginal tax rates. For these higher-income taxpayers between 2003-2011 who unaffected by the additional complexity of the capital gains schedule, the estimated ETI is 0.28. High earners' responsiveness to marginal tax rates increases over time, with taxpayers who report any self-employment income responding more than others. I apply simplified formulas in the literature that use the estimated ETI parameter in conjunction with marginal tax rates and the shape of the income distribution to estimate efficiency costs and optimal top marginal tax rates, as discussed in Section 1.7. Intuitively, higher taxpayer responsiveness generates larger distortions in the economy leading to higher efficiency costs and lower optimal top marginal tax rates.

My estimates for the average ETI for high earners of 0.15 - 0.28 imply an efficiency cost ranging from 22 cents to 45 cents per dollar of additional tax revenue collected. The estimated optimal top marginal tax rate lies between 70 percent and 82 percent. In the presence of transfer costs and fiscal externalities, these estimates serve as upper bounds on efficiency costs and lower bounds on optimal top marginal tax rates (Chetty, 2009).

I make three key contributions to the literature. First, I account for the interaction of the regular income tax and AMT schedules in the United States to provide evidence of substantial bunching around the top kink in the combined schedule, resulting in elasticities of 0.15 to 0.28. In contrast, earlier studies show no response at the top kink in the regular income tax schedule (Saez, 2010; Mortenson and Whitten, 2016). To date, no prior study has studied the combined schedule for estimating the ETI for high earners.

The second contribution that I make is methodological. I provide a unique setting that mitigates recent endogeneity concerns related to the use of bunching methods on kink points fixed in taxable income (Blomquist and Newey, 2017; Bertanha et al., 2016, 2020). Since fixed kinks at which marginal tax rates change are jointly determined with taxable income, observed taxable income is likely correlated with unobserved heterogeneity. Intuitively, it is plausible that individuals select into particular bins of the income distribution not as a result of strategic responses to marginal tax rates but because of some underlying preferences for those income levels. If this occurs, then observed bunching (or troughs) in the taxable income distribution might reflect preferences rather than strategic decision-making related to tax rates, causing bias in the estimation of the ETI of unknown direction. However, in the setting that I leverage, the top kink in the combined schedule varies for each taxpayer across taxable income generating a distribution of top kinks, as illustrated in Section 1.3.2. This unique feature of the combined schedule weakens the correlation between taxable income and unobserved heterogeneity, increasing confidence in the ability of my estimator to estimate an unbiased ETI parameter.

Third, I contribute to the small literature on the AMT by providing the only estimates on taxpayers' responses to the AMT. Previous literature in this area has specifically focused on forecasting the coverage and revenue impact of evolving AMT laws of the early 2000s

(Burman et al., 2003), its impact on average marginal tax rates (Feenberg and Poterba, 2003), and the role of the AMT as a fiscal stabilizer (Galle and Klick, 2011). However, the AMT has not been leveraged to assess taxpayer behavior and its impact on efficiency and the optimal schedule.

From a policy perspective, my results point to optimal top marginal tax rates that are higher than prevailing rates. The higher ETI for self-employed individuals confirms the previously documented relationship between the absence of third-party reporting and higher tax avoidance behavior. And a comparison of the relationship between the size of the marginal tax rate change around kinks and bunching responses suggests that a larger number of income tax brackets with smaller marginal tax rate changes across brackets will reduce taxable income responses, leading to lower efficiency costs of taxation.

1.2 Prior Literature

I contribute to the literature on estimating the elasticity of taxable income (ETI) with respect to the net-of-tax rate for high earners. The ETI measures the taxable income response of taxpayers to changes in marginal tax rates. As discussed in Section 1.1, this parameter can be a sufficient statistic for estimating efficiency costs and optimal top marginal tax rates (Feldstein, 1999) under no transfer costs (Chetty, 2009) and no fiscal externalities (Slemrod, 1998; Saez, 2004). In particular, the large share of tax revenue generated by high-income taxpayers and their greater hypothesized ability to respond to changes in tax rates makes studying the ETI of high earners extremely important. I use a bunching estimator and leverage the conjoined nature of the Alternative Minimum Tax (AMT) and the regular, federal income tax schedule to estimate the ETI of high earners to be approximately 0.15 to 0.28.

The core challenge with estimating the ETI is related to the endogeneity of tax rates, since taxable income and tax rates are jointly determined. As taxable income rises, the marginal tax rate that the taxable income is subject to increases under a nonlinear schedule. This makes it difficult to disentangle variation in marginal tax rates from variation in taxable income. Prior literature has predominantly used two methods to address this endogeneity concern. The first approach to estimating the ETI for high earners leverages tax reforms that introduce plausibly exogeneous changes in marginal tax rates. The major, federal tax reforms that have been studied in the literature include the Economic Recovery Tax Act (ERTA) of 1981, the Tax Reform Act (TRA) of 1986, the Omnibus Reconciliation Acts (OBRA) of 1990 and 1993 and the American Taxpayer Relief Act (ATRA) of 2012.[7] I compare the estimates from some of the seminal studies using tax reforms in Figure 1.[8] Panel A sorts these estimates by publication year of the study and shows that studies using tax reforms have found a wide array of estimates, ranging from 0 to 1.7, with more recent studies finding lower estimates of the ETI for high-income taxpayers. Panel B sorts the studies by the median year of analysis considered in each study. It shows how the ETI estimates related to tax reforms in the 1980s were higher than those that were introduced later. It is possible that structural ETI was higher in the 1980s due to features of tax audit system or the specific aspects of tax reforms in this time-period, or that earlier studies did not sufficiently account for the endogeneity of

[7] ERTA 1981 and TRA 1986 reduced the top marginal tax rate from 70 percent to 50 percent, and from 50 percent to 28 percent, respectively. OBRA 1990 and 1993 increased the top marginal tax rate from 28 percent to 31 percent, from 31 percent to 39.6 percent. ARTA 2012 increased the tax rate from 35 percent to 39 percent.

[8] If a study has multiple estimates for the ETI of high earners, I average the estimated ETI.

marginal tax rates. Details on the studies represented in Figure 1 are provided in Table A.1 of Appendix A.

Initial estimates using the tax reforms approach tend to be high (Lindsey, 1986; Feldstein, 1995). Lindsey (1986) and Feldstein's (1995) identification strategies rely on secular growth rates of real income being the same for the groups being compared. If these growth rates vary across groups due to non-tax related reasons, then taxable incomes of taxpayers in high-income groups would be different from taxpayers in low-income groups across time, even in the absence of tax changes. This differential income growth is well documented. Saez and Zucman (2020) find that between 1980-2018, the national income share of the top one percent grew by 2 percent per year, compared to an annual, average growth rate of 0.2 percent for the bottom 50 percent of the income distribution. The higher secular growth in the income share of high-income groups would bias the estimate of the ETI for high earners in Lindsey (1987) and Feldstein (1995) upwards, plausibly accounting for the high estimates found in these studies. To deal with the issue of secular income growth, most tax reform studies conducted after 1995 controlled for time trends and exploited instrumental variables to disentangle variation in tax rates from variation in taxable income, producing smaller estimates (Gruber and Saez, 2002; Saez, 2003).[9] As shown in Panel A of Figure 1, the estimated ETI of high-income taxpayers is lower, in the range of 0 to 1 from year 1997 onwards.

The second approach to estimating the ETI attempts to avoid the identification issues caused by differential secular income growth rates by using cross-sectional income tax data

[9] For a detailed discussion of other identification issues related to studies using tax reforms, review Saez, Slemrod and Giertz (2012).

and employing bunching methods. This approach involves overlaying the observed taxable income distribution across a stable, income tax schedule. Observed bunching in this distribution around kinks in the tax schedule plausibly reflects strategic taxable income responses of taxpayers, with taxpayers locating on the side of the kink where the marginal tax rate is lower. The excess mass in the distribution captures this strategic response and is compared to the magnitude of the change in marginal tax rates at the kink to estimate the ETI. However, while studies using tax reforms have found a wide range of estimates, bunching methods have found no taxable income response at the top kinks of the income tax schedule in the United States (Saez 2010; Mortenson and Whitten, 2016). In Panel A of Figure 1, I compare estimates for the ETI of high earners in the US with estimates from bunching studies conducted using Danish tax data. It is notable that unlike in the US, bunching estimates are non-zero for Danish data and in the range of 0.1 to 0.3. In fact, recent estimates for the ETI of high earners in China, not included in the figure, stand at 0.41 (He et al., 2018).

Panel A

Panel B

Figure 1: Historical Estimates of the ETI for High-Income Taxpayers

Notes: Panel A illustrates estimates of the ETI of high-income taxpayers in the prior literature, sorted by Publication year. Panel B contains the same estimates sorted by the median year of the analysis sample used by each study. Studies are divided into four types: tax reforms (non-bunching), bunching studies in the US, bunching studies in Denmark, and estimates obtained in this study.

There are two potential reasons for the difference between estimates of the observed ETI for high earners in the United States and in other countries. First, it is possible that high-income taxpayers in the United States simply do not respond to the top kink in the income tax schedule, as compared to their global counterparts, due to reasons including a lack of salience of the top kink and low structural elasticities. It is also plausible that earlier studies in the US do not consider relevant features of the income tax code when measuring taxpayer bunching responses, resulting in measurement error that introduces downward bias in these estimates. In this paper, I argue that the federal, regular income tax schedule used by previous bunching studies is insufficient to map the effective tax schedule that applies to high-income individuals. At the federal level, high earners are likely subject to both the regular income tax and the Alternative Minimum Tax (AMT) schedules. The AMT has its own marginal tax rates and allowable deductions. Taxpayers separately calculate their income tax liabilities on the regular income tax and the AMT schedules and are liable for the higher of the two taxes. The effective schedule, therefore, is the upper envelope of the interaction of the two individual schedules. I discuss the structure of the combined schedule in more detail in Section 1.3.

By considering taxpayer behavior along the combined schedule, I provide evidence of previously undetected bunching at the top kink of the combined schedule, in contrast to studies using the top kink in the regular income tax schedule, resulting in higher estimates of the ETI for high-income taxpayers as shown in Figure 1. My estimates of 0.15 to 0.28 are more in line with bunching studies conducted in other countries. To the best of my knowledge, this is the first paper to study taxpayer responses to kinks in the combined schedule, specifically in relation to the AMT. Earlier literature has forecasted the coverage and revenue impact of evolving AMT laws of the early 2000s (Burman et al., 2003), assessed the AMT's impact on average marginal tax rates (Feenberg and Poterba, 2003), and studied the role of the AMT as a fiscal stabilizer (Galle and Klick, 2011). However, the AMT has not been leveraged to assess taxpayer behavior and its impact on efficiency costs of taxation.

I also provide a unique setting that mitigates endogeneity concerns related to the use of bunching methods on kink points fixed in taxable income. By providing a unique setting where the location of the top kink in the personal income tax schedule varies across taxpayers, I am able to disentangle variation in marginal tax rates from variation in taxable income to better address endogeneity concerns related to traditional bunching estimators. Earlier bunching studies use tax schedule kinks that are fixed in taxable income in a given tax year. For a single budget set, variation in tax rates across the budget set occurs with variation in taxable income as well as with variation in preferences. The correlation of taxable income and underlying preferences makes it challenging to distinguish the taxable income elasticity from unobserved heterogeneity (Blomquist and Newey, 2017; Bertanha et al., 2016). Intuitively, it is impossible to know if an individual chooses to locate at a kink because of tax rate variation or due to underlying preferences. The variation in the location of the top kink in the combined schedule across high-income taxpayers however, generates multiple budget sets, limiting exposure to such selection bias by delinking variation in marginal tax rates from variation in taxable income.

The next section provides a detailed overview of the AMT focusing on its features that interact with the regular income tax schedule to give rise to the combined, effective personal income tax schedule at the federal level in the United States.

1.3 Conceptual Framework for the Combined Income Tax Schedule

In this section, I assess the coverage of the combined income tax schedule and unpack specific features of the regular income tax and AMT schedules that give rise to the top, intersection kink in the combined schedule. I show how the location of the top kink in the combined schedule is misaligned with the top kink in the regular income tax schedule, potentially creating a downward bias in earlier estimates of the ETI for high-income taxpayers that only looked at the top kink in the regular schedule. I also discuss how the variation in the location of the top kink in the combined schedule can be used to address endogeneity concerns associated with the earlier use of bunching estimators on fixed kink points and provide evidence for this variation.

The AMT reduces the ability of high-income taxpayers to shelter income from taxation with the use of deductions. The AMT and the regular income tax schedules function in parallel to each other. Taxpayers calculate their income tax liability using both the regular income tax form (Form 1040) as well as the AMT form (Form 6251). Once taxpayers have calculated personal income tax liabilities based on both schedules, they are liable to pay the higher of the two amounts represented by the upper envelope of the interaction of the two schedules.

The number of taxpayers who are subject to the upper envelope of the combined AMT-regular tax schedule increases at higher income levels. For example, I find that approximately 0.03 percent of the population of taxpayers with real adjusted gross income (AGI) less than $50,000 face the combined AMT-regular schedule. On the other hand, 47 percent (60 percent) of taxpayers with real AGI greater than $200,000 ($300,000) face the

AMT. Table 1 provides details on the fraction of taxpayers who are subject to the combined schedule by real 2007 AGI brackets. The table also provides this breakdown for the subpopulation that submitted Form 6251, the form used to report AMT liability. Since taxpayers submitting this form already expect to be subject to the combined schedule, the fraction of taxpayers who are subject to the AMT, conditional on submitting Form 6251 is close to 100 percent at high income levels.

Table 1: Fraction of Taxpayers Facing the Combined AMT-Regular Tax Schedule

Real (2007) AGI Brackets in '000s of $	**% of taxpayers facing combined schedule**	**% of taxpayers facing combined schedule, conditional on submitting Form 6251**
less than 50	0.03	1.64
50 to 100	0.35	8.36
100 to 200	3.11	23.56
200 to 300	33.09	62.63
300 to 400	63.05	94.97
400 to 500	68.11	97.58
more than 500	57.2	93.5

Studying the taxable income response of high earners around the top kink in the regular income tax schedule without accounting for the AMT and the presence of the combined tax schedule can affect bunching-based estimates of their ETI through two channels. First, the top kink in the combined schedule does not systematically align with the top kink in the regular income tax schedule. Studying bunching behavior only around the latter will introduce measurement error and bias estimates of the ETI downward, because the top kink in the regular schedule does not affect taxpayers who are subject to the combined schedule. For these taxpayers, strategic decision-making occurs around the top kink in the combined schedule.

Second, taxpayers on the margin are incentivized to locate on the side of the kink offering the lower marginal tax rate. In fact, as shown by Chetty et al. (2011), the utility loss associated with larger changes in marginal tax rates can justify higher adjustment costs for taxpayers to relocate on the side of the kink offering the lower marginal tax rate. The top kink on the combined schedule provides a more substantial jump in marginal tax rates relative to the regular schedule at high income levels. Specifically, the marginal tax rate at the top kink in the combined schedule increases from 28 percent to approximately 39 percent (35 percent) between 1993-2002 (2003-2011). Compare this to changes at the top kink in the regular schedule, where the marginal tax rate increases from 36 percent to approximately 39 percent between 1993-2002, and from 33 percent to 35 percent between 2002-2011. The top kink in the combined schedule therefore, becomes a valuable device for assessing taxpayer responsiveness to changing marginal tax rates. Below, I discuss the features of the regular schedule, the AMT schedule, and their interaction that results in misalignment of kinks between the regular and combined schedules. I also detail the larger changes in marginal tax rates at the top kink in the combined schedule relative to the regular schedule.

1.3.1 Tax Brackets, Marginal Tax Rates, and Taxable Income

The AMT differs from the federal, regular income tax schedule in three distinct ways that are related to taxable income brackets, marginal tax rates, and the definition of taxable income.[10] First, the regular income tax and AMT schedules contain taxable income brackets of different sizes. The regular income tax schedule had five brackets between 1993-2001, and then six

[10] The detailed legislative history of the AMT is provided in Table A.2 of Appendix A.

brackets between 2002-2012. In contrast, the AMT schedule contains two statutory taxable income brackets. However, a fixed deduction provided by the AMT is phased out at high income levels, causing the AMT to have four distinct effective taxable income brackets. Second, both schedules exhibit different marginal tax rates corresponding to each taxable income bracket. As an example, for tax year 2000, the differences in taxable income brackets and corresponding marginal tax rates for married joint filers are provided in Table 2.

Table 2: Taxable Income Brackets and Marginal Tax Rates in Year 2000

Regular Taxable Income (MFJ)	**Tax Rates**
$0 - $43,850	15%
$43,850 - $105,950	28%
$105,950 - $161,450	31%
$161,450 - $288,350	36%
$288,350 and above	39.6%

AMT Taxable Income (MFJ)	**Tax Rates**
$0 - $105,000	26%
$105,000 - $161,000	32.5%
$161,000 - $285,000	35%
$285,000 and above	28%

From 1993-2001, the marginal tax rates in the regular income tax schedule increase from 15 percent in the lowest bracket to 39.6 percent in the highest bracket. From 2002 to 2011, marginal tax rates range from 10 percent in the lowest bracket to 35 percent in the highest bracket. In comparison, The AMT has a non-graduated schedule in terms of effective marginal tax rates. In 1993, the Omnibus Budget Reconciliation Act (OBRA) altered the AMT schedule by eliminating a flat marginal tax rate of 24 percent and introducing a two-tiered schedule, with statutory tax rates of 26 percent and 28 percent. The Act also provided a fixed AMT deduction of $45,000 to married joint filers and $33,750 to single filers. These exemption amounts are phased-out at higher taxable income levels. For example, in year 2000, this phaseout begins at $105,000 for married joint filers and $78,750 for single filers. In the phaseout range, every additional dollar of taxable income reduces the fixed deduction by 25 cents leading to effective marginal tax rates that are 1.25 times the statutory marginal tax rates. The fixed deduction completely phases out at taxable income of $285,000 ($213,750) for married joint filers (single filers) in year 2000, creating an effective AMT schedule consisting of four distinct marginal tax rates: 26 percent, 32.5 percent at the point where the exemption phaseout begins, 35 percent where the 28 percent statutory rate begins and exemption phaseout continues, and 28 percent, the point where the fixed deduction is completely phased out.

Third, taxable income is defined differently on the two schedules. The regular income tax schedule does not tax all earned income. Instead, it allows taxpayers to subtract certain deductible consumption and excludable income items from their total earned income for taxation purposes. The residual income forms the tax base on which prevailing tax rates are applied. While a discussion of all the exemptions is beyond the scope of this paper, some of the excluded income items include portions of retirement income, certain types of scholarship income, interest gained from municipal bonds and charitable donations received. As compared to excluded income, deductible consumption expenses that favor certain uses of a taxpayer's income include charitable contributions, state and local taxes paid, real estate taxes paid, interest paid on home mortgage, medical expenses, business expenses and miscellaneous expenditure. High-income taxpayers disproportionately use these excludable income and deductible consumption items that are subject to favorable tax treatment. For example, in fiscal year 2010, taxpayers with incomes below $50,000 used 8.8 percent of all medical deductions, 1.4 percent of all state and local tax deductions, and 2.8 percent of mortgage interest deductions. Compare these utilization rates to those of taxpayers with incomes above $100,000, for whom the shares of these deduction amounts were 49.3 percent, 85.6 percent, and 78.3 percent, respectively.[11] The regular income tax code also provides a fixed standard deduction that can be used by taxpayers for whom the above deduction amounts are less than the standard deduction. Prior to the Tax Cuts and Jobs Act (TCJA) 2017, the regular tax schedule also allowed for personal exemptions for each member of the family.

On the other hand, the AMT disallows major deductions such as personal exemptions, the standard deduction, and important itemized deductions such as the state and local tax (SALT) deduction, and miscellaneous deductions used primarily by business owners.[12] By redefining taxable income, the AMT causes a larger part of earned income to be counted as taxable that is otherwise sheltered from taxation on the regular schedule. However, the AMT provides a substantial fixed deduction that keeps low- to middle-income taxpayers out of the AMT.

1.3.2 Interaction of the Regular Income Tax and AMT Schedules

In Panel A of Figure 2, I illustrate the regular income tax schedule using tax rules prevailing in year 2000. Marginal tax rates increase at each kink in the schedule, represented by the change in slope at the kink points. For example, the marginal tax rate in the lowest taxable income bracket is 15 percent, while the marginal tax rate in the highest bracket is 39.6 percent. The length of each interval between kink points depends on the size of income tax brackets. Marginal tax rates and the length of income tax brackets are fully contingent on the prevailing tax law, and common to all taxpayers. In contrast, the starting point of the tax

[11] Estimates computed using the Joint Committee on Taxation's (JCT) "Estimates of Federal Tax Expenditures for Fiscal Years 2011-2015".
[12] The AMT also partially disallows medical and dental deductions, accelerated depreciation, and deductions on home mortgage interest on non-primary property, among others.

schedule along pre-tax income represented by the x-intercept is determined by the total amount of allowable regular income tax deductions that a taxpayer claims and therefore, this parameter varies across taxpayers.

Panel B of Figure 2 provides a similar representation for the AMT schedule with corresponding tax brackets and effective marginal tax rates. The x-intercept of the AMT schedule is equal to the sum of the fixed AMT deduction and the regular tax deductions allowed by the AMT. At higher income levels, deductions under the regular tax schedule are on average lower than those under the AMT, by design. Therefore, Figure 2 relates to a high-income/high-deduction type taxpayer with $x\ intercept_{AMT} < x\ intercept_{regular}$. Note that on average, $x\ intercept_{AMT} > x\ intercept_{regular}$ for a low-income/low-deduction type taxpayer. The differences across taxpayers in the amount of deductions taken on the regular income tax and AMT schedules generates variation in the location of the point at which the two tax schedules intersect. This variation is the key reason for the misalignment of the top kinks on the combined schedule and the regular income tax schedule. Further, as I explain in Section 1.4.4, this variation in the location of the intersection kink disentangles variation in marginal tax rates from variation in taxable income, severing the link between taxable income and unobserved heterogeneity and mitigating a key endogeneity concern associated with the use of bunching estimators on fixed kink points.

Figure 3, Panel A brings together the regular income tax schedule with the AMT schedule for a high-income/high-deduction type taxpayer.[13] Taxpayers pay the higher of the two

[13] High earnings do not automatically translate into higher deductions. However, high earners disproportionately use larger deduction items such as state and local taxes, mortgage interest deduction and medical deductions (JCT Estimates, 2011-2015), which are fully or partially offset by the AMT, leading to high

personal income taxes and therefore, the combined income tax schedule is the upper envelope of the interaction of the two piecewise linear tax functions. The upper envelope of the combined schedule is shaded in gray. The point at which the AMT and the regular tax schedules interact is the intersection kink of the combined schedule.

The case for a low-income/low-deduction type taxpayer is different. The substantial, fixed deduction provided by the AMT shifts the AMT function to the right of the zero pre-tax income point. Across 1993 to 2011, the fixed deduction is as low as $45,000 and as high as $74,450 for married joint filers. This ensures that low- and middle- income taxpayers are only subjected to the regular income tax schedule. In general, this holds true if allowable deductions under the regular tax schedule are less than the fixed deduction provided by the AMT. Such a scenario for a hypothetical low-earner/low-deduction type taxpayer is illustrated in Panel B of Figure 3. For these taxpayers, the regular income tax schedule continues to be the effective tax schedule. This is potentially one reason for prior studies detecting bunching responses for low-income taxpayers when using the regular income tax schedule, but not for high earners who are in fact, subject to the combined schedule. In this paper, I focus on the high-income/high-deduction type of taxpayer responding to the combined schedule in Panel A of Figure 3 to estimate the ETI of high earners in the United States.

earners having lower deductions on the AMT schedule relative to the regular income tax schedule, on average.

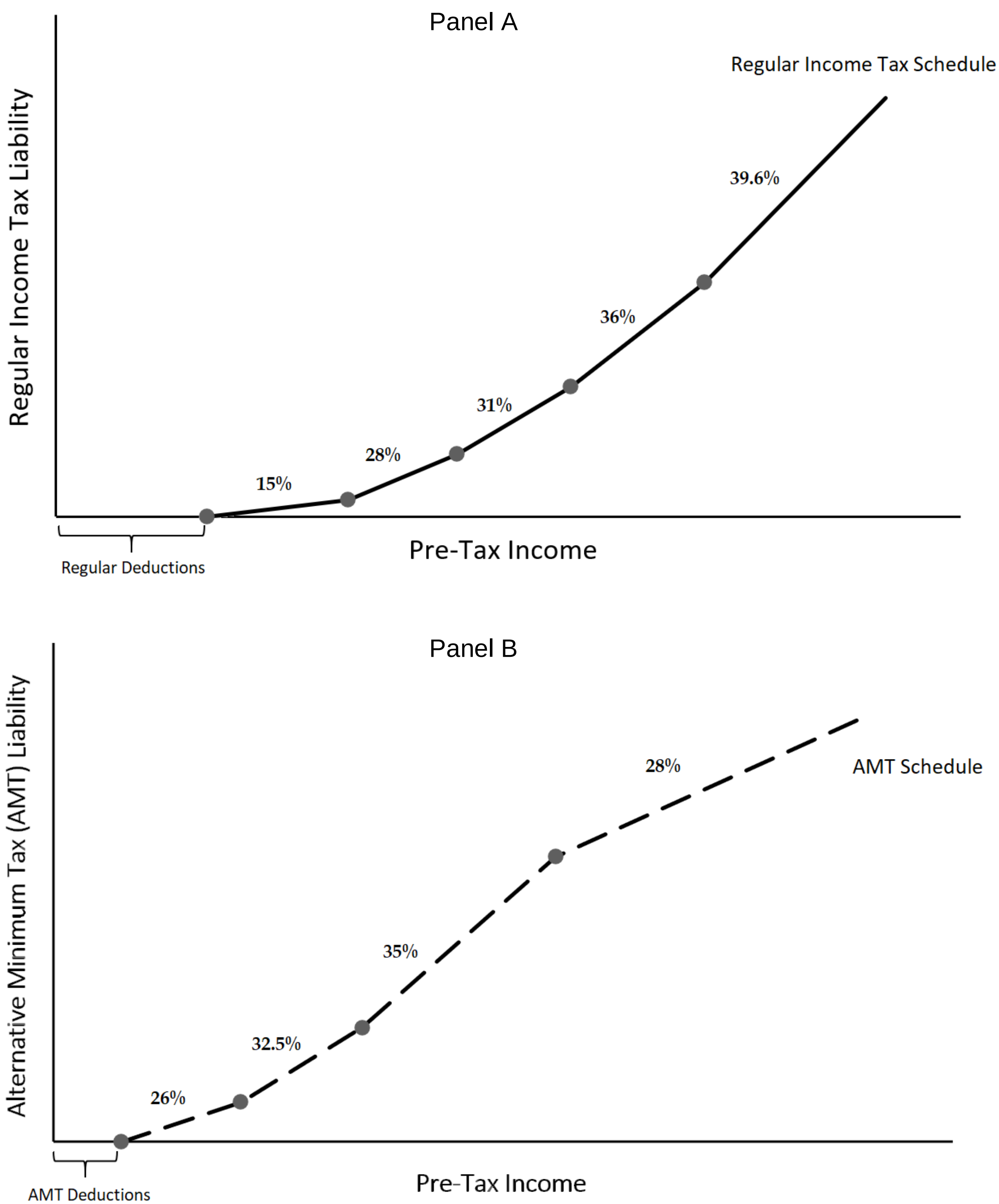


Figure 2: The Regular Income Tax and the AMT Schedules

Notes: Panel A illustrates the regular, federal income tax schedule for a hypothetical taxpayer. The slopes and the length of line segments in the piecewise linear function are based on marginal tax rates and the size of income tax brackets as provided in the tax code for year 2000. The x-intercept is determined by the amount of allowable deductions claimed by the taxpayer under the regular schedule. Panel B illustrates the AMT schedule. This piecewise function corresponds to marginal tax rates and income tax brackets on the AMT schedule. The x-intercept is determined by the amount of allowable deductions claimed under the AMT.The figures are not drawn to scale.

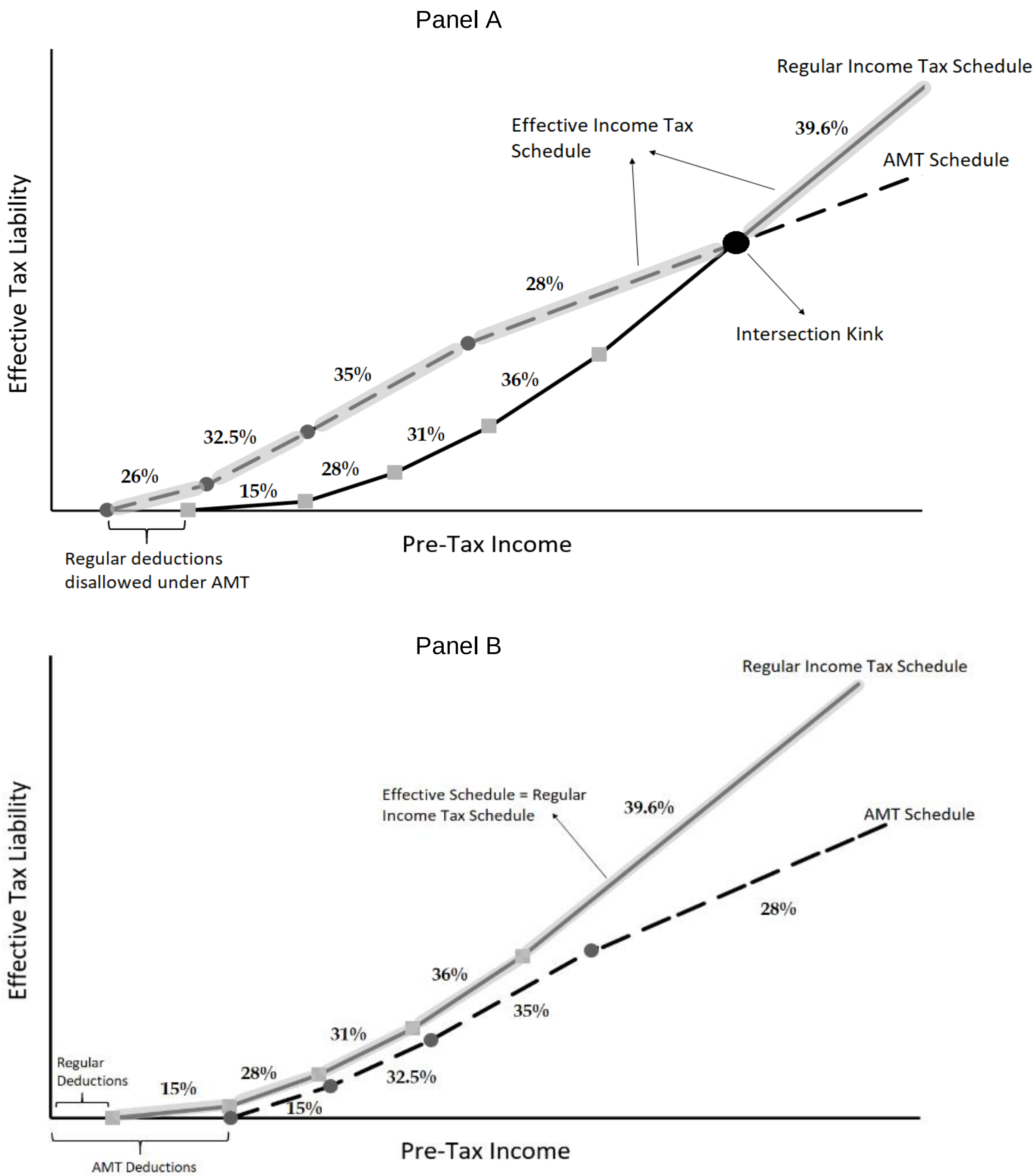


Figure 3: The Combined Schedule

Notes: Panel A is a representation of the combined schedule for a hypothetical high-income/high-deduction type taxpayer. Taxpayers pay the higher of the two taxes, leading to the effective schedule being the upper envelope of the combined schedule highlighted in grey. Panel B illustrates the combined schedule for a hypothetical low-income/low-deduction type taxpayer. For such a taxpayer, deductions on the AMT are, on average, greater than deductions on the regular tax schedule, leading to the AMT function being shifted further to the right relative to the regular schedule. Since taxpayers pay the higher of the two taxes, the regular tax schedule continues to be the effective schedule for such a taxpayer.

Variation in the location of the top kink in the combined schedule is driven by variation in the difference in the x-intercepts of the two schedules, with the latter depending on the difference in the amount of deductions allowed under the regular tax and AMT schedules. Specifically, one could imagine a range of differences in the x-intercepts, only one of which is illustrated in Figure 3, Panel A, generating a range of intersection kinks. In Figure 4, I provide the observed distribution of intersection kinks along regular taxable income to illustrate the variation in the location of the intersection kinks. The figure disaggregates the overall,

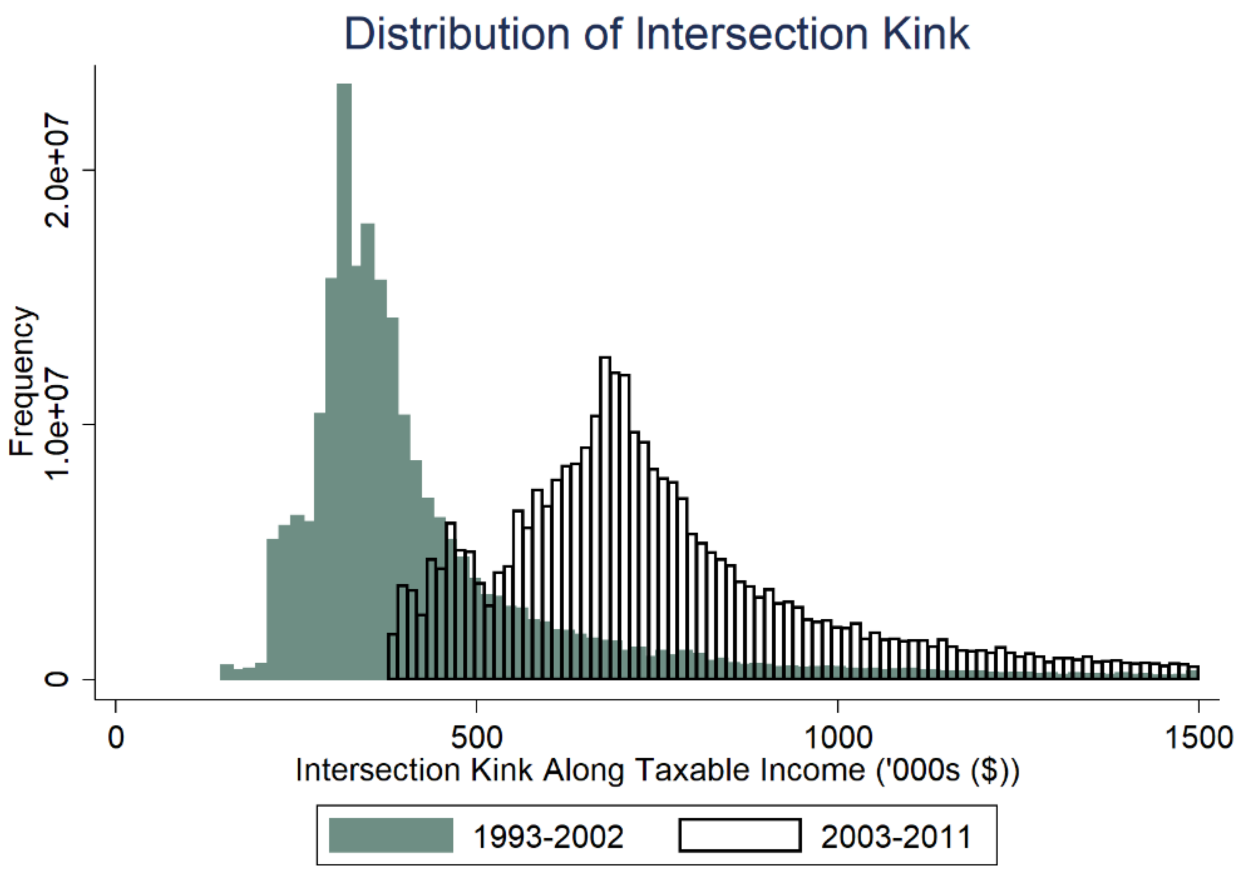


Figure 4: Distribution of the Intersection Kink Relative to Regular Taxable Income

Notes: The location of the itnersection kink in the combined schedule varies across taxable income, unlike kinks in the individual schedules that are fixed in taxable income. The bimodal distribution is divided in two, with the shaded distribution representing the time period 1993-2002, and the unshaded distribution representing the time period 2003-2011. Tax reforms of 2003 followed by annual increases in AMT exemption amounts shifted the underlying AMT schedule to the right, leading to the intersection kink also shifting to the right in the combined schedule. This causes the intersection to appear on average, at higher taxable income levels between 2003-2011.

bimodal distribution into two separate distributions corresponding to time periods 1993-2002 and 2003-2011. Changes made to the tax code through increases in AMT fixed deduction amounts from 2003 onwards increasingly shifted the AMT function to the right, shifting intersection kinks on average, to higher income levels.

In Section 1.4, I use the features of the regular income tax and AMT schedules discussed in this section to construct the two tax functions for each taxpayer in my sample from 1993-2011. For each taxpayer who is captured in Table 1 and who is subject to the combined schedule, I solve the two piecewise linear tax functions to find the top, intersection kink in the combined schedule. I use information on each taxpayer's observed taxable income and the location of the taxpayer-specific intersection kink to assess how far the individual's reported taxable income lies from the top, intersection kink. Aggregating across taxpayers, I show evidence of bunching to the left of the intersection kink where the marginal tax rate is 28 percent, as compared to approximately 38 percent to the right of the kink.

1.4 Empirical Methodology

1.4.1 Data

I use income tax return data from 1993-2011, housed at the National Bureau of Economic Research (NBER). The Statistics of Income (SOI) division of the Internal Revenue Service (IRS) has published annual samples of individual income tax returns in the form of Public Use Files (PUF) since 1960. These microdata are generated using a stratified random sample of tax filers. Sampling weights have varied and high earners face a larger sampling rate, with those at the very top of the income distribution facing an approximately 33 percent chance of showing up in the data. Since this study specifically looks at high earners, such oversampling allows me to capture greater variation in tax returns for this subpopulation.

I peg the start and end dates of the analysis time-period to the introduction of tax reforms that substantially altered the AMT. The Omnibus Budget Reconciliation Act (OBRA) of 1993 changed the AMT schedule by eliminating the flat marginal tax rate of 24 percent and introducing a two-tiered schedule, with statutory tax rates of 26 percent and 28 percent. OBRA 1993 also introduced a fixed deduction of $45,000 on alternative minimum taxable income for married joint filers and $33,750 for single filers. As discussed in Section 1.3, the phaseout of the fixed deduction creates four effective marginal tax rates: 26 percent, 32.5 percent, 35 percent, and 28 percent. I end the period of analysis at year 2011. The American Taxpayer Relief Act (ATRA) of 2012 indexed the AMT exemption amounts to inflation. To avoid this tax year with characteristics that are significantly different from those found in other tax years, I omit the year from the analysis.

I divide the sample into two time periods for the heterogeneity analysis: 1993-2002 and 2003-2011. I choose 2002 as the endpoint for the first time-period because while the AMT fixed deduction amounts were relatively stable before 2003, Congress increased the deduction amounts annually on an ad-hoc basis from 2003 onwards. These increases shifted the AMT schedule to the right along the range of pre-tax incomes, leading to the intersection kink appearing at higher income levels.

I limit the data to tax returns submitted by married joint filers and single filers, leading to a dataset containing 2.3 million observations, representing approximately 2 billion unique tax returns. Out of the total number of taxpayers filing these returns, 5.3 percent submit

Form 6251, the form used to compute AMT liability (the unweighted fraction in the data sample is 34 percent). However, this fraction increases to 24 percent for taxpayers with adjusted gross income (AGI) in real 2007 terms greater than $100,000 and to 58.5 percent for taxpayers with real AGI greater than $200,000. The IRS puts the burden of submitting the AMT form on the taxpayer. This implies that in case Form 6251 is not submitted and the IRS predicts that the taxpayer would owe AMT liability, then there is a possibility of audit. From 2006 to 2011, taxpayers also had access to an IRS-provided web tool called the AMT Assistant, which required responses to a handful of questions related to the income level and filing status of the taxpayer for the tool to make a recommendation regarding the submission of Form 6251.

I remove taxpayers who do not face the combined schedule as illustrated in Figure 3 Panel B from my analysis sample. Since these taxpayers do not face the combined schedule, the regular income tax schedule continues to be the effective schedule that applies to them. This leads to an analysis sample containing 273,856 observations representing approximately 5.9 million tax returns. Further, in line with earlier literature, I restrict the frame of the analysis to a range within which the effective kink lies. I limit the sample to individuals within $300,000 (-$150,000, +$150,000) of their effective kink. This is my analysis sample, with a total of 36,639 observations, representing approximately 1.2 million individual income tax returns.

The population median AGI for these individuals is $679,400 in real 2007 dollars, corresponding to taxpayers in the top percentile of the income distribution. The median effective, taxable income for these taxpayers is $536,600. The intersection kink for these taxpayers lies on average, at \$430,200 for time period 1993-2002, and at \$679,307 for time period 2003-2011.

1.4.2 Solving for the Top Kink in the Combined Schedule

This section discusses the methodology that I use to construct the combined schedule. Recall that the combined schedule is the upper envelope of the federal regular income tax and AMT schedules. As discussed, a taxpayer can shelter part of his or her pretax income from taxation by taking deductions under both the regular income tax and the AMT schedules. Let the pre-tax income in a calendar year for a given taxpayer be Y. Let the income sheltered from the regular income tax schedule be S_R, which is equal to D_R, the regular income tax deductions.

The AMT also allows for some income sheltering denoted by S_{AMT}. The AMT has a fixed deduction for each tax return filing category that I denote by D_{AMT}. Further, the AMT disallows a fraction of the deductions α claimed under the regular income tax. Therefore, deductions taken under the regular income tax schedule that are partially allowed under the AMT are $D_R(1-\alpha)$. If $S_R = S_{AMT}$, then $D_R = D_{AMT} + D_R(1-\alpha)$, or $\alpha = \frac{D_{AMT}}{D_R}$. If $S_{AMT} < S_R$, as is the case in Figure 5, Panel A, then the taxpayer has a unique intersection kink. If $S_{AMT} \geq S_R$, the taxpayer's combined schedule can either have two intersection kinks – one at a low level of taxable income and the other at a high level of taxable income, or no intersection kinks, as shown in Panels B and C, respectively.

To ensure that the analysis covers a unique intersection kink, I restrict the sample to taxpayers for whom $S_{AMT} < S_R$, retaining approximately 35 percent of observations. The median regular taxable income of the population of taxpayers in Panel A is \$365,700 in real 2007 dollars. For individuals in Panels B and C whom I exclude, the median real income is

$80,100. Therefore, by restricting the sample to taxpayers facing the combined schedule as in Panel A, I study the taxable income response of a group that reports a high level of taxable income and is likely to respond to the top kink on the combined schedule.

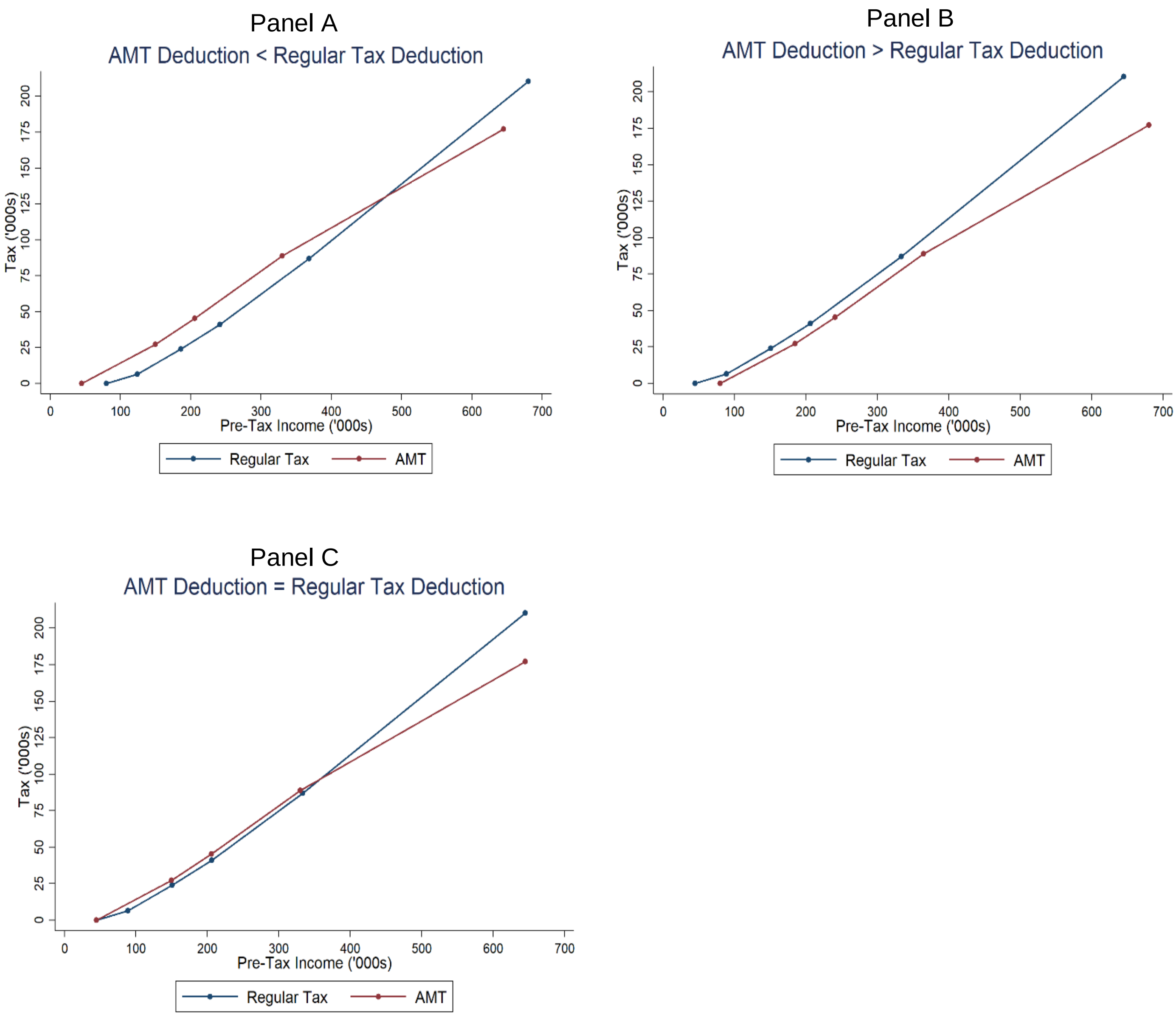


Figure 5: Relationship between Deductions and the Intersecton Kink

Notes: These figures show how the location of the intersection kink varies when the difference in deductions allowed between the regular income tax and the AMT schedules varies. The current study isolates the analysis to taxpayers who exhibit a structure of deductions and the combined schedule like the one shown in Panel A.

The two schedules are piecewise linear. The IRS data does not provide exhaustive information on deductions, so $S_{AMT} < S_R$ cannot be observed directly. However, taxable incomes on the two schedules that are observed can be used as a proxy. Let T_R be regular taxable income, defined as $T_R = Y - S_R$. Let T_{AMT} be AMT income, defined as $T_{AMT} = Y - S_{AMT}$.

Then,

$$T_{AMT} - T_R = (Y - S_{AMT}) - (Y - S_R) = S_R - S_{AMT} \tag{1}$$

The difference in taxable incomes based on the two schedules precisely equals the difference in deductions allowed under the two schedules. Therefore, I operationalize $S_{AMT} < S_R$ by using $T_{AMT} > T_R$ to find the difference between the x-intercepts of the two functions.

I solve the system of the two piecewise linear functions to find the location of the intersection kink. Because the location of the intersection kink varies across taxpayers, I standardize the location of this kink in the aggregate. I do this by estimating the distance of effective taxable income for each taxpayer from the top, intersection kink in their respective effective schedule and plotting the distribution of these differences relative to the intersection kink. For example, consider the case of three taxpayers for whom the intersection kinks are located at \$380,000, \$400,000, and \$420,000 in terms of effective taxable income, respectively. Assume that all three of these taxpayers are bunching to the left of their respective intersection kinks, with their corresponding observed incomes being

$379,000, $399,000, and $419,000, respectively. To observe this bunching behavior in the aggregate, I subtract the location of the intersection kink from their observed taxable income, leading to their taxable incomes with respect to the intersection kink being -$1,000 each. The centered distribution has the intersection kink at the $0 point, while all three of the taxpayers locate at $1,000 to the left of the centered intersection kink.

1.4.3 Estimation Method

To estimate the elasticity of taxable income, I use the traditional bunching estimator developed by Saez (2010). Saez (2010) models the behavior of taxpayers around kink points using quasi-linear utility increasing in after-tax income (consumption) and decreasing in before-tax income (effort). Income effects are assumed to be negligible. In the case of changing tax rates, taxpayers who locate to the right of the kink point in a no tax scenario would instead prefer to locate at or close to the kink point under non-linearities introduced by changing marginal tax rates. This relationship is illustrated in Figure A.1 of Appendix A. The kink in the income tax schedule generates a kink, which leads individuals with indifference curves of the type H to cluster at the kink point.

For the base estimator, I employ a simple parameterized model with a quasi-linear and iso-elastic utility function of the form:

$$u(c, k) = c - \frac{n}{1 + 1/e}\left(\frac{k}{n}\right)^{1+1/e} \tag{2}$$

where c is consumption, k is before-tax income, n is an ability parameter distributed with density $f(n)$, and e is compensated elasticity of reported income (Saez, 2010).[14] In a no-tax scenario, the marginal tax rate is constant at τ_0 throughout the cumulative distribution, denoted by $L_0(k)$. The introduction of a different marginal tax rate of τ_1 at K creates a convex kink in the budget set. Taking this kink point into account, individuals with $n \in [K/(1-\tau_0)^e, K/(1-\tau_1)^e]$ choose $k = K$ and bunch at the kink point. This leads to the fraction of the population bunching to be:

$$b = K\left[\left(\frac{1-\tau_0}{1-\tau_1}\right)^e - 1\right]\frac{l(K)_- + l(K)_+/\left(\frac{1-\tau_0}{1-\tau_1}\right)^e}{2} \tag{3}$$

This function can be solved explicitly to express e as a function of observed or empirically estimable variables. Simplification leads to (Wang et al., 2020):

$$\epsilon = \frac{b(\tau_0,\tau_1)}{K\log\left(\frac{1-\tau_0}{1-\tau_1}\right)} \approx \frac{\hat{b}}{\left|\frac{K}{W}\cdot\frac{\Delta\tau}{1-\tau_0}\right|} \tag{4}$$

where τ_0 and τ_1 are the effective marginal tax rates on either side of the intersection kink and are observed. For example, in year 2000, τ_0 is 28 percent and τ_1 is 39.6 percent. W is the binwidth chosen for binning taxpayers in effective income groups. The traditional bunching estimator uses a fixed K in taxable income. However, since the location of the intersection

[14] By using a quasi-linear utility function, I abstract from any income effects for simplicity (Gruber and Saez, 2002). This is standard in the literature and a study of income effects is beyond the scope of this paper.

kink varies along the regular taxable income spectrum, I take the weighted (population) average of the effective taxable income in the bunching region as an estimate of K.

The difference between the observed taxable income density in the presence of the kink point and the counterfactual density that would plausibly have existed in the absence of the kink point is denoted by $\hat{b}$. In other words, $\hat{b}$ quantifies excess mass, or the magnitude of bunching in the bunching region. To estimate b, Saez (2010) assumes the counterfactual density to be linear in the bunching region. In contrast, I fit a polynomial function of order p across the bunching region to estimate the counterfactual density. Dividing the range of taxable incomes relative to the intersection kink into bins of size W, I fit a polynomial of order p to the counts for each of the taxable income bins, excluding data near the kinks after estimating a regression of the form:

$$C_j = \sum_{i=0}^{p} \beta_i Z_j^i + \sum_{r=-l}^{u} \phi_r D_j + \epsilon_j \tag{5}$$

where C_j is the count of observations found in bin j, Z_j is the midpoint level of the effective taxable income in bin j, and D_j is a dummy for each bin found in the bunching region. Therefore, there are $l + u$ indicators such that $D_j = 1$ if $Z_j \in [K - l, K + u]$, where K is the location of the kink and l is the distance to the left of the kink and u is the distance to the right of the kink measured in terms of effective taxable income. I choose a polynomial of order 7 based on the joint minimization of the Akaike Information Criterion and the Bayesian Information Criterion. The counterfactual frequency of observations, $\widehat{C_j}$, is then derived using predicted counts from $\widehat{C_j} = \sum_{i=0}^{p} \hat{\beta}_i Z_j^i$, which omits the impact of the dummies $\widehat{\phi_r}$.

Using the actual and the estimated counterfactual densities, the quantity of "excess bunching" can be estimated using:

$$\hat{b} = \sum_{j=-l}^{u} \frac{(C_j - \widehat{C}_j)}{N} \tag{6}$$

where the numerator sums the difference between the number of observations in each bin of the observed density and the counterfactual density in the bunching region. The denominator N scales the excess bunching by the number of bins in the bunching region.

I further impose the constraint that taxpayers who bunch do so by reducing their taxable income, so that the number of taxpayers missing from the right of the intersection kink is equivalent to the number of individuals bunching to the left of the intersection kink (Chetty et al., 2011). I calculate the standard error for $\hat{b}$ using a parametric bootstrap procedure by drawing from the estimated vector of errors for the counterfactual estimation equation with replacement to generate a new set of counts and applying the above technique to calculate a new estimate of $\hat{b}^k$. I define the standard error of $\hat{b}$ as the standard deviation of the distribution of $\hat{b}^k s$. This ensures that the number of observations represented by the area of the counterfactual density does not exceed those in the observed distribution. Plugging in the observed marginal tax rates, binwidth, estimates of excess bunching $\hat{b}^k$, and the location of the kink point K into (4) provides my base estimates of the elasticity of taxable income.

1.4.4 Robustness to Endogeneity Concern

The present setting where the location of the top kink in the combined schedule varies across taxpayers provides a unique opportunity to mitigate recent endogeneity concerns related to the use of bunching methods on kink points fixed in taxable income (Blomquist and Newey, 2017; Bertanha et al., 2016, 2020). Since fixed kinks at which marginal tax rates change are jointly determined with taxable income, observed taxable income is likely correlated with unobserved heterogeneity. Intuitively, it is plausible that individuals select into particular bins of the income distribution not as a result of strategic responses to marginal tax rates but because of some underlying preferences for those income levels. If this occurs, then observed bunching (or troughs) in the taxable income distribution might reflect preferences rather than strategic decision-making related to tax rates, causing bias in the estimation of the ETI of unknown direction.

However, in the setting that I leverage, the top kink in the combined schedule varies for each taxpayer across taxable income generating a distribution of top kinks. This is unique feature of the combined schedule weakens the correlation between taxable income and unobserved heterogeneity, increasing confidence in the ability of my estimator to estimate an unbiased ETI parameter.

1.4.5 Parameter Selection and the Functional-Form Assumption

In this subsection, I discuss my method for selecting the bandwidth and the binwidth for my estimate for the average ETI of high-income taxpayers. I close the section with a brief discussion on how I relax the standard functional-form assumptions for the shape of the underlying counterfactual density to test the robustness of my estimate.

The choice of binwidth leads to a trade-off between noise and precision: the greater the binwidth, the less noisy and smoother the histogram; the smaller the binwidth, the noisier the histogram, since it reveals more variation in the data. I compute the optimal binwidth using a data-driven approach. I also use other binwidths for comparison as discussed in Section 1.6. For the optimal binwidth selection, I use the Freedman-Diaconis method:

$$W = 2 * IQR * n^{-1/3} \tag{7}$$

where W is the binwidth, IQR is the interquantile range of the distribution of effective taxable income, and n is the number of observed tax returns. I find W to be \$8,106.

I estimate the bandwidth and construct the bunching region using the algorithm for bandwidth-selection proposed by Bosch et al. (2020). The construction of the bunching region comprises two choices: the choice for the location of the bunching region (symmetric or asymmetric) and the length of the bunching region on either side of the kink point. Earlier methods in the literature for selecting the location and range of the bunching region have used either one of two approaches. The first approach uses a symmetric bunching region around the kink point and assesses the sensitivity of the bunching estimate to the symmetric widening or contraction of the bunching region. The second approach considers graphical evidence of bunching and pegs the lower and upper bounds of the bunching region to visually obvious starting and ending points of anomalous bunching. The sensitivity of bunching estimates is tested by varying the size and location of this bunching region. In this paper, I allow the bunching window to be defined by a data-driven procedure as described below.

The algorithm for selecting the bandwidth is as follows. Initially, the bin containing the kink point is assumed to be the excluded region as in (5), so that the excluded region becomes $(z_{-}, z_{+}) = (0,0)$, where z_i identifies taxable income bins. A local linear regression with $p = 1$ is fitted through the scatterplot of frequencies of observations in each bin versus bin identifiers that are sorted by income. However, the regression omits the impact of the excluded region: the bin containing the kink point $(z_{-}, z_{+}) = (0,0)$. I form a 95 percent confidence interval around this local linear regression line. Contiguous bins around the kink point for which the frequencies lie outside the 95 percent confidence interval form my data-driven bunching region under $(z_{-}, z_{+}) = (0,0)$. The left-most bin of this bunching region is the lower bound of the bunching region, while the right-most bin is the upper bound.

I then add one bin to either side of the excluded region, so that $(z_{-}, z_{+}) = (-1, +1)$ and repeat the process, to obtain a fresh pair of lower and upper bounds for the bunching region under $(z_{-}, z_{+}) = (-1, +1)$. I keep adding bins to either side of the excluded region such that $z_{-} \in \{-Z, (-Z+1), (-Z+2), \dots, 0\}$ and $z_{+} \in \{0, 1, \dots, Z\}$ and in process, generate a distribution of lower and upper bounds for the data-driven bunching region. I pick the modal bins from the distributions of the lower bounds and the upper bounds. The modal bin of the distribution of lower bounds serves as the lower bound of the bunching window in my analysis. Similarly, I pick the modal bin of the distribution of upper bounds to be the upper bound of the bunching window in my analysis. This process results in an asymmetric bunching window of $(-\$40{,}529, +\$16{,}212)$. While I use this bunching window for my base analysis, I also test the sensitivity of the average ETI estimate for high earners to varying choices of bandwidth in Section 1.6.

I also test the robustness of my estimate for the ETI of high-income taxpayers by using weaker assumptions for the functional form of the counterfactual density in the bunching region. I leverage the method developed by Bertanha et al. (2020) and assume that the counterfactual density belongs to the family of Lipschitz continuous functions. In the context of the counterfactual density assumed to be defined by a Lipschitz function, there exists a real number such that the line connecting the endpoints of a given bunching region has a slope which is not greater than the absolute value of this real number, known as the Lipschitz constant. This limits the magnitude of the slope of the counterfactual density in the bunching region. Such a limitation is achieved by constraining the area under the counterfactual density by the area of the observed distribution.[15] By limiting the slope of the counterfactual density in the bunching region, I establish upper and lower bounds on the size of the excess mass resulting in bounds on the estimated ETI. Results for this robustness check are provided in Section 1.5.2.

1.5 Results

1.5.1 Graphical Evidence

I find graphical evidence of clustering to the left of the top, intersection kink in the combined schedule, as shown in Figure 6, Panel A. This figure provides the weighted distribution of taxable income relative to the intersection kink for taxpayers in the sample. Note that to give an expanded view of the distribution around the intersection kink point, I plot the observed distribution within -$200,000 to +$200,000 of the intersection kink. Panels B and C provide histograms disaggregated by the time periods 1993-2002 and 2003-2011. Both periods reveal bunching responses just to the left of the intersection kink with more pronounced bunching for the latter period. With increasing AMT bunching amounts, the intersection kink

[15] I thank Nathan Seegert at the University of Utah for providing me with early access to his statistical program for identifying these bounds. The final Stata package is available under the label "bunching".

shifts to the right along the taxable income distribution. The accentuated bunching response revealed in the time period 2003-2011 arguably captures the potentially higher behavioral response at relatively higher income levels.

Figure 7 contrasts the bunching responses to the top kink in the combined schedule and the regular income tax schedule. To provide a more granular view of any potential bunching, I provide histograms with smaller binwidths of $4,000. Panel A of Figure 7 replicates Panel A of Figure 6 with a smaller binwidth, confirming bunching at the top kink of the combined schedule. Panel B of Figure 7 confirms the lack of bunching at the top kink in the regular schedule, first studied in Saez (2002).

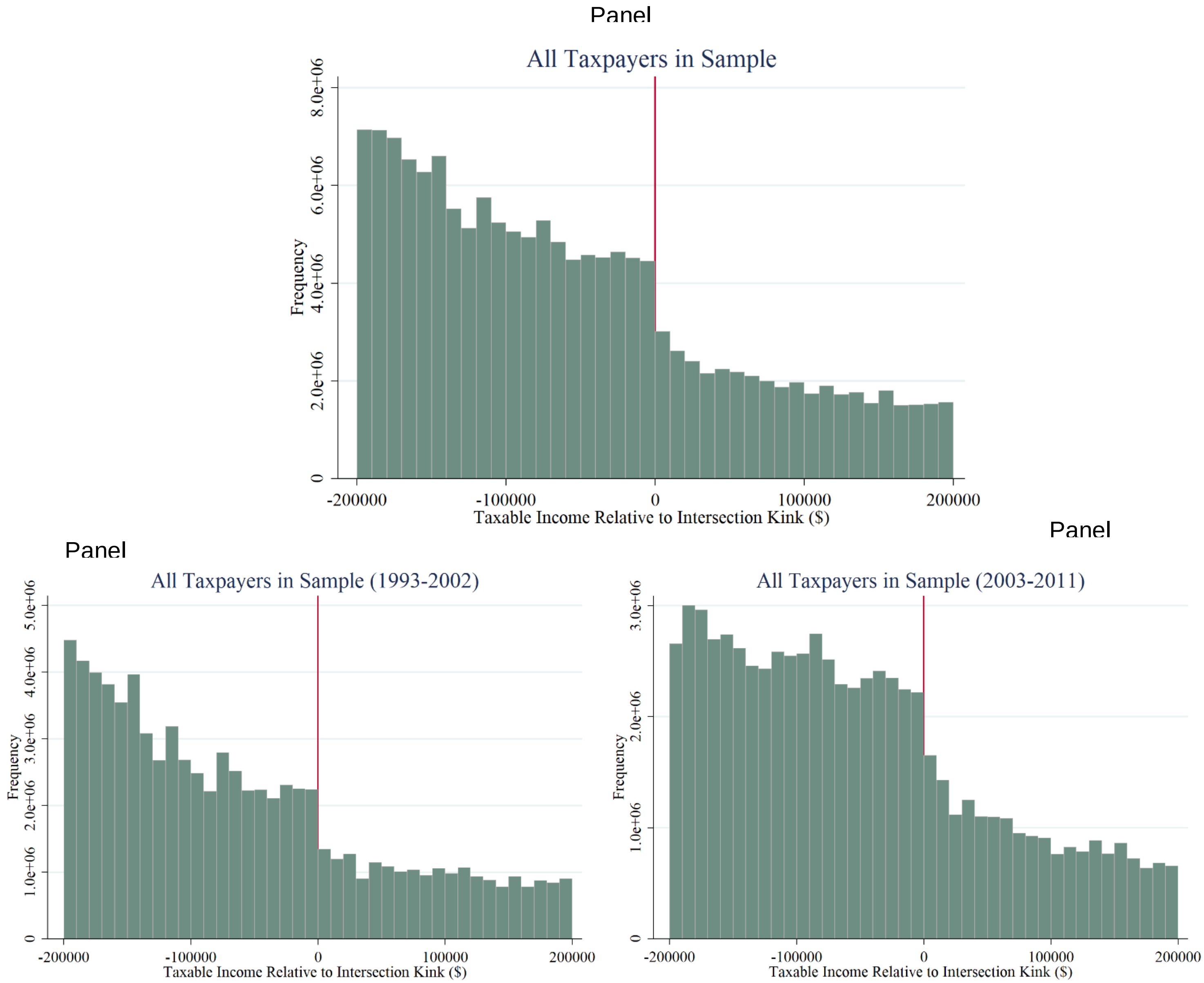
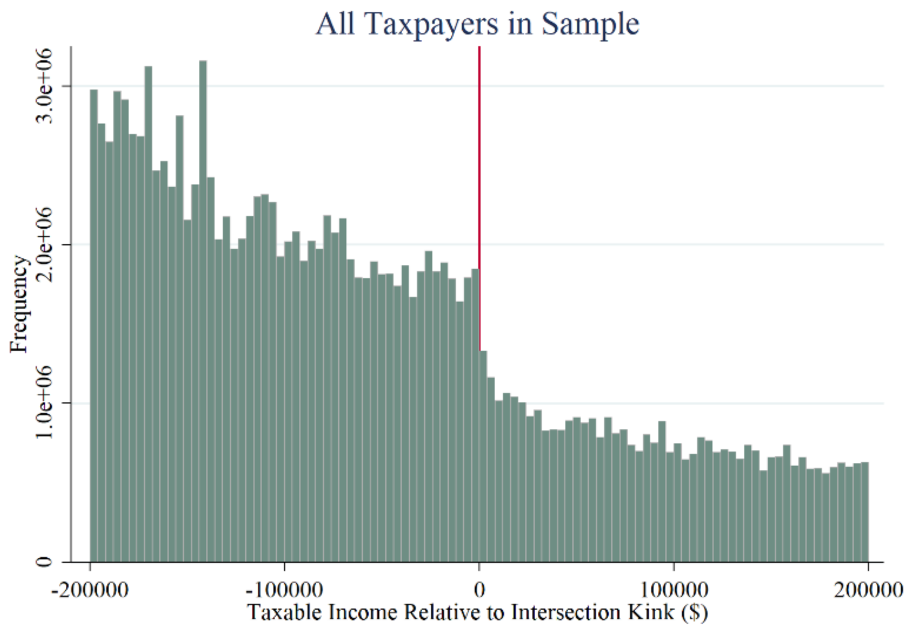


Figure 6: Graphical Evidence of Bunching

Notes: The figures show bunching of taxpayers in the aggregate around the intersection kink. Histograms are constructed with binwidths of $10,000. Panel A shows the distribution of effective taxable income relative to the intersection kink for all observations in the study sample. Panels B and C show the distributions for subpopulations disaggregated by two time periods: 1993-2002 and 2003-2011.

Panel A



Panel B

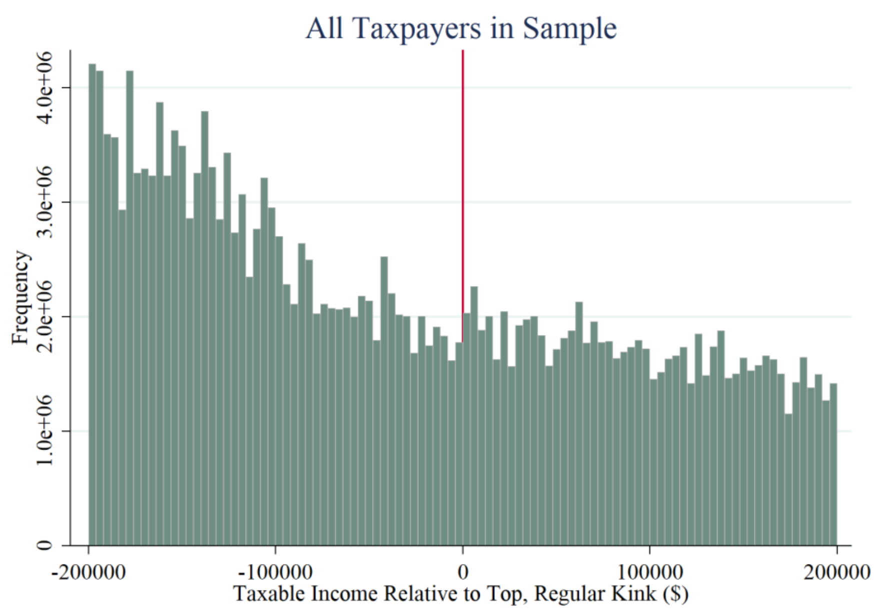


Figure 7: Bunching Responses around the Top Kinks in the Combined and Regular Schedules

Notes: The figures show bunching responses around the intersection kink in the combined schedule (Panel A) and the top kink in the regular income tax schedule when the combined schedule is not considered (Panel B). Histograms have binwidth of $4,000.

I further disaggregate the total sample into wage earners and the self-employed. Self-employed individuals are defined as taxpayers who reported any non-zero income from non-wage sources including sole proprietorships, partnerships, S-Corporations, and farming. Wage earners are those taxpayers who reported zero earnings from these sources. I will refer to taxpayers with any positive self-employment earnings as "self-employed" though this does not preclude them having wage-based income as well. Existing literature has predicted and provided evidence for other segments of the income distribution, significant avoidance behavior by self-employed individuals as compared to wage earners. Pure wage earners in the United States face third-party reporting, with their employers sending the W-2 form containing information on the employees' earnings to the IRS. The IRS uses this third-party reported information to cross-check employee-reported income and mismatches between employee- and employer-reported incomes increase the probability of audit for pure wage earners. Self-employed individuals face third-party reporting only for a fraction of their overall income corresponding to the part of their incomes that comes from wages. These taxpayers have greater flexibility in how they report self-employment income providing them with a larger margin to manipulate taxable income.

Graphical evidence in Figure 8, Panel A shows that high-income pure wage earners also cluster to the left of the intersection kink, though relatively less sharply as compared to taxpayers who have access to self-employment income (Panel B). The substantial bunching for high-income wage earners is in contrast to earlier studies that show very low bunching for wage earners in the overall taxable income distribution (Saez, 2010; Chetty et al. 2011). Two reasons possibly lead to this divergence. First, high-income wage earners have increased bargaining power, allowing them to negotiate the substitution of highly taxed monetary compensation with untaxed fringe benefits. Second, high-income managers and executives get a larger share of their earnings in the form of stocks as compared to lower-income wage earners. The realization of gains or losses on such stocks can be timed flexibly as compared to annual wage earnings.

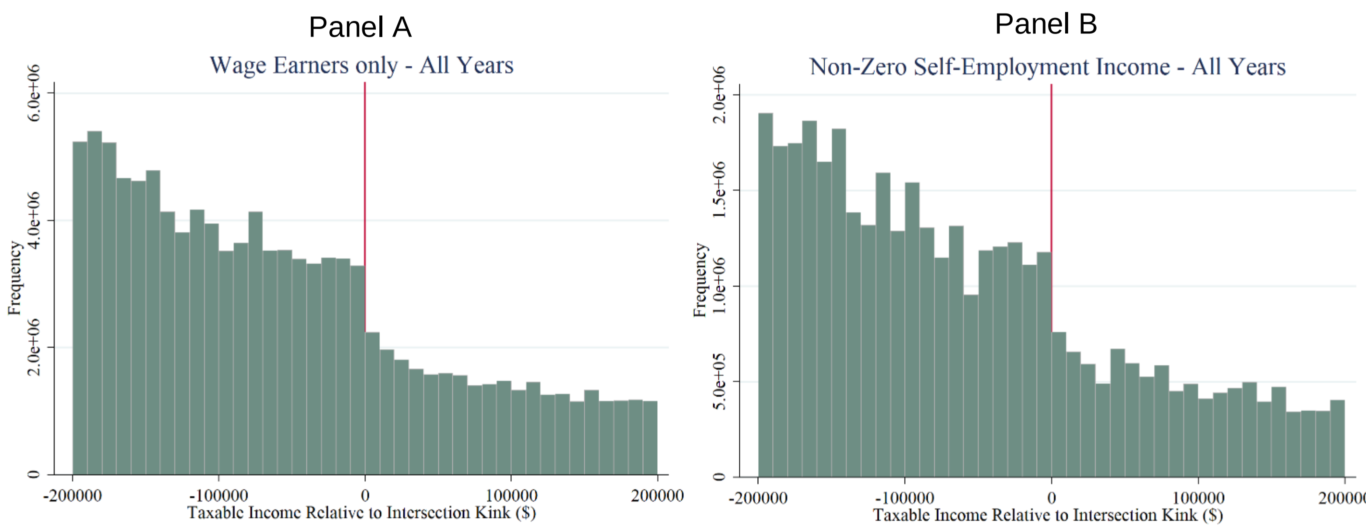


Figure 8: Bunching Responses of Wage Earners and the Self-Employed

Notes: These figures provide visual evidence of bunching around the top, intersection kink for pure wage earners (Panel A) who do not report any self-employment income, and for taxpayers with any positive self-employment earnings (Panel B). Histograms are constructed using binwidths of $10,000.

The availability of capital stock and strategic realizations of capital gains and losses provide high-income taxpayers with the ability to optimize tax sheltering. Note that between 1993 and 2011, short-term capital gains are taxed at the same rates as ordinary income and therefore, divergent strategies for tax sheltering using short-term capital gains are unlikely. However, if long-term capital gains or losses are realized strategically across time, then studying long-term capital stock activity can shed light on tax avoidance mechanisms.[16] If realizations are timed according to current and future expected tax rates, then avoidance behavior can give rise to fiscal externalities that also need to be incorporated into estimates of the ETI.

Assessing strategic behavior on the capital gains channel, however, is difficult for two reasons. First, the cross-sectional nature of publicly available IRS tax return microdata is not amenable to assessing the timing of realizations for the same taxpayers. Using panel tax data can allow for more flexibility in studying these mechanisms. Second, in the context of the interaction of the regular income tax schedule and the AMT, assessing the impact of capital gains implies overlaying a third schedule on top of the first two schedules. To avoid this problem, Saez (2010) considers all taxable income net of capital gains. While I replicate this for the main analysis, I also divide the sample into individuals reporting long-term capital gains and those who do not report such gains, and separately find graphical evidence and elasticity estimates for both groups. Thus, estimates for the subpopulation of tax returns only impacted by the combined schedule and not the capital gains schedule provides the clearest insights into taxpayer behavior around the intersection kink. Figure 9 illustrates my results. I find that individuals who report no long-term capital gains in a given year (Panel B) have a greater bunching response demonstrated by the larger excess mass to the right of the intersection kink, as compared to taxpayers who do report such gains (Panel A).

[16] Realizations of short-term capital gains do not occur in isolation from strategies related to long-term capital gains. For simplicity however, I treat all short-term capital gains as ordinary income and abstract from their effect on the ability of taxpayers to realize long-term capital gains.

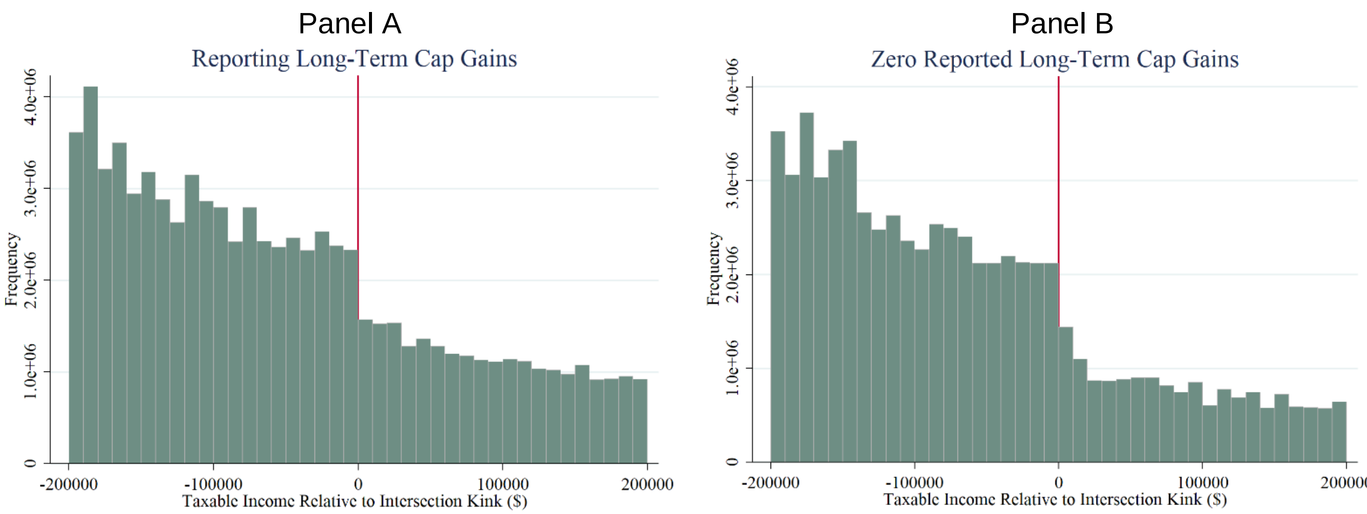


Figure 9: Bunching Responses Related to Long-Term Capital Gains

Notes: These figures provide visual evidence of bunching around the top, intersection kink for taxpayers reporting long-term capital gains (Panel A) and those not reporting such gains (Panel B). Histograms are constructed using binwidths of $10,000.

1.5.2 Elasticity Estimates

Figure 10 illustrates the observed and counterfactual distributions of effective taxable income relative to the kink point for high earners in my sample of tax returns filed between 1993-2011. The line connecting frequencies per taxable income bin represents the observed density and the smooth line running through the observed distribution represents the counterfactual density, estimated using (5). The zero point in the support of the distribution represents the recentered location of the top, intersection kink point. The vertical dashed lines represent the bounds of the bunching region. The density of observed effective taxable income around the top kink of the combined schedule provides evidence that high earners bunch to the left of the intersection kink, with the difference between net-of-tax rates between the left and the right of the intersection kink being approximately 10 percentage points.

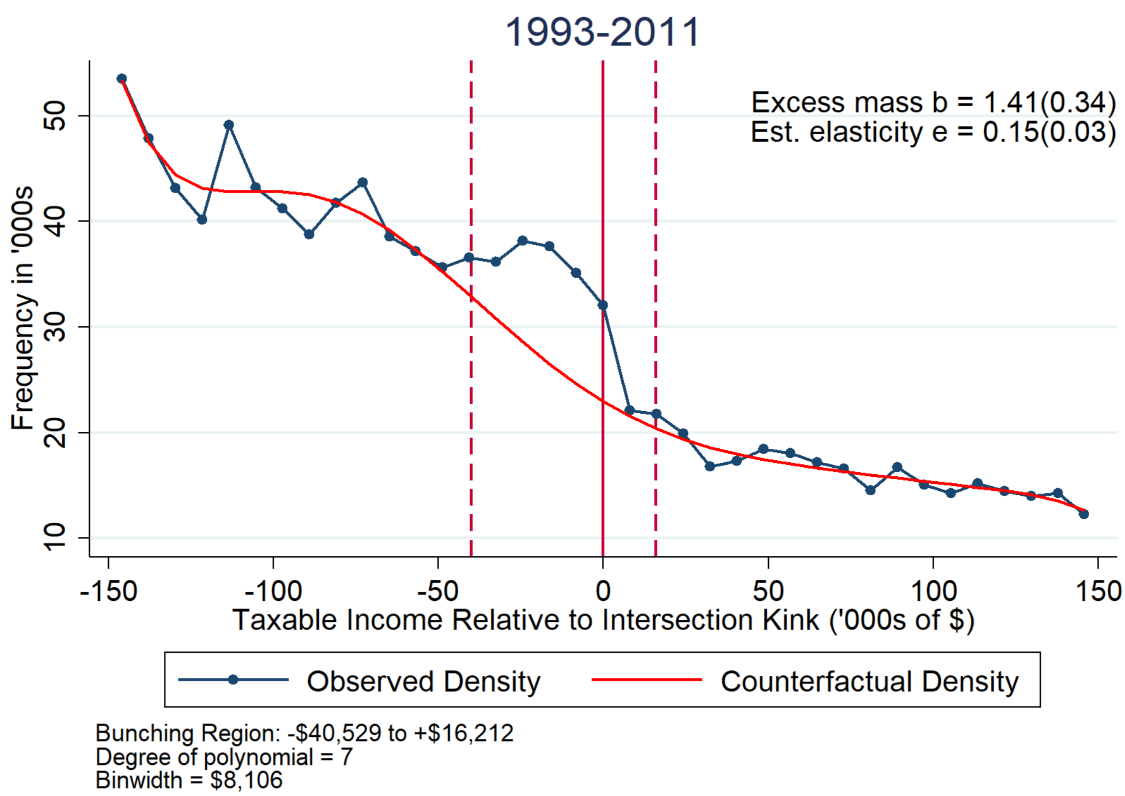


Figure 10: Distribution of Observed Versus Counterfactual Taxable Incomes

Notes: The bold, vertical line (red) represents the centered intersection kink. Dashed, vertical lines (red) represent the lower and upper bounds of the bunching region defined as -$40,529 and $16,212. Observations are binned with the optimal binwidth of $8,106. The connected line illustrates the observed distribution of taxable income. A seventh-order polynomial is used to construct the counterfactual density represented by the smooth line running through the observed taxable income distribution. Estimates for the excess mass and the ETI are provided at the top-right. Bootstrapped SEs are shown in parentheses.

Using this bunching response, I use the approach discussed in Section 1.4.3 to estimate the ETI for high earners to be 0.15, estimated precisely within a 95 percent confidence interval. This estimate is economically significant as compared to earlier estimates of close to zero for high earners in the United States obtained with the use of bunching estimators on the regular income tax schedule (Saez, 2010; Mortenson & Whitten, 2016). The difference in estimates confirms the evidence provided in figure 7. I use the approach for estimating non-parametric bounds for the average ETI estimate discussed in Section 1.4.3. The average ETI estimate of 0.15 is bounded below at 0.12 and above at 0.17.

Figure 11 disaggregates taxpayers by type of income, the presence or absence of capital gains, and time period. Panels A and B in Figure 11 show that both pure wage earners and taxpayers with some self-employment income bunch to the left of the intersection kink. However, this bunching is more pronounced for taxpayers with self-employment income. I estimate the ETI for wage earners of 0.12. In contrast, the estimated ETI is 0.24 for individuals reporting non-zero self-employment income, twice that of wage earners. The estimated elasticity for the self-employed is remarkably similar to the observed elasticity for this subpopulation in Denmark (Chetty et al., 2011).

In my analysis sample, 28 percent of taxpayers report at least some non-zero self-employment income. The remaining 72 percent only report pure wage earnings. As I discuss in Section 1.5.1, taxpayers with self-employment income do not face third-party reporting for at least some part of their incomes, creating space for tax avoidance behavior that can be more aggressive relative to the behavior of pure wage earners for whom such tax avoidance space is more limited. However, high-wage employees such as executives and managers might also have greater bargaining power as compared to low-wage employees vis-à-vis fringe benefits and access to stock options. While previous research has shown that bunching responses of wage earners for the overall population are weaker, it is plausible that these responses are non-trivial for high-income wage earners.

I also divide the analysis sample across taxpayers who do not report long-term capital gains and those who do. Approximately 43 percent of the population represented by the analysis sample does not report long-term capital gains, as opposed to 57 percent that reports non-zero long-term capital gains. Panels C and D in Figure 11 provide evidence of bunching to the left of the intersection kink for both groups. However, this bunching response is accentuated for taxpayers not reporting long-term capital gains. It is possible that for the group reporting such gains, the added complexity of the capital gains schedule combined with the interaction of the regular income tax and AMT schedules results in some of the bunching at the kink point being dispersed. Specifically, the capital gains schedule can create a wedge between the combined AMT-regular income tax schedule and the true combined schedule. Therefore, the cleanest estimate of the ETI of high earners comes from the group of individuals for whom, the combined schedule is just the upper bound of the AMT and regular income tax schedule: taxpayers who are unaffected by the long-term capital gains schedule. For this group, the estimated ETI is 0.20, as compared to 0.11 for taxpayers reporting some form of long-term capital gains.

I also explore taxpayers' differential responses to the top kink in the combined schedule across time. I separately estimate the ETI of high-income taxpayers for the time periods 1993-2002 and 2003-2011. In my sample, 58 percent (42 percent) of the population comes from the first (second) time period. The response to the top, intersection kink is greater for the latter time period as shown in Panels E and F of Figure 11. The ETI is 0.12 in time period 1993-2002 and is 0.20 in time period 2003-2011.

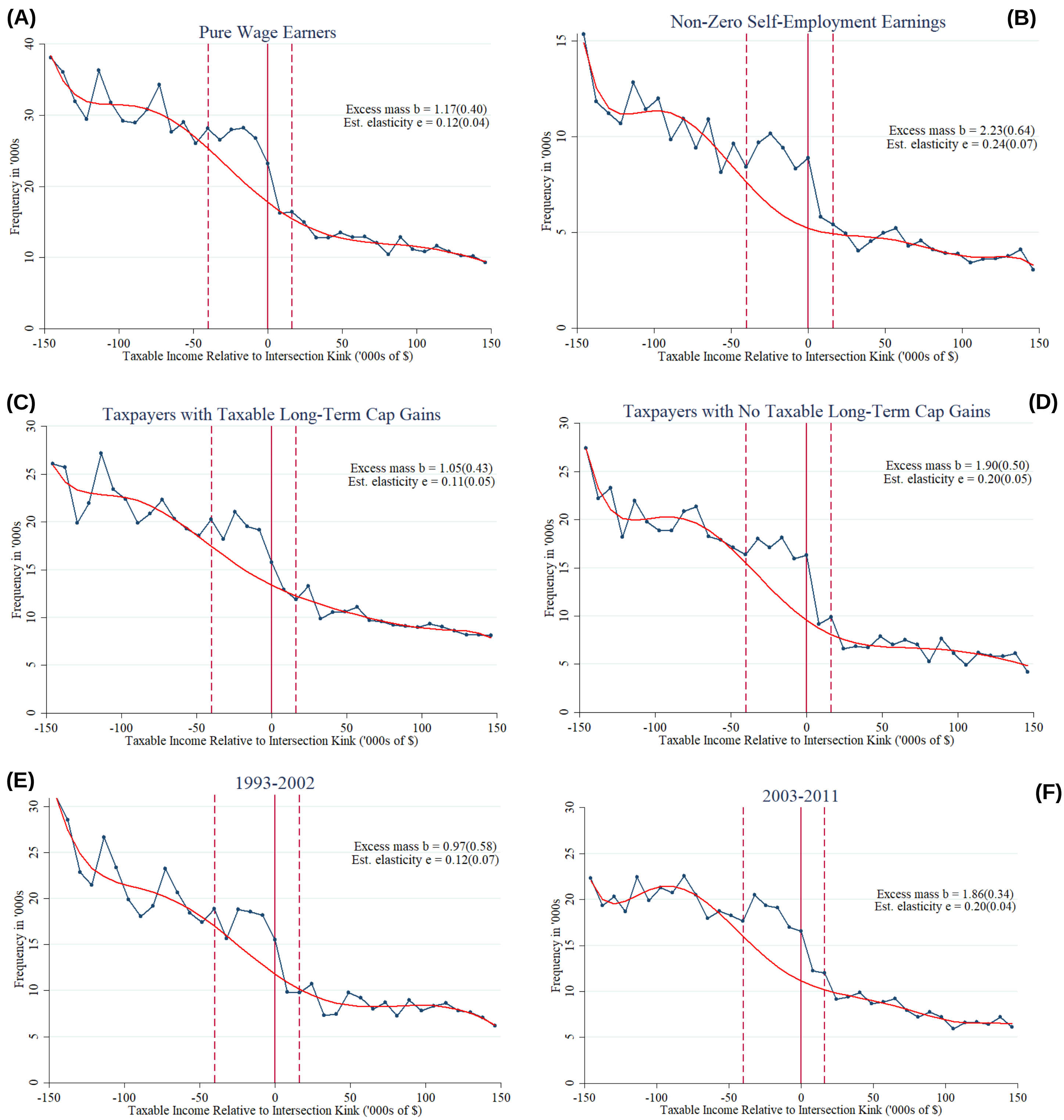


Figure 11: Distribution of Observed versus Counterfactual Taxable Income (Disaggregated)

Notes: This figure illustrates the distribution of the observed and counterfactual taxable income densities for pure wage earners (Panel A); taxpayers reporting any self-employment income (Panel B); taxpayers reporting long-term capital gains (Panel C); taxpayers reporting no long-term capital gains (Panel D); time period 1993-2002 (Panel E); and time period 2003-2011 (Panel F). The bold, vertical line (red) represents the centered intersection kink. Dashed, vertical lines (red) represent the lower and upper bounds of the bunching region defined as -$40,529 and $16,212. Observations are binned using binwidths of $8,106. A seventh-order polynomial is used to construct the counterfactual distribution. Estimates for excess mass and the ETI are provided at the top-right for each figure. Bootstrapped SEs are shown in parentheses.

There are three institutional features that can explain the difference in the elasticity estimates across the two time periods. First, key features of the AMT schedule including the fixed deduction amount and marginal tax rates remained consistent from 1993 to 2000, with a slight increase in the fixed deduction amount for years 2001 and 2002. On the other hand, Congress increased the fixed deduction amount annually from 2003 to 2011, reducing certainty around the future AMT structure. Learning effects would predict that the inconsistency of the AMT schedule from 2003-2011 would lead to taxpayers optimizing behavior around the top, intersection kink with increased informational frictions leading to reduced bunching. On the other hand, a stable policy environment allows taxpayers to gradually learn how best to optimize their economic and taxpaying behavior. This learning is also a function of the diffusion of information about avoidance strategies across taxpayers about a tax policy (Chetty, Friedman & Saez; 2013). Such diffusion of information plausibly slows down when policies change rapidly. Similarly, adjustment and search costs can also hamper optimization of real labor supply and tax strategies (Chetty, Friedman, Olsen, and Pistafferi, 2011). In terms of real outcomes, adjustment costs of switching jobs (extensive margin) or altering hours worked (intensive margin) in response to rapidly changing marginal tax rates are plausibly prohibitive (Gelber, Sacks & Jones; 2020). Tax strategies to maximize taxpayer utility can also take time to devise and implement. For example, income-shifting across time by design will be observed in future time periods. On the other hand, tax sheltering such as a higher use of charitable contributions and mortgage interest deductions might be constrained in the current time period due to contractual obligations and housing market effects, respectively. If this is the case, then I would expect to find bunching behavior to be more pronounced between 1993-2002 when the AMT and regular income tax policies remained largely stable over time, and less so between 2003-2011 when policies changed more frequently.

Second, with the fixed deduction increasing substantially over the time period 2003-2011, the intersection kink shifted to the right relative to 1993-2002, thereby affecting individuals with higher earnings. Since higher-income taxpayers have enhanced ability to change their behavior in response to changing marginal tax rates at the top intersection kink, it is possible that the bunching response would be higher in the second time period. Third, improvements in tax technology[17] used by taxpayers across time can reduce optimization frictions, leading to increased bunching in later years. Observed bunching responses shown in Panels E and F suggest that the aggregate effects of the second and third institutional features outweigh learning effects related to the first institutional feature, leading to the estimated ETI in the second time period being higher than in the first time period.

Table 3 summarizes the average ETI estimate for high earners and the estimate ETI for different subpopulations. The average ETI estimate is 0.15. The estimate is higher at 0.25 for self-employed individuals, and lower at 0.12 for pure wage earners. For taxpayers unaffected by the complexity of the long-term capital gains schedule and therefore, corresponding to the cleanest estimates, the estimated ETI is 0.20. These estimates are statistically significant at the 99 percent confidence level, except for the estimate on taxpayers reporting long-term capital gains, for whom the estimate is statistically significant at the 95 percent confidence level.

[17] "Tax technology" here refers to digital tools that allow for fast and flexible planning of annual income flows to minimize tax liability; gradual improvement of abilities of tax accountants and advisors; and an increase in information flow regarding tax avoidance strategies and easier access to such information via the internet.

Table 3: Summary of Elasticity Estimates for High Earners

Years	MTR Change (%)	All Filers (1)	Self-employment Income (2)	Wage earners only (3)	Positive Long-Term Cap Gains (LTCG) (4)	Non-positive LTCG (5)
1993-2011	28 to 37.3	0.15***	0.25***	0.12***	0.11**	0.20***
		(0.04)	(0.07)	(0.04)	(0.05)	(0.05)
1993-2002	28 to 39.5	0.12*	0.25**	0.08	0.11	0.14
		(0.07)	(0.11)	(0.07)	(0.09)	(0.11)
2003 - 2011	28 to 35	0.20***	0.26***	0.19***	0.13***	0.28***
		(0.04)	(0.07)	(0.04)	(0.04)	(0.06)

Notes: The table presents estimates of the ETI for high-income taxpayers based on bunching evidence around the top kink in the combined schedule as described in Sections 1.3 and 1.5.2 in the text. The time period 1993-2002 covers two tax acts (OBRA 1993 and EGTRRA 2001) and the time period 2003-2011 covers JGTRRA. The marginal tax rate change relates to the tax rates on either side of the top kink in the combined schedule. For 1993-2011 and 1993-2002, the top marginal tax rates are calculated as the average of the top tax rate in the years during those time periods, weighted by the number of years for which a tax rate applied. The subpopulation with self-employment income (column 2) is defined as individuals who reported any earnings from sole proprietorships, partnerships, S-Corporations, or farming. Wage earners only (column 3) are defined as individuals who did not receive any such self-employment income. The subpopulation reporting capital gains (column 4) is defined as individuals who reported any non-zero long-term capital gains between 1993-2011. Non-positive long-term gains (column 5) relate to individuals who did not report any long-term capital gains. Bootstrapped SEs are reported in parentheses.

As compared to the average ETI estimate of 0.15, the ETI estimates tend to be lower in the first time period. The average ETI estimate for the time period 1993-2002 is 0.12, as compared to 0.20 between 2003 and 2011. Trends within the two time periods are similar to those found in the entire time period: self-employed individuals respond more than wage earners, and the estimated ETI for individuals not reporting long-term capital gains is higher than the average elasticity estimate. Between 1993-2002, only the estimates for the average ETI and the ETI estimate for the self-employed are statistically significant at the 90 percent and 95 percent confidence levels, respectively. The time period 2003-2011 provides evidence of much higher responsiveness of high earners to the top kink in the combined schedule. The average ETI estimate is 0.20, with the self-employed continuing to respond more than wage earners. In fact, the highest estimate comes from high earners not reporting long-term capital gains, for whom the estimated ETI is 0.28. This is the cleanest estimate for the time period 2003-2011, given that it avoids the additional complexity of the capital gains schedule. All estimates for time period 2003-2011 are statistically significant at the 99 percent confidence level. I illustrate these results in Figure 12, together with the confidence intervals corresponding to each estimate.

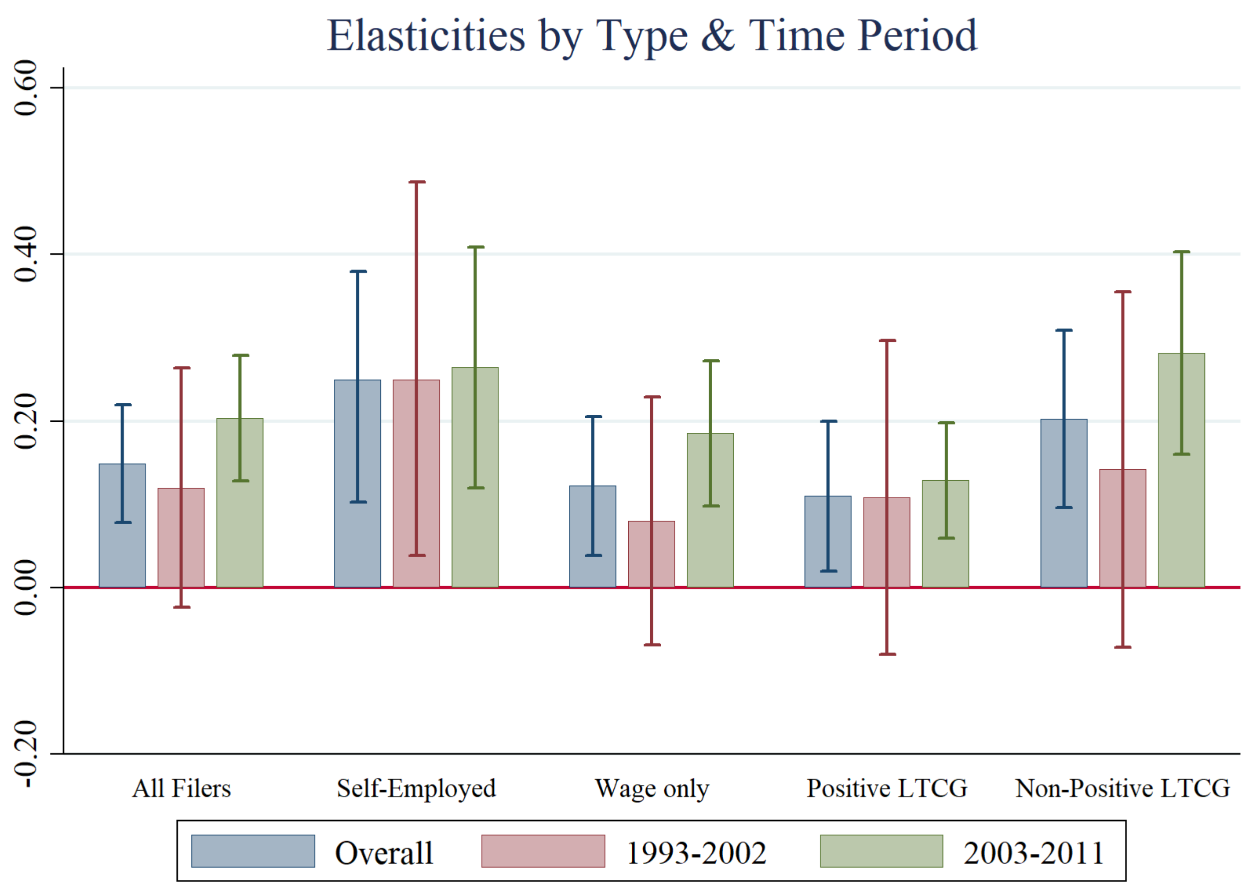


Figure 12: Elasticity Estimates for High Earners by Population Type

Notes: The figure presents the ETI estimates for high income taxpayers, disaggregated by type and time-period. The first bar in each case is related to the overall time-period; the second bar corresponds to 1993-2002; and the third bar is related to 2003-2011. The 95 percent confidence interval for each estimate is also indicated. The self-employed are defined as individuals who reported any earnings from sole proprietorships, partnerships, S-Corporations, or farming. Wage only individuals comprise taxpayers who did not receive any self-employment income. The subpopulation labeled "positive long-term capital gains" (LTCG) is defined as individuals who reported any non-zero long-term capital gains. Individuals with non-positive long-term gains (Non-Positive LTCG) are defined as taxpayers who did not report any long-term capital gains.

In Section 1.7, I use the average estimated ETI of 0.15 for the overall sample, and the cleanest estimates from the most recent time period of 0.28 to estimate efficiency costs and optimal top marginal tax rates.

1.6 Sensitivity to Model Parameters

In this section, I test the sensitivity of my average ETI estimate to the choice of the binwidth and the bandwidth. Choosing the binwidth involves a tradeoff between noise and bias. Recall that I use an IRS-provided sample of income tax return data. For this sample, the fraction of high earners in the population is low. This is illustrated by Figure A.2 in Appendix A, which shows the distribution of regular taxable income, truncated at $10,000 and $1 million. It resembles a Pareto distribution, with a thin right-tail with a smaller fraction of individuals at higher income levels providing for fewer observations to use when estimating the ETI. Further, my analysis window comprises the distribution of taxpayers with taxable incomes relative to the intersection kink point. This window comprises a further subsample of the data. Choosing a binwidth that is too small risks generating noisy estimates, with inference significantly affected by increased variance.

While I use a data-driven binwidth of $8,106 for my estimate of the average ETI for high earners, I assess the sensitivity of this estimate to varying binwidths. To do so, I pick binwidths between $500 and $12,000, reconstruct the counterfactual density and find estimates for the ETI corresponding to each binwidth. I plot these estimates and their 95 percent confidence intervals in Figure 13. It is reassuring to see that the average ETI estimate for high earners remains stable over the range of binwidths considered.

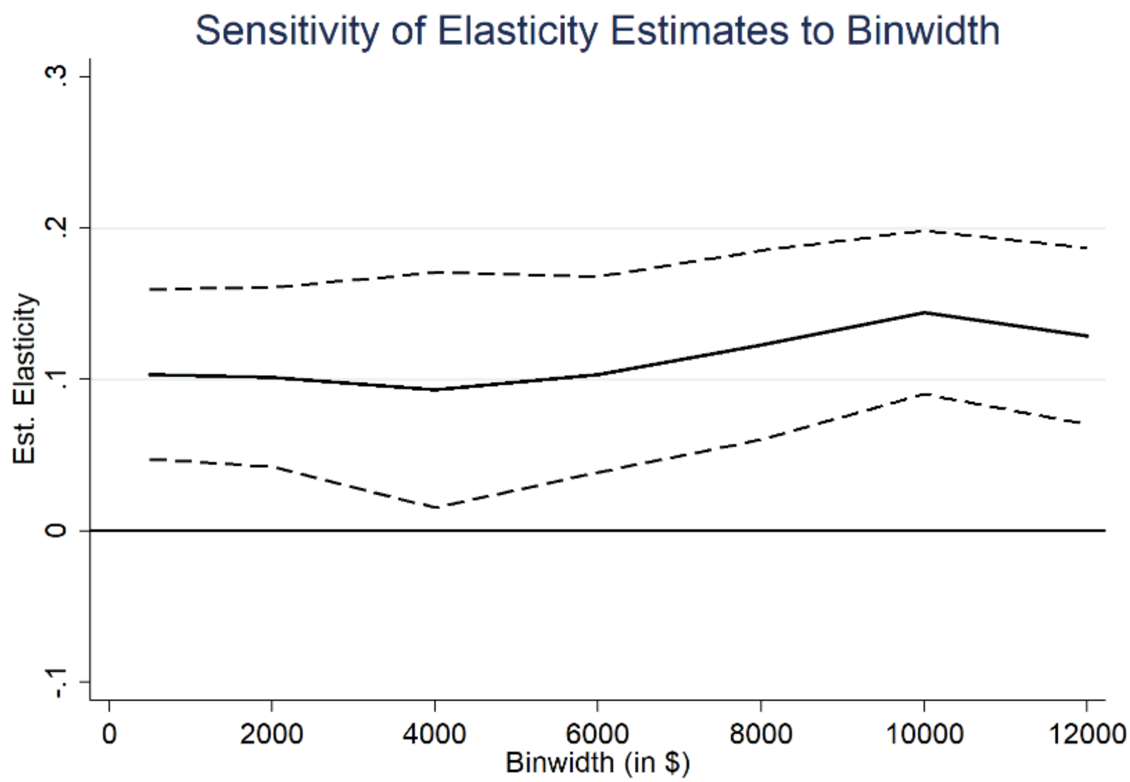


Figure 13: Sensitivity of Elasticity Estimates to Binwidth

Notes: This figure presents the sensitivity of the estimated ETI for high earners to the choice of binwidth. The solid line represents the estimated ETI corresponding to each binwidth at $2,000 intervals. The dashed lines capture the confidence interval related to the estimated ETI across binwidth size.

The choice of bandwidth relates to the decision regarding the size of the bunching region. Unlike regression discontinuity designs where a smaller bunching region minimizes bias by ensuring that the treatment and control groups on either side of a given cutoff are similar to each other, bunching estimators leverage the manipulation itself around the cutoff point. The cutoff point in the current context is the top, intersection kink in the combined schedule. If the bandwidth is too narrow, then some bunching that represents strategic decision-making by taxpayers might lie outside the assumed bunching window and not be captured, leading to the ETI to be underestimated. On the other hand, a wide bandwidth might a) capture part of the distribution where no manipulation due to the kink point is taking place leading to the

ETI being biased in an unknown direction, or b) capture manipulation related to other parts of the tax schedule, such as other kink points leading to overestimation of the elasticity parameter.

Besides using the algorithm developed by Bosch et al. (2020), I test the sensitivity of the average ETI estimate for high earners to different bandwidths. I do this for the entire time period, where the marginal tax rate on either side of the top kink in the combined schedule was 28 percent on the left, and 37.5 percent (weighted average across years) on the right of the kink point. I hold the binwidth constant at $8,106 to assess the sensitivity of my estimates exclusively to changes in the bandwidth. Visual evidence suggests that taxpayers' bunching response is captured within two bins to the right of the kink point. Therefore, I only extend the bandwidth to the left of the kink point for the sensitivity analysis presented here.

Table 4 shows the results of this sensitivity check. A smaller bunching window cuts into visually observed bunching, plausibly excluding excess mass outside the assumed bunching window, leading to the ETI parameter being underestimated. Compared to the average ETI estimate of 0.15 obtained with the use of the base bandwidth, the estimated ETI is 0.12 when a smaller bandwidth is used as shown in the first row of Table 4. Increasing the bunching window, however, does not affect the average elasticity estimate. This holds true if there is no strategic bunching in other parts of the distribution within the analysis window, or if other bunching cancels out in the aggregate. In such a case, increasing the bunching region would not impact the estimation of the ETI parameter.

Table 4: Sensitivity of Elasticity Estimates to Bandwidth

Years	**MTR Change**	**Binwidth ($)**	**Bunching region ($)**	**All Filers**
1993-2011	28% - 37.5%	8,106	-32,423, +16,211	0.12***
				(0.03)
			-40,529, +16,211	0.15**
				(0.04)
			-48,634, +16,211	0.15***
				(0.04)

Notes: The table presents the estimates of the ETI for high-income taxpayers under varying bandwidth sizes. The marginal tax rate on either side of the top kink in the combined schedule is 28 percent on the left, and 37.5 percent (weighted average across years) on the right. The binwidth is held constant at $8,106. Bandwidths to the left of the top kink in the combined schedule increase in absolute terms from -$32,423 to -$48,634. The estimated ETI corresponding to each bandwidth is provided in the last column (column 5). Bootstrapped SEs are shown in parentheses.

1.7 Efficiency Cost and the Optimal Top Marginal Tax Rate

The elasticity of taxable income with respect to the net-of-tax rate provides key insights into the responsiveness of taxpayers to changing marginal tax rates. It also serves as a core parameter for estimating the efficiency cost of taxation and for conducting welfare analyses. In fact, assuming no externalities and market failures and under negligible income effects, the elasticity parameter serves as a sufficient statistic for estimating efficiency costs of taxation (Feldstein, 1999). I return to the tenability of these assumptions at the end of this section.

Simplifying the model in terms of behavioral and mechanical costs of taxation, Saez, Slemrod and Giertz (2012) discuss how the literature has evolved to show that the marginal deadweight burden (MDB) or marginal excess cost of funds (MECF) is equal to $1 - dB/dR$, where dB is the extra amount of utility lost over and above additional tax revenue collected through a tax increase. dR is overall change in tax revenue due to a tax increase. This translates to:[18]

$$MECF(\tau, \varepsilon, \alpha) = \frac{1 - \tau}{1 - \tau - \varepsilon.\alpha.\tau} \tag{8}$$

where τ is the prevailing top marginal tax rate, ε is the elasticity of taxable income, and α is the Pareto parameter. Greater responsiveness of taxpayers to higher marginal tax rates corresponds to higher values for the ETI parameter, ε. A larger behavioral change is economically more distortionary than a small change. Therefore, efficiency costs increase in ε. Similarly, as the marginal tax rate τ increases, the loss of social utility to the taxpayer at the margin increases, leading to higher efficiency costs to the economy. The right-tail of the income distribution can be shown to be Pareto distributed. The Pareto parameter, α, estimates the thickness of the right tail. A thicker (thinner) right tail corresponds to a lower (higher) α. The thinner the tail and higher the Pareto parameter, the higher the efficiency costs. This is because with a thinner tail, the loss of social utility on the margin is greater than inframarginal revenue gains in the right tail due to marginal tax rate increases.

I use my estimates for the average ETI for high earners corresponding to the overall sample, and for the sample unaffected by the additional complexity of the capital gains schedule in the second half of my analysis time-period to estimate efficiency costs using the above formula. I assume the α parameter to be equal to 1.5 based on the analysis of the US income distribution conducted by Piketty and Saez (2003). Similar to Saez et al. (2012), I

[18] See Saez et al. (2012) for the detailed simplification of the formula.

assume the average top state income tax rate to be 5.9 percent, the Medicare payroll tax rate to be 2.9 percent, and the average sales tax rate to be 2.3 percent. The weighted average of the top marginal tax rate for my analysis period is 37.3 percent. Considering the deductibility of state income taxes from the federal income tax schedule and the deductibility of the Medicare payroll tax from both state and federal income tax schedules, I estimate the average aggregate top marginal tax rate to be 44.8 percent. This leads to the following marginal excess cost of funds:

$$MECF(0.448, 0.15, 1.5) = \frac{1 - 0.448}{1 - 0.448 - (0.15 * 1.5 * 0.448)} \approx 22\%$$

This estimate for the efficiency cost of income taxation suggests that an additional dollar of income tax collected generates an efficiency cost of 22 cents. Similarly, using the ETI parameter for taxpayers with no long-term taxable capital gains in the more recent time period covering years 2003 to 2011, where the overall top marginal tax rate is 42.5 percent due to a lower federal top marginal tax rate of 35 percent, I estimate the efficiency cost to be 45 cents for an additional dollar collected in tax revenue:

$$MECF(0.425, 0.28, 1.5) = \frac{1 - 0.425}{1 - 0.425 - (0.28 * 1.5 * 0.425)} \approx 45\%$$

By combining efficiency costs with social welfare weights, I can use the estimated taxpayer behavioral response to conduct welfare analyses. The estimated ETI is a key ingredient in the estimation of optimal top marginal tax rate. Building on the subliterature on optimal top marginal tax rates initiated by Mirrlees (1971), Diamond and Saez (2011) show that under the mechanical and behavioral effects on revenue of increasing taxation and under a quasi-linear utility function increasing in consumption but decreasing in effort, the optimal top marginal tax rate can be represented by:

$$\tau^* = \frac{1 - \bar{g}}{1 - \bar{g} + (\alpha * \varepsilon)} \tag{9}$$

where $\bar{g}$ is the weighted average of the social marginal weights (g_i) for high-income taxpayers. The social marginal weight g_i can be thought of as the social marginal value of providing an additional dollar of consumption to individual i. Under a Rawlsian social welfare function where social marginal weights are concentrated at the bottom of the income distribution, $\bar{g} \approx 0$. Under a utilitarian social welfare function with concave utility functions, the social marginal value of consumption for high earners decreases rapidly at the top of the income distribution, also approaching zero. Therefore, (9) simplifies to:

$$\tau^*(\alpha, \varepsilon) = \frac{1}{1 + (\alpha * \varepsilon)} \tag{10}$$

I plug the Pareto parameter of 1.5 and my two main estimates of 0.15 and 0.28 for the ETI of high earners into (10) and find corresponding optimal top marginal tax rates of 82 percent and 70 percent, respectively:

$$\tau^*(1.5, 0.15) = \frac{1}{1 + (1.5 * 0.15)} \approx 82\%$$

$$\tau^*(1.5, 0.28) = \frac{1}{1 + (1.5 * 0.28)} \approx 70\%$$

The range of estimated optimal top marginal tax rates coincides with emerging research on this question (Diamond and Saez, 2011; Saez, Slemrod and Giertz, 2012; Piketty, Saez and Stantcheva, 2014).

The formulae used to estimate efficiency costs and optimal top marginal tax rates here make relatively strong assumptions related to externalities, long-term responses to taxation, and the type of costs associated with taxation. For example, increasing marginal tax rates along the income tax schedule can cause individuals to shift their income across tax bases in search of lower-taxed income streams. It is also possible that if income realized at a future point in time is taxed at a non-zero rate that is different from the current rate, then taxpayers' retiming of income gains can create a wedge between short-run and long-term elasticities. This wedge will affect the estimation of efficiency costs in the current time period. Such fiscal externalities (Saez et al., 2012) lead to some of the efficiency costs of income taxation to be recouped on other tax schedules or across time, leading to higher optimal top marginal tax rates.

A similar argument can be made for classical externalities. Individuals avoiding taxes via charitable giving or increased mortgage interest deduction amounts can generate externalities for other economic agents, reducing the efficiency cost of income taxation. Further, Chetty (2009) argues that if the costs of taxation are not purely real resource costs but instead include transfers to other agents as well – say, via tax penalties imposed for illegal tax avoidance or evasion that are redistributed to other agents – then the ETI parameter is not sufficient for estimating efficiency costs.

The study of these externalities is beyond the scope of this paper. However, in the presence of externalities that can offset efficiency costs, my estimates for the ETI of high-income taxpayers suggest a lower bound on the efficiency cost of 22 cents per dollar of additional tax revenue collected. This implies a lower bound on the optimal marginal tax rate of 70 percent, much higher than the prevailing, effective top marginal tax rate at the federal level.

1.8 Conclusion

The standard bunching approach to measuring the ETI for high earners in the United States has been to construct the federal regular income tax schedule, overlay the distribution of taxable income across it, and then use taxpayer bunching responses around kink points to estimate the elasticity. This paper argues that the regular income tax schedule is not the correct schedule for estimating the elasticity for high earners. High-income taxpayers face an effective tax schedule that is the upper bound of the interaction of the piecewise linear regular income tax and AMT schedules. This combined schedule is what taxpayers respond to when optimizing taxpaying behavior and therefore should form the underlying tax schedule used in bunching studies for higher earners. The use of the combined schedule for analysis of high earners' behavior resolves the inconsistency between previous elasticity estimates that found substantial responsiveness amongst low-income taxpayers but no responsiveness amongst high-income taxpayers (Saez, 2010; Mortenson and Whitten, 2016)

I characterize this combined schedule and highlight its properties. The combined schedule allows me to capture larger bunching responses at its top kink . The combined schedule contains its own kink points that do not necessarily align with kinks in the regular income tax schedule when the latter is considered in isolation. In fact, the intersection kink in the combined schedule – the point where the regular income tax and AMT schedules intersect – provides a novel device for measuring taxpayer response. The intersection kink presents a larger change in the marginal tax rate, plausibly generating stronger taxpayer responses as compared to top kinks in the regular income tax schedule. Further, the location of the intersection kink varies for each taxpayer as compared to kinks in the regular income tax schedule that are fixed in taxable income. This variation provides me with estimates for the ETI that mitigate endogeneity concerns affecting earlier studies that use bunching methods on fixed kink points. By using the variation in the location of the top, intersection kink in the combined schedule, I disentangle variation in taxable income from variation in underlying preferences, increasing confidence in the ability of my estimator to capture the true ETI parameter.

Using publicly available IRS taxpayer microdata from 1993-2011, I find that the average ETI estimate for high earners in the United States is 0.15, as compared to earlier estimates in the literature that were close to zero. For the sample unaffected by the complexity of the capital gains schedule, the estimated elasticity is 0.20 – rising to 0.28 for the time period 2003-2011 when annual changes in the tax code shifted intersection kinks for taxpayers to higher parts of the income distribution. Self-employed individuals respond twice as much as wage earners, with an estimated elasticity of 0.24. However, wage earners also reveal non-trivial responsiveness to tax rates, with an estimated ETI of 0.12. This sheds light on the increased ability of taxpayers at the top of the income distribution to alter work hours, or to convert monetary compensation to fringe benefits. Back of the envelope calculations reveal efficiency costs bounded above at 45 cents per dollar of additional tax revenue collected and the estimated optimal top marginal tax rate bounded below at 70 percent.

The current analysis creates a range of avenues for future work. In the context of the United States, further analyses should examine the interaction of the regular and AMT schedules together with the capital gains schedule. Future work should also consider the dynamic responses of taxpayers by using tax panels available at the IRS, to a) unpack the mechanisms underlying individuals' responses to the combined schedule across time, and b) to shed light on short-term versus long-term responses to the combined schedule. Beyond the United States, this paper expounds the importance of considering details of the tax code that give rise to effective schedules with characteristics including kinks that are different from the primary tax schedule being considered. Such under-the-hood work is necessary for identifying the true incentive structure faced by taxpayers, when estimating taxpayer responsiveness to such incentives.

From a policy perspective, my results point to optimal top marginal tax rates that are higher than prevailing top marginal tax rates. The higher estimated ETI for self-employed individuals confirms the previously documented relationship between the absence of third-party reporting and higher tax avoidance behavior. And a comparison of the relationship between bunching responses and the size of the marginal tax rate change around kinks suggests that a larger number of income tax brackets with smaller marginal tax rate changes across brackets will reduce taxable income responses leading to lower efficiency costs of taxation.

2. Tax Reforms and Voting Behavior in the United States*

Ali Abbas†, David Albouyᴪ, and Michael Lovenheim⊥

Abstract

Do taxpayers vote in their economic self-interest? We examine the effect of changing income tax burdens on voting behavior across the United States. Such analyses are hindered by the endogeneity of tax burdens to tax policy and secular changes in income growth and party preference. To mitigate concerns about estimation bias, we use a novel simulated instrumental variable approach in conjunction with survey, administrative, and voting data at the presidential and House levels for the years 2010 to 2020. By accounting for secular trends in baseline demographics and isolating changes in tax burdens that arise purely due to variation in tax policy from changes caused by demographic shifts, we estimate that an increase in tax burdens by about half a standard deviation increases the vote share for the Republican party by one to six percentage points. This relationship is strongest, both statistically and in terms of magnitude, for presidential elections. For House elections, we find suggestive, but not definitive evidence that this relationship holds. Our analysis shows that contrary to popular belief, taxpayers continue to vote in their economic self-interest.

* We are grateful to Dirk Foremny, Nancy Chau, and Benjamin Leyden for their comments. This research also benefited from comments received at seminars at the Economics Department and the Dyson School of Applied Economics and Management at Cornell University. A special thanks to the participants at the National Tax Association Annual Conference 2020 for their valuable feedback.

† Charles H. Dyson School of Applied Economics & Management, Cornell University. Email: aa2378@cornell.edu
ᴪ Department of Economics, University of Illinois Urbana-Champaign. Email: albouy@illinois.edu
⊥ The Department of Economics, Cornell University. Email: mfl55@cornell.edu

2.1 Introduction

“Why people vote against their economic self-interest”, The Economist (2018) begins to explain, as it attributes the idea of voting against one’s self-interest to a lack of credibility of political candidates. “It was cultural anxiety that drove white, working-class voters to Trump”, claims The Atlantic (2017). The idea that voters increasingly vote against their economic self-interest is best captured by Thomas Frank in his New York Times bestseller (2004): “There is no bad economic turn a conservative cannot do unto his buddy in the working class, as long as cultural solidarity has been cemented over a beer.” In fact, as we discuss below, “self-interest” spans a range of objectives that a voter might wish to maximize.

The political economy literature has explored various economic and non-economic determinants of voters’ decisions at the voting booth. In this paper, we consider an important and understudied economic channel of voter decision-making: taxation that directly affects taxpayers’ disposable income. We study income tax burdens that include capital gains for taxpayers. Specifically, we provide evidence on the impact of federal taxation on voter choices in the United States, leveraging tax reforms between 2010 and 2020 including the American Taxpayer Relief Act of 2012 and the Tax Cuts and Jobs Act of 2017. This time period consists of two presidential election cycles and six House election cycles. We use voting data at the county level, administrative tax data from approximately 3,200 counties in the United States, and income and demographic data from the American Community Survey (ACS).

There are two key challenges of estimating a causal relationship between party vote shares and tax burdens across geographic units such as counties. First, observable and unobservable confounding factors can bias the estimated impact of tax burdens on party vote shares. An example of such an observable confounder is the share of college-educated in the population. As we show in Section 2.6.1, college share is positively correlated with incomes and therefore tax liabilities, and negatively correlated with Republican party vote shares. Moreover, the baseline level of college share affects how vote shares evolve with rising college shares. These secular trends in baseline demographics can result in misestimation of the relationship between tax burdens and party vote shares. We eliminate the effect of time-invariant unobservables by using a first-differencing empirical model, and control for both changes in demographic characteristics and linear time trends in the relationship between baseline demographics and party vote shares to control for key secular trends.

Second, we want to estimate the impact of changes in tax policy on party vote shares, mediated by changing tax liabilities. However, aggregate federal tax burdens at the county-level at any point in time are a function of both tax policy as well as county-level demographics, such as age, marital status, and number of dependents. To isolate the effect of tax policy on party vote shares, we use a novel simulated instrumental variable (IV) approach. Using the National Bureau of Economic Research's (NBER) tax simulator (TAXSIM), we simulate federal tax liabilities across our units of analysis, *holding baseline demographics constant*. The simulated tax variable partials out the effect of changing demographics and provides us with changes in tax burdens that are purely a consequence of tax policy changes.

We estimate the causal link between federal income taxation and voting choices, and find that as tax liability per capita increases by $1,000 between election cycles, the Republican vote share increases by one to six percentage points within our units of analysis. This result is strongest – both statistically and in terms of magnitude – for presidential elections as compared to House elections. Our key findings suggest that substantial heterogeneity notwithstanding, taxpayers on average continue to vote with their wallets.

We make two key contributions in this paper. First, we provide the first examination in the literature of how actual changes in tax burdens affect voting behavior. Our work fits into the literature that assesses the retrospective determinants of voting behavior. Prior work in this area has considered the impact of fiscal expansions in the form of higher supply of public goods or transfers (Brender & Drazen, 2008), economic shocks such as changes in employment prospects (Brunner, Ross, & Washington, 2011), wealth shocks such as winning high-stakes lotteries during election years (Bagues & Esteve-Volart, 2016), federal outlays (Levitt & Snyder Jr, 1997), or outright cash transfers (De La O, 2013; Manacorda, Miguel, & Vigorito, 2011; Zucco Jr, 2013). The literature on the retrospective link between government policies and voters' decisions at the voting booth is silent on the direct impact of taxation resulting from tax cuts or hikes. We plug this gap in the literature. Our analysis is also different from other work that considers prospective determinants of voting behavior such as campaign promises (Levy, 2020), since we leverage the impact of observed changes in tax burdens on voting behavior, not just what's promised by political candidates during election campaigns.

Second, we contribute to the literature on the use of simulated instrumental variables to study government programs, taxation and fiscal transfers. Previous work that has used such instruments include the literature on healthcare (Currie & Gruber, 1996; Cutler & Gruber, 1996; Kalíšková, 2015), and taxpayer responses to tax and transfer programs (Dahl &

Lochner, 2012; Gruber & Saez, 2002; Moffitt & Wilhelm, 2000). Our paper is the first, to the best of our knowledge, to leverage a simulated tax instrument to the political economy literature on voting behavior. As we explain in Section 2.4, the use of this instrument is critical to our identification strategy and avoids earlier pitfalls in studying the causal link between federal taxation and voting behavior in the United States.

2.2 Prior Literature

The hypothesis that politicians reward their political base through government redistribution is not new. This hypothesis is commonly associated with the "machine" politics of nineteenth century urban polities. However, researchers have generally hypothesized that such redistribution happens through targeted spending – namely, "pork-barrel" politics – rather than through the income tax system. Cox and McCubbins (1986) create a game-theoretic model whereby "Politicians will adopt strategies in which they invest little (if at all) in opposition groups, somewhat more in swing groups, and more still in their support groups." Dixit and Londregan (1996) formulate a more general model, whereby swing voters may receive greater transfers than core support groups depending on how apolitical (or greedy) voters are along the ideological spectrum.[19] Thus, it is far from obvious that politicians will reward their base. It is even possible that if most voters are rather ideological and polarized, politicians may even create redistributive schemes that penalize their base in their pursuit of swing voters.

[19] McCarty (2000) develops a legislative bargaining model with a president's veto powers, showing that the latter is effective in allocating spending toward districts favored by the president.

The empirical research on political parties and redistribution has offered a number of related findings, although none, to our knowledge, cover targeting through tax reforms. Using a panel from 1982 to 2000, Larcinese et al. (2006) find that U.S. states that heavily supported the incumbent president in past presidential elections tend to receive more in federal funds, while marginal and swing states are not rewarded. Further, examining federal spending, Berry et al. (2010) find that U.S. House districts whose Representatives are of the same party as the U.S. President receive 5 percent more in funding from high-variation federal programs. Albouy (2013) finds that states with Senate delegations in the majority party receive greater federal grants and defense spending. Overall, these studies do support the hypothesis that politicians redistribute spending towards their constituents, although these studies say little about taxes.[20]

2.3 Data

We use data from three sources: the Statistics of Income (SOI) division at the Internal Revenue Service (IRS); the American Community Survey (ACS); and voting data on presidential and House elections from David Leip's Atlas.

SOI Income Tax Data: The SOI provides annual income tax data at the county level. The range of variables provided by the SOI was expanded substantially in 2010. Prior to 2010, the IRS published county level tax data on six key variables, including the number of tax returns, the number of tax exemptions taken by taxpayers, number of taxpayers within

[20] Examining taxes in a panel of U.S. states, Reed (2006) finds that overall tax burdens are higher when Democrats control the state legislature, although governors have little effect.

different bands of adjusted gross income (AGI), wages and salaries, dividends, and interest. From 2010 onwards, this list was expanded to cover approximately 70 variables, including among others, the number of dependents, business or professional net income, net capital gains, and itemized deductions such as state and local income taxes, mortgage interest payment, and real estate taxes. This expansion provides us with the means to capture the details of income taxation in detail at the country level from 2010 to 2018.

ACS Income and Demographic Data: The ACS is conducted by the Census Bureau of the US federal government. The ACS collects information on demographic, housing, social, and economic characteristics of the population. We use the 1-in-100 representative ACS 1-year public use microdata samples (PUMS) from 2010 to 2020 and use this data with the National Bureau of Economic Research's (NBER) tax simulation model (TAXSIM) to simulate tax liabilities for each observation in the sample.

TAXSIM requires inputting of 22 variables, including temporal and geographic identifiers such as year and state, filing status, age of tax filers, the number of dependents and children, incomes, and expenses that can be deducted against income for federal tax purposes. The types of income required for accurately simulating liabilities include salaries and wages, dividends, interest received, short- and long-term capital gains, pensions, social security benefits, and unemployment compensation. Types of expenses include mortgage interest payments, property taxes paid, and childcare. Using this data, TAXSIM estimates individual-level federal and state tax liabilities.

Observations in the ACS are identified at the Public Use Microdata Area (PUMA) level. PUMAs are survey-specific areas containing populations of 100,000 or more. Since PUMA definitions were changed for the 2010 decennial census and applied to ACS data from 2012 onwards, we crosswalk all PUMAs for data from 2012-2020 to PUMA definitions as of 2010 for comparability across years. Crosswalks were acquired from the Missouri Census Data Center (MCDC) Geographic Correspondence Engine (Geocorr).

Further, there is no uniform relationship between PUMAs and counties. A PUMA can represent one county or be partitioned into multiple counties. Conversely, multiple PUMAs can add up to form one county. Lastly, PUMA-county mapping can also be complex when multiple PUMAs can map on to multiple counties. Since we combine IRS SOI income tax data that is identified at the county level with estimated tax data simulated from the ACS at the PUMA level, we construct a max(PUMA, County) identifier. This identifier aggregates information up to the larger (in terms of population) of the PUMA and county levels for different geographic areas, in essence, partitioning the United States into a set of 1,132 unique max(PUMA, County) units. By targeting the higher of the two, we are also able to reduce noise in our estimates of the relationship between federal income taxation and voting behavior. Details of the construction of the simulated income tax variable using ACS data are provided in Appendix B.1.

Presidential and House Elections Data: We obtained voting data for presidential and House races from David Leip's Atlas. Presidential election data for the 50 states and Washington D.C. spans three electoral cycles with winners in parentheses: 2012 (Democrat), 2016 (Republican) and 2020 (Democrat). House election data for the 50 states covers six electoral cycles: 2010, 2012, 2014, 2016 and 2020.

The voting data includes information on the total number of votes cast per county, and the number of votes cast for Democratic, Republican, and other nominees. To convert presidential and House voting data from the county to the max(PUMA, County) level, we crosswalk the data using the MCDC county-PUMA crosswalk, maintaining 2010 PUMA boundaries. Once we have the total number of votes cast and the number of votes cast for Republican candidates by max(PUMA, County), we estimate the share of votes cast for Republican candidates.

We merge SOI income tax data, ACS-based simulated income tax data, and voting data for presidential and House elections at the max(PUMA, County) level to generate a dataset containing 12,452 observations, with 1,132 pairs between max(PUMA, County) and years.

Demographic Controls: We control for a parsimonious set of demographic variables describing education level (percentage of the population with a college degree), racial diversity (percentage of population that is non-white), age (percentage of population that is 65 years of age or above), and sex (percentage of population that is female). We use ACS public use data provided at the PUMA level, and re-express this information at the max(PUMA, County) level. We also control for base year population density. Population density is defined as the total population in a max(PUMA, County) divided by the land area in square miles. Data for the numerator and denominator are obtained from the decennial census of 2010 conducted by the Census Bureau.

The average max(PUMA, County) has 16 percent of individuals who are non-white, is 51 percent female, with 20 percent college graduates and 14 percent of individuals who are 65 years of age or above. However, the Table 5 shows that there is wide variation around these means.

Table 5: Summary Statistics for Demographic Controls

Variable	# of Max(PUMA, County)	Mean	Std. Dev.	Min	Max
College Grads (%)	1,132	20.4	6.7	9.7	60.6
Non-White (%)	1,132	16.1	14.3	0.5	77.8
Age > 65 Years (%)	1,132	14.3	3.3	6.3	36.2
Female (%)	1,132	50.6	1.3	40.2	53.8

Voting data is available up till 2020, ACS data is available up till 2019, and SOI tax data is available up till 2018. For analysis of the 2020 election cycle, we inflate 2019 ACS data by the CPI-U inflation factor. Since SOI tax data is missing for both 2019 and 2020, we only focus on the reduced-form specifications using ACS data for the 2020 electoral cycle, as discussed in Sections 2.4 and 2.6.

2.4 Empirical Methodology

We estimate the impact of tax liability within max(PUMA, County) on the Republican vote share for presidential and House elections. We explore initial trends in cross-sectional data by election year, and then correct for endogeneity concerns by running first-difference, instrumental variable models.

A. Cross-Sectional Trends in Republican Vote Share and Tax Liability

We first run cross-sectional specifications for each year in our data to assess correlations between tax burdens and Republican vote shares. The cross-sectional estimating equation for both types of elections is:

$$V_{i,t} = \alpha + \beta_1 TL_{i,t} + \beta_2 X_{i.t} + \epsilon_{i,t}$$

The dependent variable $V_{i,t}$ is the Republican vote share (in percentage points) in max(PUMA, County) i and year t, where $\{t: 2010 \leq t \leq 2020\}$. The key variable of interest $TL_{i,t}$ is SOI observed income tax liability per capita. $X_{i,t}$ is a vector of controls including percentage of population with college degree, percentage non-white, percentage aged 65 or above, and percentage female. The stochastic error term is represented by $\epsilon_{i,t}$.

An issue for estimation is that income tax liability is likely correlated with observed and unobserved characteristics across max(PUMA, County) units, that are in turn, correlated with Republican vote shares. For example, if aggregate incomes are higher in Democratic-leaning states mechanically resulting in higher tax liabilities and lower Republican vote shares, then the cross-sectional specification will understate the impact of tax liability on Republican vote share. We control for observables such as college education but cannot control for time-varying or time-invariant unobservables in the cross-sectional specification.

B. Using Variation in Tax Liabilities Across Time: First-Difference Model

We mitigate bias in the cross-sectional relationship above by measuring the relationship between changes in Republican vote share and income tax liability within max(PUMA,County) units over time. We use a first-difference (FD) specification to eliminate the effect of any time-invariant confounding factors. There are two estimation concerns with the FD approach. First, secular changes in income growth and party preference can confound our estimated relationship between Republican vote shares and tax liabilities. For example, changes in college education are correlated with increasing income levels, and therefore, with increasing tax liabilities under the same tax policy parameters. However, an increase in college education is also correlated with decreasing Republican vote shares, with the baseline level of college education affecting how this trend affects the evolution of Republican vote shares. To control for such secular changes, we use a model of the form below, where we control for a linear time trend and interact it with a parsimonious set of baseline demographic controls:

$$V_{i,t} = \alpha + \beta_1 TL_{i,t} + \beta_2 X_{i.t} + \beta_3 X_{i,t0} + \gamma t + \delta X_{i,t0} * t + \epsilon_{i,t}$$

The dependent variable $V_{i,t}$ and key variable of interest $TL_{i,t}$ are the same as in the cross-sectional specification. The additional vector $X_{i,t0}$ comprises baseline demographic controls. For House elections, $X_{i,t}$ also contains a variable capturing incumbency. The time trend is represented by variable t. Taking first-differences, we obtain:

$$\Delta V_{i,t} = \alpha + B_1 \Delta TL_{i,t} + B_2 \Delta X_{i,t} + \beta_3 X_{i,t0} + \eta_{i,t}$$

The above specification captures the impact of changes in tax liabilities on changes in Republican vote shares, controlling for changes in demographic characteristics and secular trends in demographic shifts and party preference.

C. Endogeneity of Tax Burden: FD Model with a Simulated Instrumental Variable (IV)

The second estimation issue that we address as a key value add of our study, is the impact of demographic shifts on tax liability. We want to measure the impact of tax liabilities *as a result of changing tax policy* on the Republican vote share. $\Delta TL_{i,t}$ above combines changes in tax liability due to policy changes and changes that occur due to demographic shifts across time. These demographic changes include changes in age, marital status, labor force participation, and the number of dependents, among other things that directly feed into the tax liability function. Migration can also change the demographic composition of our units of analysis. Such composition shifts cause tax liability to be endogenous to tax policy. To mitigate this endogeneity concern that can bias results in an ambiguous direction, we use a simulated instrumental variable strategy. The simulated instrumental variable isolates the effect of policy reforms from the effect of demographic shifts on tax liabilities in our units of analysis. We sketch the decomposition as follows:

$$TL^1 - TL^0 = \tau^1(I^1) - \tau^0(I^0)$$

$$= [\tau^1(I^0) - \tau^0(I^0)] - [\tau^1(I^1) - \tau^1(I^0)]$$

TL^t indicates tax liability at a given point in time t. $\tau^t(I^t)$ indicates tax liability due to tax policy τ^t at time t with demographic composition I^t at time t. In the second equation, the first expression in parentheses is the change in tax liability purely due to tax policy, holding baseline demographics constant. The second expression in parentheses is the effect of pure demographic shifts. We use NBER's TAXSIM to estimate $\tau^1(I^0)$.

For any given year t and for each max(PUMA, County) i, the simulated income tax liability per capita variable is defined by $\tilde{\tau}_{i,t}$. This variable simulates the income tax liability in max(PUMA, County) i in year t, where $t = [2010, 2020]$, holding base year 2010 demographics constant. The simulated variable captures the portion of tax liabilities in our units of analysis that are solely a result of tax policy, independent of demographic shifts across time. We predict the part of observed tax liability that arises purely due to tax policy changes with the help of the simulated tax variable, isolate it and use it in our instrumental variable approach. The following are the reduced-form and first-stage specifications using the simulated tax variable.

Reduced-Form:

$$\Delta V_{i,t} = \alpha + B_1 \Delta\tilde{\tau}_{i,t} + B_2 \Delta X_{i,t} + \beta_3 X_{i,t0} + \eta_{i,t}$$

$\Delta\tilde{\tau}_{i,t}$ measures the change in tax liability within max(PUMA, County) i across years t that is solely a result of changes in tax policy. The other variables are the same as in our first-difference OLS estimator.

First-Stage:

$$\Delta TL_t = \alpha + \gamma_1 \Delta\tilde{\tau}_{i,t} + \gamma_2 \Delta X_{i,t} + \gamma_3 X_{i,t0} + v_{i,t}$$

The first-stage specification regresses the change in observed tax liability on the simulated tax (policy) variable, with the same controls as in the reduced-form specification. The IV estimator that isolates and estimates the effect of the change in policy-driven income tax liability per capita on Republican vote share is then $\frac{\beta_1}{\gamma_1}$.

We also estimate the change in tax liability that is purely due to demographic shifts by subtracting $\tilde{\tau}_{i,t}$ from $TL_{i,t}$, to generate the variable $\tilde{I}_{i,t}$, or:

$$\tilde{I}_{i,t} = TL_{i,t} - \tilde{\tau}_{i,t}$$

The reduced-form and first-stage specifications incorporating the pure demographic tax variable then become:

Reduced-Form:

$$\Delta V_{i,t} = \alpha + B_1 \Delta\tilde{\tau}_t + \beta_2 \Delta\tilde{I}_t + B_3 \Delta X_t + \beta_4 X_{t0} + \epsilon_t$$

First-Stage:

$$\Delta TL_t = \alpha + \gamma_1 \Delta\tilde{\tau}_t + \gamma_2 \tilde{I}_t + \gamma_3 \Delta X_t + \gamma_4 X_{t0} + \epsilon_t$$

In these additional specifications, we control for this demographic tax liability variable in addition to our other demographic controls to assess the impact on the sign and magnitude of the key variable of interest.

We estimate the relationship between tax burden and Republic vote share using two long panels for 2012-2016 and 2016-2020 that include three presidential elections held in 2012, 2016, and 2020. With $T = 2$, these results are equivalent to running fixed effects models, and so we estimate the relationship between the variation within max(PUMA, County) around the means for the variables in the specifications. We also estimate the above specifications for a stacked panel, comprising 2012, 2016, and 2020, where $T = 3$.

For House elections, we consider short panels in addition to similar long panels as above, and a stacked panel. Short panels capture electoral cycles 2010-2012, 2012-2014, 2014-2016, 2016-2018, and 2018-2020, with elections held every two years between 2010 and

2020. Our primary results are for long panels for both presidential and House elections (2012-2016 and 2016-2020). This is because of two reasons. First, it is not clear how taxpayers respond to tax changes during midterm elections. Second, the two key tax reforms between 2010 and 2020 were the ATRA 2012 and TCJA 2017. The ATRA 2012 took effect in 2013, and its short- and long-term effects should be captured by the long panel for 2012-2016. Similarly, for TCJA which took effect in 2017, the long panel for 2016-2020 should capture short- and long-term effects on voting behavior.

2.5 Key Tax Reforms

2.5.1 American Taxpayer Relief Act of 2012

The American Taxpayer Relief Act (ATRA) of 2012 took effect on January 1, 2013 and prevented many of the tax cuts introduced during the Great Recession of 2007-2009 from lapsing. However, the Act increased taxes on high-income earners through a suite of changes, and increased taxes on the rest of the income distribution by allowing a payroll tax reduction to expire.

High-income taxpayers were hit by the following: an increase in the top marginal tax rate from 35 percent to 39.6 percent, increase in the top capital gains tax rate from 15 percent to 20 percent, increase in the top estate tax rate from 35 percent to 40 percent, reinstatement of the limitation on itemized deductions (Pease) and the personal exemption phaseout (PEP), and the introduction of a Medicare tax of 0.9 percent and a Net Investment Income Tax (NIIT) of 3.8 percent, among others. On the other hand, the Alternative Minimum Tax (AMT) phaseout thresholds were increased and indexed to inflation, preventing bracket creep from bringing middle income earners into the AMT net. For the overall income distribution, the expiration of the payroll tax deduction increased payroll taxes from 4.2 percent to 6.2 percent, raising aggregate income taxes by an estimated $500 billion in 2013, affecting approximately 90 percent of US households who would have seen their tax bills increase by $3,500, on average.[21]

Looking at the immediate impact of ATRA 2012, income tax liability per capita in the raw SOI data increased for all but 226 out of the 1,132 max(PUMA, County) units between 2012 and 2013, while based on our simulated tax variable, tax liability increased in all units.

2.5.2 Tax Cuts and Jobs Act of 2017

The Tax Cuts and Jobs Act (TCJA) of 2017 took effect on January 1, 2018 and significantly reduced overall tax burdens. The TCJA decreased marginal tax rates for four out of the six income tax brackets by one to four percentage points. And while TCJA eliminated personal exemptions, it simultaneously compensated for this reduced ability of taxpayers to shield income from taxation by approximately doubling the standard deduction. The Child Tax Credit (CTC) was temporary doubled from $1,000 to $2,000, with a $500 credit provided to children under 17 years of age who were ineligible for the $2,000 CTC.

For high earners, the impact of the TCJA on tax liabilities was statutorily ambiguous, and empirical varied. The Act increased gift and estate tax exemptions, decreasing tax liabilities. It also substantially increased the AMT phaseout threshold, drastically limiting the reach of this parallel tax structure for middle and high earners. However, the TCJA brought about a

[21] https://www.taxpolicycenter.org/briefing-book/what-did-american-taxpayer-relief-act-2012-do

key change in the tax code by capping the State and Local Tax (SALT) deduction at $10,000. SALT provides high earners the ability to deduct state-level property, income, and sales taxes paid from federal taxable income, reducing their tax bills. The SALT limitation significantly increased the potential tax bills for high earners, especially in Democratic-leaning states, which had high income and property tax rates and routinely availed of the full SALT amount prior to 2018.

As a result of TCJA 2017, income tax liability per capita in the raw SOI decreased for all but 87 units, while our simulated variable suggests that the policy decreased taxes in 372 units and increased them in 760 units.

2.6 Results

2.6.1 Presidential Elections

We begin by estimating the cross-sectional and first-difference specifications for the long panel for 2012-2016. Results for the cross-sectional specification are provided in Table 6. The first column shows estimates of the relationship without controls; the second row includes demographic controls; and the third row weights observations by population size. Across our units of analysis, a $1,000 increase in SOI tax liability per capita in 2012 (2016) results in a decrease in Republican vote share of 1.86 (3.01) percentage points. However, the sign is reversed when we control for demographic controls in column (2). This suggests that in the cross-sectional analysis for any given year, tax liabilities are higher in units that vote more Democratic, as compared to units that vote more Republican in both 2012 and 2016.

Table 6: Presidential – Cross-Sectional (2012 & 2016)

	(1)	(2)	(3)	(4)	(5)	(6)
SOI (2012)	-1.86***	1.99***	1.43***			
	(0.24)	(0.32)	(0.32)			
SOI (2016)				-3.01***	0.73***	0.55**
				(0.22)	(0.26)	(0.26)
Constant	60.23***	83.56***	104.70***	67.07***	86.02***	84.13***
	(0.82)	(14.97)	(21.26)	(0.86)	(14.69)	(19.45)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1132	1132	1132	1132	1132	1132
Adjusted R^2	0.06	0.37	0.55	0.18	0.53	0.66
Weighted	No	No	Yes	No	No	Yes

Notes : * p<0.10, ** p<0.05, *** p<0.01. LHS variable is Republican Vote Share (pp). for respective year Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. Weighted by population, where indicated. Controls include population density (2012) and shares of college-educated, non-white, aged over 65, female for respective year. Heteroskedasticity-robust SEs.

The first-differencing OLS approach controls for these confounding factors except for the impact of changing demographics on changing tax burdens. Table 7 shows that within our units of analysis, an increase in tax liability per capita of $1,000 between 2012 and 2016 results in a decrease in Republican vote share by 3.28 percentage points. Even after controlling for secular trends by baseline demographics, the relationship remains negative at the 99 percent confidence level. But once we mitigate the endogeneity of tax burdens due to demographic shifts by using our simulated tax variable holding 2010 demographics constant, we find that an increase in tax burdens within our units of analysis between 2012 and 2016 results in increasing Republican vote shares. These reduced-form results are provided in Table 8. Columns (3)-(6) provide additional results for specifications where we also control for tax liability that is a result of purely demographic changes. Columns (2) and (4) are our preferred results, but we also show results with observations weighted by population in columns (3) and (6).

Table 7: Presidential – FD OLS (2012-2016)

	(1)	(2)	(3)
SOI	-3.28***	-0.77***	0.10
	(0.31)	(0.29)	(0.40)
Constant	3.44***	-5.33	-21.28**
	(0.22)	(5.72)	(8.80)
Controls	No	Yes	Yes
Observations	1130	1130	1130
Adjusted R^2	0.11	0.43	0.46
Weighted	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is Presidential Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2012-2016. Weighted by base year population, where indicated. Controls include delta shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Table 8: Presidential – FD Reduced-Form (2012-2016)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	-2.85***	2.31***	1.39**	-3.70***	2.86***	1.24
	(0.43)	(0.57)	(0.71)	(0.46)	(0.64)	(0.77)
Simulated (Composition)				-1.30***	0.56**	-0.15
				(0.29)	(0.26)	(0.41)
Constant	3.91***	-5.33	-22.25**	4.65***	-5.06	-21.91**
	(0.37)	(5.71)	(9.13)	(0.39)	(5.69)	(9.11)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1130	1130	1130	1130	1130	1130
Adjusted R^2	0.03	0.44	0.47	0.05	0.44	0.47
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is Presidential Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2012-2016. Weighted by base year population, where indicated. Controls include delta shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

We also regress SOI tax liability per capita on our simulated instrumental variable to assess how well the simulated IV explains the SOI tax variable. These first-stage results are shown in Table 9. The F-stat > 40 for all specifications, suggesting that our IV is a strong one. The resultant IV estimate suggests that as tax liability per capita in our units of analysis increases by $1,000, the Republican vote share increases by 5.92 percentage points (Table 10). The magnitude is smaller when we weight the observations by population size. However, we do not see a good reason for preferring our weighted results to those of the unweighted specifications.

Table 9: Presidential – FD First Stage (2012-2016)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	0.85***	0.39***	0.46***	0.96***	0.42***	0.59***
	(0.07)	(0.09)	(0.11)	(0.07)	(0.08)	(0.12)
Simulated (Composition)				0.16***	0.03	0.13**
				(0.04)	(0.04)	(0.06)
Constant	-0.13***	-1.87**	2.67	-0.22***	-1.85**	2.38
	(0.05)	(0.84)	(2.02)	(0.05)	(0.84)	(1.89)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1132	1132	1132	1132	1132	1132
Adjusted R^2	0.29	0.43	0.57	0.31	0.43	0.57
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is SOI Per Capita. Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2012-2016. Weighted by base year population, where indicated. Controls include delta shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Table 10: Presidential – FD IV (2012-2016)

	(1)	(2)	(3)	(4)	(5)	(6)
SOI	-3.34***	5.92***	3.03**	-3.86***	6.84***	2.15
	(0.50)	(2.00)	(1.53)	(0.49)	(2.20)	(1.34)
Simulated (Composition)				-0.69**	0.36	-0.41
				(0.29)	(0.38)	(0.34)
Constant	3.47***	5.35	-30.79***	3.78***	7.18	-27.38***
	(0.30)	(8.48)	(11.39)	(0.29)	(9.11)	(10.59)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1130	1130	1130	1130	1130	1130
Adjusted R^2	0.11	0.15	0.40	0.11	0.06	0.43
Weighted	No	No	Yes	No	No	Yes

Notes: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. LHS variable is Presidential Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in \$1000s per capita. FD model using data from 2012-2016. Weighted by base year population, where indicated. Controls include delta shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Controlling for the share of the population with a college degree is especially important in controlling for secular trends in college education that affect both tax liabilities and Republican vote share, with college graduates increasingly leaning Democrat. Figure 14 illustrates this increasingly stark trend across time.

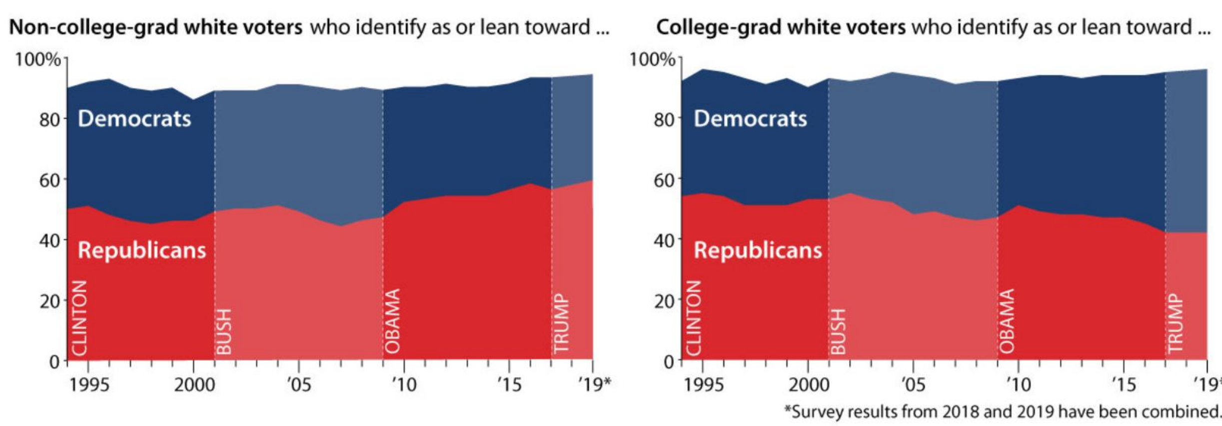


Figure 14: A Growing Education Gap (Source: Pew Research Center | Jacob Turcotte/Staff

We also provide a visualization of the impact on naïve estimates of not controlling for this key variable in Figures 15 and 16. Figure 15 shows the raw scatter plot and linear line of best fit for the change in Republican vote share versus change in simulated tax burdens caused purely by policy reforms. The downward, negative trend is what we see in our reduced-form results (Table 8, columns 1 and 4). The confounding effect is illustrated in Figure 16, where changes in college share and tax liability are positively correlated. This confirms the trend observed in Figure 14 – that college education and Republican vote share are negatively related. Thus, not including changes in college share and secular trends in baseline college share bias our estimates downwards. This is corrected by including the change and trend in college share in the specification, together with other controls. The residualized plot that corresponds to our reduced-form estimate in Table 8, column (2) is illustrated in Figure 17.

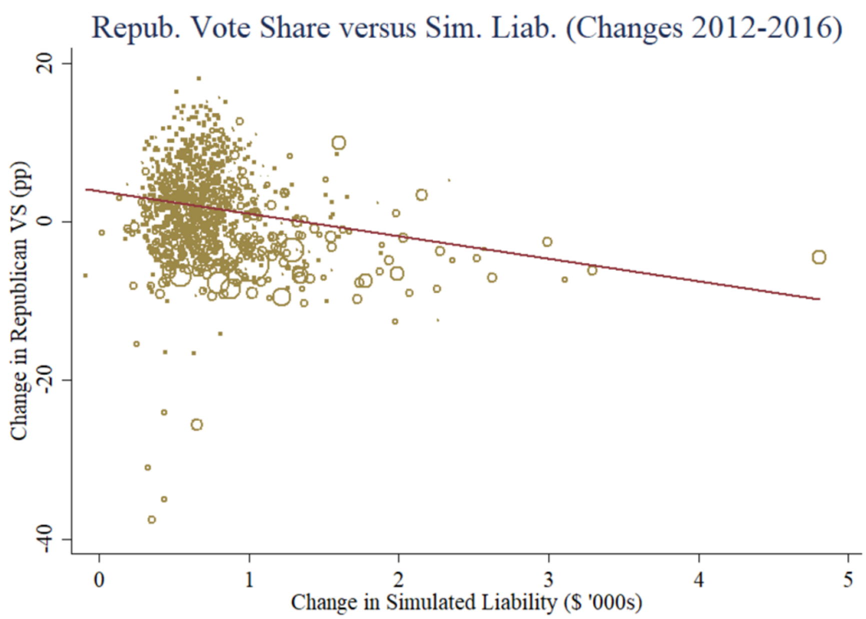


Figure 15: Scatter Plot of Republican Vote Share and Simulated Liability Per Capita

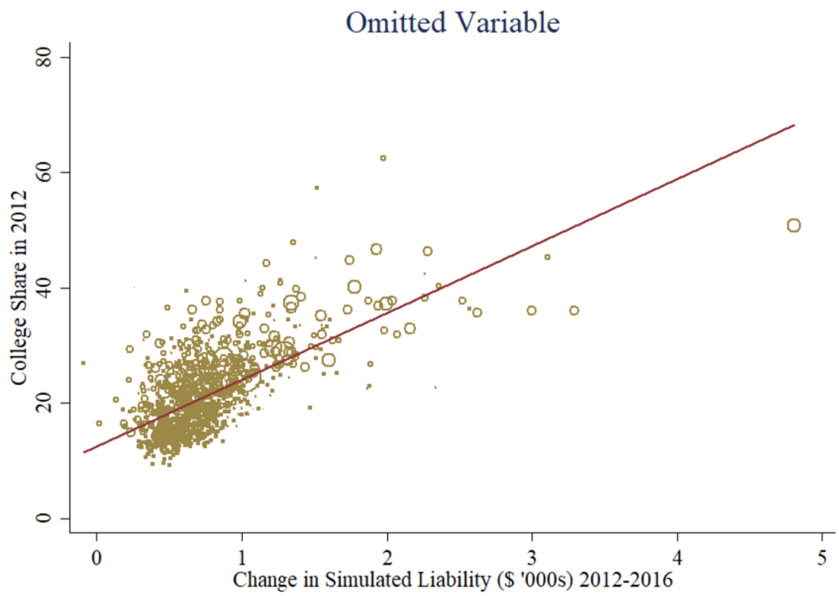


Figure 16: Relationship between College Share and Simulated Liability Per Capita

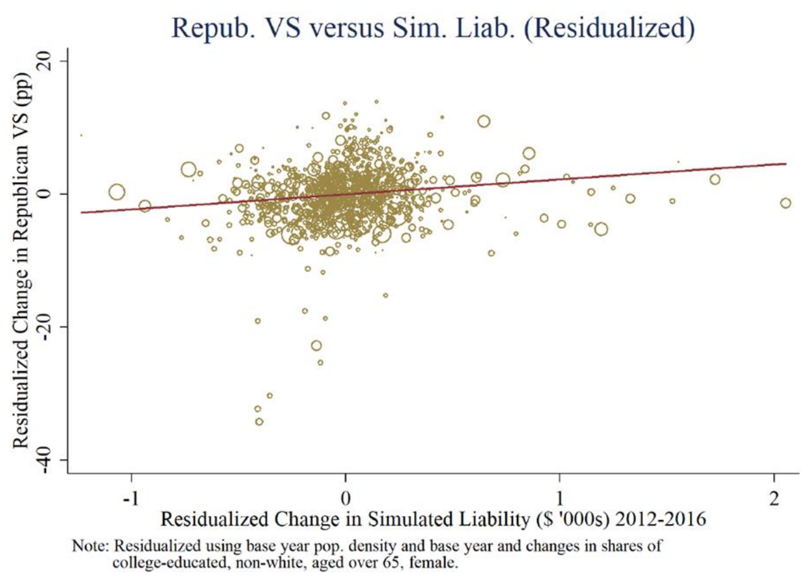


Figure 17: Residual Scatter Plot Republican Vote Share versus Simulated Liability Per Capita

For the long panel comprising years 2016 and 2020, we can only provide results for the reduced-form specification. This is due of the unavailability of SOI tax data for 2020, which precludes us from running the FD OLS or FD IV for 2016-2020. However, we can still use our simulated tax variable to obtain reduced-form estimates for this time period, since ACS data is available up till 2019, and we inflate this to 2020 estimates. These results are provided in Table 11. The relationship holds across the two time periods, albeit with a lower magnitude and a lower confidence level. As indicated in column (2), as simulated tax liability per capita increases by $1,000, the Republican vote share increases by 0.95 percentage points at the 95 percent confidence level.

Table 11: Presidential – FD Reduced-Form (2016-2020)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	0.61	0.95**	0.67	0.12	1.04**	0.99
	(0.49)	(0.40)	(0.62)	(0.48)	(0.42)	(0.65)
Simulated (Composition)				-0.54***	0.08	0.39***
				(0.08)	(0.10)	(0.15)
Constant	0.65***	20.05***	31.61***	1.61***	19.65***	26.91***
	(0.17)	(3.30)	(7.30)	(0.23)	(3.31)	(7.04)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1131	1131	1131	1131	1131	1131
Adjusted R^2	0.00	0.21	0.20	0.05	0.21	0.21
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is Presidential Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2016-2020. Weighted by base year population, where indicated. Controls include delta shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Our results for the stacked panel are similarly affected by our inability to incorporate 2020 SOI data. However, the reduced-form results shown in Table 12, column (2) suggest that as tax liability per capita increases by $1,000, Republican Vote share between 2012 and 2020 increases by 0.93 percentage points. This result is statistically significant at the 99 percent confidence level.

Table 12: Presidential – FD Reduced-Form (2012-2020)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	-0.50*	0.93***	-1.65***	-2.05***	1.70***	0.60
	(0.28)	(0.29)	(0.39)	(0.30)	(0.35)	(0.39)
Simulated (Composition)				-0.95***	0.34***	0.92***
				(0.09)	(0.11)	(0.17)
Constant	1.58***	8.02**	5.12	3.13***	7.41**	-0.76
	(0.14)	(3.15)	(5.13)	(0.21)	(3.13)	(4.87)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	2261	2261	2261	2261	2261	2261
Adjusted R^2	0.00	0.28	0.27	0.04	0.28	0.29
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is Presidential Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2012-2020. Weighted by base year population, where indicated. Controls include delta shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

2.6.2 House Elections

We first consider the three main (non-midterm) electoral cycles between 2010 and 2020: 2012, 2016, and 2020. Similar to the responses we see from voters voting in presidential elections, the FD OLS results provided in Table 13 reveal a negative relationship between observed SOI tax liability per capita and Republican vote share for time period 2012-2016.

Table 13: House – FD OLS (2012-2016)

	(1)	(2)	(3)
SOI	-3.92***	-1.25**	-0.72
	(0.49)	(0.61)	(0.62)
Constant	5.57***	-9.09	-11.73
	(0.44)	(14.16)	(17.59)
Controls	No	Yes	Yes
Observations	1131	1131	1131
Adjusted R^2	0.04	0.13	0.14
Weighted	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in \$1000s per capita. FD model using data from 2012-2016. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

In our reduced-form specifications with the simulated tax policy variable (Table 14), we control for secular trends by baseline demographics and isolate the impact of changing tax liabilities that are a result of pure tax policy changes (column 2). For House-level analysis, we also control for incumbency. We find that as tax liability per capita increases by \$1,000, Republican vote share increases by 0.66 percentage points. However, this result is not statistically significant at the 90 percent confidence level. This seems to suggest the effect of federal tax policy on voting behavior flows primarily through the presidential election channel, with estimates based on House elections being too noisy to ascertain an estimated relationship with a high level of confidence. The first-stage results are the same as in Table 9, since we cover the same time period here (2012-2016). Putting the reduced-form and the first-stage results together, our IV estimate in Table 15, column 2 suggests that as tax liability per capita increases by $1,000, Republican vote share increases by 1.69 percentage points. This result is not statistically significant at the 90 percent confidence level.

Table 14: House – FD Reduced-Form (2012-2016)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	-4.23***	0.66	0.58	-5.03***	0.64	1.36
	(0.81)	(1.02)	(1.41)	(0.88)	(1.18)	(1.57)
Simulated (Composition)				-1.21*	-0.02	0.79
				(0.65)	(0.67)	(0.87)
Constant	6.73***	-7.38	-14.53	7.42***	-7.39	-16.22
	(0.68)	(14.11)	(17.81)	(0.76)	(14.10)	(17.61)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1131	1131	1131	1131	1131	1131
Adjusted R^2	0.02	0.13	0.14	0.02	0.13	0.14
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2012-2016. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Table 15: House – FD IV (2012-2016)

	(1)	(2)	(3)	(4)	(5)	(6)
SOI	-4.97***	1.69	1.26	-5.26***	1.51	2.27
	(0.97)	(2.62)	(3.11)	(0.91)	(2.80)	(2.70)
Simulated (Composition)				-0.37	-0.07	0.46
				(0.64)	(0.62)	(0.80)
Constant	6.09***	-4.21	-17.92	6.26***	-4.60	-21.63
	(0.57)	(14.62)	(21.86)	(0.58)	(14.66)	(20.20)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1131	1131	1131	1131	1131	1131
Adjusted R^2	0.04	0.11	0.13	0.04	0.12	0.11
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2012-2016. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

For time period 2016-2020, we can only provide reduced-form estimates using the simulated tax policy variable, since SOI tax data is unavailable for 2020. These are shown in Table 16, column 2. The direction of the coefficient on the simulated tax policy variable switches as compared to presidential elections, and we see that as tax liability increases, Republican vote share decreases. With the Republican party being the incumbent party in Congress in 2016, this suggests that taxpayers in units of analysis where taxes increased impose a penalty on the party at the voting booth, reducing Republican vote share. However, the results continue to be too noisy to form reliable estimates.

Table 16: House – FD Reduced-Form (2016-2020)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	0.48	-0.49	1.80	3.26*	0.42	1.99
	(1.84)	(1.68)	(2.17)	(1.78)	(1.77)	(2.33)
Simulated (Composition)				3.10***	0.76	0.23
				(0.33)	(0.52)	(0.60)
Constant	-1.72***	-15.13	1.31	-7.23***	-19.01	-1.48
	(0.57)	(15.23)	(19.70)	(0.86)	(15.50)	(20.57)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1126	1126	1126	1126	1126	1126
Adjusted R^2	-0.00	0.18	0.20	0.07	0.18	0.20
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2016-2020. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

We also consider short panels for years 2012-2014 and 2016-2018. The first time period should capture the impact of the ATRA 2012 and the second time period should capture the impact of the TCJA 2017, albeit within shorter time periods. These shorter time periods therefore, should provide us with an understanding of short-term changes in voting behavior. These shorter panels include midterm election cycles in 2014 and 2018. We find that the short-term changes in voting behavior in House elections are similar to longer term patterns that we detected by dropping information on midterm elections. The Republican vote share increase between 2012-2014 when the ATRA 2012 increased tax liabilities for high-income taxpayers and for others through the payroll tax increase (Table 17), while the Republican vote share decreases between 2016-2018 when the incumbent Republican-controlled House decreased taxes in general but had a potentially adverse effect on high income taxpayers in Democratic controlled areas (Table 18). However, these estimates are not statistically significant at the 90 percent confidence level.

Table 17: House – FD Reduced-Form (2012-2014)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	-0.84	0.80	1.22	-0.70	0.89	1.56
	(0.88)	(1.30)	(1.89)	(0.91)	(1.40)	(2.00)
Simulated (Composition)				0.53	0.16	0.60
				(0.64)	(0.67)	(0.91)
Constant	3.77***	-24.48*	-8.76	3.67***	-24.37*	-9.17
	(0.68)	(13.97)	(21.86)	(0.69)	(14.05)	(21.82)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1130	1130	1130	1130	1130	1130
Adjusted R^2	-0.00	0.05	0.05	-0.00	0.05	0.05
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2012-2014. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Table 18: House – FD Reduced-Form (2016-2018)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	0.94	-0.18	3.07*	0.94	-0.20	3.39**
	(1.17)	(1.53)	(1.71)	(1.17)	(1.55)	(1.71)
Simulated (Composition)				0.00	-0.03	0.79
				(0.56)	(0.57)	(0.68)
Constant	-5.24***	15.92	22.99	-5.24***	15.93	21.09
	(0.61)	(14.22)	(17.28)	(0.61)	(14.22)	(17.23)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1131	1131	1131	1131	1131	1131
Adjusted R^2	-0.00	0.02	0.03	-0.00	0.02	0.03
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2016-2018. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

First-stage results for House Elections between 2012-2014 and 2016-2018 are provided in tables 19 and 20, respectively. The first-stage indicates that the simulated tax policy variable is a strong IV. The IV results that combine the reduced-form and first-stage results are provided in tables 21 and 22, replicating the direction and statistical significance of the reduced-form estimates, with minor changes in the magnitude.

Table 19: House – FD First Stage (2012-2014)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	1.15***	0.94***	1.26***	1.15***	0.94***	1.30***
	(0.11)	(0.10)	(0.11)	(0.12)	(0.10)	(0.11)
Simulated (Composition)				0.02	0.00	0.06
				(0.03)	(0.03)	(0.04)
Constant	-0.35***	0.98**	3.25***	-0.35***	0.98**	3.21***
	(0.07)	(0.43)	(1.04)	(0.07)	(0.43)	(1.03)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1132	1132	1132	1132	1132	1132
Adjusted R^2	0.47	0.51	0.77	0.47	0.51	0.77
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is SOI Per Capita. Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2012-2014. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Table 20: House – FD First Stage (2016-2018)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	1.18***	0.83***	1.26***	1.19***	0.85***	1.28***
	(0.20)	(0.19)	(0.22)	(0.20)	(0.20)	(0.22)
Simulated (Composition)				0.07***	0.05*	0.05
				(0.03)	(0.03)	(0.05)
Constant	-0.51***	2.29***	4.48***	-0.50***	2.27***	4.37***
	(0.09)	(0.60)	(1.42)	(0.09)	(0.60)	(1.41)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1132	1132	1132	1132	1132	1132
Adjusted R^2	0.29	0.37	0.59	0.29	0.38	0.59
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is SOI Per Capita. Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2016-2018. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Table 21: House – FD IV (2012-2014)

	(1)	(2)	(3)	(4)	(5)	(6)
SOI	-0.73	0.85	0.96	-0.61	0.94	1.20
	(0.75)	(1.39)	(1.47)	(0.77)	(1.49)	(1.53)
Simulated (Composition)				0.54	0.16	0.53
				(0.64)	(0.67)	(0.89)
Constant	3.52***	-25.31*	-11.90	3.46***	-25.30*	-13.03
	(0.44)	(14.09)	(22.15)	(0.45)	(14.12)	(22.10)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1130	1130	1130	1130	1130	1130
Adjusted R^2	0.00	0.05	0.04	0.00	0.05	0.04
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in \$1000s per capita. FD model using data from 2012-2014. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Table 22: House – FD IV (2016-2018)

	(1)	(2)	(3)	(4)	(5)	(6)
SOI	0.80	-0.22	2.44*	0.79	-0.23	2.65*
	(0.97)	(1.85)	(1.43)	(0.96)	(1.82)	(1.40)
Simulated (Composition)				-0.05	-0.02	0.66
				(0.56)	(0.56)	(0.72)
Constant	-4.84***	16.42	12.03	-4.85***	16.46	9.47
	(0.29)	(15.09)	(18.38)	(0.29)	(15.02)	(18.55)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	1131	1131	1131	1131	1131	1131
Adjusted R^2	-0.00	0.02	0.01	-0.00	0.02	0.01
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in \$1000s per capita. FD model using data from 2016-2018. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

We also consider the effect of changes in tax liability per capita across electoral cycles using a stacked panel for House elections. This panel comprises years for which we have SOI tax data: 2010 to 2018. The FD OLS estimates are provided in Table 23. As compared to presidential election results, House elections reveal a positive relationship between SOI tax liability per capita and Republican vote shares, though we expect these estimates to be biased due to secular trends in baseline demographics and confounding of observed tax burdens due to demographic shifts.

Table 23: House – FD OLS (2010-2018)

	(1)	(2)	(3)
SOI	0.95***	2.00***	1.29*
	(0.36)	(0.47)	(0.72)
Constant	-1.20***	-0.16	-1.86
	(0.10)	(2.89)	(3.81)
Controls	No	Yes	Yes
Observations	4523	4523	4523
Adjusted R^2	0.00	0.04	0.02
Weighted	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2010-2018. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

We use the simulated tax policy variable in reduced-form specifications, with estimates provided in Table 24. Our preferred, unweighted estimate for which we also control for secular trends by baseline demographics indicates that as income tax liability per capita increases by $1,000, Republican vote share in House elections increases by 1.96 percentage points. This result is statistically significant at the 99 percent confidence level. The result holds when we also control for changes in tax liability per capita that arise purely due to demographic shifts. First-stage results shown in Table 25 confirm a strong IV across specifications, resulting in an IV estimate of an increase in Republican vote share of 3.66 percentage points for every $1,000 increase in tax liability per capita in Table 26.

Table 24: House – FD Reduced-Form (2010-2018)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	2.27***	1.96***	3.44***	2.42***	2.12***	3.76***
	(0.51)	(0.63)	(0.89)	(0.52)	(0.64)	(0.93)
Simulated (Composition)				0.70*	0.77**	1.73***
				(0.38)	(0.38)	(0.55)
Constant	-1.86***	1.33	0.61	-1.92***	0.93	-1.66
	(0.20)	(2.87)	(3.78)	(0.21)	(2.89)	(3.80)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	4523	4523	4523	4523	4523	4523
Adjusted R^2	0.00	0.04	0.02	0.00	0.04	0.03
Weighted	No	No	Yes	No	No	Yes

Notes: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in $1000s per capita. FD model using data from 2010-2018. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Table 25: House – FD First Stage (2010-2018)

	(1)	(2)	(3)	(4)	(5)	(6)
Simulated (Policy)	0.70***	0.53***	0.78***	0.72***	0.56***	0.81***
	(0.07)	(0.06)	(0.12)	(0.07)	(0.06)	(0.13)
Simulated (Composition)				0.13***	0.11***	0.18***
				(0.02)	(0.02)	(0.04)
Constant	-0.03	0.76**	2.44***	-0.04*	0.70*	2.20***
	(0.02)	(0.38)	(0.77)	(0.02)	(0.39)	(0.72)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	4528	4528	4528	4528	4528	4528
Adjusted R^2	0.19	0.29	0.42	0.21	0.30	0.43
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is SOI Per Capita. Unit of analysis is Max(PUMA,County). All liabilities are in \$1000s per capita. FD model using data from 2010-2018. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

Table 26: House – FD IV (2010-2018)

	(1)	(2)	(3)	(4)	(5)	(6)
SOI	3.25***	3.66***	4.43***	3.34***	3.80***	4.64***
	(0.87)	(1.27)	(1.61)	(0.85)	(1.24)	(1.63)
Simulated (Composition)				0.28	0.35	0.88
				(0.38)	(0.39)	(0.54)
Constant	-1.75***	-1.48	-10.18*	-1.78***	-1.76	-11.85*
	(0.21)	(3.18)	(6.05)	(0.20)	(3.18)	(6.23)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	4523	4523	4523	4523	4523	4523
Adjusted R^2	-0.01	0.04	-0.01	-0.01	0.04	-0.01
Weighted	No	No	Yes	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. LHS variable is House Republican VS (pp). Unit of analysis is Max(PUMA,County). All liabilities are in \$1000s per capita. FD model using data from 2010-2018. Weighted by base year population, where indicated. Controls include delta incumbency, shares of college-educated, non-white, aged over 65, female; and baselines for the demographics. SEs are clustered at Max(PUMA,County) level.

2.7 Concluding Remarks

In this paper, we provide the first evidence of its kind on the causal link between changing federal tax burdens and voting behavior in presidential and House elections in the United States, in the time period 2010 and 2012. We show that rising tax burdens generate substantial electoral gains for the Republican party, the party that has over the last four decades emerged as the biggest proponent of tax cuts. We identify the causal effect of changing federal tax burdens on voting behavior by isolating the effect of pure policy-induced changes in tax burdens from changes that arise due to demographic shifts. Such demographic shifts that affect tax burdens are caused by changing characteristics of residents in our units of analysis over time, such as marital status, age, job type, and number of dependents, and by the changing composition of residents themselves through migratory flows. To isolate the pure tax policy-induced change in tax burden, we use a simulated instrumental variable strategy that predicts tax burdens based on prevailing tax policy between 2010 and 2020, holding baseline 2010 demographics constant. The two key tax policy reforms during this time period that we leverage are the American Taxpayer Relief Act of 2012 and the Tax Cuts and Jobs Act of 2017.

We further show that it is necessary to control not just for changes in tax burdens due to demographic shifts to eliminate confounders that affect the estimation of the relationship between federal tax burdens and voting behavior, but also to control for secular trends in baseline demographics. These secular trends can bias estimates downward. Our results show that within our units of analysis, a one-half standard deviation increase in federal income tax liability per capita results in a one to six percentage point gain in vote share for the Republican party. These results are more robust for presidential elections relative to House elections.

Our approach contributes to the political economy literature on taxation and voting behavior. In particular, we add a missing piece in the subliterature on retrospective determinants of voting behavior. Prior work has considered the impact of economic shocks, wealth shocks, fiscal outlays and cash transfers, among others. We address the core challenges with examining the causal link between federal tax burdens and voting behavior with a novel identification strategy. In the process, we also provide evidence on how taxpayers continue to vote in their economic self-interest, on average. This runs counter to the emerging popular belief that taxpayers primarily vote based on cultural values, groupthink, religion, and race.

3. Multidimensional Measurement of Sectoral Performance: Evidence from Public Schools in Pakistan*

Abstract

I develop a tool for measuring the multidimensional performance of the public sector in the spirit of multidimensional measures of poverty, and apply it to the case of public education in Pakistan. The framework allows fiscally constrained policymakers and relevant development practitioners to measure a sector's resource base, follow it over time, and optimize targeting of resources. The measure's decompositional properties provide for easy identification of the sources of deprivation along various dimensions and across subgroups, such as geographical areas and subsectors. In an application to the public education sector in Sindh province, Pakistan, I show that 27 percent of public schools are multidimensionally deprived and the weakest dimensions are physical infrastructure and facilities. Single-sex, rural schools, where instruction is in the native Sindhi language contribute the most to the overall deprivation measurement. Such identification permits efficient allocation of policy attention. Targeting public resources to these weak links can generate the biggest bang for the buck. This is especially valuable in resource-constrained, developing countries. The measure allows policymakers to glean critical sectoral information from the din of administrative and survey data.

* I am grateful to David Foster, Ravi Kanbur, Arnab Basu, Nancy Chau, Shanjun Li, and Ariel Ortiz-Bobea for excellent comments and feedback. I also thank the Reform Support Unit of the Sindh provincial Government, Pakistan for data support and assistance. Any opinions and conclusions expressed herein are those of the author.

3.1 Introduction

Fiscal constraints in developing countries present a dual challenge. On the one hand, developing countries require significant investments across public and private sectors to catch up with industrialized countries. Such investments are difficult to make under a highly constrained resource base. Due to a lack of such investments, governments in developing countries have a hard time creating future fiscal space, leading to a vicious cycle of low investments and poor public sector health and performance. Given the scarcity of resources, it becomes imperative that the state has access to low-cost analytical tools to allow it to identify sectoral dimensions and geographic regions that are the most deprived, and to funnel public investments into these subgroups and dimensions to optimize limited public investments.

An important strand of development literature over the past decade has experimented with the use of measures that capture multiple dimensions when assessing the level of deprivation of an individual or a group of individuals. For example, in the measurement of poverty, it is now widely accepted that a unidimensional approach focusing on income levels is insufficient. $n - 1$ other dimensions such as health, education and nutrition, among others, factor into the "well-being" of an individual – leading to the well-documented multidimensional approach to the measurement of poverty. The Alkire-Foster measure of multidimensional poverty developed by Sabina Alkire and James Foster (2011) is such an approach.

Alkire and Foster apply their method to measuring aggregate multidimensional poverty, factoring in a number of "functionings" (dimensions), weights associated with these dimensions indicating their relative importance, and cutoffs – particular measures beyond which an individual is considered poor in a given dimension. The choice of dimensions, weights and cutoffs is flexible. This flexibility allows for a process of democratic deliberation and consensus-formation, leading to choices that are reflective of a country or region's context, and its ethical and normative standpoints.

Further, the Alkire-Foster method provides a powerful tool for policymakers to make intra- and inter-country comparisons, as well as comparisons across different subgroups via decompositional properties built into the tool. I contribute to this literature by expanding the use of the Alkire-Foster method which studies poverty levels with the use of individual-level data, to analyzing the health of a sector with the use of sectoral, unit-level data. More specifically, for this paper, I build a multidimensional measure of sectoral deprivation, called the Multidimensional Sectoral Deprivation Index ($MSDI$). The method can be used to measure sectoral health at the aggregate level, which can also be decomposed for better targeting of resources.

As proof-of-concept, I consider the public education sector, develop a methodology for assessing its readiness, and apply the method to the case of Sindh province in Pakistan. Broadly, countries are on either one of two, broad trajectories in the education sector: those in the developed world where educational infrastructure is robust – with some variation – and focus has shifted to the provision of sophisticated pedagogical improvements, school-based nourishment programs and the use of high-end technology for assistive learning; and those in low-resource, education-poor developing countries where issues related to poor school infrastructure, low enrolment, teacher absenteeism, and poor learning quality are pervasive.

For schools on the former trajectory of education growth, indicators such as the availability of a school building are not very informative. For these schools, the more relevant indicators are class atmosphere, consensus and cooperation amongst teachers, and positive reinforcement of students (Nordenbo et al., 2010). On the other hand, for schools on the latter trajectory, when there exists a lack of qualified teachers and/or high levels of teacher absenteeism, and one-teacher schools, consensus and cooperation amongst teachers becomes a second-order issue.

While development studies as those documented in Glewwe et al. (2012) provide substantial internal validity and shine a light on critical inputs, policymakers need as part of their toolkits, ways of measuring education sector health when external validity is not well established. Given that different factors affect school outcomes differentially, with heterogeneity of impact across regions, capturing a diverse set of dimensions is critical to measuring the quality of overall educational infrastructure in a geographical region.

Section 3.2 presents the conceptual framework for the construction of $MSDI$, together with its properties of decomposability and a comparison with alternate approaches. In Section 3.3, I will discuss the context in which I apply this tool, including the source and nature of the data used, and the choice of dimensions, weights, and cutoffs. Section 3.4 provides results of the application of the $MSDI$, followed by Section 3.5 where I conduct a sensitivity analysis of these empirical results. Section 3.6 concludes the discussion.

3.2 Methodology

3.2.1 Construction of the Multidimensional Sectoral Deprivation Index (MSDI)

In constructing the $MSDI$, I borrow from the framework developed by Alkire and Foster (2007). For the purpose of this paper, I will restrict the analysis to the use of the M_0 measure. M_0 is known as the adjusted poverty headcount ratio, and it provides an index of multidimensional poverty. In my setting, the $MSDI$ is the analog of the M_0, and serves as the multidimensional sectoral deprivation index. In this section, I will formally discuss censoring of the characteristics matrix, cutoffs, weights for each dimension, and the calculation of the simple $MSDI$.

We begin with a basic set up. $i = 1,2, \ldots, n$ indexes sectoral units, while $j = 1,2, \ldots, d$ indexes specified dimensions. I set up a sectoral unit achievement matrix X, with each element represented by ij, or sectoral unit i's performance on dimension j. Sectoral units can comprise bus and train stations in the transportation sector, factories in the manufacturing sector, hospitals in the health sector, or schools in the public education sector.

For the purpose of identifying "deprivation" within a given dimension, I specify a cutoff vector z, with z_j serving as the cutoff for each dimension j. For each x_{ij}, in achievement matrix X, I replace the value of x_{ij} with a 0 when $x_{ij} \leq z_j$, and with a 1 when $x_{ij} > z_j$. This transforms the achievement matrix X into the deprivation matrix g^0, with each g^0_{ij} indicating whether sectoral unit i is deprived in dimension j. This serves as the first round of *censoring*, in that, it suppresses the *level* of deprivation and exclusively focuses on a binary indicator for the *presence* of deprivation of a sectoral unit within a given dimension.

From g^0, I construct a column vector c' of deprivation counts, with ${c_i}' = \sum_{j=1}^{d} g_{ij}^0$. Each element of c' provides the count for dimensions in which sectoral unit i is deprived. However, given that all dimensions might not hold the same relative importance in contributing to the deprivation of a given sectoral unit, based on empirical evidence and/or normative considerations, a flexible weighting scheme is used for the d dimensions, defined by the vector $w = [w_1\ w_2\ \ldots w_d]$. These weights do not necessarily have to sum to 1 but are normalized for convenience. Using these weights, I construct a deprivation score vector c for each sectoral unit, as $c_i = \sum_{j=1}^{d} w_j g_{ij}^0$.

The next step is to identify a given sectoral unit as being either multidimensionally deprived, or non-deprived. An identification function $\rho_k(x_i; z)$ takes a value of 1 if a sectoral unit is multidimensionally deprived, or 0 if it is not. The *intersection* approach implies that a sectoral unit be considered multidimensionally deprived only if it is deprived in all dimensions, so that $\rho_k(x_i; z) = 1$ if $c_i = 1$, or 0 otherwise. So even a sectoral unit which is deprived in $d - 1$ dimensions will be captured as being multidimensionally non-deprived. On the other extreme, the *union* approach would imply that $\rho_k(x_i; z) = 1$ if $c_i \geq 0$, and 0 otherwise. In this case, a sectoral unit which is deprived in at least one dimension will be identified as being multidimensionally deprived.

While both approaches have their merits in different settings, given that a number of factors combine to optimize the performance of a given sector, an intermediate approach appears to be more suitable. This approach uses a cutoff k with $k \in [0,1]$, above which, a sectoral unit is termed as multidimensionally deprived. As with the dimensional cutoff z, the choice of k is flexible, as discussed in the following sections. In this case, the identification function $\rho_k(x_i; z) = 1$ if $c_i \geq k$, and 0 otherwise.

Using the dual cutoff-identification approach, I construct the censored deprivation matrix $g^0(k)$, with $g_{ij}^0(k) = \rho_k(x_i; z) * g_{ij}^0$. Thus, if a sectoral unit is multidimensionally non-deprived, then its deprivation in all individual dimensions is suppressed to 0. This is an important step and allows the $MSDI$ to focus on the extent of deprivation of deprived sectoral units, and not be affected by changes in the deprivation level of non-deprived sectoral units. Similarly, the vector of censored deprivation scores is constructed using $c_i(k) = \sum_{j=1}^{d} w_j g_{ij}^0 (k)$.

Aggregating the censored deprivation scores, I calculate the $MSDI$ as the mean of the censored deprivation score vector:

$$MSDI = \frac{1}{n} \times \sum_{i=1}^{n} c_i(k)$$

The $MSDI$ can also be expressed as the product of the deprivation incidence H (fraction of sectoral units that are deprived) and the deprivation intensity (A), or the average deprivation score among deprived sectoral units:

$$MSDI = H \times A = \frac{q}{n} \times \frac{1}{q} \sum_{i=1}^{q} c_i(k)$$

Another interpretation of MSDI is that it provides the share of weighted deprivations experienced by the deprived divided by the maximum possible deprivations that could be experienced if all sectoral units were deprived in all dimensions.

$$MSDI = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d} w_j g_{ij}^0(k)$$

It is important to note that for any specified weighting and cutoff scheme, the $MSDI$ satisfies decomposability, replication invariance, symmetry, deprivation focus, weak and dimensional monotonicity, nontriviality, normalization, and weak rearrangement, discussed in detail in Alkire-Foster (2011).

3.2.2 Decomposability

Subgroup Decompositions

In this paper, I will exploit the $MSDI's$ decomposability property extensively to unpack heterogeneity in deprivation across subgroups and dimensions. The ability to do so follows from the flexibility afforded by the $MSDI$ for both subgroup as well as dimensional decomposability. Here, I briefly discuss the mechanics of the tool's decomposition property, and how this affects the analysis in the following sections.

Subgroups can include different administrative regions such as states, provinces, counties, districts, villages, and cities, as well as groupings based on classifications, for example, in the case of the education sector, urban-rural, school gender, and primary medium of instruction, among others. I index each subgroup by $s = 1, \ldots, m$, with the population share of the subgroup given by $p^s = n^s/n$. Further, we can divide achievement matrix X into its different constituent subgroups, each indexed by X^s. By repeating the process outlined in the previous section, I compute the $MSDI$ for each of these constituent achievement sub-matrices ($MSDI(X^s)$). So the overall $MSDI$ can be expressed as:

$$MSDI(X) = \sum_{s=1}^{m} p^s * MSDI(X^s)$$

A noticeable property of this expression is that it is additive. Using this property, the contribution of each subgroup to overall sectoral unit deprivation can be calculated as follows:

$$D_s^0 = p^s \frac{MSDI(X^s)}{MSDI(X)}$$

Where,

$$\sum_{s=1}^{m} D_s^0 = 1$$

The contribution of a given subgroup s depends both on its population share, as well as its $MSDI$ as a fraction of overall $MSDI$, with $\frac{\partial D_s^0}{\partial p^s} > 0$ and $\frac{\partial D_s^0}{\partial MSDI(X^s)} > 0$. If sectoral deprivation is distributed uniformly across the population of sectoral units, then $p^s = D_s^0$, implying that the population share of the subgroup will be equal to the subgroup's share of aggregate sectoral deprivation. In reality, there will be heterogeneous distribution of deprivation burden across subgroups. Therefore, cases where $p^s < D_s^0$ allow me to pinpoint subgroups for which, the contribution to sectoral deprivation is disproportionately higher than the subgroup population size. This provides a useful policy device to pinpoint stragglers and devise more effective, targeted policies to improve overall sectoral health and performance.

The cardinality of the measure is useful in comparing different subgroups – for example, geographic regions – as well as comparing dimensional contributions. The $MSDI$ of a sectoral unit is simply the $MSDI(x)$ of a submatrix which is a singleton and is equivalent to the sectoral unit's censored deprivation score. Similar to the overall $MSDI$, the censored deprivation score of each sectoral unit provides a cardinal ranking of sectoral units along the deprivation spectrum. Meaningful information can be gleaned by comparing which dimensions the sectoral units are deprived in. The next section provides further details.

Dimensional Decompositions

The $MSDI$ can also be used to decompose a dimension's contribution to the overall sectoral deprivation level. Without loss of generality, the additive nature of the $MSDI$ allows it to be expressed as the weighted sum of each dimensional censored headcount ratio $h_j(k)$, where $h_j(k) = \sum_{i=1}^{n} g_{ij}^0(k)$. Intuitively, the censored headcount ratio of each dimension j is the proportion of the population of sectoral units that is identified as deprived, and further, the fraction that is deprived in dimension j. Therefore, the $MSDI$ can be expressed as:

$$MSDI = \sum_{j=1}^{d} w_j * h_j(k)$$

Under the restriction $w_1 = \cdots = w_d$, this expression collapses to:

$$MSDI = w \sum_{j=1}^{d} h_j(k)$$

If weights are not uniform, then the contribution of each dimension to overall sectoral deprivation not only depends on each dimension's censored headcount ratio, but also the weights associated with it. More formally,

$$\lambda_j^0(k) = w_j \frac{h_j(k)}{MSDI}$$

Two dimensions can have the same censored headcount ratio, $h_j = h_{-j}$. However, if $w_j > w_{-j}$, then $\lambda_j^0 > \lambda_{-j}^0$. In the case of uniform weights, equal h implies the same dimensional contribution to overall sectoral deprivation.

Dimensional decomposition provides a tool to policymakers to focus their attention on dimensions that are contributing disproportionately to overall sectoral deprivation, as compared to the weights associated with them. This allows policymakers to target dimensions that are acting as weak links in the system. Further, dimensional decompositions can be combined with subgroup decompositions, allowing policymakers to focus on the performance of specific dimensions within a subgroup. Under highly constrained resources, the ability to do this is critical for allocating and utilizing taxpayer money most efficiently.

3.2.3 Comparison to Alternative Approaches

Other approaches to ranking the health of educational units such as schools include the use of production efficiency techniques such as data envelopment analysis (DEA) and stochastic frontier analysis (SEA). These methods can handle multiple inputs and outputs to establish a production frontier to measure technical and allocative efficiency of individual sectoral units. Loosely, the distance of a unit from the frontier provides a measure of the unit's relative inefficiency. A representation of SFA and DEA-based frontiers is provided in Figure 18.

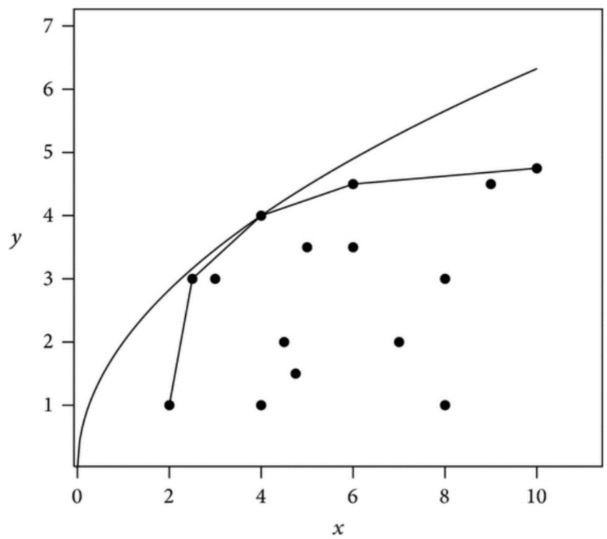


Figure 18: Data Envelopment Analysis versus Stochastic Frontier Analysis (Source: Aparicio et al., 2014)

However, these methods come with several shortcomings, and might not be best aligned with the aim of this analysis, which is to focus on the most low-resource sectoral units and construct a ranking which is unaffected by high-resource units. Firstly, as a non-parametric approach which does not require assumptions regarding the functional form of the frontier, DEA is a valuable analytical tool and useful in supporting practical decision-making in situations such as reducing inefficiencies of sectoral units. However, the approach requires distribution and production assumptions, which if inaccurate, can generate a bias over the frontier. Since the DEA models in current use provide only a limited range of production assumptions, they are hard to test.

Small unit bias has also been observed in prior studies using DEA, with units that are small appearing to be relatively less inefficient, systematically. Further, the number of units on the efficient frontier is found to be an increasing function of the number of input and output variables (Berg, 2010). In terms of our study, another issue which arises with ranking sectoral using DEA is that the approach uses observed observations with input-output bundles to form the frontier. Thus, using the example of the public education sector in Sindh province, Pakistan, ranking of schools using this approach would be solely based on schools in Pakistan on the frontier, with benchmarking used to evaluate potential changes in the bundle of inputs. However, given that low-resource, poorly performing countries such as Pakistan aspire to reach international standards, without comparable data on schools – and their input-output bundles which have reached such levels under a similar context, the frontier used for ranking schools might not be the correct one to use.

Worth noting is that DEA analyses are sensitive to the selection of inputs and outputs, similar to the $MSDI$. Eventually, the tool to be used should be aligned with both the existing sectoral capital in a given location, as well as feasibility for data collection and administrative rollout.

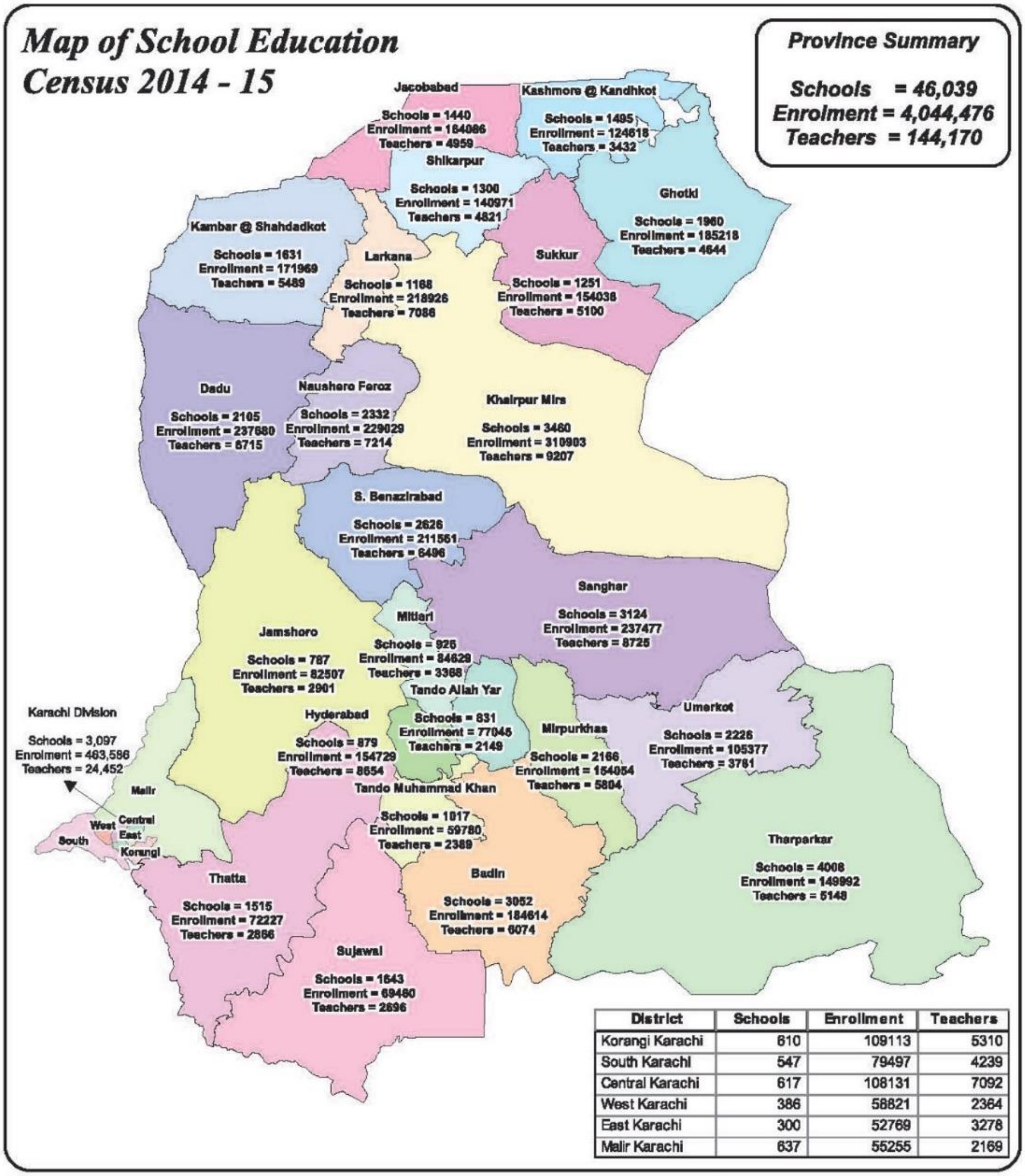
3.3 Application: Public Education in Sindh, Pakistan

3.3.1 Overview of Pakistan's Education Sector Performance

Over the last decade, Pakistan's economy has shown significant growth. Real GDP has increased at a rate close to four percent since 2010. However, the country's expenditure on education has stagnated, staying at less than two percent of GDP. Under such conditions, Pakistan did not meet its objective of providing universal primary education by 2015 under the Millennium Development Goals (MDG). In fact, close to one-third of all primary school-age children in Pakistan remained out of school (UNESCO, 2015). Even for those who were enrolled in primary school at least once, approximately 38 percent dropped out (UNDP HDR, 2015). For the stayers, basic numeracy and language skills remained lower than the grade level in which they were enrolled (Andrabi et al., 2013). Dysfunctional schools[22], a dearth of necessary school infrastructure, and teacher absenteeism adversely impact quality of learning at schools (Dundar et al., 2014). The overall performance in the education sectors also masks significant heterogeneity across provinces, and across urban and rural districts within provinces.

Sindh province, with a population of 42.4 million[23], is the second-largest province of Pakistan, and is a particularly resource-constrained province that faces large deficits in public service delivery. The Annual School Census (2014-15) put the number of public schools in the province at 46,071. With 1.08 schools per 1,000 inhabitants, the province has one of the densest public schooling systems in the world. But while there are a large number of schools on paper, many of these schools do not function in reality. Approximately 15 percent of schools in rural areas have either been closed for six months or more, have no students enrolled in them, or do not have teachers assigned to them, according to the ASC, leading to the phenomenon known as "ghost schools".

[22] A school being functional refers to schools being open with teachers and students registered at the time of the Annual School Census (ASC) of 2014-15.
[23] The population of Sindh is roughly one-quarter of Pakistan's total population.



District	Schools	Enrollment	Teachers
Korangi Karachi	610	109113	5310
South Karachi	547	79497	4239
Central Karachi	617	108131	7092
West Karachi	386	58821	2364
East Karachi	300	52769	3278
Malir Karachi	637	55255	2169

Figure 19: Sindh Province with its Districts

Across schools that are functional in rural Sindh, 57 percent of schools only have one teacher assigned to them. Annual Status of Education Report (ASER) Pakistan's 2015 survey reveals that teacher absenteeism hovers around 12 percent for public primary schools. In terms of physical infrastructure, a quarter of schools in Sindh either do not have a school building, or even when a school building exists, it lacks access to facilities such as drinking water, electricity, functioning bathrooms and boundary walls.

Along with poor infrastructure and an endemic shortage of teachers, rural Sindh also has low student enrollment rates. According to results from the Pakistan Social and Living Standards Measurement Survey (PSLM 2014-2015), only 61 percent of all Sindhi children ages 6-10 are enrolled in school at the primary level. The net enrollment rate is 73 percent in urban areas in Sindh province, compared to 77 percent in all of urban Pakistan. The net enrollment rate drops to 52 percent in rural areas in Sindh, compared to 63 percent in all of rural Pakistan.

Students' learning levels in Sindh province correspond to the inadequate investment and inputs in public education. According to the ASER Pakistan's 2015 survey, only 24 percent of Grade 3 students can read words in English, while only 19 percent of Grade 5 students can read full sentences. For Math, learning outcomes are slightly better: 32 percent of Grade 3 students can subtract, while 33 percent of Grade 5 students can perform division. For both subjects, boys outperform girls by six percentage points. These poor learning outcomes can also be partially explained by the fact that on average, only 17 percent of the students' mothers and 44 percent of their fathers have attained at least primary schooling (ASER

2015). The institutional structure of the educational system in Sindh is detailed in Appendix C.1.

3.3.2 Data

The Annual School Census (ASC) collects information on all public schools in the province, consolidated via the Sindh Education Management Information System (SEMIS). Details on the history and mechanics of SEMIS are provided in Appendix C.2. These schools include primary, elementary, middle, secondary, and higher secondary schools; boys, girls and mixed schools; schools located in urban and rural areas; and schools that use either English, Urdu or Sindhi as the primary medium of instruction. Data is collected on school characteristics including functionality, ownership status, infrastructure, classroom equipment, additional facilities such as labs and playgrounds, teachers, students, School Management Committees (SMC)[24], and SMC funding.

The ASC data used for this paper comes from the comprehensive survey conducted by the Sindh Government's Reform Support Unit (RSU) during FY 2014-15 and contains 46,071 unique schools. Of these schools, 10,625 schools are boys-only schools, 7,069 are girls-only schools, and 28,377 are co-educational schools. 41,364 of these schools are in rural areas, while only 4,707 are in urban areas. 41,721 contain primary-level classes, 1,788 contain middle-level classes, 538 contain elementary-level classes, 1,729 contain secondary-level classes and 295 contain higher secondary-level classes. These are not mutually exclusive

[24] School Management Committees (SMC) are community platforms for parents, teachers and other community members to foster a dialogue on the status of schools and education at the local (village/neighborhood) level. SMCs are recognized by the provincial government, with each SMC related to a primary school receiving PKR 22,000 (approximately USD 200) annually. Executive body members are elected by the community, and are responsible for infrastructure funding, temporary hiring of additional teachers and augmenting transportation options to bring children to school through the use of publicly formulated School Improvement Plans (SIPs).

categories, since many schools can have all levels of education, while others are restricted to specific levels, for example, primary-level schooling. A majority of schools in our data use Sindhi as the primary medium of instruction.

Indicators on which data is collected have largely remained unchanged over the past few years. Therefore, while I utilize an annual cross-section of the ASC data for testing and recalibrated the $MSDI$, the analysis can be expanded to cover multiple years. Importantly, a similar Annual School Census with analogous data collection tools is conducted in Punjab province, the largest province of Pakistan, with an estimated population of 100 million people, as well as in Khyber-Pukhtunkhwa (KPK) and Balochistan provinces. This generates potential for the application of the tool both across time and provinces.

3.3.3 Choice of Dimensions: Public Education in Sindh

In this subsection, I expand on the framework established above and provide the rationale for the choice of dimensions related to the public education sector in Sindh, Pakistan. In terms of choosing dimensions for the $MSDI$, I engage intensively with existing development literature to identify dimensions which have a positive impact on educational outcomes. Firstly, I look at impact evaluations across developing countries by using reviews of studies exploring the impact of educational inputs and processes on school outcomes. While these individual studies can have strong internal validity, their external validity in a region with a large number of schools in culturally, geographically, climatically, and linguistically distinct sub-regions is weak. Therefore, relying on individual studies or studies conducted in distinct regions to extrapolate over the functioning of all schools in Sindh province requires strong assumptions related to the generalizability of results of these experiments and quasiexperiments. However, the rich corpus of these studies provides evidence on the range of dimensions that have an impact on key schooling outcomes broadly, under relatively similar conditions.

I focus my attention on Glewwe et al. (2012) and the International Initiative for Impact Evaluation's (3ie) systematic review of "the impact of education programs on learning and school participation in low- and middle-income countries" (2016). Glewwe et al. review a large number of articles and working papers (total of 9,000), concentrating on 79 papers that are relevant to a developing country context, and which use econometric techniques (such as randomized controlled trials, regression discontinuity design, difference-in-differences, matching, and ordinary least square analysis) to assess the impact on student educational outcomes of school infrastructure and pedagogical supplies, teacher and principal characteristics, and/or school organization. Of these, a total of 43 papers are classified as "high-quality", if they used robust identification strategies. The three key areas that seem to work broadly are:

1. School infrastructure and pedagogical materials (electricity, roof/wall/floor, desks/tables/chairs, blackboard, textbooks, library, computers, etc.)
2. Teacher and principal characteristics (education, training, experience, sex, subject knowledge, and ethnicity)
3. School organization (pupil-teacher ratio, teaching methods, teacher absence, homework assignment, student assessment methods, teacher contract, expenditure per pupil, etc.)

Two key outcomes that researchers have looked at are student test scores, and time spent by students in school. The focus in this paper is on the inputs that most affect these two outcomes. Restricting the sample to the 43 shortlisted, high-quality studies, Glewwe et al. find the following results:

Table 27: Indicators and their impacts on key student outcomes

Sr. #	Indicator	Impact on Student Test Scores	Impact on Student Time in School
1	Desks/tables/chairs	Positive	N/A
2	Blackboards	Positive	N/A
3	Libraries	Positive	N/A
4	Roof/wall/floor	Positive	Positive
5	Teacher knowledge	Positive	N/A
6	Teacher training	Positive	Negative
7	Teacher presence	Positive	N/A
8	School meals	Positive	N/A
9	Hours of school day	Positive	N/A
10	Tutoring	Positive	N/A
11	Textbooks/workbooks	Mixed	Positive
12	Computers lab	Mixed	N/A
13	Electricity	Mixed	N/A
14	Teacher education level	Mixed	Positive
15	Teacher experience	Mixed	Mixed
16	Student-teacher ratio	Mixed	Positive
17	Multi-grade teaching	Mixed	N/A
18	Contract teacher	Mixed	N/A

Therefore, amongst the above-mentioned indicators, those related to basic infrastructure appear to have mostly positive impacts on key student outcomes, while other teacher and school organization-related indicators have positive, or mixed results.

Complementary to the above-mentioned review, the International Initiative for Impact Evaluation's (3ie) systematic review synthesizes evidence from 216 education-related programs covering approximately 16 million children across 52 lower and middle-income countries. Besides considering a wider range of studies, the studies reviewed by 3ie follow robust methodologies, and assess impact of inputs and processes on a range of outcomes, such as math scores, language, enrollment, attendance, dropout and cognitive outcomes. The meta-analysis found that the following factors seemed to work, or were promising: merit-based scholarships, school-feeding, cash transfers, public-private partnerships, community-based monitoring, remedial education, new schools and infrastructure, structured pedagogy, and extra time in school. The impacts of other factors were less clear.

Recent literature from Pakistan – mostly gathered as part of Learning and Educational Achievement in Punjab Schools (LEAPS) – shows that school quality affects the demand for schooling and evaluates the magnitude of this relationship. LEAPS data includes both public and private schools. Since the public schooling system is free, parents' demand for better schools can be proxied by their willingness to pay for private schools (via changes in fee) for improved schooling services. Andrabi et al. (2017) find that the equivalent impact of a one standard deviation increase in an index of basic infrastructure (rooms, chairs, blackboards, and other non-building material) leads to a 0.07 standard deviation increase in fees (PKR 55), while a one standard deviation increase in advanced infrastructure (such as a library, fans, or computer facility) leads to a 0.17 standard deviation (Rs 141) increase. Similarly, Carneiro et al. (2013) find that school attributes such as facilities, teacher attributes (the proportion of female teachers, the proportion of teachers with a university degree and the proportion of teachers with at least 3 years of experience), and permanent classrooms affect parents' willingness to pay. There is heterogeneity in these effects, both across boys and girls schools, as well as along different parts of respective distributions.

Thus, there is a noticeable consensus that emerges from these meta-studies. Improvements in basic infrastructure, reduction in teacher absenteeism, improvement in teacher quality, and increase in community engagement and monitoring seem to positively impact key student outcomes. These are, therefore, promising dimensions that should be included in any index for measuring the health of the education sector.

In the best-case scenario where these studies have strong external validity, improvements along any of these dimensions should lead to better outcomes. In the worst-case scenario where these studies have poor external validity, there is no guarantee that these dimensions will lead to better outcomes. To hedge for this, I take a conservative approach in measuring deprivation by considering basic schooling inputs provisions. These include, among others, the provision of shelter, a teacher being present in the classroom, and chairs for all students to sit on.

Critically, basic indicators that are seemingly relevant in the context of developing countries, and in Pakistan's case specifically, are covered under the data collection exercises of provincial governments in Pakistan. Consolidating indicators having a positive impact on school and student outcomes, and comparing these with data available from Sindh province, I finalize a set of 23 indicators, clustered under five disparate dimensions. These dimensions include school status; infrastructure and facilities; teachers; classrooms; and community engagement. Table 28 provides further details. For example, teacher qualification is an indicator within the teacher dimension of school resources, while availability of blackboards is an indicator within the classroom dimension of school resources.

Table 28: Description of Dimensions, Indicators, and Cutoffs

	Dimensions				
	School Status	**Infrastructure and Facilities**	**Teachers**	**Classrooms**	**Community Engagement**
Indicators	Functional	Building structure	Teacher qualification	# of Classrooms	School Management Committee (SMC) functional
		Boundary wall	Teacher experience	Room utilization	SMC funds
		Electricity	Total # of teachers	Students per classroom	
		Fans	Student-teacher ratio	Blackboard	
		Facilities Index		Chairs for students	
		Toilets		Desks for students	
		Student-functional toilet ratio		Chairs for teachers	
		Water		Desks for teachers	

While I have engaged existing literature to establish relevant indicators, I have to omit some indicators that the literature suggests should be added, but for which, the Sindh government does not collect information. For example, in terms of teacher quality, it is worthwhile knowing what the teachers were doing at the time of the data collection visit. While such intensive data is collected for impact evaluations conducted in Sindh province by other non-governmental and/or multilateral organizations such as the World Bank, this data is not collected by the government during the ASC. There is a tradeoff between using either of these two kinds of datasets. While richer data from these impact evaluations can be used to cover more indicators when assessing school deprivation levels in select districts of Sindh province, a focus on a limited number of subgroups prevents the tool's subgroup decompositional abilities to be exploited optimally. Moreover, of interest are trends in school deprivation across time, for which such intensive data is unavailable. Standardized data collected regularly under the ASC allows for the establishment of a set of basic indicators which can be used flexibly for subgroup, as well as intertemporal sectoral health tracking.

3.3.4 Choice of Weights and Cutoffs

This section details the selection of weights accorded to each indicator, as well as the cutoff below which a sectoral unit – school, in this application – is marked as being deprived. It is important to highlight the flexible nature of the selection of weights associated with each dimension. The selection of these weights depends on empirical findings, normative and political economy considerations, as well as participation in and implementation of international accords.

For simplicity, I begin the application of the $MSDI$ on schools in Sindh using uniform weights. While this approach allows me to control for variation in weights to focus on other sources of variation such as dimensional and subgroup impacts on the $MSDI$, the use of uniform weights is a first step. After preliminary analysis, I allow for a more flexible approach containing minimal restrictions on weights, by simulating the distribution of the $MSDI$ over a range of weights in Section 3.5. I begin the analysis by giving weights equal to 1/23 to each of the twenty-three indicators.

The two stages of cutoff selection include setting cutoffs z_j for each dimension j, so that school i can be considered as deprived in that dimension, based on the school's performance on that dimension, and then selecting a cutoff k for the multidimensional deprivation of school i based on the overall deprivation score – taking into account all dimensions.

Selection of cutoffs is based on two key concerns: (i) the nature of responses to different questions for corresponding indicators (binary, categorical, continuous), raising practical restrictions on the selection of cutoffs, and (ii) the motivation for inclusion (or exclusion) of responses in the definition of deprivation. For indicators with binary or categorical responses, I have taken a conservative approach, and allowed the identification function to take a value of 1 whenever the response is clearly indicative of deprivation. For example, for the indicator *school functionality*, both temporary and permanent closure of a school are considered to deprive the school. Conversely, for continuous responses such as number of teachers, student-teacher-ratio, among others, I have used existing literature to guide my selection process. Details on cutoffs z_j are provided in Table 29.

I choose conservative dimensional floors as thresholds. However, while some of them are straightforward (school is closed; no teachers in the school; SMC is not functional; less than one fan per classroom; more students than desks), others such as a student-teacher ratio (STR) of 50, or classroom-room ratio of 0.3 are more arbitrary, benchmark cutoffs.

Table 29: Description of Dimensions, Indicators, and Cutoffs

Dimension	Indicator	Weight	Cutoff
School status	Functional	1/23	School is temporarily or permanently closed
Infrastructure and Facilities	Building structure	1/23	Building structure appears to be dangerous for occupants
	Boundary wall	1/23	The wall is either absent, or dangerous for passersby
	Electricity	1/23	The school is not connected to the grid, or does not receive any electric supply from the grid
	Fans	1/23	Less than one fan per classroom in the school
	Facilities Index	1/23	Facilities include water pump, computer/science/physics/chemistry/biology/home economics labs, library, playground, medical first aid equipment and sports equipment. Cutoff: schools have access to less than three facilities
	Toilets	1/23	Bathroom facility is unavailable
	Student-functional toilet ratio	1/23	More than 30 students to every functional toilet in the school
	Water	1/23	Drinking water is unavailable in the school
Teachers	Teacher qualification	1/23	School-level average teacher qualification is less than required for grade being taught
	Teacher experience	1/23	School-level average teaching experience is less than 5 years
	Total # of teachers	1/23	There are no teachers in the school
	Student-teacher ratio	1/23	Student-teacher ratio is higher than 50 students to a teacher
Classrooms	# of Classrooms	1/23	There are no classrooms in the school
	Room utilization	1/23	Classroom-Room ratio at the school is less than 0.3 (there is less than one classroom to every three rooms in the school)
	Students per classroom	1/23	More than 40 students per classroom
	Blackboard	1/23	Some classrooms in the school do not have blackboards
	Chairs for students	1/23	There are more students than chairs so some students do not have access to a chair
	Desks for students	1/23	There are more than three students to each desk
	Chairs for teachers	1/23	There are more teachers than chairs (so some teachers do not have access to a chair)
	Desks for teachers	1/23	There is more than one teacher to a desk (so some teachers do not have a desk)

Community Engagement	School Management Committee (SMC) functional	1/23	SMC is not functional
	SMC funds	1/23	SMC funds were not disbursed to the respective SMC in FY 2014-15

Choices of weights and cutoffs are not trivial. Variation in k can generate significant variation in the $MSDI$ measure. A low k can lead to a high measurement for overall deprivation, while a high k can lead to a low measurement for deprivation. However, the key is to establish rules for weights and cutoffs and use them over at least a few years. This allows for comparisons across time related to both the overall $MSDI$, as well as performance within subgroups and dimensions. Frequent switching of these parameters makes it difficult to compare results over time, limiting the ability of policymakers to assess sectoral health.

A similar issue arises with varying weights over time. Arguably, one could use data-driven techniques and estimate weights using regression-based methods or factor analysis. However, the correlations uncovered by these methods are sensitive to a given time period, as well as the level of geographic aggregation. If estimated weights vary across time, then it would not make sense to maintain the same weights longitudinally. In this case, comparing sectoral health across years is akin to comparing apples with oranges. Similarly, if estimated weights vary by geographic aggregation as they plausibly do, then intra-region comparisons using the same region-level weights faces a similar issue.

The variable choice of cutoffs z, k, and weights w provides a valuable opportunity to policymakers, politicians, and citizens of a country to evaluate, assess, and impose their preferences and priorities on the evolution of schools in the country. This process of consensual calibration of the $MSDI$ allows for the index to be reflective of contextual expectations and aspirations. For simplicity, I take $k = 0.5$, which combined with the uniformity of the distribution of weights implies that for a school to be deemed multidimensionally deprived, it must be deprived in at least 13/23 indicators. The sensitivity of this measure is assessed in Section 3.5.

3.4 Discussion of Results

This section provides summary statistics and results for the application of the multi-dimensional sectoral deprivation index (MSDI) to public schools in Sindh in FY 2014-15.

3.4.1 Descriptive Statistics

ASC data reveals that 13 percent of surveyed schools were found to be either temporarily or permanently closed. In terms of infrastructure and facilities, approximately one-third of school buildings appeared to be hazardous for occupants. Further, one-half of schools either did not have a boundary wall, or the boundary wall was hazardous for passersby.

I also find that 62 percent of schools were not connected to the electricity grid, and 47 percent of schools had more classrooms than fans. In a province where summers are long and hot – temperatures frequently cross a hundred degrees Fahrenheit – the lack of electricity connections and fans in classrooms can cause acute discomfort to students and teachers during hot school hours. 46 percent of schools did not have access to a bathroom – students would be forced to defecate in other locations, such as open fields. Even for schools where there were functioning bathroom facilities available, 87 percent of schools had more than 30 students per functioning washroom. One-half of schools did not have access to drinking water.

Table 30: Summary Statistics

Sr. #	Indicators	Deprivation Matrix		Censored Deprivation Matrix	
		Mean	SD	Mean	SD
1	School is temporarily or permanently closed	13%	34%	46%	50%
2	Building structure appears to be dangerous for occupants	29%	45%	63%	48%
3	The wall is either absent, or dangerous for passersby	50%	50%	86%	35%
4	The school is not connected to the grid, or does not receive any electric supply from the grid	62%	48%	94%	23%
5	Less than one fan per classroom in the school	47%	50%	54%	50%
6	Washroom facility is unavailable	46%	50%	89%	31%
7	More than 30 students to every functional toilet in the school	87%	34%	96%	19%
8	Drinking water is unavailable in the school	51%	50%	88%	32%
9	School-level average teacher qualification is less than required by statutes for grade being taught	17%	37%	14%	35%
10	School-level average teaching experience is less than 5 years	7%	25%	9%	29%
11	There are no teachers in the school	13%	34%	46%	50%
12	Student-teacher ratio is higher than 50 students to a teacher	29%	45%	59%	49%
13	There are no classrooms in the school	16%	36%	44%	50%
14	Classroom-Room ratio at the school is less than 0.3 (there is less than one classroom to every three rooms in the school)	0%	6%	0%	5%
15	More than 40 students per classroom	38%	48%	56%	50%
16	Some classrooms in the school do not have blackboards	28%	45%	43%	49%
17	There are more students than chairs (so some students do not have access to a chair)	99%	8%	100%	4%
18	There are more than three students to each desk	77%	42%	95%	22%
19	There are more teachers than chairs (so some teachers do not have access to a chair)	29%	45%	72%	45%
20	There is more than one teacher to a desk (so some teachers do not have a desk)	50%	50%	84%	37%
21	Facilities include water pump, computer/science/physics/chemistry/biology/home economics labs, library, playground, medical first aid equipment and sports equipment. Cutoff: schools have access to less than three facilities	97%	17%	100%	2%
22	SMC is not functional	18%	38%	54%	50%
23	SMC funds were not disbursed to the respective SMC in FY 2014-15	39%	49%	74%	44%

In 17 percent of schools, the average teacher qualification was less than what is required by prevailing statutes. An undertrained teacher would be one teaching elementary school but being untrained; teaching middle school but only having a Primary Teaching Certificate (PTC) or less; teaching secondary school but having a Certificate of Teaching (CT) or less; or teaching secondary or higher secondary school without having at least a Bachelor's degree.

Teachers while underqualified, have been holding positions for a relatively long tenure, on average. 93 percent of schools have teachers with an average tenure of greater than or equal to five years. 13 percent of schools were open but did not have a teacher present. For approximately one-third of schools, the student-teacher-ratio (STR) is greater than 50 students to a teacher. I also look at the student-to-classroom (STC) ratio, to get a sense of how crowded classrooms are in Sindh province, and find that approximately 38 percent of schools have STC of 40.

Out of all schools, 16 percent do not have formal classroom. Thus, typical classroom activities would take place in makeshift classrooms, under palm trees or tin foil structures. In terms of classroom facilities, I find that more than 28 percent of schools have at least one classroom without a blackboard – a key medium of instruction. Almost all schools have at least one classroom with more students than chairs, so some students sit on the floor, while 77 percent of schools have at least some classrooms where more than three students use one study desk. These desks at the primary level are small, so more than three children per desk implies that some students do not have access to desk space for books and stationery. On the other hand, 29 percent of schools have some teachers who do not have access to chairs, while half of the schools have at least some teachers who do not have access to tables.

Auxiliary facilities include water pumps; computer, science, physics, chemistry, biology, and home economics labs; libraries; playgrounds; medical and first aid equipment; and sports equipment. I construct an index of these facilities, with deprivation on the index indicated by schools having access to less than three out of these eleven facilities. I find that 97 percent of schools are deprived in terms of extra facilities. As discussed earlier, there is empirical evidence that these facilities have a positive and significant impact on student outcomes. Therefore, this is a high level of deprivation and an obvious target for policymakers.

Besides school-level inputs, the participation of parents and the broader monitoring of school administration by the community are seen in the literature to positively influence school outcomes. Therefore, I focus on School Management Committees (SMC), which comprise teachers, parents and other community members, and are responsible for developing School Improvement Plans (SIP), hiring temporary teachers, and ensuring that transportation is provided to students so that they can get to school. I find that 18 percent of SMCs in the province (all schools are expected to have one) are not functional. Further, SMC funding was not disbursed to approximately two-fifths of all SMCs.

3.4.2 Preliminary Results

Based on the selected dimensions, cutoffs, and weights associated with each dimension, the overall *MSDI* for public schools for FY 2014-15 in Sindh clocks in at 0.17. Further, approximately 27 percent of schools in Sindh are deprived. On the other hand, the average deprivation score of deprived schools is 0.64. Thus, the deprivation intensity across deprived schools is high.

Further, I exploit the decomposition properties of the $MSDI$, and perform decompositions across subgroups such as the six administrative divisions and 28 administrative districts,[25] as well as decompositions across school classifications such as location (rural/urban), gender (boys/girls/mixed), and medium of instruction (Sindhi/English/Urdu). I also implement decompositions across each of the dimensions considered, to explore their contribution to overall school deprivation in Sindh province. I augment this with a division-wise, dimensional decomposition, to see how these contributions change across divisions of Sindh province. Table 31 provides results of the decomposition across divisions. Note that whenever the percentage contribution of a division to the overall $MSDI$ is greater than the population share (of schools) of the division, then the division contributes disproportionately to the $MSDI$. Fitting this criterion should raise a red flag and serve as a target for policymakers.

Table 31: Division-wise Decomposition Statistics

Division	**Sample Size**	**HC Ratio (Poverty Incidence)**	**Poverty Intensity (Average Deprivation Score)**	**Multidimensional Sectoral Deprivation Index (MSDI)**	**Population Share (%)**	**Percentage contribution to MSDI (%)**
Hyderabad	12,760	0.27	0.63	0.17	27.7%	27.2%
Karachi	3,099	0.06	0.61	0.04	6.7%	1.5%
Larkana	7,037	0.34	0.63	0.22	**15.3%**	**19.3%**
Mirpurkhas	8,411	0.39	0.65	0.25	**18.3%**	**27.2%**
SBA at Nawabshah	8,086	0.23	0.64	0.15	17.6%	15.1%
Sukkur	6,678	0.18	0.62	0.11	14.5%	9.6%

[25] In descending order, administrative sub-units of Pakistan include federal, provincial, district, tehsil/town, union, mauza.

I find that the highly urbanized division of Karachi contributes 6.7 percent to the overall school population in the province but contributes only 1.5 percent to the deprivation level of schools in the province. Similarly, Sukkur division contains 14.5 percent of schools in the province, but its contribution to the school deprivation score is 9.6 percent. On the other end of the spectrum, the eastern, lower riparian division of Mirpurkhas contains 18.3 percent of schools in the data, but its contribution to the overall school deprivation score is 27.2 percent. These results are also illustrated in Figure 20. Divisions above, and to the left of the 45-degree line contribute disproportionately more to the *MSDI*.

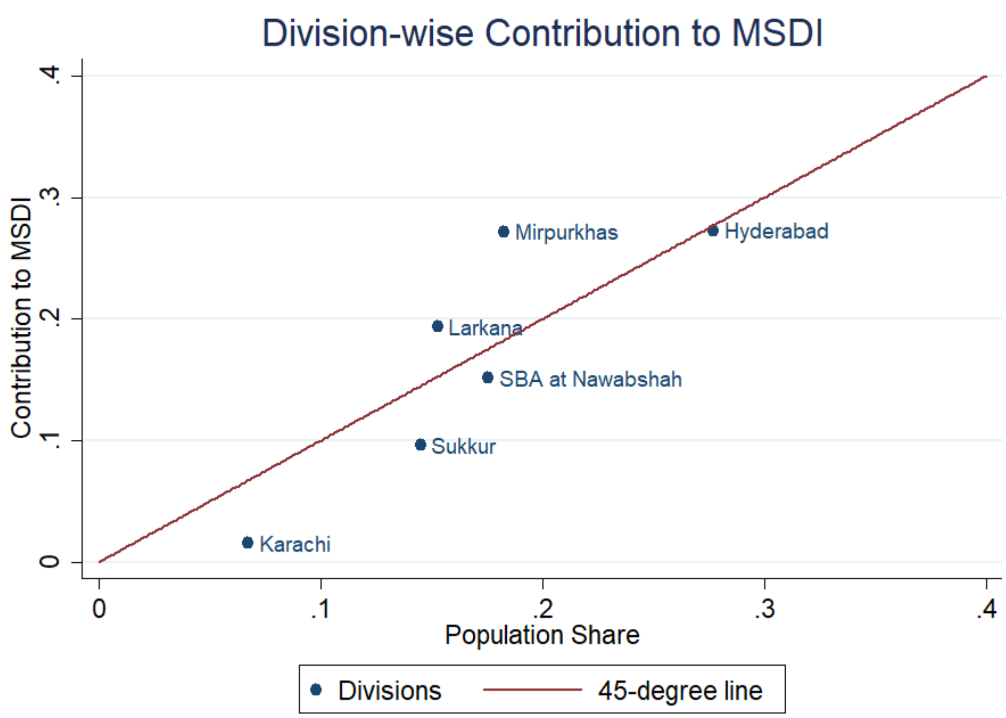


Figure 20: Division-wise Contribution to MSDI

Geographic regions can be further disaggregated for more targeted resource allocation. Table 32 disaggregates key *MSDI* statistics from the division to the district level.

Table 32: District-wise Decomposition Statistics

District	Sample Size	HC Ratio (Poverty Incidence)	Poverty Intensity (Average Deprivation Score)	Multidimensional Sectoral Deprivation Index (MSDI)	Population Share (%)	Percentage contribution to MSDI (%)
Badin	3,052	0.22	0.59	0.13	6.6%	5.0%
Central Karachi	617	0.04	0.59	0.02	1.3%	0.2%
Dadu	2,106	0.28	0.63	0.18	**4.6%**	**4.7%**
East Karachi	302	0.04	0.59	0.02	0.7%	0.1%
Ghotki	1,961	0.27	0.61	0.17	4.3%	4.1%
Hyderabad	881	0.06	0.59	0.04	1.9%	0.4%
Jacobabad	1,440	0.37	0.59	0.22	**3.1%**	**4.0%**
Jamshoro	788	0.24	0.66	0.16	1.7%	1.6%
Kambar-Shahdadkot	1,631	0.31	0.65	0.20	**3.5%**	**4.1%**
Kashmore	1,495	0.54	0.65	0.35	**3.2%**	**6.7%**
Khairpur Mirs	3,460	0.14	0.65	0.09	7.5%	4.1%
Korangi Karachi	610	0.05	0.57	0.03	1.3%	0.2%
Larkana	1,171	0.09	0.58	0.05	2.5%	0.8%
Malir Karachi	637	0.12	0.63	0.07	1.4%	0.6%
Mirpur Khas	2,169	0.30	0.64	0.19	**4.7%**	**5.2%**
Mitiari	926	0.15	0.64	0.09	2.0%	1.1%
Naushero Feroze	2,333	0.18	0.64	0.12	5.1%	3.4%
Sanghar	3,126	0.29	0.64	0.19	**6.8%**	**7.4%**
Shaheed Benazirabad	2,627	0.20	0.64	0.13	5.7%	4.3%
Shikarpur	1,300	0.35	0.65	0.23	**2.8%**	**3.8%**
South Karachi	547	0.06	0.63	0.04	1.2%	0.3%
Sujawal	1,644	0.50	0.64	0.32	**3.6%**	**6.7%**
Sukkur	1,257	0.14	0.61	0.08	2.7%	1.4%
Tando Allah Yar	831	0.16	0.63	0.10	1.8%	1.1%
Tando Muhammad Khan	1,017	0.23	0.59	0.14	2.2%	1.7%
Tharparkar	4,009	0.46	0.66	0.30	**8.7%**	**15.4%**
Thatta	1,515	0.38	0.67	0.26	**3.3%**	**4.9%**
Umerkot	2,233	0.36	0.64	0.23	**4.8%**	**6.5%**
West Karachi	386	0.05	0.60	0.03	0.8%	0.2%

Districts Dadu, Thatta, Mirpur Khas, Tharparkar, Sanghar, Jacobabad, Shikarpur, Umerkot, Kashmore, Kambar-Shahdadkot, and Sujawal contribute disproportionately to the overall deprivation score. Amongst these, Tharparkar, Sujawal, and Kashmore are of particular concern. Population shares and percentage contribution to overall $MSDI$ for these three districts are, 8.7%, 3.6% and 3.2%, and 15.4%, 6.7% and 6.7%, respectively.

These results are further explored in Figure 21. Divisions Hyderabad, Larkana, and Mirpurkahs appear to be the worst-performing divisions in terms of school-level resources, with a large fraction of their constituent districts lying above the 45-degree line. Note that Hyderabad, in terms of aggregate $MSDI$, did not raise a red flag in the division-wise analysis. This reveals that there is substantial variation in school-level deprivation across districts within Hyderabad division.

On the other extreme is division Karachi, which is also the urban, commercial and financial center of Sindh province. All districts of division Karachi lie well below and to the right of the 45-degree line, indicating a much lower contribution to the provincial $MSDI$ as compared to the districts' contribution to the set of schools. Divisions SBA at Nawabshah and Sukkur both appear to perform better than their counterparts, but not as well as division Karachi.

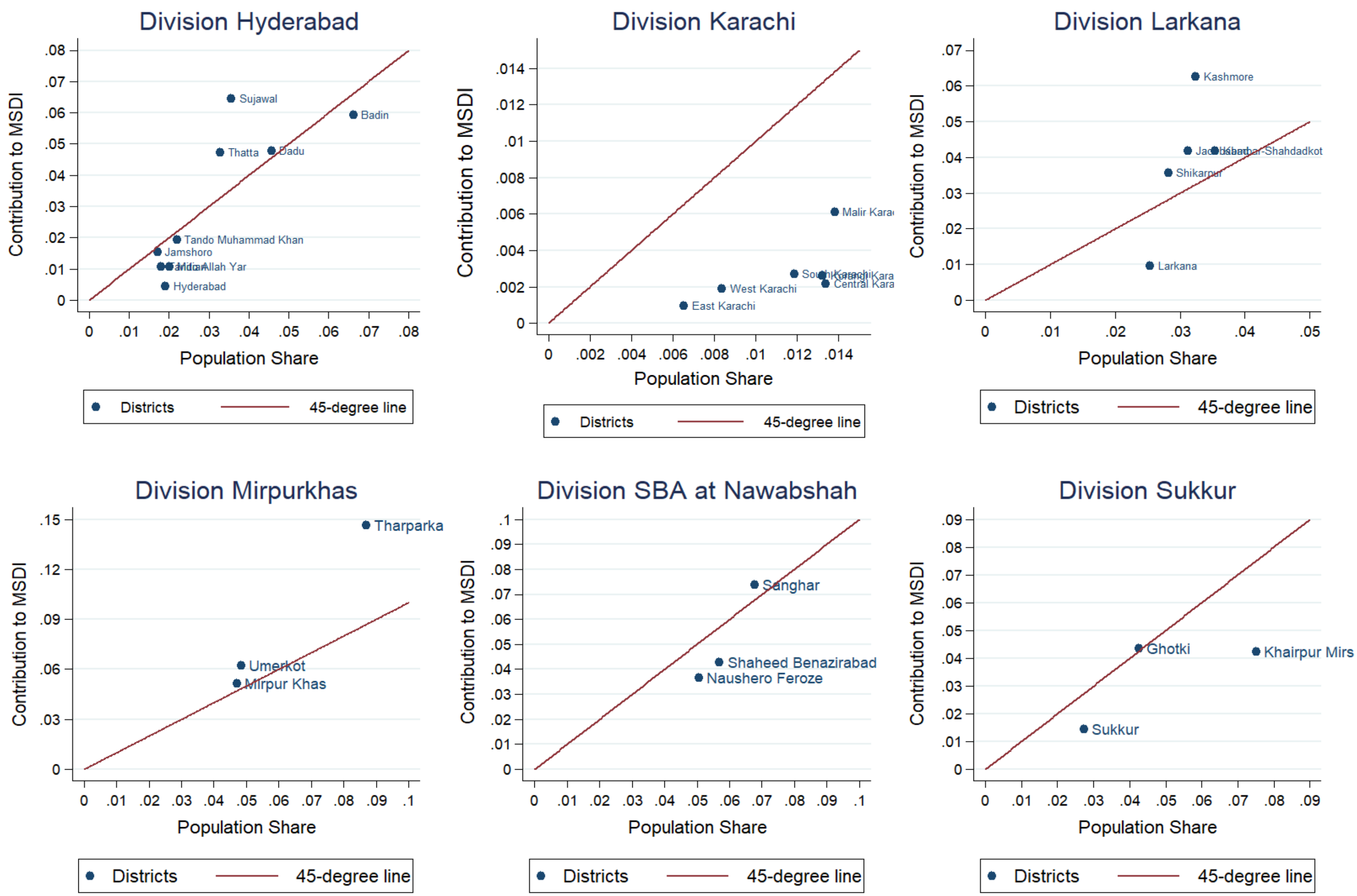


Figure 21: District-wise Contribution to MSDI (by Division)

Decompositions by location, gender and medium of instruction reveal interesting patterns. Urban schools comprise 10.2 percent of all schools in the province, but their contribution to the overall $MSDI$ score is 2.6 percent. Therefore, rural schools contribute disproportionately (97.4 percent) to school deprivation in the province. Further, single-gender schools contribute disproportionately to the overall score on the $MSDI$, as compared to mixed-gender schools. Boys-only and girls-only schools comprise 23.1 percent and 15.3 percent of all schools, respectively. However, their respective contributions to the overall $MSDI$ score are 29.7 percent and 19.7 percent, respectively. 61.6 percent of schools are mixed-gender. However, their contribution to the overall deprivation score stands at 50.6 percent.

In terms of medium of instruction, Urdu and English medium schools comprise approximately one-tenth of schools in Sindh. However, their contribution to the $MSDI$ score is even lower, at 3.5 percent. On the other hand, most schools are Sindhi-medium, and contribute 96.3 percent to the overall deprivation score. These results are provided in Table 33 and are illustrated in Figure 22. Sindhi-medium, single-gender, rural schools lie below and to the right of the 45-degree line, suggesting that they are the least resourced as compared to their subgroup counterparts.

Table 33: Other Subgroup-wise Decomposition Statistics

Classification	**Subgroup**	**Sample Size**	**HC Ratio Poverty Incidence**	**Poverty Intensity (Average Deprivation Score)**	**Multidimensional Sectoral Deprivation Index (MSDI)**	**Population Share (%)**	**Percentage contribution to MSDI (%)**
Location	Urban	4,707	0.07	0.61	0.04	10.2%	2.6%
	Rural	41,364	0.29	0.64	0.19	**89.8%**	**97.4%**
Gender	Boys	10,625	0.33	0.66	0.22	**23.1%**	**29.7%**
	Girls	7,069	0.33	0.66	0.22	**15.3%**	**19.7%**
	Mixed	28,377	0.23	0.61	0.14	61.6%	50.6%
Medium	Urdu	3,246	0.07	0.60	0.04	7.0%	1.8%
	Sindhi	41,274	0.29	0.64	0.18	**89.6%**	**96.5%**
	English	1,551	0.14	0.64	0.09	3.4%	1.7%

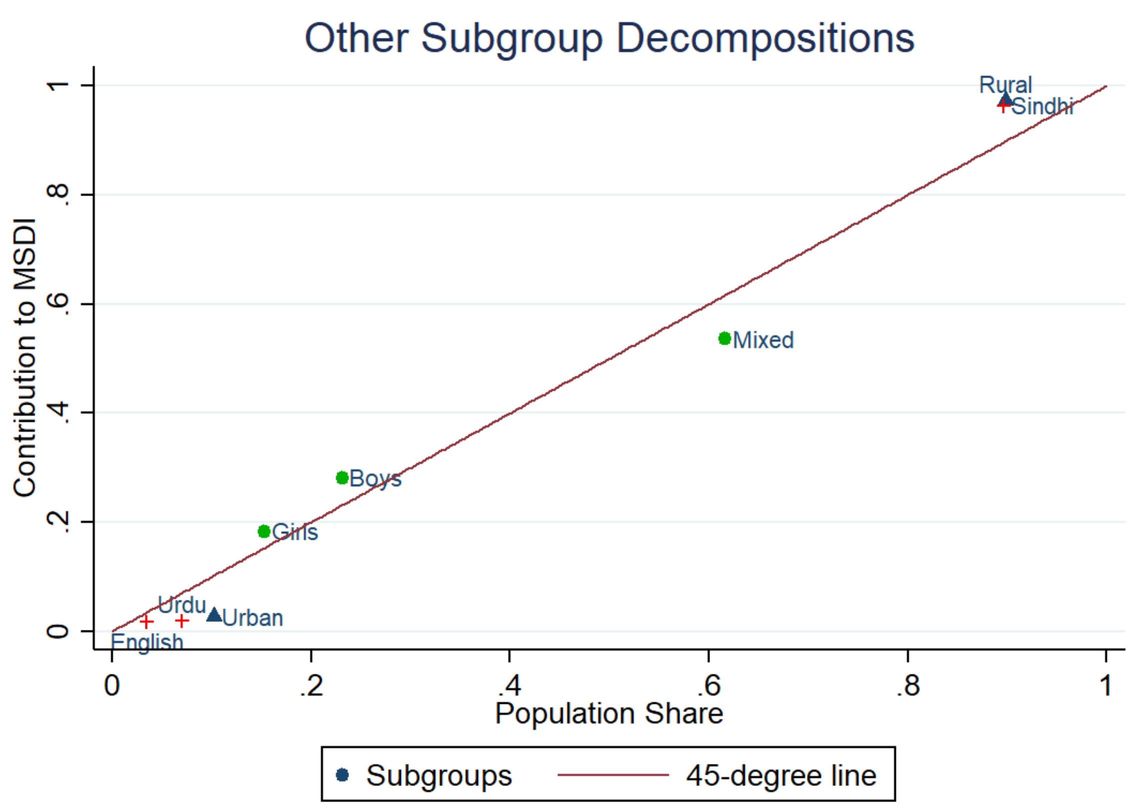


Figure 22: Other Subgroup-wise Contribution to MSDI

While subgroup decompositions provide geographical targets for enhanced fiscal and policy interventions, dimensional decompositions shine a light on areas within schools that are the worst resourced and causing the highest level of deprivation. Given that each of the 23 indicators are equally weighted, the dimensions that contain the highest number of indicators are likely to make the highest contribution to the overall $MSDI$ score, as compared to other dimensions. Therefore, a more informative approach is to study the contribution of each indicator within these dimensions to the overall $MSDI$ score. Figure 23 provides a snapshot of the performance of schools on each indicator.

I find that school status; teacher qualification, experience and number of teachers; number of classrooms, room utilization, and blackboards; and functionality of SMCs contribute disproportionately less to the overall $MSDI$ score. Conversely, boundary walls, electricity, facilities index, toilets, student to functional toilet ratio and drinking water; chairs for students, desks for students and teachers; and SMC funds disbursement contribute disproportionately more to overall deprivation. Specific statistics can be found in Appendix C.3, Table C.1. Such a decomposition can be replicated for each division to generate dimensional maps providing a powerful visual tool for targeted policymaking.

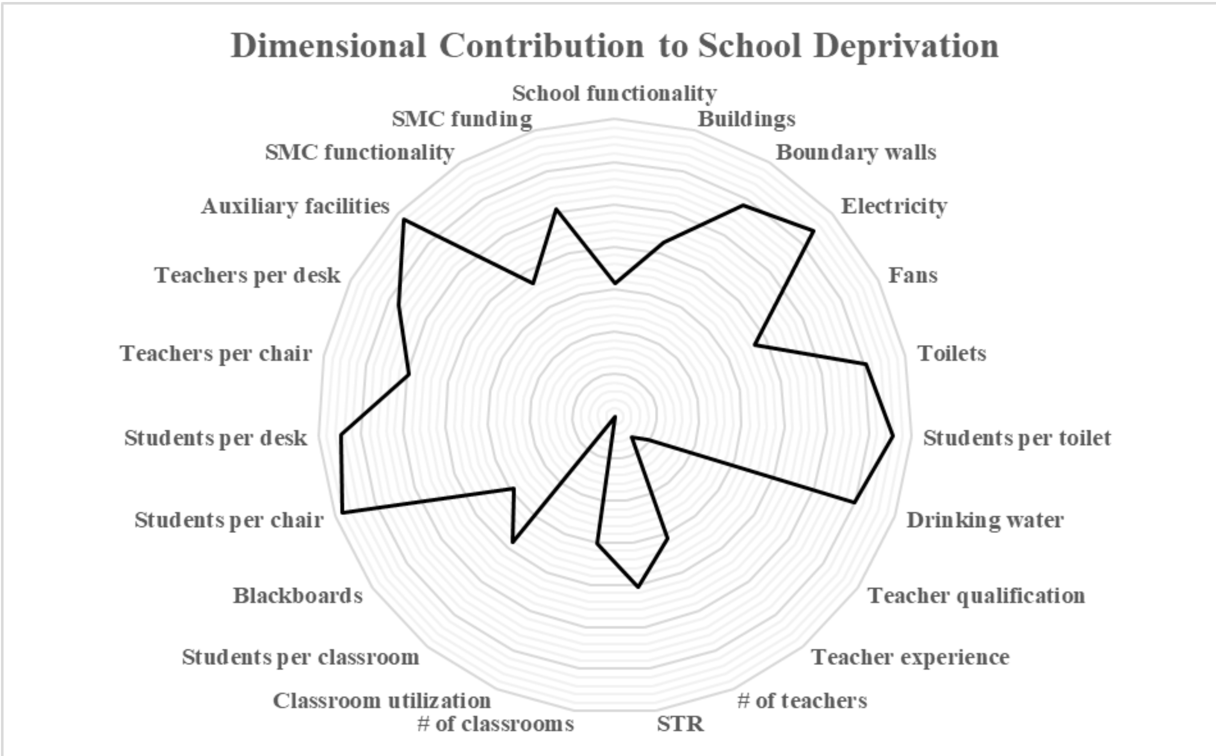


Figure 23: Contribution to MSDI per Indicator

Table 34 compares dimensional breakdowns across divisions. For example, I find that the contribution of school status to the overall score is the lowest in Karachi (0.1 percent) while its contribution from Mirpurkhas (1 percent) is the highest. Similarly, community engagement contributes the most to the overall score from Mirpurkhas (2.4 percent) as compared to Karachi (0.1 percent). These results are illustrated in Figure 24.

Table 34: Dimensional Contribution to MSDI (by Division)

		Dimensions				
	Division	School Status	Infrastructure and Facilities	Teachers	Classrooms	Community Engagement
1	Hyderabad	0.9%	12.8%	2.4%	8.9%	2.3%
2	Karachi	0.1%	0.7%	0.1%	0.5%	0.1%
3	Larkana	0.5%	8.9%	1.5%	6.7%	1.6%
4	Mirpurkhas	1.0%	12.2%	2.5%	9.2%	2.4%
5	SBA at Nawabshah	0.5%	6.8%	1.3%	5.1%	1.5%
6	Sukkur	0.3%	4.3%	1.0%	3.2%	0.8%

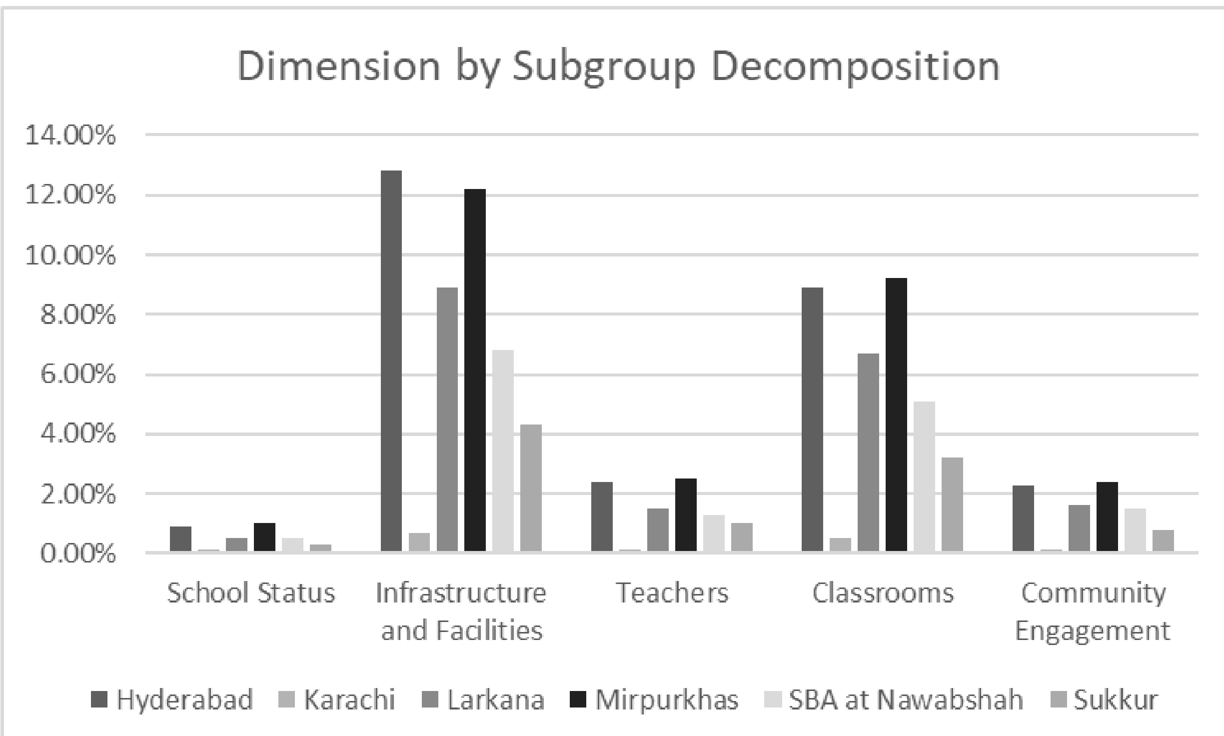


Figure 24: Dimensional Contributions to MSDI (by Division)

3.5 Sensitivity Analysis

Recall that the deprivation score vector c contains scores for each sectoral unit given by $c_i = \sum_{j=1}^{d} w_j g_{ij}^0$. Using constant weights for convenience for each of the indicators reduces this to $c_i = \frac{1}{23}\sum_{j=1}^{d} g_{ij}^0$. However, in practice, these weights can vary, based on the relative importance of the set of indicators used. To assess how the $MSDI$ score will change based on variation in these weights, I run a Monte Carlo simulation and generate a distribution of the overall $MSDI$ score.

For each round of the simulation, I draw a random weight for each indicator with $w_i \sim U(0,10)$, without replacement. I then normalize the weights by dividing each weight by the sum of all twenty-three randomly drawn weights. Thus, each random normalized weight is $W_i \in [0,1]$ and $\sum_{j=1}^{23} W_j = 1$. Using these simulated weights, I calculate a new deprivation score c_i for each school in the application, and then follow the process explained in Section 3.2.1 to calculate the respective sample $MSDI$. I repeat the process a total of 500 times, leading to a distribution comprising 500 $MSDI$ scores. I then compare the mean of this distribution with the $MSDI$ score that I obtained using uniform weights to provide a measure of sensitivity to the results of the analysis with constant weights.

I also conduct the same sensitivity analysis for the fraction of deprived schools H, and the intensity of deprivation A. For each of these constituents of the $MSDI$, I compare the means of their distributions with the values for H and A which I obtained using uniform weights. I repeat the process by also altering the overall deprivation score cutoff k, with $k \in \{0.25, 0.5, 0.75\}$. Recall that deprivation scores c_i that are less than k are suppressed to 0.

Figure 25 provides the results of the sensitivity analysis. Each column of graphs corresponds to a particular k value, while each row of graphs takes as its independent variable the $MSDI$, H, and A statistics, respectively. The graphs have been smoothed using an efficient Epanechnikov kernel. The solid vertical lines indicate the measurement of the statistic based on uniform w_i's. The dotted vertical lines are the mean statistics obtained from the Monte Carlo simulation. It is obvious that for different values of k, the statistics obtained under the strong assumption of uniform weights are very close to the simulated means of the statistics under varying weights. However, for $k = 0.25$, the means of the simulated A and H are marginally different from their measurements under uniform weights. However, these differences are economically small.

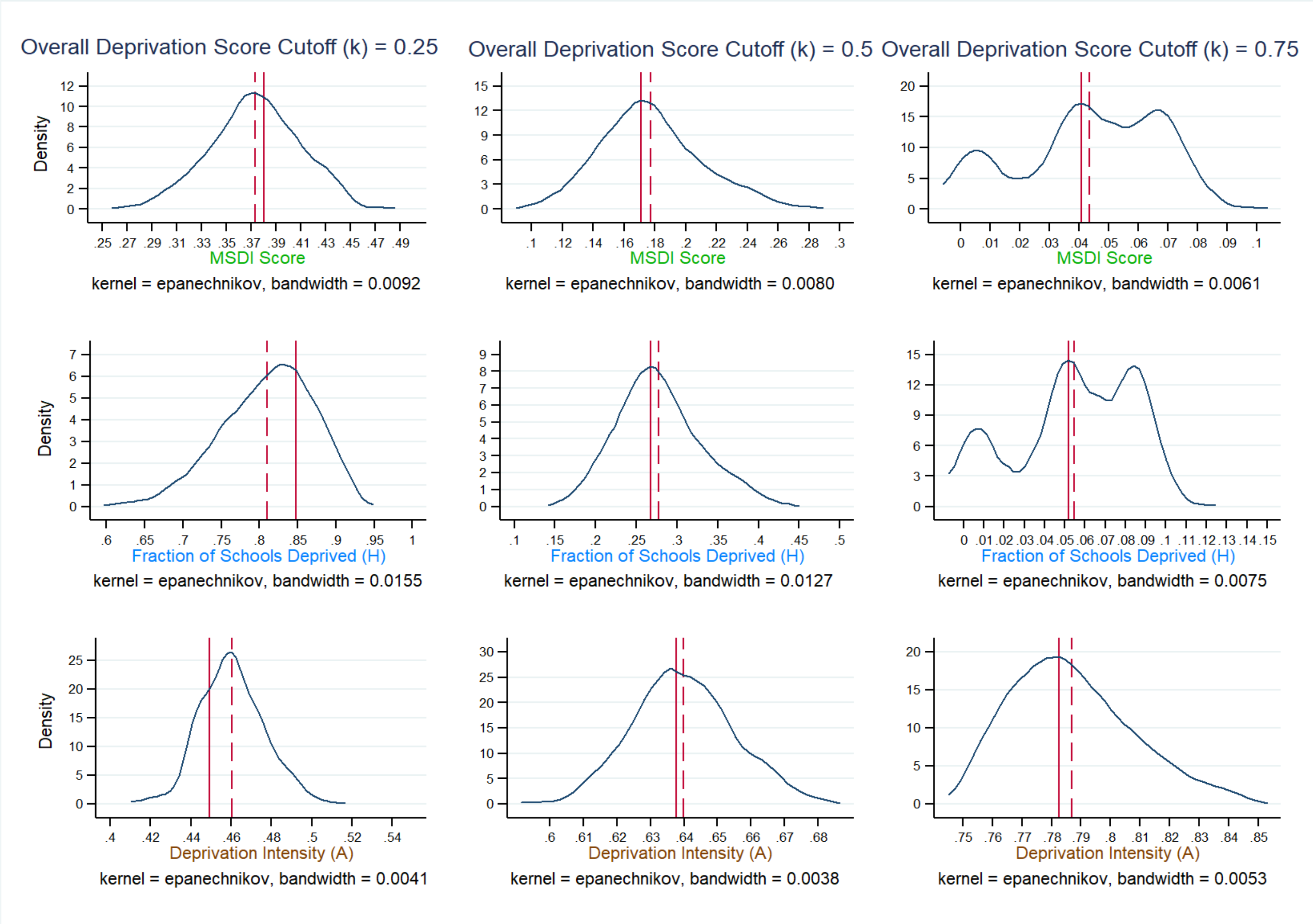


Figure 25: Sensitivity Analysis

The estimated statistics, however, are highly sensitive to the choice of the overall deprivation score cutoff k. Moving from $k = 0.25$ to $k = 0.75$, the mean of the simulated $MSDI$ scores ranges from 0.37 to 0.045. Similarly, the fraction of deprived schools ranges from 81% to 5.5%. This underlines the debate between choosing either an identification function $\rho_k(x_i; z)$ under the intersection or union approaches on the one hand, and a more flexible intermediate identification function on the other, as discussed in section 3.2.1. The choice of an extreme identification function can understate (intersection) or overstate (union) multidimensional deprivation substantially. Therefore, the results of the simulation with varying k cutoff scores suggest that the intermediate approach would be preferred. As pointed out earlier, even with high sensitivity of the measure to k, the use of a consistent k across regions and time allows for strong cardinal comparisons to gauge improvements or worsening of the $MSDI$. Further, gradually reducing k and focusing on subgroups that persist in the fraction of those being deprived allow for clearer identification of the worst-performers.

3.6 Conclusion

The development of the Multidimensional Sectoral Deprivation Index ($MSDI$) provides a unique method of synthesizing a large amount of information that policymakers face on a regular basis, when assessing sectoral health and resource base. The wide range of indicators related to sectoral inputs, processes and outcomes can make cross-regional and intertemporal comparisons difficult. The MSDI allows policymakers to glean useful information from the din of survey and administrative data.

For example, in the case of public education in Sindh province, should policymakers target those districts that are resourced poorly on some subset of infrastructure and school facilities rather than those districts which are resourced poorly on others? Should they be concerned more about districts performing poorly on indicators such as on teacher qualification, as compared to how they are performing on student-teacher ratios? How do policymakers reconcile these variations in education sector health related to school, classroom, student and teacher indicators? Given that policymakers in developing country contexts have access to scarce resources, it is important that they have a robust tool that allows them to measure sectoral health, unpack what they learn, and compare sectoral health and the resource base over time. The $MSDI$ serves as one such tool.

The $MSDI$ can be exploited by researchers and policymakers to focus on the most deprived sectoral units, with deprivation measured via the inclusion of a large number of indicators across a few key dimensions, to construct a single index of deprivation scores. The parameters leveraged by the tool can also be adjusted based on emerging research on what inputs have the largest impact on the performance of sectoral units, as well as on evolving normative concerns about the rights of citizens. However, the tool's capabilities are optimized when these parameters are held stable for a period of time to make longitudinal analysis possible. Thus, the $MSDI$ has high potential for entering the toolkit of policymakers and academics, when targeting sectoral units on the right-tail of the deprivation distribution.

During the development of the tool, I piloted its application to the case of public education in Sindh province, Pakistan, by using the Annual Census of Schools (ASC) from FY 2014-15. Results reveal a high level of deprivation, with subgroup and dimensional decompositions allowing for deeper insights into the regions, classifications and dimensions that are contributing the most to school deprivation in Pakistan.

There is substantial scope for expansion in this research agenda. Some avenues for further research include expanding data fed to the tool for inter-provincial, cross-country and intertemporal analyses; standardizing these scores based on community aspirations and expectations; and including other subgroups such as privates schools and madrassahs (religious schools) in the case of Pakistan. Another potential area of research is to assess the impact of policy interventions on the $MSDI$ by generating $MSDI$ scores with stable parameters over a period of time, and conducting reduced-form impact evaluations using the $MSDI$ as the outcome variable.

Bibliography

Albouy, David 2013 "Partisan Representation in Congress and the Geographic Distribution of Federal Funds" Review of Economics and Statistics, 95(1): 127-141

Alkire, Sabine, and James Foster, 2007. "Counting and Multidimensional Poverty Measurement". OPHI Working Paper 7, University of Oxford (2008).

Alkire, Sabine, and James Foster, 2011. "Counting and Multidimensional Poverty Measurement", Journal of Public Economics 95 (2011) 476–487

Allik, Mirjam, Leyland, Alastair, Ichihara, Maria, and Ruth Dundas, 2019. "Creating small-area deprivation indices: a guide for stages and options", Journal of Epidemiology and Community Health, 2020;74:20-25.

Angrist, Joshua D., and Alan B. Krueger, 1992. "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables With Moments From Two Samples", Journal of the American Statistical Association, 87,328-336.

Andrabi, Tahir, Das, Jishnu, and Asim I. Khwaja, 2013. "Students Today, Teachers Tomorrow: Identifying Constraints on the Provision of Education", Journal of Public Economics, Volume 100, Pages 1-14

Andrabi, Tahir, Das, Jishnu, and Asim I. Khwaja, 2017. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets", American Economic Review, 107 (6): 1535-63.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, T. G. Vishwanath and Tristan Zajonc, 2008. "Pakistan - Learning and Educational Achievements in Punjab Schools (LEAPS): insights to inform the education policy debate.".

Annual Status of Education Report Pakistan, 2016. "ASER Pakistan 2016", South Asian Forum for Education Development, Idara-e-Taleem-o-Aagahi

Aparicio, Juan, Lopez-Espin, Jose J., Martinez-Moreno, Raul, and Jesus T. Pastor, 2014. "Benchmarking in Data Envelopment Analysis: An Approach Based on Genetic Algorithms and Parallel Programming," Advances in Operations Research, vol. 2014, Article ID 431749, 9 pages, 2014. https://doi.org/10.1155/2014/431749

Bagues, M., & Esteve-Volart, B. (2016). “Politicians’ Luck of the Draw: Evidence from the Spanish Christmas Lottery.” *Journal of Political Economy*, 124(5), 1269–1294.

Barber, Michael, 2013. “The Good News from Pakistan”, Reform

Berg, Sanford V., 2010. “Water Utility Benchmarking: Measurement, Methodology, and Performance Incentives”. London, UK: IWA Publishing, 2010. 172 p. ISBN: 9781843392729.

Berry, Christopher R., Barry C. Burden, and William G. Howell. 2010. “The President and the Distribution of Federal Spending.” American Political Science Review, 104(4): 783-799

Best, Michael C., Brockmeyer, Anne, Kleven, Henrick J., and Mazhar Waseem, 2015. “Production vs Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan”, Journal of Political Economy 123(6), 1311-1355

Brender, A., & Drazen, A. (2008). “How do Budget Deficits and Economic Growth Affect Reelection Prospects? Evidence from a Large Panel of Countries.” *American Economic Review*, 98(5), 2203–20.

Bruce, Donald, and Xiaowen Liu, 2014. “Tax Evasion and Self-Employment in the US: A Look at the Alternative Minimum Tax”, Internal Revenue Service, Statistics of Income Division Publication

Brunner, E., Ross, S. L., & Washington, E. (2011). “Economics and Policy Preferences: Causal Evidence of the Impact of Economic Conditions on Support for Redistribution and Other Ballot Proposals.” *Review of Economics and Statistics*, 93(3), 888–906.

Burman, Leonard E., Engler, Alex, Khitatrakun, Surachai, Nunns, James R., and Sarah Armstrong, 2017. “A Synthetic Income Tax Return Data File: Tentative Work Plan and Discussion Draft”, Tax Policy Center, Urban Institute and The Brookings Institution

Carneiro, P., Das, J., & Reis, H., 2013. “Parental valuation of school attributes in developing countries: Evidence from Pakistan”. Working Paper. Retrieved from: https://www.econstor.eu/bitstream/10419/142399/1/dp9960.pdf

Chetty, Raj, Friedman, John, Olsen, Tore, and Luigi Pistaferri, 2011. “Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records”, Quarterly Journal of Economics, 126, 749–804

Cox, Gary W., and Matthew D. McCubbins. 1986. "Electoral Politics as a Redistributive Game." Journal of Politics 48: 370-89

Currie, J., & Gruber, J. (1996). "Health Insurance Eligibility, Utilization of Medical Care, and Child Health." *The Quarterly Journal of Economics*, *111*(2), 431–466.

Cutler, D. M., & Gruber, J. (1996). "Does Public Insurance Crowd Out Private Insurance?" *The Quarterly Journal of Economics*, *111*(2), 391–430.

Dahl, G. B., & Lochner, L. (2012). "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit." *American Economic Review*, 102(5), 1927–1956.

Dahl, Molly, Schwabish, Jonathan, DeLeire, Thomas, and Timothy Smeeding, 2012. "e Earned Income Tax Credit and Expected Social Security Retirement Benefits Among Low-Income Women", Congressional Budget Office, Working Paper 2012-06

De La O, A. L. (2013). "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico." *American Journal of Political Science*, 57(1), 1–14.

Dixit, Avinash and John Londregan. 1996. "The Determinants of Success of Special Interests in Redistributive Politics" The Journal of Politics, 58(4): 1132-1155

Dundar, Halil, Beteille, Tara, Riboud, Michelle, and Anil Deolalikar, 2014. "Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities", Washington, DC: World Bank Group.

Evans, W., & Fitzgerald, D. (2017). "The Economic and Social Outcomes of Refugees in the United States: Evidence from the ACS." (No. w23498). Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w23498

Fancy, Haider, and Jamila Razzaq, 2017. "Accountability in Education in Pakistan", Global Education Monitoring Report, UNESCO

Feenberg, Daniel R., and James M. Poterba, 2004. "The Alternative Minimum Tax and Effective Marginal Tax Rates." National Tax Journal 57(2), 407–427

Fleenor, Patrick, and Andrew Chamberlain, 2005. "Backgrounder on the Individual Alternative Minimum Tax (AMT)", Tax Foundation

Frank, T. (2007). *What's the Matter with Kansas? How Conservatives Won the Heart of America.* Metropolitan Books.

Glewwe, Paul, Hanushek, Eric, Humpage, Sarah and Renato Ravina, 2012. "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010," Working Papers 120033, University of Minnesota, Center for International Food and Agricultural Policy.

Green, Emma. (May 9, 2017). "It Was Cultural Anxiety That Drove White, Working-Class Voters to Trump" The Atlantic.
https://www.theatlantic.com/politics/archive/2017/05/white-working-class-trump-cultural-anxiety/525771/

Gruber, Jon, and Emmanuel Saez, 2002. "The Elasticity of Taxable Income: Evidence and Implications", Journal of Public Economics, 84(1), 1-32.

Hamermesh, Daniel S., Frazis, Harley, and Jay Stewart, 2005. "Data Watch: The American Time Use Survey", Journal of Economic Perspectives, Volume 19, Number 1, Winter 2015, Pages 221–232

Inoue, Atsushi, and Gary Solon, 2010. "Two-Sample Instrumental Variables Estimators", The Review of Economics and Statistics 2010 92:3, 557-561

Institute for Statistics, 2015. "A Growing Number of Children and Adolescents are Out of School as Aid Fails to Meet the Mark", UNESCO, Policy Paper 22, Fact Sheet 31

Institute of Social and Policy Sciences, 2012. "Eighteenth Constitutional Amendment: Federal and Provincial Roles and Responsibilities in Education", Islamabad, Pakistan

Internal Revenue Service, Statistics of Income (SOI) Division, "Design Changes to the SOI Public Use File (PUF)" (https://www.irs.gov/pub/irs-soi/14rfpufredesignrecommen.pdf)

Internal Revenue Service, Statistics of Income (SOI) Division, 2010. "SOI Products and Services" (https://www.irs.gov/pub/irs-soi/12otproductswinbul.pdf)

Jaap Scheerens, 2004. "The quality of education at the beginning of the 21st century", UNESCO, 2005/ED/EFA/MRT/PI/43

Jahan et al., 2015. "Human Development Report 2015", United Nations Development Programme

Kalíšková, K. (2020). "Tax and Transfer Policies and the Female Labor Supply in the EU." Empir Econ 58, 749–775 (2020). https://doi.org/10.1007/s00181-018-1544-2

Khan, Ayaz M., and Munawar S. Mirza, 2011. "Implementation of Decentralization in Education in Pakistan: Framework, Status and the Way forward", Journal of Research and Reflections in Education, Vol.5, No.2, pp 146 -169

Kleven, Henrik J., Knudsen, Martin B., Kreiner, Claus T., Pedersen, Soren, and Emmanuel Saez, 2011. “Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Demark”, Econometrica, Vol. 79, No. 3, 651–692

Kleven, Henrik J., and Mazhar Waseem, 2013. “Using Notches to Uncover Optimization Frictions and Structural Elasticities: Evidence from Pakistan”, Quarterly Journal of Economics, 128, 669–723, 2013

Larcinese, Valentino, Leonzio Rizzo, and Cecilia lesta. 2006. "Allocating the U.S. Federal Budget to the States: The Impact of the President." Journal of Politics 68: 447-56.

Leip, D. *Dave Leip's Atlas of U.S. Presidential Elections*. http://uselectionatlas.org (date).

Leiserson, Greg, and Jeffrey Rohaly, 2008. “The Individual Alternative Minimum Tax: Historical Data and Projections”, Tax Policy Center, Urban Institute and The Brookings Institution

Levitt, S. D., & Snyder Jr, J. M. (1997). “The Impact of Federal Spending on House Election Outcomes.” *Journal of Political Economy*, 105(1), 30–53.

Lévy, A., Colange, C., Fischel, W., Gombin, J., Jeantet, Q.W., Muñoz, M., Moscona, J., & Passarelli, F. (2020). “Testing the Self-Interested Voter Hypothesis Evidence from the Macron victory in France.” Unpublished.

Manacorda, M., Miguel, E., & Vigorito, A. (2011). “Government Transfers and Political Support.” *American Economic Journal: Applied Economics*, 3(3), 1–28.

McCarty, Nolan M. 2000. "Presidential Pork: Executive Veto Power and Distributive Politics." American Political Science Review 94:117-29

Moffitt, R. A., & Wilhelm, M. O. (2000). “Taxation and the Labor Supply Decisions of the Affluent.” In J. Slemrod (Ed.), *Does Atlas Shrug? Economic Consequences of Taxing the Rich* (pp. 193–234). New York: Russell Sage Foundation.

National Bureau of Economic Research, 2017 “US Individual Income Tax Public Use Sample Documentation” (http://users.nber.org/~taxsim/gdb/)

Nordenbo et al., 2010. “Input, Process, and Learning in Primary and Lower Secondary Schools: A Systematic Review Carried Out for The Nordic Indicator Workgroup (DNI)”, Danish Clearinghouse for Educational Research, Copenhagen

Pakistan Bureau of Statistics, 2015. “Pakistan Social and Living Standards Measurement Survey (PSLM 2014-15)”, Government of Pakistan

Pissarides, Cristopher, and Guglielmo Weber, 1989. “An Expenditure-Based Estimate of Britain’s Black Economy”, Journal of Public Economics 39(1), 17–32.

Planning Commission, 2013. “Pakistan Millennium Development Goals”, Government of Pakistan

Reed, Robert W. “Democrats, Republicans, and Taxes: Evidence that Political Parties Matter” Journal of Public Economics, 90: 725-750

Saez, Emmanuel, 2010. “Do Taxpayers Bunch at Kink Points?”, American Economic Journal: Economic Policy 2 (August 2010): 180–212

Saez, E., Matsaganis, M., & Tsakloglou, P. (2012, January). “Earnings Determination and Taxes: Evidence from a Cohort-based Payroll Tax Reform in Greece.” *The Quarterly Journal of Economics*, 127(1), 493-533

Saez, Emmanuel, Slemrod, Joel, and Seth H. Giertz, 2012. “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review”, Journal of Economic Literature, 50:1, 3–50

Scheerens, Jaap, Luyten, Hans, and Jan van Ravens, 2011. “Perspectives on Educational Quality”, SpringerBriefs in Education, DOI: 10.1007/978-94-007-0926-3_2

School Education Department, 2016. “Sindh Education Profile 2014-15”, Government of Sindh

School Education Department, 2017. “Sindh Education Profile 2015-16”, Government of Sindh

Snilstveit et al., 2016. “The Impact of Education Programmes on Learning and School Participation in Low- and Middle-Income Countries”, International Initiative for Impact Evaluation (3ie)

The Christian Science Monitor. (Oct 27 2020). “As college grads flee the GOP, political ‘diploma divide’ grows.” https://www.csmonitor.com/USA/Politics/2020/1027/As-college-grads-flee-the-GOP-political-diploma-divide-grows

The Economist. (June 5, 2018). “Why people vote against their economic interests.” https://www.economist.com/democracy-in-america/2018/06/05/why-people-vote-against-their-economic-interests

U.S. Census Bureau, 2017. “American Time Use Survey User’s Guide”, (https://www.bls.gov/tus/atususersguide.pdf)

Waseem, Mazhar, 2017. “Using Movement of Exemption Cutoff to Estimate Tax Evasion: Evidence from Pakistan”, Review of Economics and Statistics (Revise & Resubmit)

Weber, Mike (Presentation), 2009. “Individual Income Tax Return Panel Data at SOI: What SOI Has and How You Can Use It”, IRS SOI (https://www.bus.umich.edu/otpr/papers/Weber.pdf)

Zucco Jr, C. (2013). “When Payouts Pay Off: Conditional Cash Transfers and Voting Behavior in Brazil 2002–10.” *American Journal of Political Science*, 57(4), 810–822.

Appendix

Appendix A: Chapter 1

Table A.1: Previous Estimates of the Elasticity of Taxable Income for High Earners

Author(s)	**Data (Years)**	**Tax Change**	**Income Definition**	**Methodology**	**Elasticity Estimate**
Panel A: Regression-based, Difference-in-differences, Instrumental Variable					
Lindsey (1986)	1979-1984	ERTA 81	Taxable Income	DiD	1.7
Feldstein (1995)	1985, 1988	TRA 1986	Taxable Income	DiD	1.25 - 2.14
Sammartino and Weiner (1997)	1989-1995	OBRA 1990 and OBRA 1993	AGI	DiD	Close to zero permanent AGI response
Carroll (1998)	1989-1995	OBRA 1991 and OBRA 1993	Taxable income	Regression-based	0.4
Moffitt and Wilhem (1998)	1983, 1989	TRA 1986	AGI	DiD; FE; IV	0 to 2 (depending on instruments)
Goolsbee (2000)	1991-1995	OBRA 1993	Taxable Income	Fixed Effects and First Differencing	Short-run > 1; Long-run: 0 - 0.4
Gruber and Saez (2002)	1979-1990	Changes across pairs of years	Taxable income	IV Approach	0.57
Saez (2003)	1979-1990	Bracket creep 1979-1981	Taxable Income	Instrumental Variable Approach	0.3 (not statistically significant)
Saez (2017)	2011-2015	2013 top bracket tax rate increase	Taxable income	Comparison of income shares	Short-run: 1.16; medium-run: 0.26
Panel B: Bunching					
Saez (2010)	1988-2004	Income Tax Kinks	Taxable income	Bunching	0.006 to 0.031
Mortensen and Whitten (2016)	1996-2016	Income Tax Kinks	Taxable Income	Bunching	0 (no response)
Chetty et al. (2011) ***	1994-2011	Income Tax Kinks	Taxable Income	Bunching	0.02
Kleven et al. (2011) ***	2007-2008	Income Tax Kinks, Audit Probability	Taxable income	Bunching, Experimental	Self-employed: 0.16; Stock income: 2.24
Kleven et al. (2014) ***	1991-2010	Danish preferential foreigner tax scheme	Annualized Earnings	Bunching	0.3

****** Studies indicated with asterisks are based on income tax data from Denmark*

Table A.2: Legislative History

Policy Instrument (Year)	Description
Tax Reform Act of 1969 (P.L. 91-172)	Introduced the "add-on" minimum income tax of 10% in excess of an exemption of $30,000.
Excise, Estate, and Gift Tax Adjustment Act of 1970 (P.L. 91-614)	Allowed deduction of the "unused regular tax carryover" from the base for the minimum tax.
Revenue Act of 1971 (P.L. 92-178)	Imposed minor provisions regarding foreign income.
Tax Reform Act of 1976 (P.L. 94-455)	Raised rate of minimum income tax to 15% and lowered exemption to $10,000 or half of regular taxes.
Tax Reduction and Simplification Act of 1977 (P.L. 95-30)	Reduced minimum tax preference for intangible costs of drilling oil and gas wells.
Revenue Act of 1978 (P.L. 95-600)	Introduced AMT alongside minimum income tax and moved certain itemized deductions and capital gains to AMT. AMT had graduated rates of 10%, 20%, and 25%, and an exemption of $20,000.
Economic Recovery Tax Act of 1981 (P.L. 97-34)	Lowered AMT rates to correspond with reductions in rates of regular income tax.
Tax Equity and Fiscal Responsibility Act of 1982 (P.L. 97-248)	Repealed "add-on" minimum tax. Made AMT rate a flat 20% of AMT income after exemptions of $30,000 for individuals and $40,000 for joint returns.
Deficit Reduction Act of 1984 (P.L. 98-369)	Made minor changes concerning investment tax credit, intangible drilling costs, and other items.
Tax Reform Act of 1986 (P.L. 99-514)	Raised AMT rate to 21%. Made high-income taxpayers subject to phase-out of exemptions. Increased number of tax preferences. Allowed an income tax credit for prior year AMT liability.
Revenue Act of 1987 (P.L. 100-203)	Made technical corrections related to Tax Reform Act of 1986.
Technical and Miscellaneous Revenue Act of 1988 (P.L. 100-647)	Made technical corrections related to Tax Reform Act of 1986.
Omnibus Budget Reconciliation Act of 1989 (P.L. 101-239)	Made further technical amendments.
Omnibus Budget Reconciliation Act of 1990 (P.L. 101-508)	Raised AMT rate to 24%.
Energy Policy Act of 1992 (P.L. 102-486)	Changes regarding intangible costs of drilling oil and gas wells.
Omnibus Reconciliation Act of 1993 (P.L. 103-66)	Introduced graduated AMT rates of 26% and 28%. Increased exemption to $33,750 for individuals and $45,000 for joint returns. Changed rules about gains on stock of small businesses.
Taxpayer Relief Act of 1997 (P.L. 105-34)	Changes regarding depreciation and farmers' installment sales.

Tax Technical Corrections Act of 1998 (P.L. 105-206)	Adjusted AMT for new capital gains rates.
Tax Relief Extension Act of 1999 (P.L. 106-170)	Changed rules about nonrefundable credits.
EGTRRA (2001)	Tax Cuts and No change in AMT
2006	Introduction of calculator
American Taxpayer Relief Act of 2012	Indexes to inflation the income thresholds for being subject to the tax
2001-2012	Changes in Exemption Rates

Table A.3: Exemption Rates Across Time

Years	**Individual tax rate**	**Married filing jointly ($)**	**Single or head of household ($)**
1986–1990	21%	40,000	30,000
1991–1992	24%		
1993–2000	26% / 28%	45,000	33,750
2001–2002		49,000	35,750
2003–2005		58,000	40,250
2006		62,550	42,500
2007		66,250	44,350
2008		69,950	46,200
2009		70,950	46,700
2010		72,450	47,450
2011		74,450	48,450
2012		78,750	50,600
2013		80,800	51,900
2014		82,100	52,800
2015		83,400	53,600
2016		83,800	53,900
2017		84,500	54,300
2018		86,200	55,400

Table A.4: Exemption Rates and Phase-Out in the Early 2000s

Status	**Single**	**Married filing jointly**	**Married filing separately**	**Trust**	**Corporation**
Tax Rate: Low	26%*	26%*	26%*	26%*	20%*
Tax Rate: High	28%*	28%*	28%*	28%*	20%*
High Rate Starts (2012 and earlier)	$175,000	$175,000	$87,500	$175,000	n/a
High Rate Starts (2013)	$179,500	$179,500	$89,750	$179,500	n/a
Exemption in 2009	$46,700	$70,950	$35,475	$22,500	$40,000
Exemption in 2010	$47,450	$72,450	$36,225	$22,500	$40,000
Exemption in 2011	$48,450	$74,450	$37,225	$22,500	$40,000
Exemption in 2012	$50,600	$78,750	$39,375	$22,500	$40,000
Exemption in 2013	$51,900	$80,800	$40,400	$23,100	$40,000
Exemption phase-out starts at (2012 and earlier)	$112,500	$150,000	$75,000	$75,000	$150,000
Exemption phase-out starts at (2013)	$115,400	$153,900	$76,950	$76,950	$150,000
No more exemption in 2009 at	$299,300	$433,800	$216,900	$165,000	$310,000
No more exemption in 2010 at	$302,300	$439,800	$219,900	$165,000	$310,000
No more exemption in 2011 at	$306,300	$447,800	$223,900	$165,000	$310,000
No more exemption in 2012 at	$314,900	$465,000	$232,500	$165,000	$310,000
No more exemption in 2013 at	$323,000	$477,100	$238,550	$165,000	$310,000
Long-term capital gains rate	15%	15%	15%	25%	20%

** For income within the exemption phase-out, marginal tax rates are effectively multiplied by 1.25, which changes 20% to 25%, changes 26% to 32.5%, and changes 28% to 35%.*

Taxpayer Bunching Behavior in Response to Kinks in the Income Tax Schedule

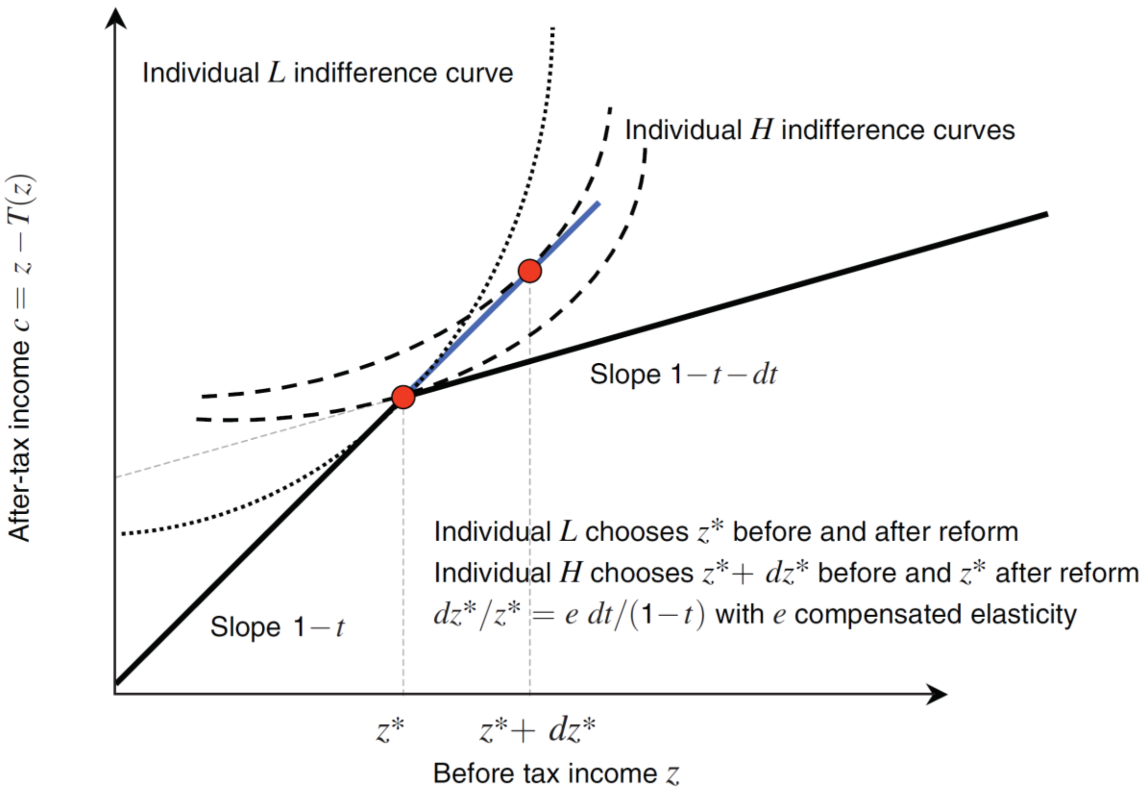


Figure A.1: Impact of Kinks in the Tax Schedule on Taxpayer Behavior

Notes: The effect of a change in the marginal tax rate represented by a kink in the budget set on taxpayer behavior. At the kink, the tax rate increases by t to dt above income level z*. Individual L who chooses z* before the reform stays at z* after the reform. Individual H chooses z* after the reform and was choosing z* + dz* before the reform. Source: Saez (2010)

Distribution of Regular Taxable Income

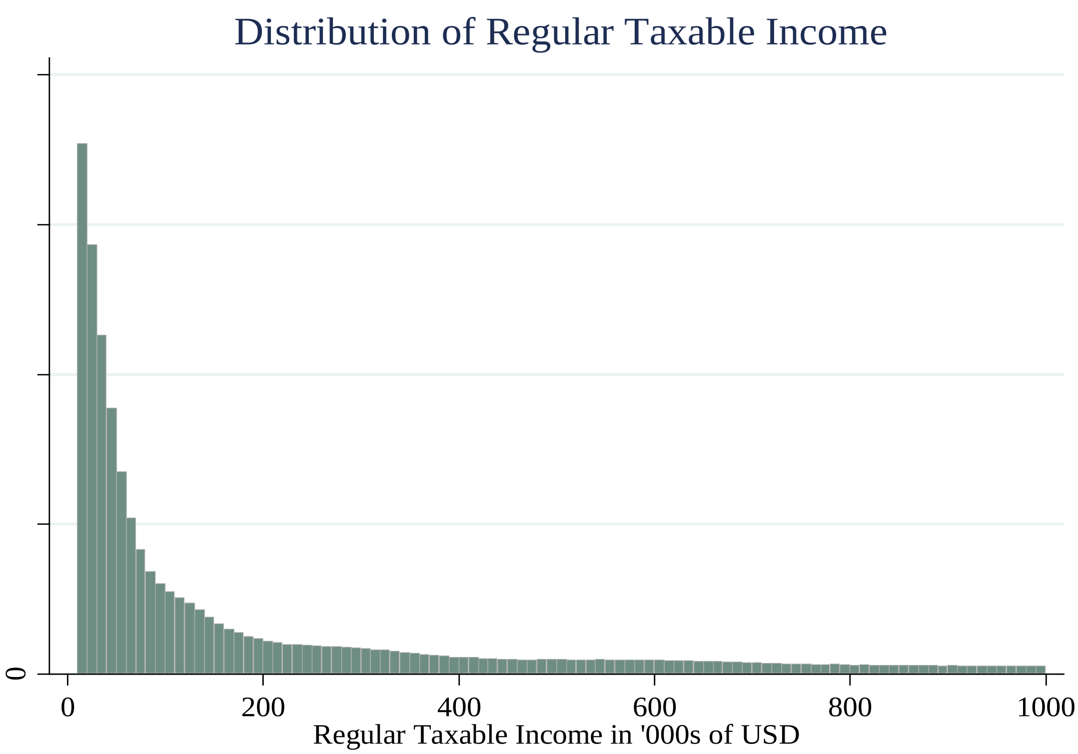


Figure A.2: Distribution of Regular Taxable Income for the Time Period 1993-2011.

Notes: Binwidth is $10,000. The distribution is truncated below at $10,000 and above at $1 million.

Appendix B: Chapter 2

Appendix B.1: Protocol for Simulating Taxes with TAXSIM and ACS Data

The ACS provides us with the following income variables that we input into TAXSIM: wages and salary; other income; public assistance income; self-employment income; interest, dividends and net rental income; pensions; and social security income.[26] We also use property tax information contained in the ACS PUMS and estimate mortgage interest payments to predict itemized deductions for individual taxpayers. For estimating the mortgage interest deduction (MID), we assumed that 60 percent of payments made on mortgages are interest (Evans and Fitzgerald, 2017). We assume individuals with selfcare and independent living difficulties to be dependents, and individuals under the age of 17 to be children eligible for the Child Tax Credit (CTC). All observations are weighted by sample weights provided in the ACS PUMS.

To ensure that our tax simulation method using TAXSIM with the ACS data is robust, we construct an ACS-based tax liability variable for each year using demographic characteristics and tax rules prevailing in that year. We do this for all years in our data for which we have both ACS and SOI data (2010-2018), and compare the performance of our ACS-based estimated liability across max(PUMA, County) against the observed tax liabilities across these units provided by the SOI. For individual years between 2010 and 2018, our ACS-based estimates explain approximately 93 percent of the variation in SOI tax data, according to the R-squared provided by annual, cross-sectional regressions of SOI tax variable on the ACS-

[26] However, the ACS does not provide information on unemployment insurance, childcare expense, and short- and long-term capital gains. This can lead to income being underestimated for taxation purposes.

based estimated tax variable. The table below provides cross-sectional coefficients for each year and the corresponding R-squared value. All coefficients are statistically significant at the 99 percent confidence level.

Table B.1: Correlation between Observed and Simulated Income Tax Liabilities Across Max(PUMA, County)

Year	Coeff.	R-sq
2010	0.74	0.94
2011	0.79	0.94
2012	0.86	0.94
2013	0.80	0.93
2014	0.87	0.92
2015	0.87	0.93
2016	0.81	0.94
2017	0.86	0.93
2018	0.87	0.93

Our estimated ACS-based tax variable underestimates true tax liability, with the gradient of the linear fit being between 0.74 in 2010 to 0.87 in 2014, 2015, and 2016. We want to ensure that our method retains variation across max(PUMA, County), but *on average*, provides a 1:1 mapping of the ACS-based tax variable on to the SOI tax variable. To correct for the systematic underestimation, we compute the difference between what ACS-based liabilities should be in a 1:1 mapping for a given observation, and the predicted value of the line of best fit for that observation. We add these individual differences to every observation to ensure that the gradient in the relationship between the SOI tax variable and the ACS-based tax variable is the same. We replicate this process for the simulated instrumental variable, which is generated by holding 2010 demographics constant.

Appendix C: Chapter 3

Appendix C.1: Institutional Structure and Decentralization of Education in Pakistan

In recent years, political, fiscal and administrative roles and responsibilities related to the education sector have been decentralized in Pakistan. Prior to 2000, education was a primarily federal subject, with certain administrative responsibilities being delegated to the divisional, provincial or district levels. The first significant move towards decentralization did not substantially affect the share of responsibilities between the federal and the lower levels. Rather, it affected the distribution at the lower levels, between the provincial and the district levels[27], by making the district, instead of the provincial level the operational tier of governance. This change was promulgated by the provinces via the provincial promulgation of the Local Government Ordinance (LGO) in 2001.

The LGO followed the broader Devolution Plan of 2000 and devolved various responsibilities from the provincial to the district level. At the district level, the *Nazim* served as the chief executive, and the administration of the district as well as the police reported to him. In terms of the district's administration, the District Coordination Officer (DCO) was appointed to coordinate the workings of 12 different departments, one of which was education. Each of these departments was headed by an Executive District Officer (EDO), who reported to the DCO. The education department was to be headed at the district level by the EDO-Education (Khan et al., 2011). The education department was reflected at the provincial level as well, and was headed by the minster of education, with executive control

[27] In descending order, administrative sub-units of Pakistan include federal, provincial, district, tehsil/town, union, mauza.

vested in the education secretary. Following the promulgation of LGO, planning, monitoring and evaluation of education, as well as disbursement of salaries and the management of teaching and non-teaching staff was transferred to the district level. The provincial government is still responsible for the creation or elimination of different educational posts and positions. Further, the provincial government is responsible for devising the provincial plan for education, as well as to provide backstopping and technical capabilities to districts via specialized units within the provincial education department. An example is the Reform Support Unit within the Education and Literacy Department of Sindh province.

The next big step in the decentralization story was the enactment of the 18^{th} Constitutional Amendment of 2010, a further transfer of responsibilities was made, this time from the federal to the provincial levels. The amendment removed shared responsibilities – such as curriculum, syllabus, planning, policy centers of excellence, standard of education, and Islamic education – between the provincial and federal levels from the concurrent list, and made them the sole responsibility of the provinces, thereby enhancing significantly, the strategic and policy-related responsibilities of the province. Following the amendment, the federal ministry of education underwent a number of name changes, with federal authorities finally settling on the Ministry of Federal Education and Professional Training in 2014. The ministry is primarily responsible for providing technical, vocational and professional skills and training, and collaborates with other organizations to create sponsorships and scholarships for students. Departments reporting to the ministry include the National Vocational and Technical Training Commission, National Commission for Human Development, National Education Foundation and the National Education Assessment System.

In terms of fiscal decentralization, the 18th Amendment required the federal government to provide provinces funding for education via the National Finance Award (NFA) – a formula that divides the revenue pie across the four provinces. However, tax revenues generated at the provincial level meet the major chunk of funding required for education at the provincial level. Certain projects and reforms are funded by loans, grants or assistance of bilateral and multilateral agencies. Specifically, districts are funded by own-revenue, provincial non-earmarked block grants, and ad-hoc education grants provided the federal level. It is important to note that the DCO is the principal accounting officer, and all funds flow through his or her office.

Appendix C.2: Sindh Education Management Information System

The Sindh Education Management Information System (SEMIS) is a derivative of the National Education Management Information System (NEMIS). NEMIS was established in 1990 at the federal level, with the assistance of the United Nations. The purpose of NEMIS was to collect key information on education indicators across the country, and to serve as the national education data repository. Starting 1992, NEMIS has produced the Pakistan Education Statistics, a key national report annually.

By the end of 1993, funding for NEMIS started diminishing, leading provinces to look inwards in terms of setting up semi-autonomous bodies that could play the role of NEMIS at the local level. In Sindh, this led to a collaboration between the Government of Sindh (GoS) and the World Bank, and the establishment of the SEMIS in 1994. The partnership ended in 1996, with SEMIS being shifted from development to non-development funding, under the supervision of the provincial education department.

In 2006, following the devolution of power in Pakistan, the Sindh government decided to consolidate provincial education data under the Reform Support Unit (RSU) housed in the Education and Literacy department of Sindh province. The RSU was a new semi-autonomous body, which was also charged with monitoring and evaluation, and policy implementation of educational reforms in the province, and started with seed funding of PKR 50 million, or approximately USD 800,000. The RSU included three separate wings: SEMIS wing, monitoring and evaluation wing, and the policy wing. The SEMIS wing was also responsible for conducting the Annual School Census (ASC), a universal survey of all public schools in the province. With the development of province-level EMIS structures, the NEMIS took on the role of consolidating data from the provinces and the special regions and developing national level statistics and reports.

In its current form, the RSU compiles data on all schools via the Annual School Census, links this with the EMIS, and maintains a GIS tool which users can use to point-and-click and go to statistics related to individual schools. Each school is identified by a unique SEMIS code. New codes are assigned after a review process of the school, and old codes have been periodically revised to streamline the coding system.

Appendix C.3: Tables for Additional Empirical Results

Table C.1: Dimensional Breakdown of Deprivation

Dimensions	Indicators	Censored HC Ratio	Percentage Contribution	Dimension-wise Contributions	Mean Contribution per Indicator
School status	Functional	0.13	2.5%	2.5%	2.5%
Infrastructure and Facilities	Building structure	0.21	4.1%	47%	6%
	Boundary wall	0.29	5.8%		
	Electricity	0.32	6.4%		
	Fans	0.19	3.8%		
	Facilities Index	0.35	6.9%		
	Toilets	0.30	6.0%		
	Student-functional toilet ratio	0.33	6.7%		
	Water	0.29	5.8%		
Teachers	Teacher qualification	0.05	1.0%	8%	2%
	Teacher experience	0.03	0.7%		
	Total # of teachers	0.13	2.5%		
	Student-teacher ratio	0.18	3.6%		
Classrooms	# of Classrooms	0.14	2.8%	34%	4.5%
	Room utilization	0.00	0.0%		
	Students per classroom	0.19	3.9%		
	Blackboard	0.14	2.8%		
	Chairs for students	0.35	6.9%		
	Desks for students	0.32	6.5%		
	Chairs for teachers	0.21	4.3%		
	Desks for teachers	0.26	5.2%		
Community Engagement	School Management Committee (SMC) functional	0.15	3.0%	8%	4%
	SMC funds	0.23	4.5%		

Appendix C.4: Additional Figures

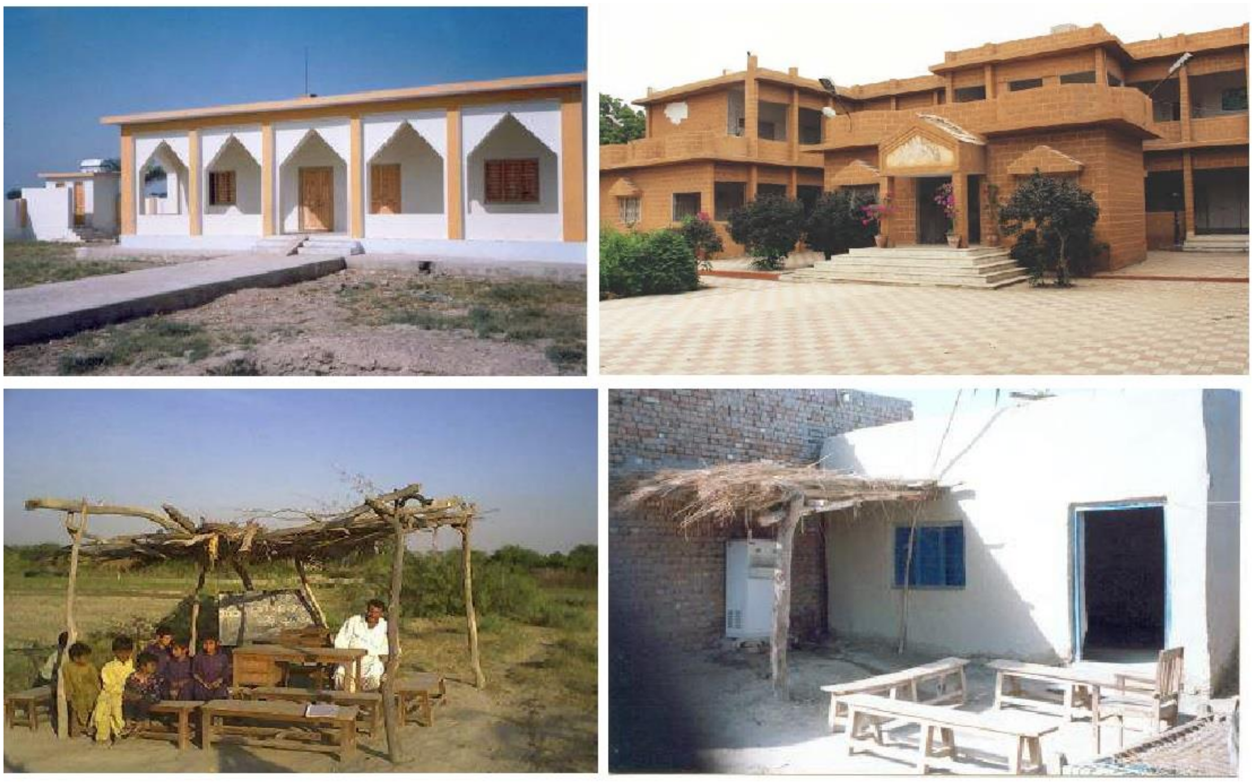


Figure C.1: The Varying State of Schools in Sindh Province

25th Annual School Census 2015-16 پنجويهون ساليانو اسڪول شماري

Sindh Education Management Information System سنڌ ايجوڪيشنل مينيجمينٽ انفارميشن سسٽم

Reform Support Unit ريفارم سپورٽ يونٽ

Education & Literacy Department - Government of Sindh

تعليم ۽ خواندگي کاتو، حڪومت سنڌ

Page - 1

SEMIS Code سيمس ڪوڊ	

Proforma for Primary Schools

GPS Coordinates

Lon (E)	
Lat (N)	

D D D   M M.M M M M

Reference Date | **31st October 2015**

1. School's Basic Information اسڪول جي بنيادي معلومات

a. Taluka تعلقو / Town ٽائون		b. Union council يونين ڪائونسل	c. Tappa تپو
d. Deh ديھ	e. N.A	f. P.S	g. Area علائقو / Village ڳوٺ
h. School name with prefix اسڪول جو نالو			
i. Address ايڊريس/ پتو			j. Phone: فون

2. Information about Data Provider اسڪول جي معلومات ڏيندڙ جا ڪوائف

a. Name نالو		b. CNIC شناختي ڪارڊ نمبر	— —
c. Designation 1=HM; هيڊ ماستر 2=Incharge; انچارج 3=Teacher; استاد 4=Other; ٻيا عهدو		d. Gender جنس	1=Male; مرد 2=Female; عورت
e. Contact No. (Tele/Mob) موبائل نمبر		f. Email-Address اي ميل ايڊريس	

3. School Status اسڪول جي حيثيت

Question سوال	Responses جواب	Code ڪوڊ
a. Type of School اسڪول جو قسم	☐ Viable; هلندڙ جوڳو ☐ Non-Viable; نہ هلندڙ جوڳو	
b. Detail Status تفصيلي حيثيت	1=Functional; فعال 2=Temporary Closed; عارضي بند 3=Permanent Closed; مستقل بند	
c. If 'Temporary Closed' OR 'Permanent Closed', then write the month & year of closure *[This option applied on both condition Viable & Non-Vialbe]*		
d. If 'Temporary Closed' OR 'Permanent Closed', then write code for major reason *[This option applied on both condition Viable & Non-Viable]*	1=Non availability of Teacher; 2=No Population / No Enrollment; 3=School ceases to function long time ago and no record available for this school (Not in existence); 4=Due to litigation; 5=Due to Law and order situation; 6=Any Other Specify ____________	

4. Location هنڌ	1=Urban; شهري 2=Rural; ٻهراڙي	
5. Level درجو	1=Primary; پرائمري 9=Primary with Permission to run middle classes; پرائمري اسڪول مڊل ڪلاس هلائڻ جي اجازت نامي سان گڏ	

6. Is this school has approved Schedule New Expenditure (SNE)? ها جي لاءِ مقرر ٿيل، ڪم ڪندڙ ۽ خالي اسامين جي تعداد لکو	Code		SNE Staff ايس اين اي عملو	Sanctioned Post مقرر ٿيل	Working Post ڪم ڪندڙ	Vacant Post خالي
	1=Yes; ها 2=No; نہ		Teaching تدريسي عملو			
			Non-Teaching غير تدريسي عملو			

7. Administration اسڪول انتظاميہ	1=TEO Male; ٽي اي او ميل 2=TEO Female; ٽي اي او فيميل 3=DO Local Bodies; ڊي او لوڪل باڊيز 4=BOC; بي او سي 5=Other specify__________; ٻيا	
8. Gender (Sex) جنس	1=Boys School; ڇوڪرن جو اسڪول 2=Girls School; ڇوڪرين جو اسڪول 3 = Mixed School; مخلوط اسڪول	

9. Medium [Multi Tick Allowed] and medium wise enrollment as on reference date. 31st Oct 2015 اسڪول جي ميڊيم جي نشاندهي ڪريو، هڪ کان وڌيڪ ميڊيم جي نشاندهي ڪري سگهجي ٿو.	Urdu Medium اردو ميڊيم ☐		Sindhi Medium سنڌي ميڊيم ☐		English Medium انگلش ميڊيم ☐	
	Enrollment داخلا		Enrollment داخلا		Enrollment داخلا	

10. Shift شفٽ	1 = Morning; صبح 2 = Afternoon; ٻيپهري 3 = Both Shifts; صبح ۽ ٻيپهري	
11. Is this a Campus School? ڇا هيءَ ڪيمپس اسڪول آهي for yes write how many schools are merged? ها جي لاءِ ضم ٿيل اسڪولن جي تعداد لکو	1=Yes; ها (No. of Merged Schools;__________) 2=No; نہ	

Figure C.2: Sindh ASC Data Collection Form for Primary Schools (Sample)