From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2024

Behavioral Effects of Taxes and Transfers

Agersnap, Ole

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01hx11xj619

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Behavioral Effects of Taxes and Transfers

Ole Agersnap

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Department of

Economics

Advisers: Owen Zidar and Henrik Kleven

May 2024

© Copyright by Ole Agersnap, 2024.

All Rights Reserved

Abstract

This dissertation comprises four chapters within public economics, which are united by a common thread: They investigate aspects of how taxes and transfers influence human behavior.

Chapter 1, co-authored with Amalie Sofie Jensen and Henrik Kleven, studies the effects of welfare generosity on international migration using reforms of immigrant welfare benefits in Denmark. The first reform lowered benefits for non-EU immigrants by about 50 percent, with no changes for EU immigrants. The policy was later repealed and reintroduced. In a quasi-experimental research design, we find sizable effects: the benefit reduction reduced the net flow of immigrants by about 5,000 people per year, and the subsequent repeal of the policy reversed the effect.

Chapter 2, co-authored with Owen Zidar, estimates the effect of capital gains taxation on realizations at the state level and develops a framework for determining federal revenue-maximizing rates. We find that the elasticity of revenues with respect to the tax rate over a 10-year period is –0.5 to –0.3. Our long-run estimates yield revenue-maximizing capital gains tax rates of 38 to 47 percent.

Chapter 3 studies the effects of complex tax systems. In a model where taxpayers can acquire costly knowledge to reduce their tax burden, I show that a complex tax system can act as a sorting device similar to second-degree price discrimination, where more elastic taxpayers will invest in more tax knowledge. I prove that if elasticities are increasing with income, tax complexity can allow the government to raise higher tax revenues at no efficiency cost. However, a simulation exercise demonstrates that complexity benefits high earners and exacerbates inequality.

Chapter 4, co-authored with Julie Brun Bjørkheim, studies a complex tax system in Norway. Using register data on business owners, we demonstrate that many taxpayers make accounting decisions that cause them to pay higher taxes, and we quantify the size of this overpayment at the individual level. We show that overpayment tends to be larger for women, the less wealthy, and immigrants. We validate the predictions of the theoretical model from Chapter 3 by showing that failure to optimize is associated with a lower estimated tax elasticity.

Acknowledgements

I would like to thank my advisers, Owen Zidar and Henrik Kleven, for their guidance and mentorship. Co-authoring with them taught me so much about the entire research process from a vague idea to final publication. They helped me understand the importance of storytelling and graphical evidence in economic research, and their straightforward feedback helped me learn how to find the right angle on a research idea and make people interested in what you have to say.

I am also thankful to my other coauthors, Amalie Sofie Jensen and Julie Brun Bjørkheim, who have been invaluable sources of inspiration and feedback, and who have done great work on the research projects they have each been part of. I could not have written this dissertation without them. Additionally, although my co-authored work with Oscar Erixson does not form part of this dissertation, I am extremely grateful for his willingness to work with a completely inexperienced graduate student at a whim, for teaching me so much, having many inspiring conversations, and making my third-year paper possible. Similarly, I warmly thank Annette Alstadsæter for her openness to work with me after I emailed her out of the blue, for inviting me to Norway on multiple occasions, and for many great conversations. I am deeply grateful to her and to the Skatteforsk Centre for Tax Research at the Norwegian University of Life Sciences for data access and generous financial support for my work.

My deepest personal gratitude and love goes to my partner, Nandita, who has been by my side during the most stressful portion of the PhD program. She has been my rock when I have been languishing over the difficulties of the job market, and she has been surprisingly willing to listen to me ramble on about tax complexity and elasticities on so many occasions. Conversations with her have sparked ideas that were crucial for parts of the final two chapters of this dissertation. She has been a light for me during the darkest times – I don't know how I wold have made it through the last years of my PhD program without her. I hope that in the future, I can provide her with as much support during her struggles as she has done for me.

Sidst, men ikke mindst, er jeg taknemmelig for min familie. Esben, Susanna, Valdemar og Viola; Rikke, Ronni, Anna og Sara; Morten, Nicoline, Asbjørn og Johanne – at tilbringe tid med dem har været et frirum, hvor jeg har kunnet lægge min forskning til side et stykke tid uden at skulle præstere. De har beredvilligt lagt hus til, når jeg har levet min nomadetilværelse i Danmark

i ferierne. Også mange tak til mine venner fra madklubben, som jeg ville ønske, jeg havde haft mulighed for at se oftere, men som jeg altid har haft glædelige gensyn med.

Den største tak går til min mor og min far, som altid har været stolte af mig og har støttet mig, selv om jeg er flyttet langt væk. Min far ville så gerne have set mig modtage min doktorgrad her i maj, men nåede desværre ikke at opleve det. Jeg er sikker på, at han ville være glad og stolt over, at der fortsat er en dr. Agersnap i familien.

Til far.

Contents

Abstract . . . iii

Acknowledgements . . . iv

1 The Welfare Magnet Hypothesis: Evidence From an Immigrant Welfare Scheme in Denmark 1

1.1 Policy Experiment and Data . . . 4

1.1.1 Policy Experiment . . . 4

1.1.2 Data . . . 6

1.1.3 Migration Patterns . . . 6

1.2 Is There a Welfare Magnet Effect? . . . 7

1.2.1 Within-Country Strategy . . . 7

1.2.2 Cross-Country Strategy . . . 12

1.3 Conclusion . . . 14

Appendices 21

1.A Supplementary Figures and Tables . . . 21

1.B Asylum Applications . . . 29

2 The Tax Elasticity of Capital Gains and Revenue-Maximizing Rates 31

2.1 Data on Capital Gains Taxation and Realizations . . . 34

2.2 Methods . . . 36

2.2.1 Estimating the Annual Effects of Capital Gains Tax Changes . . . 36

2.2.2 Discussion of Alternative Specifications and Semi-Elasticities. . . . 38

2.2.3 Estimating effects over multiple years . . . 38

2.3 Capital Gains Tax Changes and Realizations at the State-Level 40
2.4 Federal Revenue-Maximizing Rates 42
2.4.1 A Simple Model of State-Level Capital Gains Realizations 43
2.4.2 From State-Level Realizations to the Federal Revenue-Maximizing Tax Rate 43
2.4.3 Estimating the Federal Revenue-Maximizing Tax Rate 44
2.4.4 Policy Implications 47
2.5 Conclusion 48

Appendices **50**

2.A Appendix Figures 50
2.B Data Appendix 73
2.C On Elasticities and Semi-Elasticities 76
2.C.1 Empirical Estimates Using a Semi-Log Specification 77
2.D Estimation Using Simulated Data 80
2.D.1 Simulation with constant elasticity 80
2.D.2 Simulation with time-varying elasticity 82
2.E Full Model 84

3 Tax Complexity as Price Discrimination: Theory and Simulation **88**

3.1 First-best case: Optimal individualized tax rates 91
3.2 Complexity as second-degree price discrimination 93
3.3 Simulation 103
3.4 Discussion and Conclusion 106

4 Complexity and Tax Overpayment: Evidence from Norway **110**

4.1 Institutional Background 113
4.2 Data 115
4.2.1 Quantifying Tax Mistakes 116
4.3 Tax overpayment: Empirical evidence 121
4.4 Linking overpayment and elasticity 130
4.5 Conclusion 134

Chapter 1

The Welfare Magnet Hypothesis: Evidence From an Immigrant Welfare Scheme in Denmark[1]

Do generous welfare benefits act as magnets for low-skilled immigrants? This is a classic debate among economists and policy makers, but there is virtually no evidence on the question. The existing literature provides correlational evidence consistent with the welfare magnet hypothesis. For example, Borjas (1999) shows that immigrant welfare recipients in the U.S. tend to be clustered in high-benefit states, while Boeri (2010) shows that low-skilled immigrants in the EU are more likely to locate in high-benefit countries. These patterns are suggestive, but it remains an open question if they reflect a causal relationship or if they are driven by confounding factors correlated with benefit levels. To obtain casual evidence, we need variation in welfare benefits that is plausibly orthogonal to other factors driving location choices.

We argue that Denmark provides an ideal setting for studying this question. First, Denmark has one of the most generous welfare systems in the world — benefit rates are even higher than in the other Nordic welfare states — making it a potential welfare magnet. Second, motivated

[1]This chapter was co-authored with Amalie Sofie Jensen and Henrik Kleven and was published in *American Economic Review: Insights* in December of 2020 (Agersnap, Jensen and Kleven, 2020). Materials from this paper have been presented at The Economic Policy Research Network Conference, Copenhagen, 2019, and the 76th Annual Congress of the IIPF, 2020. We thank Mette Foged, Bo Honoré, Ilyana Kuziemko, Helena Skyt Nielsen, Torben Tranæs, Owen Zidar, and three anonymous referees for comments.

by welfare magnet concerns and general anti-immigration sentiments, Denmark has experimented with immigrant welfare schemes that sharply reduce benefits to certain foreign immigrants. In June 2002, shortly after the formation of a new government supported by a far-right, anti-immigration party, Denmark introduced a welfare scheme that reduced benefits by up to 50% for immigrants from outside the EU.[2] The welfare scheme was controversial and widely debated. It was repealed in 2012 following the election of a center-left government, and then reinstituted in 2015 after the return of a center-right government.

The Danish government has been active in disseminating information about the welfare scheme to potential immigrants. For example, around the reinstitution of the scheme in 2015, the government ran an ad campaign in Lebanese newspapers informing refugees of the reduced benefits.[3] These ads, illustrated in Figure 1.5, highlight the 50% benefit reduction at the top and suggest that the scheme was motivated largely by a desire to reduce the inflow of refugees.

To study the impact of the welfare scheme on migration flows, we consider two difference-in-differences strategies. The first strategy is a *within-country* approach based on comparing immigration flows to Denmark from outside the EU (treatments) and from inside the EU (controls) around the three reforms.[4] The findings are striking: The immigration flows of the two groups evolve in parallel during the twenty years leading up to the 2002 reform, diverge sharply after the introduction of the immigrant welfare scheme in 2002, converge again following the repeal of the scheme in 2012, and diverge once more when the scheme is reintroduced in 2015. We find that the scheme reduced the net flow of immigrants by almost 5,000 people per year, corresponding to an elasticity of migration flow with respect to benefits equal to 1.3. The second strategy is a *cross-country* approach based on comparing non-EU immigration flows in Denmark and a synthetic control country constructed from the other Nordic countries. Consistent with the within-country approach, the cross-country approach features large and sharp migration effects around each of the three reforms.

To further underpin our interpretation of the data, we decompose the immigration effects by

[2]To be precise, the scheme applied to immigrants from outside the EU *and* the four member countries of the European Free Trade Association (EFTA), namely Iceland, Liechtenstein, Norway, and Switzerland.

[3]Lebanon hosts a number of large refugee camps and provides a common transit for Middle Eastern refugees headed to Europe.

[4]To avoid confounding effects of the EU enlargements in the post-reform period, coming from "inside the EU" is defined based on the member countries in the pre-scheme year, 2001.

type of residence permit: asylum permits, family permits, and work/study permits. Only those coming on asylum or through family relations are treated by the welfare cuts. Consistent with this, we show that the effects are driven entirely by asylum- and family-based immigration; the effect on those coming for work or study is a precisely estimated zero. This implies that our estimates are not biased by time-varying unobservables that affect all non-EU immigrants relative to EU immigrants. Any threat to identification must come from time-varying unobservables that vary both by region of origin (EU vs non-EU) and by immigrant type within region (asylum/family vs work/study).

It is worth highlighting two points on interpretation. First, the effects should be interpreted as capturing location decisions *conditional on migration*. Since Denmark is just one small country, we would not expect the decision to emigrate from, say, Afghanistan to be affected by the Danish welfare system. Rather, it is the decision by an Afghan migrant to locate in Denmark instead of, say, Sweden or Germany that is affected by the Danish welfare system. Second, the presence of sizeable welfare magnet effects may make it tempting for governments to introduce immigrant welfare schemes like the Danish one, and in fact several countries have introduced or are discussing the introduction of related policies.[5] Specifically, to the extent that the net fiscal impact of low-income immigration is negative, it may be individually optimal for countries to lower immigrant welfare benefits.[6] However, such policies impose negative fiscal externalities on other countries and are, in general, not socially optimal from a global perspective. This tension between local and global welfare when setting benefits for low-income immigrants is analogous to the tension that arises when setting taxes for high-income immigrants (see Kleven et al. 2019). Our findings suggest that the issues surrounding tax competition and the risk of a "race to the bottom" may be equally relevant for welfare policy.

Our paper contributes to an empirical literature estimating welfare magnet effects, for example Blank (1988), Borjas (1999), Dodson (2001), Gelbach (2004), Kaushal (2005), Fiva (2009), De Giorgi

[5]In the U.S., the welfare reform act of 1996 (PRWORA) denied non-citizens eligibility for welfare benefits. Special immigrant welfare schemes have also been passed in Canada (in 2014), Germany (in 2016), and Austria (in 2019, though the law was ruled unconstitutional and overturned later that same year). Countries where such welfare schemes have been proposed, but not yet implemented, include Finland, France, Latvia, Lithuania, The Netherlands, and Switzerland.

[6]We do not estimate the net fiscal impact of the immigrant welfare reform in this paper. Even if low-income immigrants collect welfare benefits (and pay little tax) in the first years after arrival, the net fiscal impact may be positive in the long run (including intergenerational effects).

and Pellizzari (2009), Boeri (2010), and Razin and Wahba (2015). Much of the literature has focused on migration responses to state-level variation in welfare benefits in the U.S., but the estimates vary greatly across studies and no consensus has been reached. Our main contribution is to provide some of the first quasi-experimental evidence on the existence of welfare magnet effects, and to show that these effects can be sizeable. The welfare scheme that we use for identification has been studied in two existing papers (Rosholm and Vejlin 2010; Andersen, Dustmann and Landersø 2019), but they focus on a different question: the impact of lowering welfare benefits on immigrant outcomes (such as employment, earnings, crime, and children) conditional on locating in Denmark. Finally, while we are interested in the effects of welfare benefits on immigration, there is a recent literature studying what is essentially the reverse of our question: the effect of immigration on preferences for redistribution and voting outcomes (Alesina, Miano and Stantcheva 2018; Alesina, Murard and Rapoport 2019; Dustmann, Vasiljeva and Damm 2019). The introduction of a special immigrant welfare scheme in Denmark provides *prima facie* evidence that immigration may shape redistributive preferences and policy. Importantly, our quasi-experimental approach will not be confounded by such reverse causation channels.

The rest of the paper is organized as follows. Section 1.1 describes the policy experiment and data, section 1.2 presents our results on welfare magnet effects, while section 1.3 concludes.

1.1 Policy Experiment and Data

1.1.1 Policy Experiment

We use the immigrant welfare scheme in Denmark as a quasi-experiment for studying welfare magnet effects. The scheme was announced in January 2002, passed in parliament in June 2002, and took effect from July 2002. Only immigrants coming from outside the EU and the four EFTA countries (Iceland, Liechtenstein, Norway, and Switzerland) are subject to the scheme. For simplicity, we will refer to treated immigrants as coming from "outside the EU". Under the scheme rules, welfare benefits are much lower than the standard, native benefit rate. The largest cut applies to married couples with children for whom the maximum cash benefit is reduced by 50% due to the scheme. For other family types, the benefit drop is somewhat lower. In order for immigrants on scheme benefits to transition to the higher native benefits, they have to stay in Denmark for at least 7

years. The scheme was repealed in January 2012 and reintroduced in September 2015. Figure 1.6 in the appendix illustrates the policy experiment, comparing cash benefits for scheme immigrants to cash benefits for non-scheme immigrants and natives over time. The figure also shows that benefits in the other Nordic countries have been smooth during the time period studied.

This benefit variation forms the basis of our two difference-in-differences designs: (i) a within-country design comparing immigration from outside the EU to immigration from inside the EU into Denmark, and (ii) a cross-country design comparing non-EU immigration in Denmark to non-EU immigration in the other Nordic countries.[7] It is worth noting that cash benefits are higher in Denmark than in the other Nordic countries *even under the scheme rules*. This difference is partly offset by housing benefits, which are higher in the other countries than in Denmark. Still, even accounting for housing support, welfare benefits for Danish scheme immigrants are similar to the benefit levels of other Nordic countries and higher than the benefit levels outside the Nordics. This does not invalidate the experiment by making Denmark a strictly preferred destination. With idiosyncratic variation in non-welfare preferences for location, there will be immigrants close to the indifference margin between Denmark and alternative destination countries. It is these marginal immigrants that may respond to the scheme by changing their destination country.

Aside from immigrant benefit reform, there have been many changes to immigration regulation during the period we study. As documented in Table 1.2, most of these changes have tightened the rules for non-EU immigrants seeking asylum-based or family-based residence, posing a threat to our empirical approach. This concern is strongest for the 2002 reform where several regulatory changes were implemented at the same time. The most important of these changes was arguably the "24-Year Rule" according to which, in order to obtain marriage-based residency, both spouses must be at least 24 years of age. To avoid any confounding effects of the 24-Year Rule, our main specification restricts the sample to immigrants of at least 30 years of age.[8] The other regulatory changes in 2002 cannot be directly controlled for, but they are likely to have had a more limited impact. Ultimately, the strength of our empirical design relies on the presence of three separate welfare reforms, where the last two reform episodes are much less affected by regulation. Specifically,

[7] To avoid confounding effects of the EU enlargements in the post-reform period, coming from "inside the EU" is defined based on the member countries in the pre-scheme year, 2001.

[8] We do not cut the sample exactly at age 24 due to the possibility of intertemporal substitution around the age cutoff, i.e. spouses below the age of 24 who postpone moving to Denmark until they have turned 24. We will investigate the robustness of our results to different age cutoffs.

the 2012 repeal of the immigrant welfare scheme (which we find almost exactly reverses the effect of the 2002 introduction of the scheme) was hardly confounded by regulatory changes.[9]

1.1.2 Data

The analysis is based on administrative data from Denmark covering the full population from 1980 to 2017.[10] We combine several registers, linked at the individual level, to get information about immigrant status, country of origin, type of residence permit, and demographics. We are also able to link family members, which is important for correctly calculating welfare benefits. For the cross-country strategy in which we compare Denmark to other Nordic countries, we combine data from the national statistics bureaus of each country with OECD's International Migration Database over the period 1991-2017.[11]

1.1.3 Migration Patterns

Denmark has seen a strong upward trend in the number of immigrants over the last four decades. The share of first-generation immigrants in the population has increased from 3% in 1980 to over 10% in 2017.[12] As a result, Denmark now has a foreign-born population share almost as high as the U.S. (13.5%).[13] Immigrants from non-EU countries — those targeted by the welfare scheme — constitute the largest group: they accounted for 67% of all first-generation immigrants in 2001, the year before the scheme was introduced.

Figure 1.1 shows migration flows from outside the EU since 1980, in total and from specific countries. Panel A shows that — apart from spikes in 1995 and 2015 — the net flow of non-EU immigration has averaged about 8,000 people per year. A large part of the aggregate flow comes from a relatively limited set of countries, illustrated in Panel A by the series for the eight main sending countries. Panel B considers each of these eight countries separately. The most striking feature of the graph is the presence of sharp spikes in 1995 and 2015, which reflect immigration from Bosnia-Herzegovina and Syria, respectively. The Bosnia spike in 1995 is driven by the Bosnian War

[9] Marriage-based (but not asylum-based) immigration rules were changed in 2011, but the changes were rolled back in 2012, thus being in effect for less than a year. See Table 1.2.

[10] See Statistics Denmark (1980–2017).

[11] Specifically, we use data from Statistics Finland (1990–2019), Statistics Norway (1970–2020), Statistics Sweden (2000–2019*a*), Statistics Sweden (2000–2019*b*), and OECD (1990–2017).

[12] See Figure 1.7

[13] OECD, International Migration Outlook 2019.

of 1992-95 combined with a special Danish law that granted Bosnian refugees residence in Denmark. Since this supply shock occurred well before the first welfare scheme reform, it does not pose a threat to identification. It does create a large spike in the pre-trend for the treatment group, however, and to avoid this we drop immigrants from Bosnia-Herzegovina in all years. The Syria spike around 2015 is driven by the Syrian Civil War. This supply shock does pose a threat to our strategy, because it coincides with the reintroduction of the welfare scheme in 2015. While some of this spike may be interpreted as a confounding supply shock that creates excess migration to Denmark, some of it may also reflect that, conditional on leaving Syria, the choice between Denmark and alternative countries reflects the Danish welfare system. It is not *a priori* clear if any bias from the Syrian refugee crisis will be upward or downward since the excess migration happens on both sides of the scheme reintroduction. We do two things to address identification concerns related to the Syrian supply shock: (i) we run the estimations on a sample without Syrian migrants, and (ii) we carry out a cross-country analysis in which we compare Denmark to other Nordic countries, which were also affected by the Syrian supply shock but did not implement any welfare reforms. Our results are robust to both of these checks.

1.2 Is There a Welfare Magnet Effect?

1.2.1 Within-Country Strategy

Figure 1.2 presents our main results graphically. Panel A shows the net immigration flow from outside the EU between 1980 and 2017, along with a linear trend estimated using pre-scheme data between 1980 and 2001.[14] The time series evidence is striking. Immigration follows a linear trend in the two decades prior to the welfare scheme, diverges sharply from trend after the introduction of the scheme in 2002, catches up with the trend following the repeal of the scheme in 2012, and diverges once more when the scheme is reintroduced in 2015. When the scheme is first introduced, immigration falls for four years before reaching a steady state relative to the linear trend. When the scheme is repealed, it takes exactly four years for immigration to catch back up with the trend. The sharp changes around each of the three reforms and the great degree of symmetry across reforms

[14]Net immigration flow is measured by the year-to-year changes in the stock of first-generation immigrants residing in Denmark according to the Central Person Register (CPR).

provide strongly suggestive evidence of a welfare magnet channel.

Of course, the time series evidence could be biased by non-welfare determinants of migration that change over time. We therefore exploit that the scheme rules apply only to non-EU immigrants, which allows us to compare treated and untreated immigrants in a difference-in-differences (DiD) framework. The results are presented in Panel B, in which we compare net immigration flows from outside the EU (treatments) and inside the EU (controls) over time. Because the two groups trend differently in the raw data, the plotted series have been de-trended by subtracting a linear trend estimated on pre-scheme data. That is, the non-EU immigration series represent the residuals between the actual flow and the predicted flow shown in Panel A, and likewise for the EU immigration series.

The patterns in Panel B are compelling and consistent with the time series evidence in Panel A. The treatment and control series evolve in parallel for a long period prior to the scheme, diverge sharply after the introduction of the scheme, converge when the scheme is repealed, and diverge once more when it is reintroduced. Again, there is a striking degree of symmetry across the different reforms: it takes four years for the full effect of the scheme introduction to materialize, and it takes exactly four years for the scheme repeal to reverse that effect. Moreover, when the scheme is reintroduced in 2015, the impact during the two years for which we have data is similar to the two-year impacts around the other reforms. The graph suggests that, at full impact, the welfare scheme reduced non-EU immigration by about 5,000 people per year, or 4.2 percent of the pre-scheme stock of non-EU immigrants in Denmark.

Although the DiD evidence looks compelling, identification could still be compromised by the presence of time-varying immigration factors that vary by country of origin (EU vs non-EU). As discussed above and documented in Table 1.2, this concern is particularly relevant for the 2002 scheme introduction as it coincided with other changes to immigration regulation. One of the main changes — the 24-Year Rule — is not a confounder here, because the figure restricts attention to immigrants aged 30 and above. But the other changes are not controlled for in the graph. The symmetry of the DiD effects across reforms is particularly informative in this light. The later scheme reforms did not coincide with such regulatory changes — the 2012 repeal in particular is hardly confounded by regulation — and still produced similar effects.

To investigate the threat from time-varying unobservables more broadly, we exploit that our

data include information about residence permits. This allows us to check if the changes around each reform are driven by the types of immigrants who are treated by the welfare scheme. There are four main types of residence permits: asylum, family, work, and study. Those coming on asylum and family permits are potentially eligible for welfare benefits and therefore treated by the scheme, while those coming for work or study are ineligible for welfare benefits and therefore untreated. Figure 1.3 shows immigration flows for the different types of residence permits between 1997-2017.[15] The figure shows that the effects on non-EU immigration are driven entirely by those on asylum and family permits; if anything, immigration by those on work and study permits accelerates following the introduction of the scheme. Furthermore, the figure shows that asylum-based immigration responds more strongly than family-based immigration. This is natural: most newly-arrived refugees have very limited job opportunities and therefore no alternative to welfare benefits, making the scheme treatment particularly severe for them.[16]

One type of confounding shock remains a concern, however: exogenous supply shocks driven by war, political unrest, and the like. In practice such shocks affect only asylum-based, non-EU immigration. If these shocks coincide with the scheme reforms, changes in non-EU immigration by those coming on asylum (relative to family, work or study permits) may reflect non-welfare factors in the sending countries. As described previously, the main supply shock during the post-scheme period is the Syrian refugee crisis around 2014-16. Indeed, Figure 1.3 shows a spike in asylum-based permits relative to other permits during these years. While it is not *a priori* clear that this variation creates bias — in part because the Syrian refugee spike happens on both sides of the scheme reintroduction in 2015 — it does represent a cause for concern. In the next section we investigate this issue by comparing Denmark to similar countries that were also affected by the Syrian supply shock. In this section, we consider a more basic robustness check: dropping Syrian immigrants from the sample throughout. We come back to this below.

Another possible concern is that our measure of immigration is based on residency, which lags behind residence *applications*. It is the decision to apply for Danish residence that should respond to benefit reform, and if the timing of applications is significantly different from the timing of permits,

[15]The figure restricts attention to the period 1997-2017, because the residence permit data is available only from 1997 onwards.

[16]When the scheme was introduced in 2002, the employment rate for refugees was about 10% after one year in Denmark, 19% after two years, and 32% after 3-5 years (Andersen, Dustmann and Landersø 2019). As a result, welfare benefits are the primary source of income for almost all refugees in the first years after arrival.

the DiD evidence is less compelling than it seems. We are able to investigate this issue by using that, for asylum seekers, our data includes information about both applications and permits. We present evidence on asylum applications in the appendix, showing that the empirical patterns for applications and permits are broadly consistent.[17] This lends additional support to our empirical strategy.

To formally estimate the effect of welfare benefits on immigration flows, we collapse the micro data to the level of welfare benefit groups. These groups are defined as those relevant for benefit eligibility (see Figure 1.6): married and unmarried individuals with different numbers of children (0, 1, 2+) coming from outside or inside the EU, a total of 12 different benefit groups. We then run the following DiD regression specification:

$$Y_{gt} = \beta B_{gt} + \gamma_g + \eta_t + \nu_{gt}, \tag{1.1}$$

where Y_{gt} is the net immigration flow for group g in year t, scaled by the pre-scheme immigrant stock for the group, B_{gt} is the maximum monthly benefit for a given group and year, γ_g is a group fixed effect, and η_t is a year fixed effect. To make interpretation easier, we convert the benefit amounts into US dollars (measured in 100s), so that the DiD coefficient β captures the impact on immigration from raising monthly benefits by 100 dollars.[18] The estimation of β in equation (1.1) is based on richer variation than in the graphical analysis, because it exploits that the benefit changes for non-EU immigrants (relative to EU immigrants) differ across families depending on marital status and the number of children. We estimate (1.1) by weighted least squares using the pre-scheme stock of each group as weights.

As discussed above, immigration flows trend upwards over time, and these non-welfare trends differ across treatment and control groups. The DiD graphs were therefore adjusted for linear, group-specific pre-trends. Consistent with this, the regression analysis is also based on de-trended data. Specifically, the outcome variable Y_{gt} is residualized by a linear, group-specific trend estimated on the pre-scheme data (1980-2001). To obtain consistent standard errors from this two-step

[17] See Figures 1.8-1.9 discussed in Section 1.A. We also note that asylum applications are more volatile than residencies, especially around the Syrian shock, which reflects that a sizeable fraction of applications are rejected. This is an important reason for focusing on residencies granted as the main outcome of interest: We are ultimately interested in the effect of benefits on actual immigration, which is better measured by residencies than by applications.

[18] We use the DKK-USD exchange rate of December 31, 2018. In the scheme reform years, we use average benefits over the year based on the month in which the reform was implemented.

procedure, we use GMM to jointly estimate the pre-trends and equation (1.1), and calculate the associated standard errors.

The results are presented in Table 1.1. Panel A shows estimates of the effect on net immigration flow (inflow minus outflow), while Panel B focuses on the inflow alone and breaks the effect down by type of residence permit. The different columns show different sample restrictions in terms of age and whether Syrian immigrants are included or not. Our baseline specification in column (1) considers immigrants above age 30 and includes Syrian immigrants, corresponding to the analysis in Figure 1.2. In this specification, we find that the immigration net flow increases by 0.44% of the stock for each $100 increase in benefits. The immigrant welfare scheme reduced benefits by up to $800 (for married couples with children), implying a total effect on immigration equal to 3.5% of the stock.[19] In the table, we convert this effect into an elasticity of the immigration net flow with respect to welfare benefits.[20] This elasticity is equal to 1.29. Reading across the different columns, we see that the estimates are very robust to alternative age cutoffs and to dropping Syrian immigrants. The migration elasticity varies between 1.29 and 1.53 across the different specifications.

Turning to the effects on gross inflow and its composition in Panel B, the following insights are worth highlighting. First, the total effect on inflow equals 0.4% of the stock per $100 dollar of benefits in the baseline specification. This represents 91% of the net immigration effect shown in Panel A, i.e. almost all of the effect is driven by reduced inflow rather than by increased outflow. Second, the inflow effect is driven entirely by asylum- and family-related migration, with somewhat larger effects on the former. The effect on work- and study-related migration is a precisely estimated zero across all specifications. As discussed, this is critical for the credibility of our results since immigrants coming on work and study permits are untreated by welfare reform. Third, the results are again robust to changing the age cutoff and dropping Syrian immigrants.

To summarize, the within-country DiD analysis reveals large and strongly significant welfare

[19] This is roughly consistent with the results in Panel B of Figure 1.2. There we see a drop in immigration of about 5,000 people (at full impact), corresponding to 4.2% of the 2001 stock of non-EU immigrants.

[20] We define the elasticity of immigration net flow with respect to welfare benefits as follows

$$\varepsilon = \hat{\beta} \cdot \frac{\mathsf{E}\left[B_{g,2001}\right]}{\mathsf{E}\left[Y_{g,2001}\right]}, \tag{1.2}$$

where $\mathsf{E}[\cdot]$ denotes weighted averages across treated benefit groups, and the elasticity has been defined relative to the pre-scheme baseline in 2001. Note that the elasticity with respect to welfare benefits corresponds to an elasticity with respect to total net income if those treated by the welfare scheme (newly-arrived immigrants on asylum or family permits) have no other earnings opportunities in Denmark in the beginning.

magnet effects. The effects of the scheme reforms are sharp and driven by changes in the inflow of immigrants coming on asylum or family permits, precisely those treated by the scheme. The natural interpretation of these effects is that they represent choice of destination country, conditional on migrating. Because new arrivals in Denmark, particularly refugees, have extremely limited employment opportunities for a period of time, it is natural that they are responsive to the drastic welfare cuts implied by the Danish scheme.

1.2.2 Cross-Country Strategy

As discussed above, a potential identification concern is the presence of push factors such as war, famine, and economic crises, which increase immigration and could be correlated in time with the Danish policy experiments. The within-country approach relying on changes in non-EU (relative to EU) immigration driven by asylum and family permits (relative to work and study permits) is not immune to this concern, because the confounding push factors apply specifically to non-EU, asylum-based immigration. The main supply shock during the post-scheme period is the Syrian Civil War, which caused a massive influx of refugees to Europe between 2014-16, i.e. around the time of the reintroduction of the Danish welfare scheme. While we have shown that our results are robust to dropping Syrian immigrants from the sample, other smaller supply shocks may still create bias. In this section, we consider a more compelling way of assessing the threat from such supply shocks: a cross-country strategy comparing Denmark to other countries that experienced the same supply shocks, but did not change welfare benefits.

For this analysis, we draw on additional data from three other Nordic countries: Norway, Sweden, and Finland. We focus on the Nordic countries because they are culturally and economically similar and therefore more likely to be affected by confounding supply shocks in the same way. We use the synthetic control approach developed by Abadie, Diamond and Hainmueller (2010) to construct a "synthetic Denmark," a weighted average of the other countries. To construct the weights, we match on annual migration flows from outside the EU during the pre-scheme period, 1991-2001. More precisely, because migration flows differ in absolute levels due to differences in country size, we match on non-EU migration flows normalized by the pre-scheme stock of non-EU migrants in each country.

Our results are presented in Figure 1.4. Panel A of the figure shows raw immigration flows from

outside the EU in each of the four countries.[21] We see that the migration flows evolve similarly in the four countries throughout the pre-scheme period, 1991-2001, lending support to the parallel trends assumption. Starting in 2002, immediately following the introduction of the Danish welfare scheme, migration to Denmark begins to decline while the three other countries continue their upward trend. Migration flows to Denmark settle at a much lower level until the repeal of the Danish welfare scheme starting in 2012, after which Denmark rapidly catches up with the other three countries. After the reintroduction of the welfare scheme in 2015, we again see flows to Denmark declining compared to its Nordic neighbors.

Panel B of the figure presents findings from the synthetic control approach. The immigration flows in Denmark and the synthetic control country match closely during the pre-scheme years, and the patterns around the scheme reforms are consistent with what we saw in the within-country analysis: Relative to the control country, migration to Denmark declines after the initial reform in 2002, catches up after the repeal in 2012, and again drops following the reintroduction in 2015. This evidence alleviates concerns that the Syrian refugee crisis biases our estimates: the other Nordic countries are exposed to the same supply shock, but experience a different timing in their immigration spike. Immigration to Denmark spikes in 2014-15 — before the reintroduction of lower benefits — while it remains flat in synthetic Denmark during these years and only starts spiking in 2016. In fact, Panel A shows that the 2016 spike happens in every Nordic country except Denmark. This is consistent with a model in which refugees choose where to seek asylum based in part on benefit levels, and where the reduction in Danish welfare benefits causes some Syrian migrants to forgo Denmark in favor of other destination countries.

The findings in Figure 1.4, taken at face value, imply that Denmark's welfare scheme had a considerably larger effect on immigration than estimated from the within-country approach. Note, however, that since we are unable to restrict the sample by age in the cross-country analysis, the response also captures any effect of the 24-Year Rule around the initial 2002 reform. Additionally, if the other Nordic countries serve as substitutes for migrants who would have chosen Denmark absent the welfare scheme, the synthetic control series will overstate the true counterfactual level

[21] We note two limitations of the cross-country data. First, because we are unable to restrict migration by age for the other Nordic countries, the analysis in this section includes migrants at all ages. As a result, the patterns around the 2002 reform may be influenced by the 24-Year Rule. Second, to eliminate the pre-scheme Bosnia shock in a consistent way across countries, we have to exclude migrants from all countries of the former Yugoslavia.

of migration to Denmark. In fact, if migrants' only alternatives to Denmark were the other Nordic countries, the effect in Figure 1.4 would be roughly doubled relative to the true causal effect: Any migrant who avoids Denmark due to welfare reform would instead appear in the control group and thus be counted twice. Of course, migrants do have options outside the Nordics, so we would expect this source of bias to be less dramatic. Nevertheless, while the cross-country analysis provides clear evidence of a welfare magnet effect, and seems to rule out that our within-country results are driven by confounding supply shocks coinciding with the Danish reforms, we caution against attaching too much importance to the specific magnitudes found here. Our estimates from the previous section are better measures of the magnitude of the welfare magnet effect.

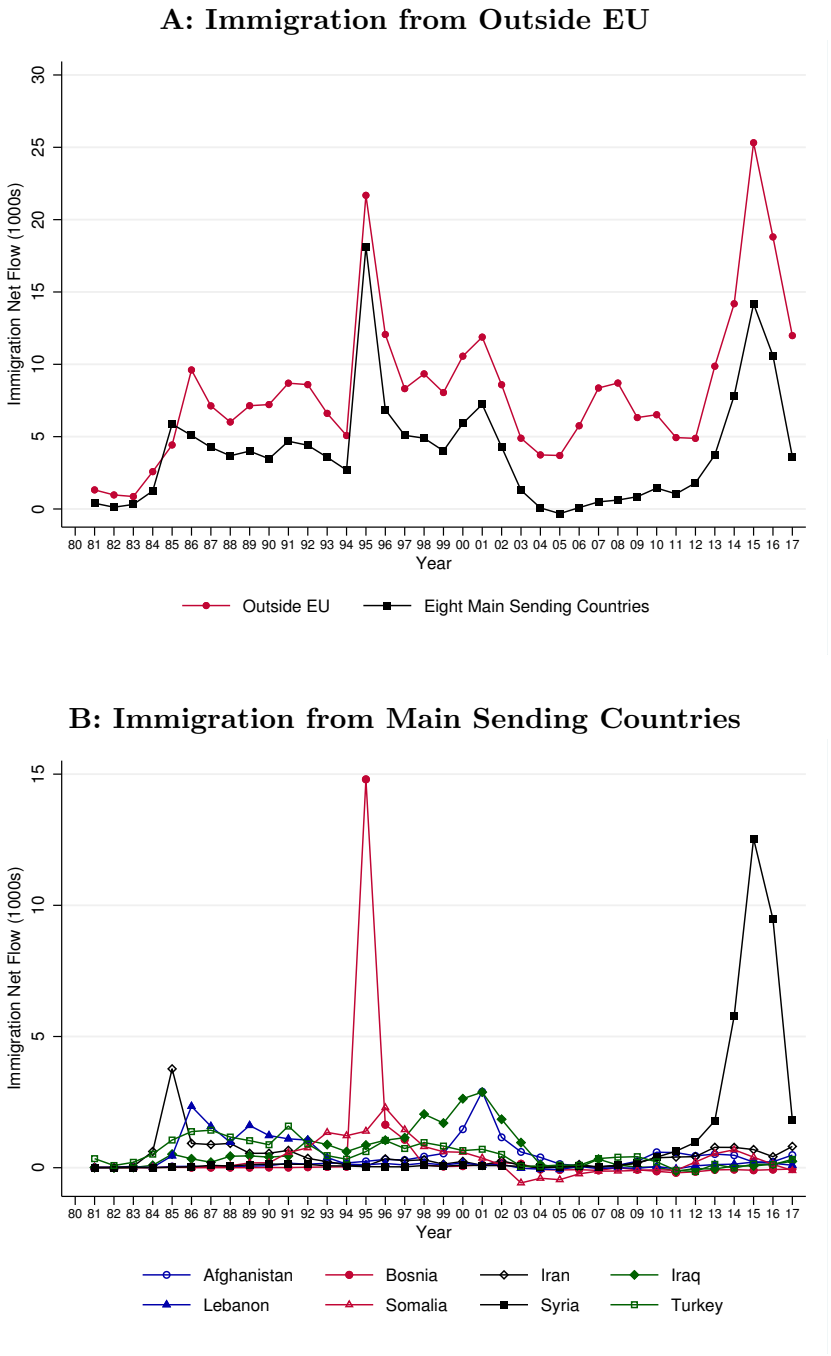
1.3 Conclusion

The possibility of welfare magnet effects has been debated for a long time, but there is virtually no causal evidence on the question. In this paper, we provide quasi-experimental evidence on such effects using a Danish welfare scheme that cut benefits by up to 50% for immigrants from outside the EU. Leveraging three scheme reforms and difference-in-differences designs, we find large and sharp effects of welfare benefits on immigration. The decision to migrate is influenced by many non-welfare factors, but our evidence suggests that, conditional on moving, the generosity of the welfare system matters for the choice of destination country.

Does our evidence imply that the immigrant welfare scheme is an optimal policy? The answer to this question depends partly on the net fiscal impact of immigration (which we do not estimate) and partly on whether we take a local or global perspective. If the net fiscal impact of low-income immigration is negative, local governments have an incentive to deter immigration by cutting means-tested welfare benefits. Assuming that governments care only about the welfare of native residents, welfare schemes like the Danish one may be optimal for individual countries. However, the allocation of immigrants across countries is essentially a zero-sum game, implying that such schemes are not globally optimal. The migrants who avoid Denmark due to the welfare scheme end up in other destination countries, imposing fiscal externalities on them. In the extreme, this may create a race-to-the-bottom in the setting of welfare benefits, similar to the race-to-the-bottom discussed in the context of tax setting (Kleven et al. 2019). Avoiding such effects may require

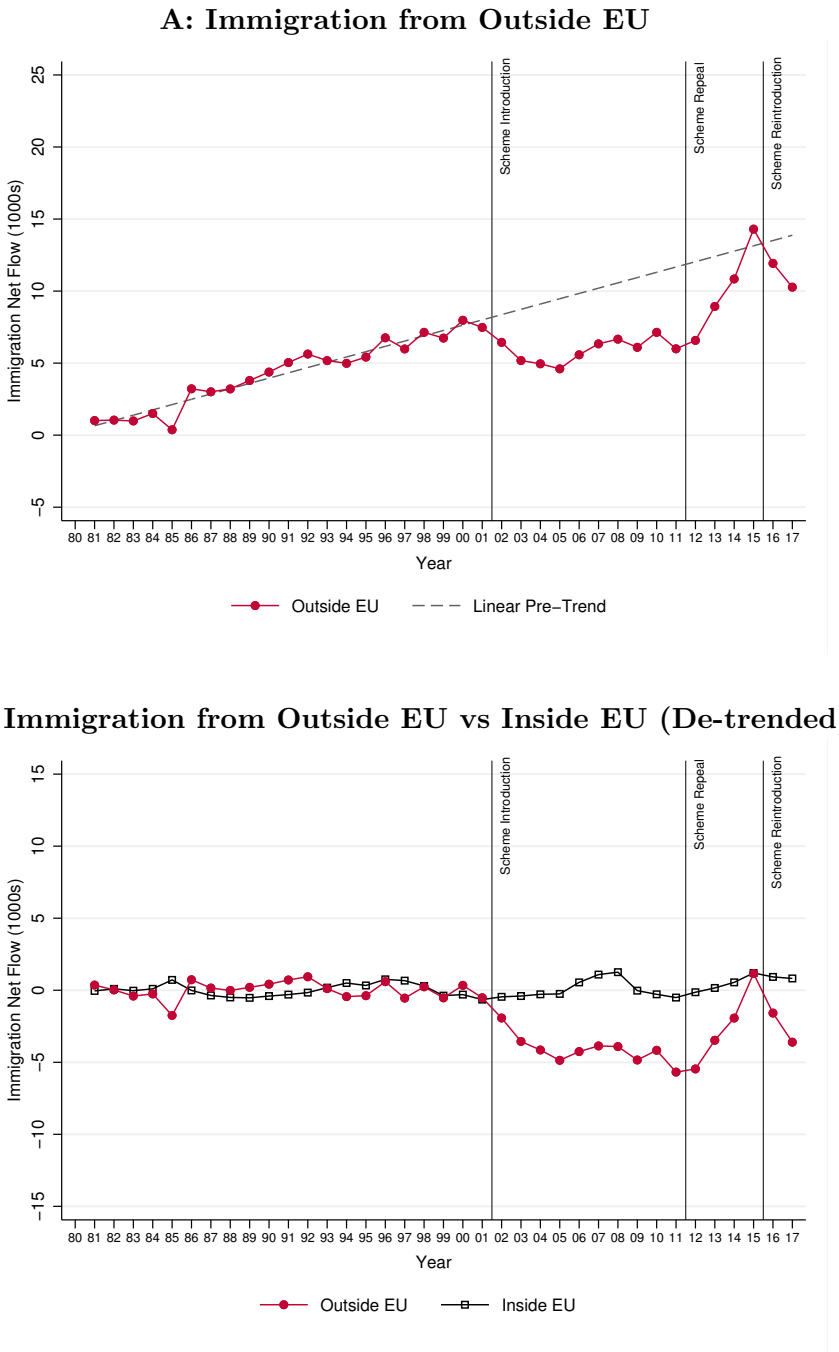
international policy coordination. In any case, while the welfare magnet effects we find may be consistent with the political motivations behind the scheme, they cannot be taken as evidence that the scheme is socially optimal.

Figure 1.1: Migration Flows 1980-2017



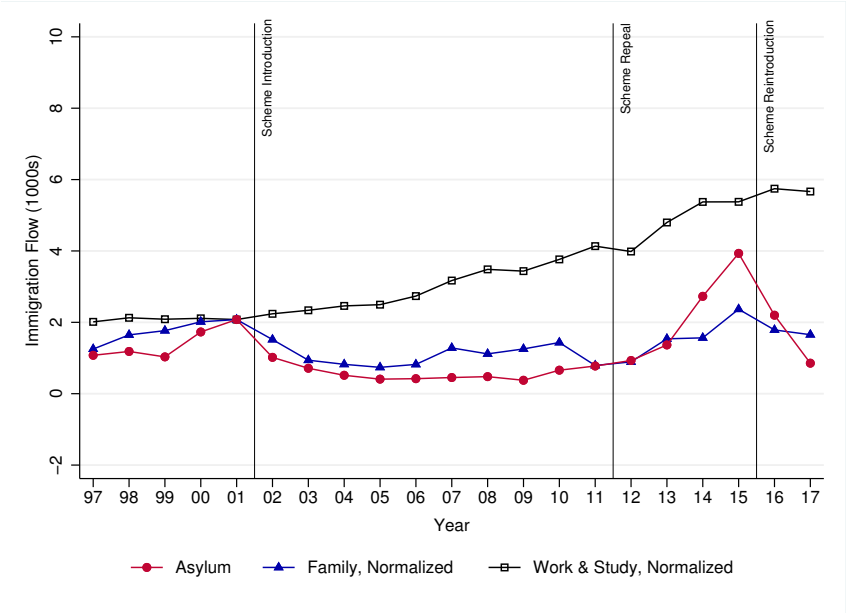
Notes: Panel A shows the net flow of immigrants in Denmark from all countries outside the EU/EFTA as well as from the eight main sending countries outside the EU/EFTA. The main sending countries are defined as those with the highest average annual net flow of immigrants over the years 1980-2017. Panel B shows the net flow of immigrants from each of the eight main sending countries separately. The annual net flow of immigrants is measured as the year-to-year change in the stock of immigrants.

Figure 1.2: Migration Responses to the Immigrant Welfare Scheme



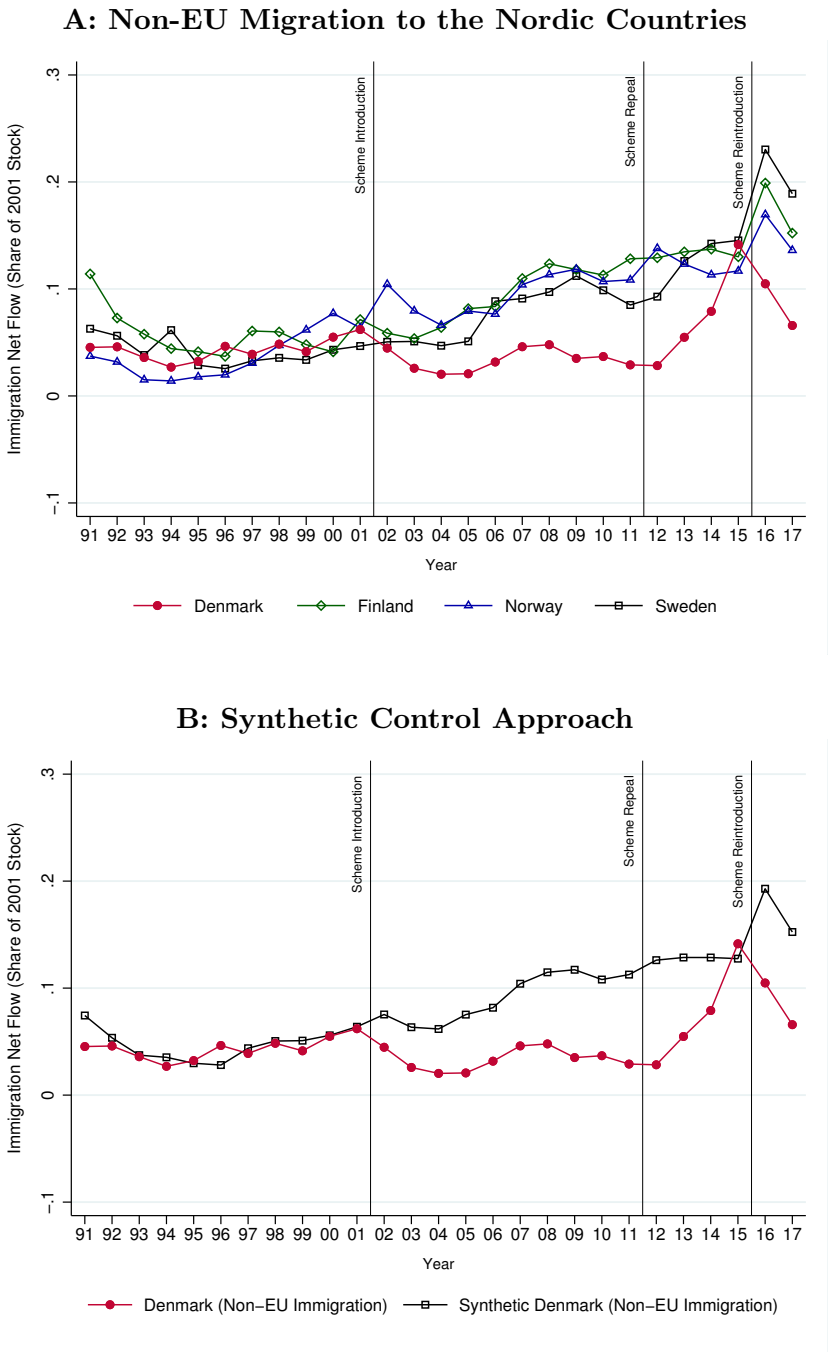
Notes: Panel A shows the net flow of immigrants in Denmark from outside the EU/EFTA over the period 1980-2017, and a linear trend estimated on the pre-scheme data period 1980-2001. Panel B compares the net flow of immigrants from outside EU/EFTA (treatment group) to the net flow of immigrants from inside EU/EFTA (control group). Each series has been de-trended using a linear, group-specific trend estimated on pre-scheme data. We use EU membership in the pre-scheme year (2001) to define the control group. The annual net flow of immigrants is measured as the year-to-year change in the stock of immigrants. The sample is restricted to immigrants who are at least 30 years of age, and immigrants from Bosnia are dropped throughout.

Figure 1.3: Immigration Inflows by Type of Residence Permit



Notes: The graph shows the inflow of immigrants in Denmark from outside the EU/EFTA by type of residence permit (asylum, family, and work/study) over the period 1997-2017. Each series has been de-trended using a linear, group-specific trend estimated on pre-scheme data (1997-2001). The sample is restricted to immigrants who are at least 30 years of age, and immigrants from Bosnia are dropped throughout.

Figure 1.4: Welfare Magnet Effects: Cross-Country Evidence



Notes: Panel A shows the annual net flow of immigrants from outside the EU/EFTA to each of the four Nordic counties, divided by the country's 2001 stock of non-EU/EFTA immigrants. Panel B retains the same data series for Denmark as Panel A, but combines the three other Nordic countries into one synthetic control country, labeled "Synthetic Denmark". This series is constructed as a weighted average of Finland, Norway and Sweden to minimize the sum of squared errors in the pre-scheme (1991-2001) immigration flows between Denmark and Synthetic Denmark. This yields the following weights: Finland 0.424, Norway 0.397, Sweden 0.179. Unlike the within-country analysis, no age restriction has been placed on migrants. To correct for the pre-scheme Bosnia shock, immigrants from the former Yugoslavia are dropped throughout.

Table 1.1: Migration Responses to the Immigrant Welfare Scheme

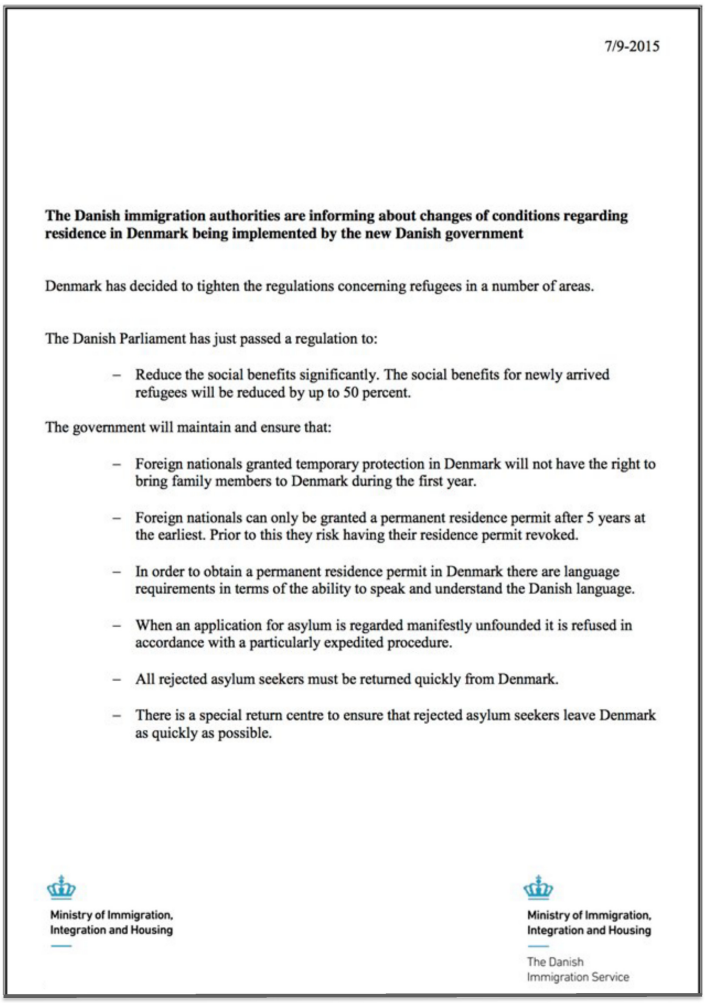
	Age ≥ 30		Age ≥ 24		Age ≥ 18	
	Syria	No Syria	Syria	No Syria	Syria	No Syria
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effects on Net Flow (Fraction of 2001 Stock)						
Effect of Benefits	0.0044***	0.0048***	0.0046***	0.0050***	0.0044***	0.0049***
(USD 100)	(0.0008)	(0.0008)	(0.0009)	(0.0009)	(0.0010)	(0.0010)
Elasticity of Migration	1.29	1.41	1.40	1.53	1.35	1.51
Flow w.r.t. Benefits	(0.23)	(0.24)	(0.27)	(0.28)	(0.30)	(0.30)
Observations	444	444	444	444	444	444
Panel B: Effects on Gross Inflow and its Composition (Fraction of 2001 Stock)						
Effect of Benefits (USD 100):						
Total	0.0040***	0.0040***	0.0042***	0.0042***	0.0045***	0.0047***
	(0.0008)	(0.0009)	(0.0012)	(0.0012)	(0.0013)	(0.0014)
Asylum	0.0029***	0.0028***	0.0031***	0.0032***	0.0032***	0.0033***
	(0.0007)	(0.0008)	(0.0011)	(0.0011)	(0.0012)	(0.0013)
Family	0.0027***	0.0027***	0.0034***	0.0033***	0.0038***	0.0038***
	(0.0007)	(0.0007)	(0.0011)	(0.0011)	(0.0012)	(0.0013)
Work & Study	0.0007	0.0007	0.0007	0.0007	0.0009	0.0009
	(0.0005)	(0.0005)	(0.0008)	(0.0008)	(0.0009)	(0.0009)
Observations	252	252	252	252	252	252

Notes: Panel A shows the effect of welfare benefits (in 100s of dollars) on immigration net flow as a fraction of the pre-scheme stock and the implied elasticity of immigration net flow with respect to benefits. Panel B shows the gross inflow of immigrants by type of residence permit as a fraction of the pre-scheme stock. The estimates of benefit effects correspond to $\hat{\beta}$ obtained from (1.1), while the estimates of elasticities correspond to ε obtained from (1.2). We estimate (1.1) by weighted least squares using the pre-scheme stock of each group as weights. The specification is run on de-trended outcome variables, i.e. the residual between the raw immigration outcome and a linear, group-specific trend estimated on pre-scheme data. The standard errors (in parentheses) are obtained by treating the pre-trend equation and equation (1.1) as a set of joint moment conditions and calculating the associated GMM standard errors. Across columns, we consider different sample restrictions by varying the age threshold and whether Syrian immigrants are included or not. Immigrants from Bosnia are dropped from the sample in all specifications.

Appendices

1.A Supplementary Figures and Tables

Figure 1.5: Ad Campaign in Lebanese Newspapers and Online, September 2015



7/9-2015

The Danish immigration authorities are informing about changes of conditions regarding residence in Denmark being implemented by the new Danish government

Denmark has decided to tighten the regulations concerning refugees in a number of areas.

The Danish Parliament has just passed a regulation to:

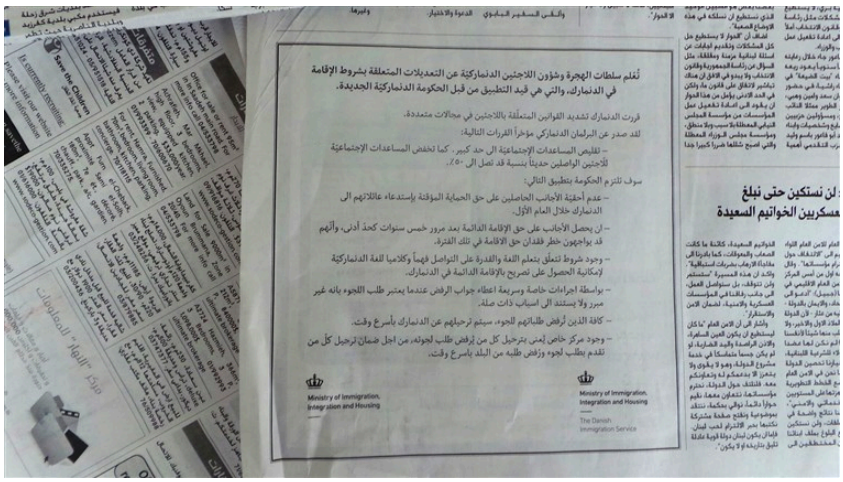
- Reduce the social benefits significantly. The social benefits for newly arrived refugees will be reduced by up to 50 percent.

The government will maintain and ensure that:

- Foreign nationals granted temporary protection in Denmark will not have the right to bring family members to Denmark during the first year.
- Foreign nationals can only be granted a permanent residence permit after 5 years at the earliest. Prior to this they risk having their residence permit revoked.
- In order to obtain a permanent residence permit in Denmark there are language requirements in terms of the ability to speak and understand the Danish language.
- When an application for asylum is regarded manifestly unfounded it is refused in accordance with a particularly expedited procedure.
- All rejected asylum seekers must be returned quickly from Denmark.
- There is a special return centre to ensure that rejected asylum seekers leave Denmark as quickly as possible.

Ministry of Immigration, Integration and Housing

Ministry of Immigration, Integration and Housing
The Danish Immigration Service



تُعلم سلطات الهجرة وشؤون اللاجئين الدنماركية عن التعديلات المتعلقة بشروط الإقامة في الدنمارك، والتي هي قيد التطبيق من قبل الحكومة الدنماركية الجديدة.

قررت الدنمارك تشديد القوانين المتعلّقة باللاجئين في مجالات متعددة.

لقد صدر عن البرلمان الدنماركي مؤخراً القرارات التالية:

- تقليص المساعدات الإجتماعية الى حد كبير. كما تخفض المساعدات الإجتماعية للأجئين الواصلين حديثاً بنسبة قد تصل الى -٥٠٪.

سوف تلتزم الحكومة بتطبيق التالي:

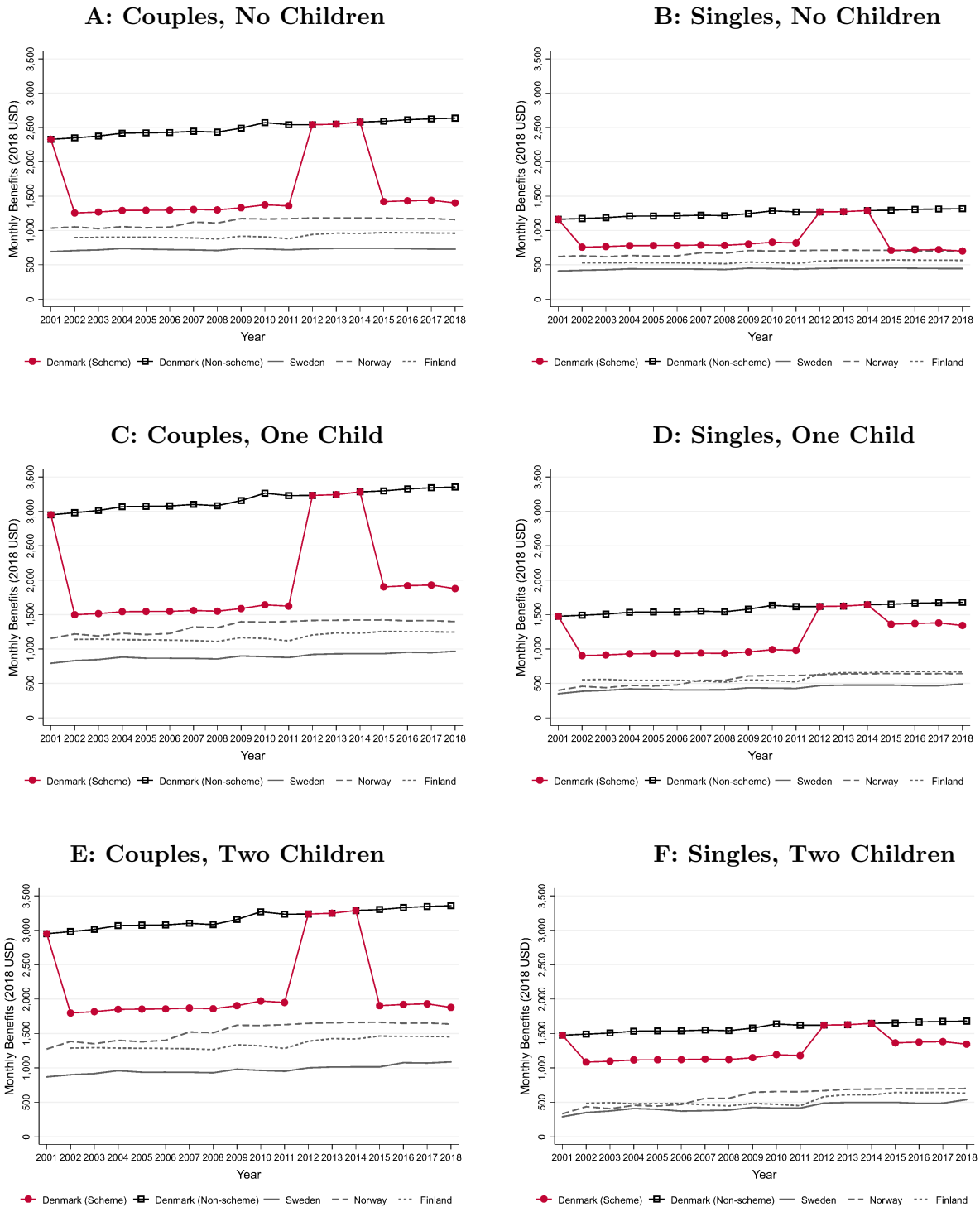
- عدم أحقيّة الأجانب الحاصلين على حق الحماية المؤقتة بإستدعاء عائلاتهم الى الدنمارك خلال العام الأوّل.
- أن يحصل الأجانب على حق الإقامة الدائمة بعد مرور خمس سنوات كحدّ أدنى، وأنهم قد يواجهون خطر فقدان حق الإقامة في تلك الفترة.
- وجود شروط تتعلق بتعلم اللغة والقدرة على التواصل فهماً وكلاميا للغة الدنماركية لإمكانية الحصول على تصريح بالإقامة الدائمة في الدنمارك.
- بواسطة اجراءات خاصة وسريعة اعطاء جواب الرفض عندما يعتبر طلب اللجوء بانه غير مبرر ولا يستند الى اسباب ذات صلة.
- كافة الذين لُرفض طلباتهم للجوء، سيتم ترحيلهم عن الدنمارك بأسرع وقت.
- وجود مركز خاص لمن يُرفض طلب لجوئه، من اجل ضمان ترحيل كلّ من تقدم بطلب لجوء ورُفض طلبه من البلد بأسرع وقت.

Ministry of Immigration, Integration and Housing

Ministry of Immigration, Integration and Housing
The Danish Immigration Service

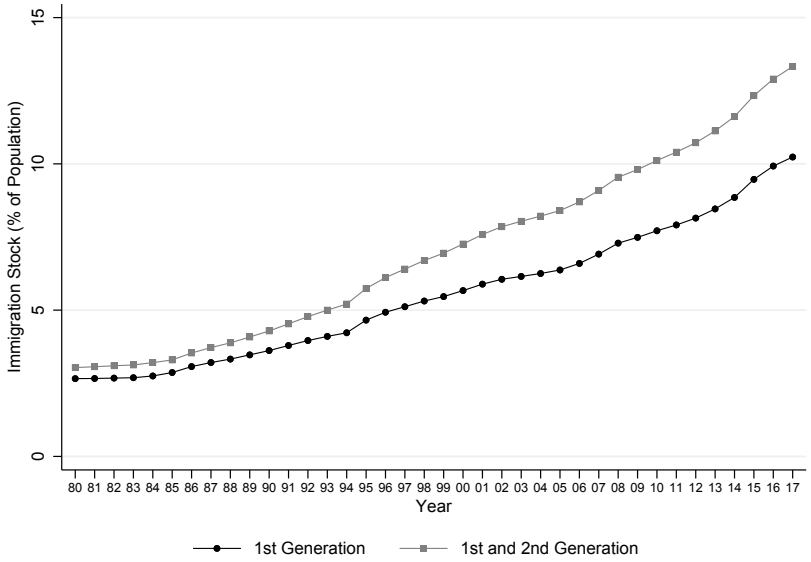
Notes: The two pictures show an ad campaign run by the Danish Government in September 2015 to inform potential immigrants about the recently (re)introduced welfare scheme, along with information about other regulatory rules. The left picture shows the campaign text in English released by the Danish Ministry of Immigration. The right picture shows the printed ad (from NBC News: https://www.nbcnews.com/storyline/europes-border-crisis/denmark-buys-ads-lebanon-newspapers-aimed-refugees-n423216). The ad campaign ran online and in Lebanese printed newspapers.

Figure 1.6: Cash Welfare in Denmark and the Other Nordic Countries 2001-2018



Notes: The figure shows maximum monthly cash benefits (in 2018 USD) for different household types in Denmark and the other Nordic countries. For Denmark, the figure shows both scheme and non-scheme benefits. Information on scheme benefits in Denmark are based on the website of STAR (The Danish Agency for Labour Market and Recruitment). The remaining benefits data are based on our own calculations using the OECD tax-benefit calculator. The calculations give benefits for non-employed households at age 40 who have been out of work for three months. For households with children, the age of the first child is set to 10 and the age of the second child is set to 8. The benefit series include only cash welfare (and not, e.g., unemployment insurance or in-kind benefits such as housing support).

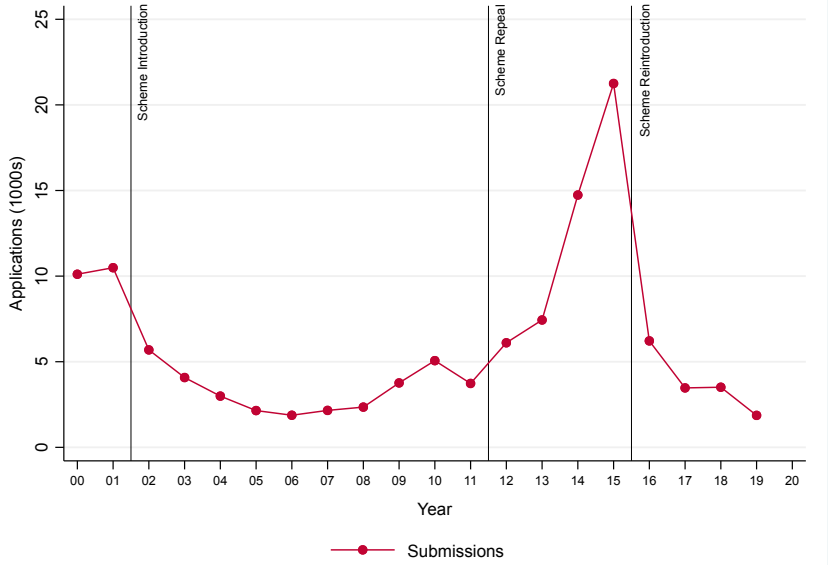
Figure 1.7: Immigrant Stock 1980-2017



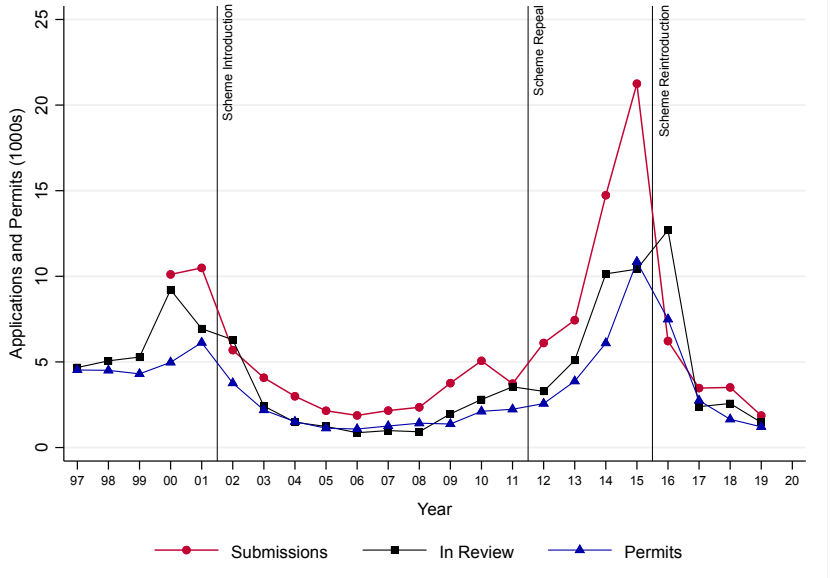
Notes: The figure shows the share of first-generation immigrants and the share of first- and second-generation immigrants in the Danish population. Definitions of first- and second-generation immigrants follow the official definitions of Statistics Denmark: a first-generation immigrant is a person who was born outside of Denmark and where neither of the parents are Danish citizens and born in Denmark. A second-generation immigrant is a person who was born in Denmark and where neither of the parents are Danish citizens and born in Denmark.

Figure 1.8: Asylum Applications from Outside EU: Annual Data

A: Application Submissions

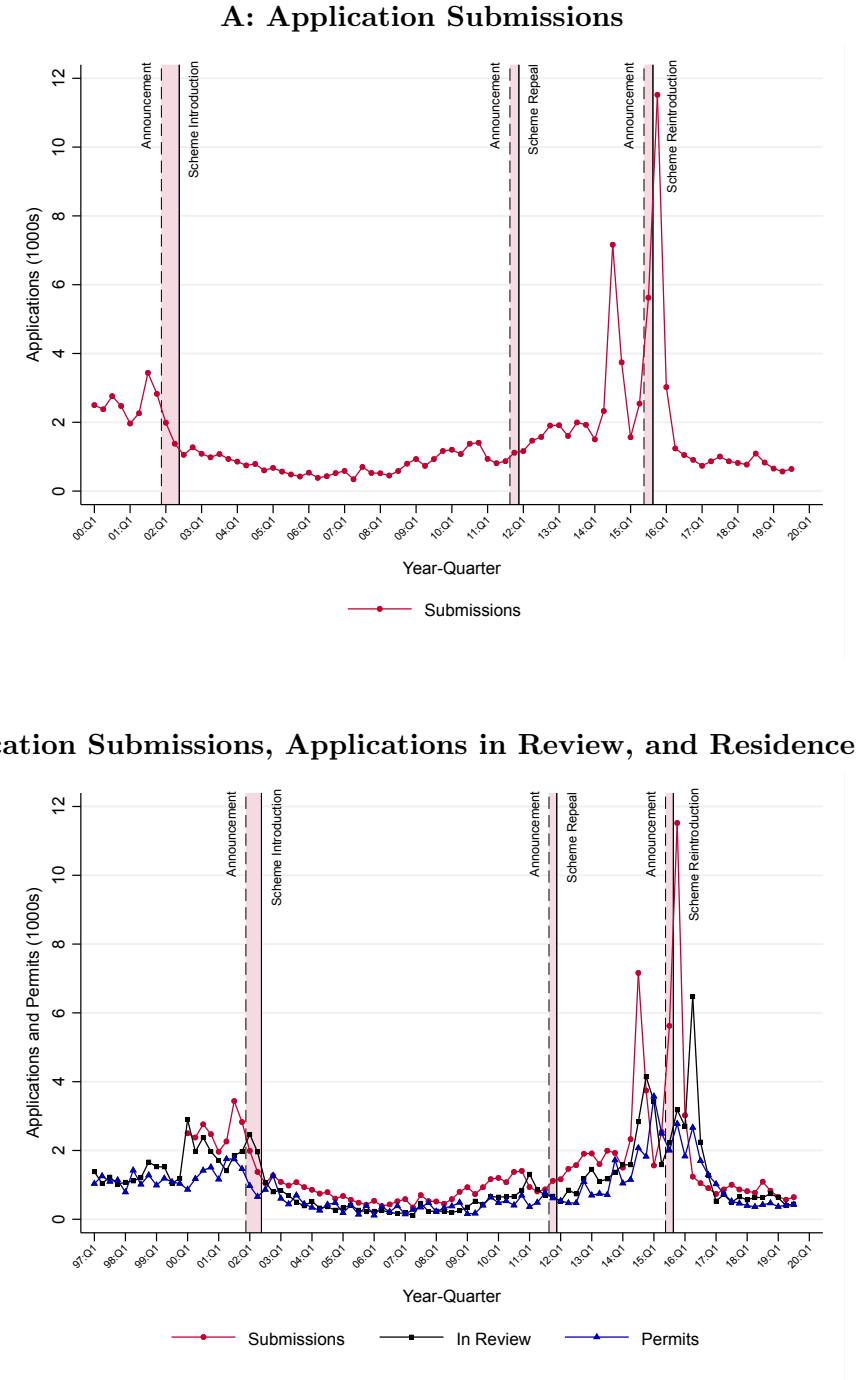


B: Application Submissions, Applications in Review, and Residence Permits



Notes: This figure shows asylum applications to Denmark at the annual level. Panel A shows applications submitted, while Panel B compares applications submitted (dots), applications registered for administrative review (squares), and residence permits granted (triangles). The solid vertical lines mark the timing of scheme reform implementations. Asylum seekers from Bosnia are dropped from the sample. The data are obtained from Statistics Denmark's online database (Statistikbanken, tables VAN5 and VAN77).

Figure 1.9: Asylum Applications from Outside EU: Quarterly Data



Notes: This figure shows asylum applications to Denmark at the quarterly level. Panel A shows applications submitted, while Panel B compares applications submitted (dots), applications registered for administrative review (squares), and residence permits granted (triangles). The solid vertical lines mark the timing of scheme reform implementations, and the dashed vertical lines mark the timing of scheme reform announcements. Asylum seekers from Bosnia are dropped from the sample. The data are obtained from Statistics Denmark's online database (Statistikbanken, tables VAN5 and VAN77).

Table 1.2: Main Changes to Immigration Regulations, 2000-2017

Date	Reform	Description	Affected group
May 2000	Affiliation requirement	Restricts marriage-based immigration to cases where the spouses' combined affiliation to Denmark is at least as strong as to any other country.	Family
June 2002	24-year rule	Marriage-based immigration restricted to couples where both spouses are at least 24 years old.	Family
June 2002	Tightening of affiliation requirement	Affiliation requirement tightened to require that spouses have a combined affiliation to Denmark stronger than to any other country.	Family
June 2002	Stricter criteria to obtain permanent residency	Required time in Denmark to apply for permanent residency extended from 3 to 7 years; tighter restrictions for convicted criminals.	Asylum and Family
June 2002	De Facto refugee concept replaced with protection status	A tightening of approval conditions for asylum seekers who do not meet critera set by international conventions.	Asylum
April 2007	Immigration test	Requirement to pass a test in Danish language and society to obtain family-based immigration.	Family
May 2010	Altered criteria to obtain permanent residency	Criteria for attainment of permanent residency changed to include requirements on employment, education etc., but required time in Denmark shortened to 4 years.	Asylum and Family
June 2011	Tightening of affiliation requirement	Affiliation requirement tightened again to require that spouses have a substantially greater affiliation to Denmark than to any other country.	Family
June 2011	Point system for marriage-based immigration	Marriage-based immigration now made conditional on obtaining a number of points based on criteria such as past employment, education and language skills.	Family

Date	Reform	Description	Affected group
May 2012	Rollback of 2011 reforms	Immigration reforms from 2011 (point system for marriage-based immigration and tightened affiliation requirement) rolled back.	Family
Feb. 2015	Introduction of temporary protection status	Certain asylum seekers fleeing conflicts or war will be granted temporary protection status, which requires periodic renewal until the conflict ends or permanent residency is attained.	Asylum
Jan. 2016	Longer waiting period for family-based migration for refugees on temporary protection status	Waiting period for refugees on temporary protection status to apply for visas for family members extended from 1 to 3 years.	Asylum and Family
Jan. 2016	Stricter criteria to obtain permanent residency	Required time in Denmark to obtain permanent residency extended from 4 to 6 years and subject to some stricter requirements on e.g. employment, language tests.	Asylum and Family
Sept. 2016	Suspension of participation in UNHCR resettlement program	Danish Government suspends participation in the UNHCR resettlement program, through which Denmark had received around 500 refugees annually until this point.	Asylum
May 2017	Longer waiting period to obtain permanent residency	Required time in Denmark to obtain residency further extended from 6 to 8 years.	Asylum and Family

Notes: This table summarizes the main changes to immigration regulation over the period 2000-2017. Many other adjustments to immigration regulation have not been included in the table, as they changed only minor details of the legislation.

1.B Asylum Applications

Our main specification uses residency as the measure of immigration. However, in the first instance, it is the decision to *apply* for Danish residency that should respond to benefit reform. Due to the lag between applying for residency and the final decision about residency, the timing of the DiD evidence in Figure 1.2 may be less compelling than it appears. The length of the lag varies by residency type and between individual cases, and in most cases we cannot observe it directly in the data. However, for asylum seekers specifically, we do observe the number and timing of applications (in addition to actual permits granted). This gives a measure of the desirability of Denmark as a destination country without any potential timing issues.

Graphical evidence on this outcome is shown in Figure 1.8 on an annual level. Panel A shows the total number of submitted asylum applications to Denmark since year 2000. We see a pattern very similar to that of the treated group in Figure 1.2 in our main analysis. The number of asylum applications declines markedly after the introduction of the welfare scheme, from around 10,000 to under 5000 within the space of a few years after the reform. It then increases again after the abolishment, and drops yet again after the reintroduction of the scheme. This pattern is thus fully consistent with what we see in our main analysis.

The series in Panel A is the best measure of asylum applications, but it is only available from 2000 onwards. Therefore, in Panel B of Figure 1.8, we add two alternative series which are available further back in time. The first of these series shows the number of applications under administrative review. This series excludes applications that were submitted, but rejected in an initial stage before undergoing a rigorous review. This series contains some lag relative to the one for total applications in Panel A. The second additional series shows the number of residence permits granted for asylum seekers. Because this occurs after the review stage, the potential lag for this series is greater still. Nevertheless, the three series track each other fairly closely.

One important point to notice from Panel B of Figure 1.8 is that the spike in asylum applications around 2015 is much larger than the spike in residence permits granted. This suggests that a sizeable part of the spike in asylum applications around this time came from individuals who did not fulfill the UNHCR's criteria for refugee status. This illustrates part of the reason why we use actual residencies rather than applications as our main outcome: Not every application leads to migration,

and while variation in applications are useful for validating the existence of behavioral responses, the economically important question relates to the impact of benefits on actual migration.

Figure 1.9 shows the same series as Figure 1.8, but on a quarterly level.[22] In this figure, we have included two vertical lines for each reform, the first indicating the announcement of the reform and the second indicating the actual implementation. Since the announcement and implementation fall within the same calendar year for each of the three reforms, this distinction was not necessary for the annual graphs. However, with the higher-frequency data considered here, it is useful to indicate both given behavior may plausibly start to react to the policy already at the point of announcement. The quarterly series are more noisy as one would expect, but the patterns remain consistent with immigration responses to the benefit reforms. We do notice that the large spike of asylum applications in 2015 occurs just after the reintroduction of the welfare scheme. This is not inconsistent with our interpretations, however, as a slight lag in the reaction to reforms is to be expected due to e.g. informational frictions.[23]

Finally, to relate and contrast these results to those obtained in our main analysis, it is worth highlighting the following points. First, the outcome variable considered here includes only one of the two types of migration treated by the welfare scheme; our main analysis also includes family-based migration. Second, the analysis here considers migration inflow, whereas our main analysis considers net flow. Third, we do not have a control group for our asylum applications outcome, so even though we observe sharp changes around the reforms, it is hard to obtain a quantitative measure of the magnitude of the welfare magnet effect from this data alone. The main purpose of this analysis is to show that the timing of changes in applications data is not too dissimilar from the timing of changes in residence data, suggesting that our estimates do capture immigration responses to welfare benefits rather than confounding effects of other immigration determinants.

[22] Unlike our main analysis where the outcome variable (net immigration flow) is observable only on an annual basis, asylum applications are observable on a quarterly basis.

[23] For instance, the Danish government's ad campaign in Lebanese newspapers (see Figure 1.5) ran in September 2015, just after the reintroduction of the welfare scheme. Its purpose was to inform potential migrants of the welfare cuts. However, while many Syrian refugees passed through Lebanon on their way to Europe, it is likely that many of those who applied for asylum in Denmark in the final quarter of 2015 would have already been further along in their journey at this point.

Chapter 2

The Tax Elasticity of Capital Gains and Revenue-Maximizing Rates[1]

The tax elasticity of capital gains realizations features centrally in U.S. fiscal policy debates. In the 1990s "capital gains tax wars," U.S. Treasury and economic officials argued that the responsiveness of realizations to capital gains tax rates was large enough that capital gains tax cuts would pay for themselves (Auten and Cordes, 1991). Others (e.g., Gravelle, 1991) asserted that the true responsiveness was much lower, so capital tax cuts would generate substantial fiscal cost. This issue has re-emerged in every presidential administration since 1990 and plays a key role in ongoing tax reform plans. For instance, this elasticity is the central parameter governing the revenue scores of Vice President Biden's plan to increase capital gains rates as well as President Trump's proposal reducing capital gains taxes.

Informing these policy debates is difficult because a wide range of estimates exist. Feldstein, Slemrod and Yitzhaki (1980), for example, estimate an elasticity with respect to the capital gains tax rate of -3.8, whereas the estimate of Burman and Randolph (1994) is -0.22. Moreover, there is limited empirical evidence in recent decades when there has been lower inflation, more widespread

[1]This chapter was co-authored with Owen Zidar and was published in *American Economic Review: Insights* in December of 2021 (Agersnap and Zidar, 2020). Materials from this paper have been presented at the NBER Business Taxation in a Federal System Meeting, 2020, and the 77th Annual Congress of the IIPF, 2021. We thank Coly Elhai and Stephanie Kestelman for providing excellent research assistance on this project. We also thank Alan Auerbach, Jerry Auten, Lily Batchelder, Tim Dowd, Patrick Driessen, Jane Gravelle, David Kamin, Paul Kindsgrab, Henrik Kleven, Ilyana Kuziemko, Robert McClelland, Jim Poterba, Natasha Sarin, Daniel Shaviro, Juan Carlos Suárez Serrato, Larry Summers, Danny Yagan, Eric Zwick, and anonymous referees as well as seminar and conference participants for helpful conversations and suggestions.

use of diversified investment vehicles, and a bigger role of pass-through firms, which have accounted for nearly half of capital gains realizations in recent years (Smith et al., 2019).

This paper estimates the effect of capital gains taxes on tax revenues and quantifies the implications for revenue-maximizing tax rates in the United States. We use a direct-projections approach and new state-level panel data on capital gains realizations and the migration of the wealthy to estimate the effects of state capital gains tax changes on realizations and location decisions. Our data, which range from 1980 to 2016, allow us to characterize responsiveness in a more recent period than most of the literature, and our direct-projections approach enables us to estimate effects over different time horizons and test for dynamic effects. We then build a simple framework to relate these state-level effects to a policy-relevant elasticity at the national level, which is the state-level realization response after removing migration effects and accounting for average state taxes and a minor aggregation adjustment term. We find that this policy-relevant elasticity of realizations with respect to capital gains tax rates over a ten year period is approximately -0.3 to -0.5 depending on the specification, and that the estimates are larger in absolute value in the short and medium run than in the long run.

We highlight three implications of these elasticity estimates. First, these estimates are well below an elasticity of one in absolute value, which indicates that capital gains tax cuts do not pay for themselves. We formally test and reject the null of an elasticity of -1.0. Second, these estimates suggest that raising capital gains tax rates by 5 percentage points (in the current regime with unlimited deferral and step-up-basis at death) would yield 18 to 30 billion in annual tax revenue, which is roughly twice the amount implied by the current approach of the Joint Committee on Taxation (JCT), which according to Gravelle (2020) currently uses an elasticity of -0.7 to score proposals.[2] Third, our long-run elasticity estimates correspond to point estimates for the revenue-maximizing capital gains tax rates of 38 to 47 percent.

Using state-level panel data provides more reforms and closer comparison groups than time-series analysis at the federal level. At the federal level, there are not only fewer reforms but also many confounding factors. Figure 2.1 plots realizations and the maximum long-run capital gains tax rate since 1980. It shows that some capital gains tax changes are associated with sizable movements

[2]Gravelle (2020) also notes that Treasury had used an estimate of -1.0 previously, but has since moved closer to the JCT's estimate. In addition, Dowd, McClelland and Muthitacharoen (2015), whose paper first appeared as a technical working paper (JCX-56-12) of joint work of the staff of the JCT and CBO, estimate an elasticity of -0.72.

in realizations, but the responses are quite unequal across reforms. It is also quite difficult to separate these movements from unrelated macroeconomic trends and asset price fluctuations. One approach is to difference out macroeconomic trends by comparing the realization series in one county with that of a similar country. In panels (b) and (c) of Figure 2.1, we implement this approach using the realization series of the U.K. around cuts to the U.S. capital gains tax rate in 1997 and 2003. These panels, however, show how precarious this approach is—it yields unstable elasticity estimates that exhibit large variance in non-tax-related country-year shocks and inherits the limitations of cross-country regressions.

Our state-level approach complements prior work by Bogart and Gentry (1995), who use state panel data from 1979 to 1990 to estimate the effect of capital gains tax rates on on state-level realizations per tax return, and ongoing work by Bakija and Gentry (2014), who use a similar approach for a longer panel from 1950-2007. Relative to this valuable work, our paper uses a different empirical approach and new data to provide new policy-relevant elasticity estimates based on a more recent sample that is closer to current conditions in terms of inflation, pass-through prevalence, and tax code. By having a better measure of location decisions of the wealthy, our approach also provides a more accurate accounting for migration effects, and thus policy-relevant realization effects, which difference out migration responses. Moreover, we use a direct-projections approach that contributes new graphical evidence on the dynamics of realizations around tax changes. Scoring capital gains tax changes requires estimating how realizations evolve over a ten-year period around the tax change, which, to the best of our knowledge, has not been done in prior work.

Much of the literature on capital gains in the 1990s and 2000s (e.g., Burman and Randolph, 1994; Auerbach and Siegel, 2000; Poterba, 2002; Auten and Joulfaian, 2004) has focused on the distinction between temporary and permanent effects, and used individual-level data to estimate these effects. The interpretation of these individual-level results, however, is complicated by factors such as strategic loss-harvesting, observations with zero realizations, and movement in and out of top income brackets depending on the timing of big realizations. While standard approaches for addressing these concerns exist (e.g., selection corrections (Heckman, 1979) to account for realization decisions and instruments for tax rates), aggregating within a state-year cell avoids these complexities and also provides a longer panel than many existing individual-level studies.

In addition, mapping estimates of micro-level responses, which are often person-weighted rather than dollar-weighted, into policy-relevant macro effects on tax revenues can be difficult.[3] Indeed, Joint Committee on Taxation (1990) highlighted similar concerns when evaluating the available literature to score reforms during the capital tax wars. Some recent promising work using bunching approaches (e.g., Dowd and McClelland, 2019; Buhlmann et al., 2020) also faces the challenge of mapping bunching responses into policy-relevant elasticities. Our state-level approach has the benefit of estimating aggregate responses, while also providing considerable variation over a long panel.

2.1 Data on Capital Gains Taxation and Realizations

Our primary outcome variable is realized capital gains by state and year from Smith, Zidar and Zwick (2020), which is the sum of short-term and long-term net realizations and is available from 1980 to 2016.[4] We inflate nominal data using CPI-U from the Bureau of Labor Statistics to measure realizations in 2018 USD. We also use data from 1980 to 2016 on the number of wealthy individuals by state and year from Smith, Zidar and Zwick (2020). Specifically, we focus on the number of individuals in the top 10% and top 1% of the national wealth distribution. Finally, some specifications use population and output data from the US Census Bureau and the Bureau of Economic Analysis.

We relate these state-level outcomes to the net-of-capital-gains-tax rate in state s in year t, which is $(1-\tau_{s,t})$, where $\tau_{s,t}$ is the maximum marginal federal and state tax rate. This variable comes from NBER TAXSIM and measures the combined effect of federal and state taxes, incorporating the deductibility of state and local taxes, the phase-out of deductions, and other state-year-specific features of the tax code. For instance, in terms of deductibility, $1-\tau_{st} \equiv 1-\tau_t^{fed}-(1-\tau_t^{fed})*\tau_{st}^{state}$.

The tax rate data are available from 1977 to 2017, which is a slightly longer range of years than the realizations and count data. We use these extra years when estimating longer-term effects. For example, although we cannot use a state tax reform in 1978 to identify the short-term impact on realizations, it can still contribute to the identification of longer-term effects.[5]

[3] For example, choosing the weights (especially for those with losses) and accounting for heterogeneous responses introduces difficulties when aggregating from micro to macro.

[4] The vast majority of realizations are long-term realizations (Figure 2.6).

[5] We generally use the terms "reform" and "tax change" interchangeably to indicate any non-zero value of

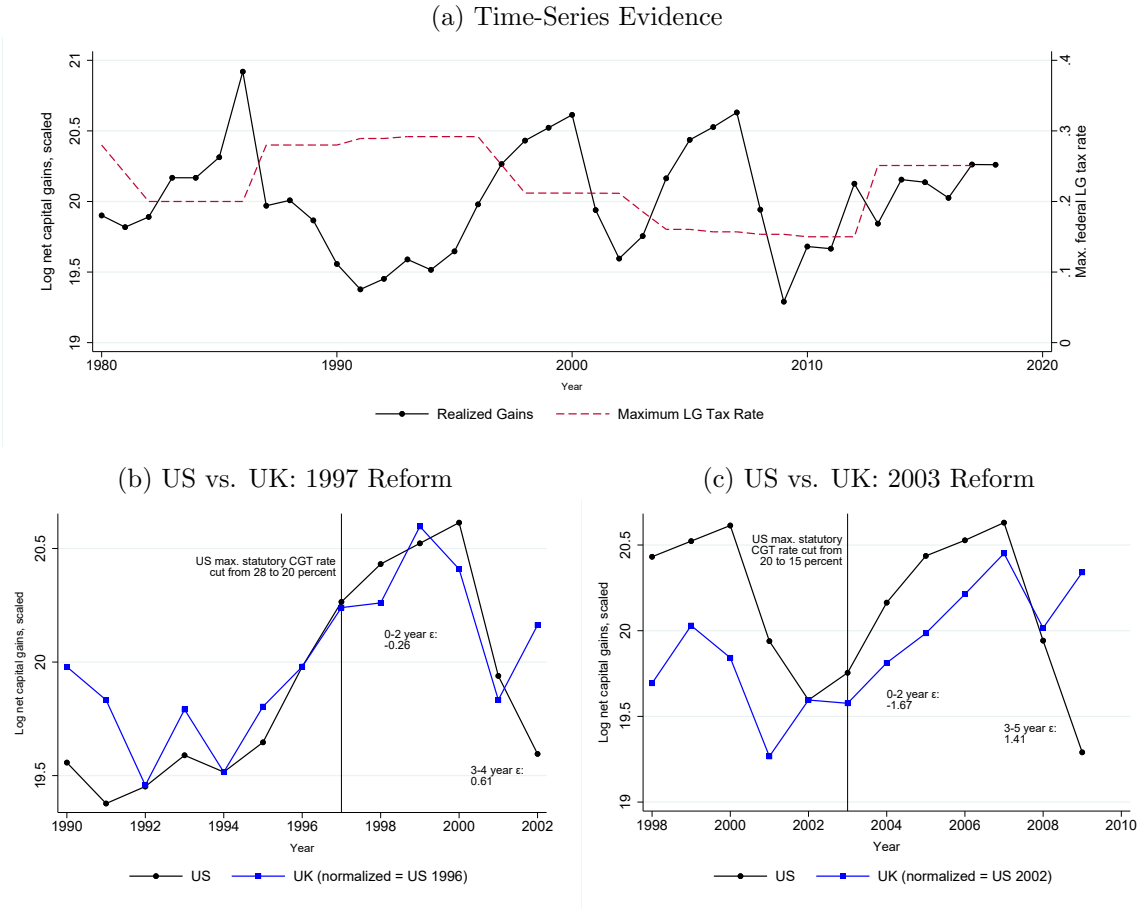


Figure 2.1: Capital Gains Realizations and Tax Rates in the United States

Notes: In all panels, net capital gains have been scaled by the ratio of a country's real GDP in 2000 to the country's real GDP in a given year (e.g. U.S. capital gains for 2005 were multiplied by the U.S. GDP in 2000 divided by the U.S. GDP in 2005). Panel (a) shows the evolution of realized capital gains and the maximum federal long-term capital gains tax rate in the US from 1980 to 2018. Panels (b) and (c) compare realized capital gains in the US to those in the UK around the time of US reforms. In all panels, net capital gains have been scaled by the ratio of a given country-year's GDP to the country's GDP in 2000. In both (b) and (c), the UK capital gains tax rate was constant throughout the period shown. The UK series has been normalized to equal the US series in the year prior to the reform. We calculate the short-term (0-2 years) and medium-term (3-4 or 3-5 years) tax elasticities provided in panels (b) and (c) by normalizing the UK series in the year before reform, calculating the average difference between the normalized UK series and the US series during the period, and dividing that difference by the difference in US log net-of-tax rates before and after the reform. This calculation gives us elasticities with respect to the net-of-tax rate, which we then multiply by $\frac{-0.22}{1-0.22}$ to convert them into elasticities with respect to the tax rate, at a tax rate of 22%. We use this tax rate to facilitate comparisons between these numbers and those given in Table 2.2, where we also use a rate of 22% to calculate tax elasticities given net-of-tax elasticities. If instead we had calculated the tax elasticities in panels (b) and (c) using the prevailing US federal tax rates prior to each reform, we would have obtained tax elasticities in panel (b) of -0.48 in the short run and 0.85 in the medium run, and in panel (c) of -1.64 in the short run and 1.28 in the medium run. We do not include the fifth year post-reform in the medium-term analysis of panel (b) because this fifth year coincides with the 2003 tax reform shown in panel (c).

Table 2.1 provides summary statistics. On average, state capital gains tax rates are 4 percent, but they range from 0 to 15 percent in our sample. Figure 2.1 plots the maximum federal tax rate over time. The current maximum capital gains tax rate is 23.8 percent. Combining both tax rates and accounting for interactions and phase-outs results in an effective keep rate of 75 percent on average, indicating that a dollar of realized capital gains amounts to 75 cents after taxes.

We find a total of 584 changes in state capital gains tax rates throughout our panel. Most of these changes are fairly small, which reflects the fact that our tax rate measure includes the effect of deductions and other minor provisions of state tax codes, so any changes to these provisions can cause the capital gains tax rate to change. The largest changes, however, are in excess of 4 percentage points (Table 2.3). In total, we have 128 state tax changes that exceed 1 percentage point in absolute value. In Figure 2.5, we provide a histogram of all changes. As we show below, our results are robust to using variation only from these larger tax changes.

We examine the relationship between changes in state capital gains tax rates and economic and policy conditions in Table 2.4 . Specifically, we regress indicators for capital gains tax increases and decreases on lags of state unemployment rates, GDP per capita, and state tax rates on personal and corporate income.[6] Most coefficients are insignificant and small, though notably, higher unemployment in the previous year is associated with a higher probability of increasing the capital gains tax rate. In our main analysis, we include specifications that condition on unemployment prior to tax reforms and do not find evidence that this relationship affects our estimates.

2.2 Methods

2.2.1 Estimating the Annual Effects of Capital Gains Tax Changes

We investigate the effects of log net-of-tax rates on log realized capital gains and on log counts of wealthy residents. We run direct projections regressions for different time horizons $h \in \{-10, -9, ..., 9, 10\}$:

$\Delta \log (1 - \tau_{s,t})$.

[6]Table 2.5 also shows that changes in state capital gains tax rates are often accompanied by changes in state personal income tax rates. We include specifications that do (as well as those that do not) control for leads and lags of changes in state tax rates on personal and corporate income.

Table 2.1: Summary Statistics on State Capital Gains Realizations and Tax Rates

	Observations	Mean	S.D.	Minimum	Maximum
Capital gains realizations					
Capital gains income ($ B)	1887	9.33	16.50	0.19	176.12
Log capital gains income	1887	15.25	1.22	12.14	18.99
Per capita capital gains income ($ 1000s)	1887	1.58	1.21	0.18	20.12
Tax variables					
State capital gains tax rate (τ_{st}^{state})	2091	0.04	0.03	0.00	0.15
Keep rate	2091	0.75	0.05	0.63	0.85
Log keep rate	2091	-0.29	0.07	-0.46	-0.16
Indicator for $\Delta\tau_{st}^{state} \neq 0$	2040	0.29	0.45	0.00	1.00
$\Delta\tau_{st}^{state}$ if $\Delta\tau_{st}^{state} \neq 0$	584	0.00	0.01	-0.08	0.09
Indicator for $\Delta\tau_{st}^{state} \geq 0.01$	2040	0.04	0.19	0.00	1.00
$\Delta\tau_{st}^{state}$ if $\Delta\tau_{st}^{state} \geq 0.01$	75	0.03	0.02	0.01	0.09
Indicator for $\Delta\tau_{st}^{state} \leq -0.01$	2040	0.03	0.16	0.00	1.00
$\Delta\tau_{st}^{state}$ if $\Delta\tau_{st}^{state} \leq -0.01$	53	-0.02	0.01	-0.08	-0.01

Notes: This table summarizes capital gains and tax variables used in our analysis, which are observed at the state-year level. We report counts and magnitudes for changes in state capital gains tax changes in three bins: non-zero changes, increases of one percentage point or more, and decreases of one percentage point or more. The sample ranges from 1980 to 2016 for realizations, and from 1977 to 2017 for tax rates. All dollar values are in 2018 dollars. We use capital gains income data from Smith, Zidar and Zwick (2020) and data on state population from the Census Bureau via FRED. Data on state and federal capital gains tax rates are from NBER TAXSIM. Keep rates ($1 - \tau_{st}$) are net of federal and state taxes on capital gains, accounting for deductibility and other provisions described in section 2.1. "S.D." refers to standard deviation. See Section 2.B for additional details.

$$y_{s,t+h} = \beta_h \Delta \log(1 - \tau_{s,t}) + \mathbf{X}'_{s,t}\mathbf{\Lambda}_h + \gamma_{s,h} + \phi_{t,h} + \varepsilon_{s,t,h} \tag{2.1}$$

where s and t index state and calendar year, $y_{s,t+h}$ is the outcome variable in year $t + h$ (log realized capital gains in our main specification), $\gamma_{s,h}$ and $\phi_{t,h}$ are horizon-specific state and year fixed effects, and $\mathbf{X}_{s,t}$ is a vector of controls. The main parameters of interest are the sequence of $\{\beta_h\}_{h=-10}^{10}$, which describe the path of realized capital gains around the tax change. The variable $\Delta \log(1 - \tau_{s,t})$ is the one-year change in the log net-of-tax rate. Using first differences helps deal with serial correlation concerns and facilitates estimating an impulse response function. To express the results as elasticities at different horizons, we normalize the coefficients to be relative to the coefficient in the year before the tax change, i.e., we define elasticities $\delta_h \equiv \beta_h - \beta_{-1}$. For example, $\delta_5 = \beta_5 - \beta_{-1}$ measures the elasticity of realized capital gains 5 years after the reform with respect to the capital gains net-of-tax rate, where the change in realizations is relative to the year before the tax event. We plot the δ_h estimates in our main figures.

In our baseline specification, we control for the vector $\mathbf{X}_{s,t}$ of leads and lags of changes in the log net-of-tax rate, i.e., $\mathbf{X}'_{s,t}\mathbf{\Lambda_h} = \sum_{r=-10,r\neq 0}^{10} \lambda_r^h \Delta \log(1-\tau_{s,t+r})$. Controlling for these other leads and lags of capital gains tax changes isolates the effect of a given tax reform. Without these controls, estimates would reflect the effect of not just the tax reform of interest, but also any other reforms occurring within the event window.[7] To check robustness, we also run specifications featuring additional controls in $\mathbf{X}_{s,t}$, including GDP in pre-reform year $t-1$, GDP-growth-bin-by-year dummies, state unemployment in pre-reform year $t-1$, and changes in state corporate and personal income taxes. Finally, we include a specification that interacts the tax change with indicators based on the size of the tax change, and report estimates for β_h coming only from larger tax reforms that exceed 1 percentage point in absolute value.

2.2.2 Discussion of Alternative Specifications and Semi-Elasticities.

To facilitate comparisons to prior estimates, we discuss the theoretical and empirical implications of using logs and semi-logs in Section 2.C. We also provide estimates using a semi-log specification, which delivers similar results. We prefer our net-of-tax formulation because it measures the relevant price governing behavior and is standard in the broader literature (Saez, Slemrod and Giertz, 2012*a*).

Compared with an event-study specification, which centers around the outcome year rather than the policy reform year, the direct projections approach accurately recovers elasticities in simulations (see Section 2.D for details).[8] In the appendix we also provide results using event-study specifications (Figure 2.11, Figure 2.16 and Table 2.7). Results are similar.

2.2.3 Estimating effects over multiple years

We extend the method in equation (2.1) to estimate the elasticity of capital gains realizations over longer time horizons. First, we consider a direct projections specification that estimates the effect of tax reforms on realizations in three-year bins, yielding estimates of the elasticity in the short

[7] In Appendix Figure 2.12, we run a specification that excludes the vector of controls for other reforms and find similar results.

[8] Since the direct-projections approach centers leads and lags on the policy reform year, it includes fewer pre-observation controls when estimating the effect of post-observation reforms or vice versa. However, it also facilitates controlling for pre-treatment conditions and handling locations with multiple events (and associated issues with adjusting standard errors).

(0-2 years), medium (3-5 years) and longer run (6-8 years):

$$y_{s,t+h} = \tilde{\beta}_h \Delta_3 \log(1-\tau_{s,t}) + \mathbf{X}'_{s,t}\tilde{\mathbf{\Lambda}}_h + \tilde{\gamma}_{s,h} + \tilde{\phi}_{t,h} + \tilde{\varepsilon}_{s,t,h}, \tag{2.2}$$

where $\Delta_3 \log(1-\tau_{s,t})$ represents the three-year change in the log net-of-tax rate (i.e., $\Delta_3 \log(1-\tau_{s,t}) = \log(1-\tau_{s,t}) - \log(1-\tau_{s,t-3})$). For each value of $h \in -9, -6, \ldots, 6, 9$, we estimate a separate instance of this regression. The controls for other reforms in this regression are also specified in 3-year bins: the vector of controls $\mathbf{X}_{s,t}$ now contains the variables $\Delta_3 \log(1-\tau_{s,t+r})$ for $r = -9, -6, -3, 3, 6, 9$. We use the notation $\tilde{\beta}$, $\tilde{\Lambda}$, etc, to distinguish the parameters in equation (2.2) from their analogues in equation (2.1).

To understand why this specification correctly identifies the average elasticity over the specified three-year periods, consider a simple example. Suppose a state changes its capital gains tax rate exactly once in year 2000. Then $\Delta_3 \log(1-\tau_{s,t})$ takes a value of zero for this state in every year except three: $t = 2000, 2001, 2002$. Now consider the regression above for $h = 0$. The variable $\Delta_3 \log(1-\tau_{s,t})$ is non-zero whenever the left-hand side variable is $y_{s,2000}$, $y_{s,2001}$ or $y_{s,2002}$. Therefore, the coefficient $\tilde{\beta}_0$ will capture the average effect of the tax reform on capital gains in these three years. Suppose instead $h = -3$. In this case, our left-hand side variable of $y_{s,t+h}$ becomes $y_{s,t-3}$. Since $\Delta_3 \log(1-\tau_{s,t})$ is still zero for all t except 2000, 2001, and 2002, $\tilde{\beta}_{-3}$ captures the effect of the tax reform on $y_{s,1997}$, $y_{s,1998}$ and $y_{s,1999}$ (the only $y_{s,t-3}$ such that $t \in 2000, 2001, 2002$). Now, define $\tilde{\delta}_0 \equiv \tilde{\beta}_0 - \tilde{\beta}_{-3}$. The parameter $\tilde{\delta}_0$ measures the difference in realizations in the periods immediately after and before the reform. In our example, $\tilde{\delta}_0$ represents the difference between average realizations in post-reform years 2000-2002 and average realizations in pre-reform years 1997-1999. In other words, $\tilde{\delta}_0$ identifies the average elasticity over a 0-2 year horizon relative to the reform year. Similarly, $\tilde{\delta}_3 \equiv \tilde{\beta}_3 - \tilde{\beta}_{-3}$ would identify the impact of the reform on the difference between average realizations in 2003-2005 and 1997-1999, thus giving us an average elasticity over a 3-5 year horizon, and so on.

We use a similar approach to estimate effects in the post period (i.e., in years 0-10) and in the long-run (i.e., years 6-10). To get a single estimate of the effect of capital gains tax reforms on

realizations in the decade following the reform, we use the following specification:

$$y_{s,t} = \beta_\ell \Delta_{11} \log(1 - \tau_{s,t}) + \mathbf{X}'_{s,t} \acute{\mathbf{\Lambda}}_h + \acute{\gamma}_{s,h} + \acute{\phi}_{t,h} + \acute{\varepsilon}_{s,t,h}, \tag{2.3}$$

where $\Delta_{11} \log(1 - \tau_{s,t}) = \log(1 - \tau_{s,t}) - \log(1 - \tau_{s,t-11})$.[9] To estimate the elasticity for the 0-10-year estimate, we take the point estimate β_ℓ from this regression and subtract off the point estimate $\tilde{\beta}_{-3}$ from (2.2). Finally, the long-run estimate for years 6-10 is implemented similarly to equation (2.3), but using a 5-year rather than 11-year difference for the right-hand side tax change variable (i.e., replacing Δ_{11} with Δ_5), and using $y_{s,t+6}$ as the outcome variable.

2.3 Capital Gains Tax Changes and Realizations at the State-Level

Figure 2.2 shows the results of our baseline specification from equation (2.1). The figure illustrates how capital gains realizations evolve before and after a change in the capital gains net-of-tax rate, controlling for other state capital tax reforms. We see no clear pre-trend: in each of the ten pre-reform years, capital gains realizations tend to be stable. We then see a jump soon after the reform, after which the point estimates decline modestly throughout the post-period. Unlike the 1986 and 2012 national reforms shown in Figure 2.1, we do not see evidence of anticipation effects preceding state capital gains tax reforms, which would have manifested as a downward spike at year -1.

On average across post-reform years 0 through 10, the point estimates directly provide an elasticity estimate of capital gains realizations with respect to the net-of-tax rate of around 3.18. The dynamics of this response are also of interest. In Figure 2.2, there is a modest downward trend over the post-period. Combining some of our individual year point estimates, the estimated short-run elasticity $\frac{1}{3}\sum_{h=1}^{3} \hat{\delta}_h$ is 3.61 (s.e. 1.22), whereas the longer run estimate $\frac{1}{3}\sum_{h=8}^{10} \hat{\delta}_h$ is somewhat lower at 2.59 (s.e. 1.42). However, we cannot reject the null that these effects are the same.

In Panel (b) of Figure 2.2, we present five robustness tests of these results: (1) controlling for pre-event state GDP, (2) controlling for pre-event state GDP growth, (3) controlling for pre-event

[9] In this specification, the vector of controls $\mathbf{X}_{s,t}$ contains variables that are still 3-year binned versions before and after the long bin: $\Delta_3 \log(1 - \tau_{s,t+r})$ for $r = -17, -14, -11, 3, 6, 9$.

state unemployment rates, (4) only using large reforms that change the capital gains tax rate by at least 1 percentage point, and (5) controlling for changes in state income and corporate tax rates. The results are remarkably similar. We also provide a range of other robustness checks in Figures 2.11-2.15, including an event study version of the analysis, a specification without controls for other capital gains tax changes in the pre- and post-reform periods, and separate analyses for small and large states.

State capital gains tax rates often move in the same direction as state income tax rates (Table 2.5). Many states treat capital gains as regular income for tax purposes, in which case the capital gains tax rate will be identical to the income tax rate. We account for these possibilities using a specification that controls for changes in personal income and corporate taxes. Panel (b) of Figure 2.2 illustrates that the elasticity estimates from this specification are very similar to the baseline.

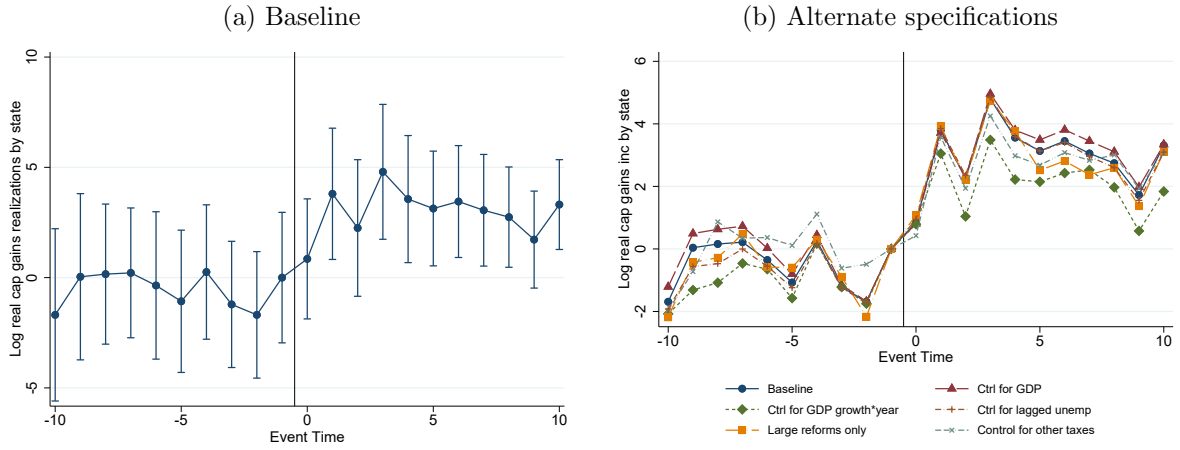
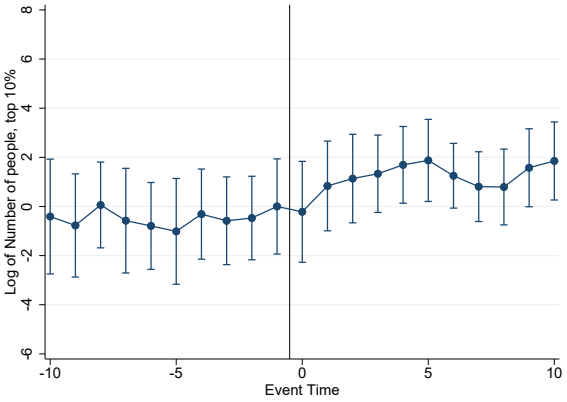


Figure 2.2: The Effect of Net-of-Tax Rate Changes on Capital Gains Realizations

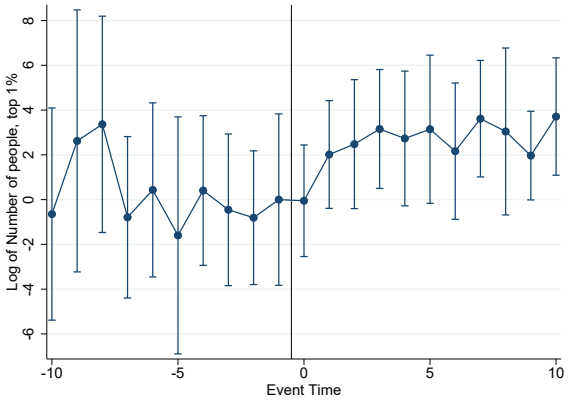
Notes: This figure presents our main results of the impact of state capital tax rate changes on capital gains realizations. The points plotted are the estimated coefficients from equation (2.1) for the impact of a change in the total (federal and state) log net-of-tax capital gains tax rate on log capital gains realizations. The alternate specifications include the following: (1) controlling linearly for GDP in the year before the reform; (2) controlling for a set of dummies interacting prior GDP growth tertiles and years, where the prior GDP growth tertile is determined using GDP growth over the most recent three years; (3) controlling for prior unemployment rates; (4) interacting the tax change variable with indicators for reforms greater or smaller than 1 percentage point, and reporting the coefficients corresponding to large changes, and (5) controlling for other state personal and corporate income tax changes. Specifically, for (4) we modify the baseline specification by estimating separate coefficients for large and small tax changes, i.e., we fit $y_{s,t+h} = \beta_h^{big} \times \mathbf{1}\left(\left|\Delta\left(1-\tau_{s,t}^{state}\right)\right| \geq 0.01\right) \times \Delta \log\left(1-\tau_{s,t}\right) + \beta_h^{small} \times \mathbf{1}\left(\left|\Delta\left(1-\tau_{s,t}^{state}\right)\right| < 0.01\right) \times \Delta \log\left(1-\tau_{s,t}\right) + \mathbf{X}_{s,t}'\mathbf{\Lambda}_h + \gamma_{s,h} + \phi_{t,h} + \varepsilon_{s,t,h}$, where β_h^{big} and β_h^{small} are the tax-change-size-specific coefficients. We report the series based on the β_h^{big} coefficients. In all series, capital gains are in real terms, and the estimated coefficients are normalized to equal 0 at time -1, i.e., we plot $\hat{\delta}_h$ as described in section 2.2.1. Standard errors are clustered at the state level.

2.4 Federal Revenue-Maximizing Rates

(a) Residents in top 10% of national wealth distribution



(b) Residents in top 1% of national wealth distribution



(c) Binned estimates of empirical and policy-relevant elasticities

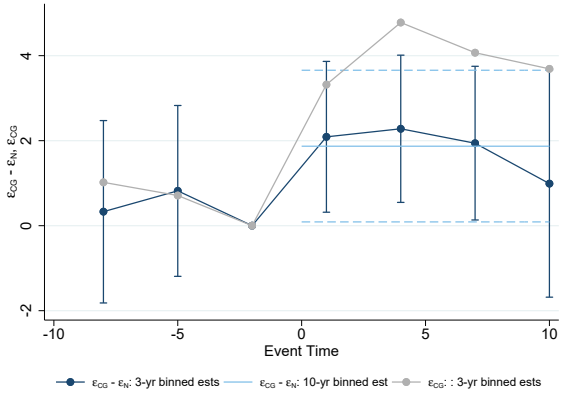


Figure 2.3: Migration Effects of State Capital Gains Taxes and Policy-Relevant Elasticity Estimates

Notes: Panels (a) and (b) of this figure show the impact of state capital gains tax reforms on migration of high-net worth residents, using the specification in equation (2.1). The outcome in panel (a) is the log of the share of residents belonging to the top 10 percent of the national wealth distribution, and in panel (b) it is the log share in the top 1 percent. The points plotted are the estimated coefficients for the impact of a one-period change in the total (federal and state) log net-of-tax capital gains tax rate on these outcome variables. In panel (c), we estimate the empirical capital gains elasticity using a specification analogous to that in equation (2.1), but using three-year bins (see section 2.2.3 for details). This estimate is represented by the grey point estimates. The dark blue point estimates represent our estimates of the policy-relevant elasticity $\hat{\varepsilon}^R = \hat{\varepsilon}^{CG} - \hat{\varepsilon}^N$ for each period. As described in section 2.2.3, $\hat{\varepsilon}^N$ is estimated using a state-year specific dollar-weighted average of the migration responses of the top 1% and top 10%. The horizontal light-blue line is our estimate of the policy-relevant elasticity using a binned specification for post-reform years 0-10 (see equation 2.3). All specifications include state and year fixed effects and controls for tax reforms in years surrounding the reform. The estimated coefficients are normalized to equal 0 at time -1, and standard errors are clustered at the state level.

In this section, we present a framework that shows how to use the state-level estimates from above to infer the policy-relevant elasticity at the national level. We then derive an estimate for the revenue-maximizing tax rate and calculate revenue effects of a hypothetical five percentage point tax rate increase.

2.4.1 A Simple Model of State-Level Capital Gains Realizations

Consider a country with states $s \in S$. Residents of state s retain a share $(1 - \tau_F - \tau_s)$ of their realized capital gains after paying federal (τ_F) and state capital gains taxes (τ_s).[10]

We can decompose total realized capital gains in state s into two terms:

$$CG_s = N_s\left(1 - \tau_F - \tau_s, \tau_{-s}\right) R_s\left(1 - \tau_F - \tau_s\right), \tag{2.4}$$

where $N_s\left(1 - \tau_F - \tau_s, \tau_{-s}\right)$ represents the number of residents in state s, and $R_s\left(1 - \tau_F - \tau_s\right)$ represents realized capital gains per resident *conditional* on residing in state s.

Using equation (2.4), we can decompose the empirical elasticity of capital gains with respect to the net-of-tax rate:

$$\varepsilon^{CG} = \frac{\partial \log CG_s}{\partial \log\left(1 - \tau_F - \tau_s\right)} = \frac{\partial \log N_s}{\partial \log\left(1 - \tau_F - \tau_s\right)} + \frac{d \log R_s}{d \log\left(1 - \tau_F - \tau_s\right)} = \varepsilon^N + \varepsilon^R. \tag{2.5}$$

Thus, ε_s^{CG} is the sum of two elasticities: a migration elasticity ε^N and a realization elasticity ε^R, which is the main object of interest and represents the "pure" per capita response of capital gains realizations to the net-of-tax rate.

2.4.2 From State-Level Realizations to the Federal Revenue-Maximizing Tax Rate

We show in Appendix Section 2.E that the federal capital gains tax rate that maximizes federal tax revenue from capital gains is:

$$\tau_F^* = \frac{1 - \bar{\tau}_S}{1 + \varepsilon^R}. \tag{2.6}$$

[10] When measuring keep rates net of federal and state taxes, we account for deductibility as described in section 2.1.

This formula resembles familiar optimal tax models (Saez, 2001; Diamond and Saez, 2011), but has an additional aggregation adjustment term $\bar{\tau}_S$ that denotes the average population-weighted state tax rate. The policy-relevant elasticity at the federal level is our estimate of the elasticity of capital gains realizations at the state level less the migration elasticity, i.e., $\varepsilon^R = \varepsilon^{CG} - \varepsilon^N$.

2.4.3 Estimating the Federal Revenue-Maximizing Tax Rate

We estimate ε^N based on the responses of those at the top of the national wealth distribution because the top groups account for essentially all capital gains (Smith, Zidar and Zwick, 2020). To do so, we fit equation (2.1) when the outcome is the share of the state population that belongs to either the top 10% or the top 1% of the national wealth distribution. Panels (a) and (b) of Figure 2.3 show the outcome of this specification for the top 10% and top 1% groups, respectively. We see fairly stable pre-trends, though somewhat noisy in the earliest pre-periods for the top 1% group. Following an increase in the net-of-tax rate, the share of residents in each of the top groups grows, and the magnitude of the response is larger for the top 1% group.

Connecting these migration responses to the theory requires weighing the following tradeoff. Focusing on the top 10% delivers a more precise estimate, but risks putting too little weight on the responsiveness of those at the top who have substantial capital gains realizations in dollar terms. Our preferred measure of migration is a combination of the two groups. Specifically, we define the outcome variable as: $\theta_s \ln N_{s,t}^{P99-P100} + (1-\theta_s) \ln N_{s,t}^{P90-P100}$, which is a dollar-weighted convex combination of the two groups, where θ_s is the state's wealth share of the top 1% relative to the top 10%.[11]

To estimate ε^R, we first estimate ε^{CG} and ε^N using our binned specifications described in section 2.2.3, with log capital gains and our migration measure from above as outcome variables. We perform this estimation at various time horizons: the main estimate covering years 0-10 after each reform, as well as three-year bins of 0-2 years, 3-5 years, 6-8 years, and 6-10 years. We then calculate $\hat{\varepsilon}_h^R = \hat{\varepsilon}_h^{CG} - \hat{\varepsilon}_h^N$ at each time horizon.

[11] θ_s is defined as

$$\theta_s = \sum_t \frac{W_{s,t}^{P99-P100}}{W_{s,t}^{P90-P100}},$$

that is, it represents the average wealth share in state s of the top 1% within the top 10% across all years in our sample. Across all states in the sample, the mean of θ_s is 0.39. The minimum value is 0.28 (Iowa), while the maximum is 0.56 (District of Columbia), with Wyoming, Nevada and New York also having values above 0.5.

Panel (c) of Figure 2.3 shows the estimates from these binned regressions graphically. The grey series shows estimates of $\hat{\varepsilon}_h^{CG}$ at different horizons h, the dark blue series shows $\hat{\varepsilon}_h^R$, and the light blue shows the average policy-relevant elasticity over a decade following each reform, $\hat{\varepsilon}_{0-10}^R$.

Table 2.2 also provides estimates of our elasticities at various time horizons. We also translate our net-of-tax elasticity estimates into an elasticity with respect to the tax rate to facilitate comparisons to prior work. Table 2.2 includes estimates from two alternative specifications: one in which we identify all elasticities from state tax reforms of at least 1 percentage point, and a second in which we control for all reforms to state income tax and corporate tax rates in a 21-year window around the capital gains tax reform in question.[12]

Overall, our baseline estimate at a 0-10 year horizon gives an elasticity with respect to the net-of-tax-rate of approximately 1.87, which translates into an elasticity of -0.53 with respect to the tax rate. The specification using only large tax changes yields an elasticity with respect to the tax rate of -0.42 and the specification controlling for other state tax changes yields an elasticity with respect to the tax rate of -0.29. While these point estimates are somewhat noisy, the point estimates are notably smaller in absolute value than the conventional elasticities used by tax analysts. Moreover, we can definitively test and reject the null hypothesis that this elasticity with respect to the tax rate equals 1 in absolute value. The p-values for the baseline, big-change-only, and other-tax-control specifications are 0.066, 0.021, and 0.31, respectively. This finding provides important evidence suggesting that capital gains tax cuts do not pay for themselves, which has been a prominent proposition in this literature (Joint Committee on Taxation, 1990; Auten and Cordes, 1991; Gravelle, 1991; Burman, 2010; Gravelle, 2020).

As in Figure 2.2, the results in Table 2.2 across all specifications show an elasticity which is somewhat bigger in absolute value in the short to medium run, and then declines in the longer run. This pattern could potentially reflect some transitory effects that are present in the short run but disappear over time. For instance, a tax cut may in the shorter run spur some individuals to realize accumulated gains to take advantage of the lower tax rate, particularly if they expect a possible reversal of the tax cut in the future. However, once many of these accumulated gains have been realized, the potential for realization of further gains is limited, and the magnitude of the

[12] Appendix Tables 2.8 and 2.9 provide additional specifications that control for pre-reform unemployment and unemployment growth. These are very similar to the results in Table 2.2.

Table 2.2: Capital Gains Elasticities and Revenue-Maximizing Tax Rates

Specification	Total elast. ε^{CG}	Policy elast. $\varepsilon^R = \varepsilon^{CG} - \varepsilon^N$	Laffer rate $\tau^* = \frac{1-\bar{\tau}_S}{1+\varepsilon^R}$	Elasticity w.r.t. tax $\varepsilon^{tax} = \varepsilon^R \cdot \frac{-0.22}{1-0.22}$	χ^2 test $\varepsilon^{tax} = -1$
Baseline					
0-10 years	3.39	1.87	0.33	-0.53	3.38
	(1.01)	(0.91)	(0.10)	(0.26)	(0.07)
0-2 years	3.32	2.09	0.30	-0.59	2.58
	(0.97)	(0.91)	(0.09)	(0.26)	(0.11)
3-5 years	4.78	2.28	0.29	-0.64	2.05
	(1.10)	(0.88)	(0.08)	(0.25)	(0.15)
6-8 years	4.07	1.94	0.32	-0.55	3.02
	(1.20)	(0.92)	(0.10)	(0.26)	(0.08)
6-10 years	3.66	1.47	0.38	-0.41	4.54
	(1.27)	(0.97)	(0.15)	(0.27)	(0.03)
Big changes only					
0-10 years	2.81	1.48	0.38	-0.42	5.31
	(1.02)	(0.89)	(0.14)	(0.25)	(0.02)
0-2 years	3.54	2.50	0.27	-0.71	1.14
	(0.97)	(0.98)	(0.07)	(0.28)	(0.29)
3-5 years	4.96	2.40	0.28	-0.68	1.54
	(1.10)	(0.92)	(0.07)	(0.26)	(0.21)
6-8 years	3.77	1.65	0.35	-0.46	4.35
	(1.19)	(0.91)	(0.12)	(0.26)	(0.04)
6-10 years	2.80	0.99	0.47	-0.28	6.48
	(1.30)	(1.00)	(0.24)	(0.28)	(0.01)
Control for other taxes					
0-10 years	2.28	1.01	0.47	-0.29	4.64
	(1.32)	(1.18)	(0.27)	(0.33)	(0.03)
0-2 years	2.38	1.25	0.42	-0.35	3.88
	(1.19)	(1.16)	(0.21)	(0.33)	(0.05)
3-5 years	3.58	1.64	0.36	-0.46	3.54
	(1.24)	(1.01)	(0.14)	(0.29)	(0.06)
6-8 years	3.32	1.40	0.39	-0.39	3.57
	(1.43)	(1.14)	(0.19)	(0.32)	(0.06)
6-10 years	2.98	1.18	0.43	-0.33	3.96
	(1.54)	(1.19)	(0.24)	(0.34)	(0.05)

Notes: This table shows estimates of main elasticities and associated revenue-maximizing rates for different specifications of equation (2.2). The 0-2, 3-5 and 6-8 year specifications use 3-year bins, the 6-10 year specification uses a 5-year bin, and the 0-10 year specification uses an 11-year bin as described by equation (2.3). The empirical elasticities $\hat{\varepsilon}^{CG}$ and $\hat{\varepsilon}^N$ are calculated as the difference between the point estimate for each specific horizon, and the point estimate for the $[-3,-1]$-year bin. Revenue-maximizing tax rates are estimated using the formula in equation (2.6), which is derived in Section 2.E. The term $\bar{\tau}_S$ adjusts for the average population-weighted state tax rate in 2016 (6.27%). All specifications control for reforms in the capital gains tax rate in three 3-year bins before and after the reform in question, and include state and year fixed effects. "Big changes only" replaces the right hand side variable of interest— the change in the log net-of-tax rate—with two separate variables, which sum changes in the log net-of-tax rate in years where the state tax rate changed by more or less than 1 percentage point, respectively. Specifically, we replace the right-hand side variable of interest – for instance, $\Delta_3 \log(1-\tau_{s,t})$ – with two variables: $\Delta_3^{big} \log(1-\tau_{s,t}) \equiv \sum_{k=0}^{2} \Delta \log(1-\tau_{s,t-k}) \times \mathbf{1}(|\Delta(1-\tau_{s,t-k})| \geq 0.01)$, which sums all the tax changes greater than 1 percentage point, and a corresponding variable that sums all smaller tax changes smaller than 1. We then report estimates for the variable that sums the big changes. The "Control for other taxes" specification includes controls for changes in state income and corporate tax rates. We control for the change in other tax taxes over the same 3-year (or 5-year or 11-year) bins that we use to identify our main point estimate in each regression, as well as three 3-year bins before and after the main period. Standard errors are clustered at the state level. Values in parentheses in columns 1-4 represent standard errors; values in parentheses in column 5 (χ^2 test) represent p-values.

behavioral effect declines in the longer run.[13] For this reason, we view our 6-10 year estimates as being the most likely to capture the long-run structural elasticity of capital gains to the net-of-tax rate, though some transitory effects might exist even beyond a 10-year horizon.

2.4.4 Policy Implications

These elasticity results have policy implications for revenue-maximizing tax rates and the revenue effects of capital gains tax reforms. Table 2.2 provides the corresponding estimates of the revenue-maximizing tax rate, $\hat{\tau}_F^*$. Our baseline elasticity estimate for the full 10-year period implies a revenue-maximizing rate around 33%, and the analogous estimates for the big-changes-only and control-for-other-tax-changes specifications are 38% and 47%, respectively. Thus, over a ten year budget window, these results suggest that a capital gains tax rate of around 40% would maximize federal capital gains tax revenues. For analysis over years 6-10, we find somewhat higher revenue maximizing rates, ranging from 38% to 47% across the three specifications.

We can use our elasticities from Table 2.2 to estimate how realizations and thus tax revenues change in response to a 5 percentage point tax increase. By the definition of these elasticities, we can estimate how realizations and thus tax revenues change for a given elasticity estimate and tax rate change. In 2017, the most recent year that our state and federal tax data covers, the average combined maximum state and federal tax rate was 27.82%.[14] We use this rate as an approximation to the average marginal tax rate. An increase of 5 percentage points in the tax rate implies that the net-of-tax rate decreases by $\frac{0.05}{1-0.2782} = 6.93\%$. We relate this percentage change in taxes to a percentage change in realizations using our baseline 0-10 year policy-relevant elasticity of 1.87 from Table 2.2, and find that realizations would have been $6.93 \times 1.87 = 13$ percent lower, for a total of $754.3 billion in realizations. To calculate counterfactual tax revenue given this realizations amount, we need to use the *average* tax rate, which differs from the marginal. In 2018, the most recent year for which we have data on federal tax revenue, total capital gains realizations were $890.6 billion (Piketty, Saez and Zucman, 2018), and federal capital gains tax revenue was $158.4 billion,

[13] This possibility seems consistent with the graphical evidence that we find when we examine the behavioral effects following tax increases and tax decreases separately, which are shown in Figure 2.13. Following a tax cut, we see a striking and immediate upward jump in realizations, and the effect grows even larger over the following 3-4 years. However, after this initial jump, the trend reverses, and the behavioral effect 7-9 years after a tax cut is much more modest. In contrast, the effect on realizations following a tax increase is much smaller and more stable over time.

[14] Figure 2.9 shows how the following calculations would change if the elasticity were measured at a different initial tax rate.

yielding an average federal tax rate of $\frac{158.4}{890.6} = 17.79\%$. We assume that the 5 percentage points tax increase also applies to this number, for a counterfactual average tax rate of $17.79 + 5 = 22.79\%$. Applying this average tax rate to the counterfactual realizations amount yields tax revenue of $0.2279 \times 754.3 = \$176.6$ billion, an increase of \$18.2 billion relative to the actual 2018 tax revenue. Repeating these steps, but using the net-of-tax elasticity estimate of 1.01 from our specification with controls for other taxes would result in an increase of \$30.3 billion in tax revenues. In sum, we estimate that a 5 percentage point tax increase would yield between \$18 billion and \$30 billion in additional revenues.

Absent any behavioral response, which corresponds to an elasticity of zero, tax revenues from this 5 percentage point tax increase would increase revenues by \$44.5 billion. Thus, depending on the specification, our estimates indicate that, at current tax rates, between a third and half of the mechanical revenue gains are lost due to behavioral effects.

Our central elasticity estimate is lower than those in most existing studies, and our estimated revenue-maximizing tax rate is correspondingly higher. In particular, it is interesting to compare our estimated elasticity to that used by the JCT to evaluate the budgetary impacts of tax reforms. They currently assume an elasticity with respect to the tax rate of 0.68 (Gravelle, 2020), substantially higher than our estimate. This elasticity would imply a revenue gain from a 5 percentage point tax increase of only \$10.6 billion, or that about three-quarters of mechanical revenue gains are lost due to behavioral effects. Appendix Figure A.1 compares our elasticity estimate, implied revenue-maximizing tax rate, and revenue impact with estimates implied by previous studies. Note that all of these estimates apply to the current capital taxation regime with unlimited deferral and step-up-basis at death.

2.5 Conclusion

This paper estimates the effect of state capital gains tax changes on realizations at the state-level. These estimates reflect the responsiveness of both capital gains realizations as well as the migration of the wealthy. These overall effects at the state-level are of interest in their own right as many states consider revenue-raising options to address budgetary pressures. We hope that estimates of how state-level realizations evolve around state capital gains tax changes may serve as useful inputs

into this process, as well as to the literature on state taxes, capital gains behavior, and migration.

We also provide policy-relevant elasticities at the federal level, which use a new framework to account for migration responses on the wealthy and other aggregation adjustments. Our main estimate is an elasticity of realizations with respect to the tax rate of -0.5 to -0.3 depending on the specification.

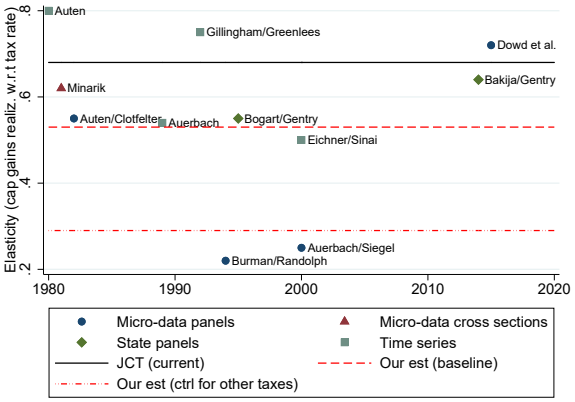
These main elasticity estimates are notably smaller than estimates used by official analysts to score federal tax reforms (Gravelle, 2020). We hope that these elasticity estimates, as well as their implications for revenue-maximizing rates and tax revenue effects, will be considered when evaluating reforms to capital gains taxation, although we note that using these estimates to evaluate large tax rate changes may require extrapolation beyond the observed variation in our sample. In addition, revenue-maximizing rates may exceed welfare-maximizing rates to the extent that capital gains taxes reduce investment below optimal levels or create lock-in effects that misallocate capital. Overall, our bottom-line finding is that raising capital gains tax rates has sizable revenue-raising potential, and that cutting capital gains tax rates has substantial fiscal cost.

Appendices

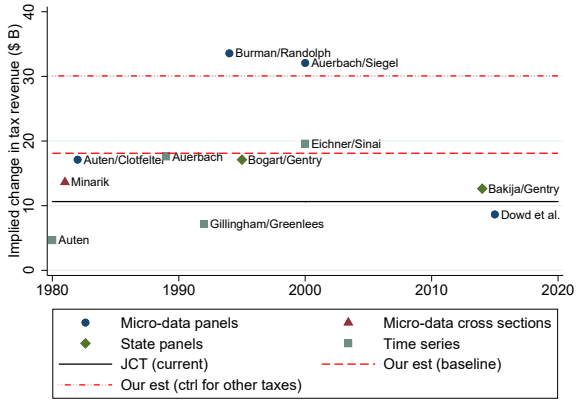
2.A Appendix Figures

Figure 2.4: Estimates from the Capital Gains Literature

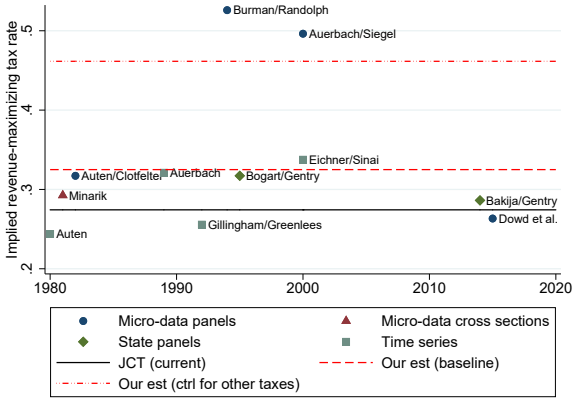
(a) Elasticities



(b) Revenue from 5 pp tax increase

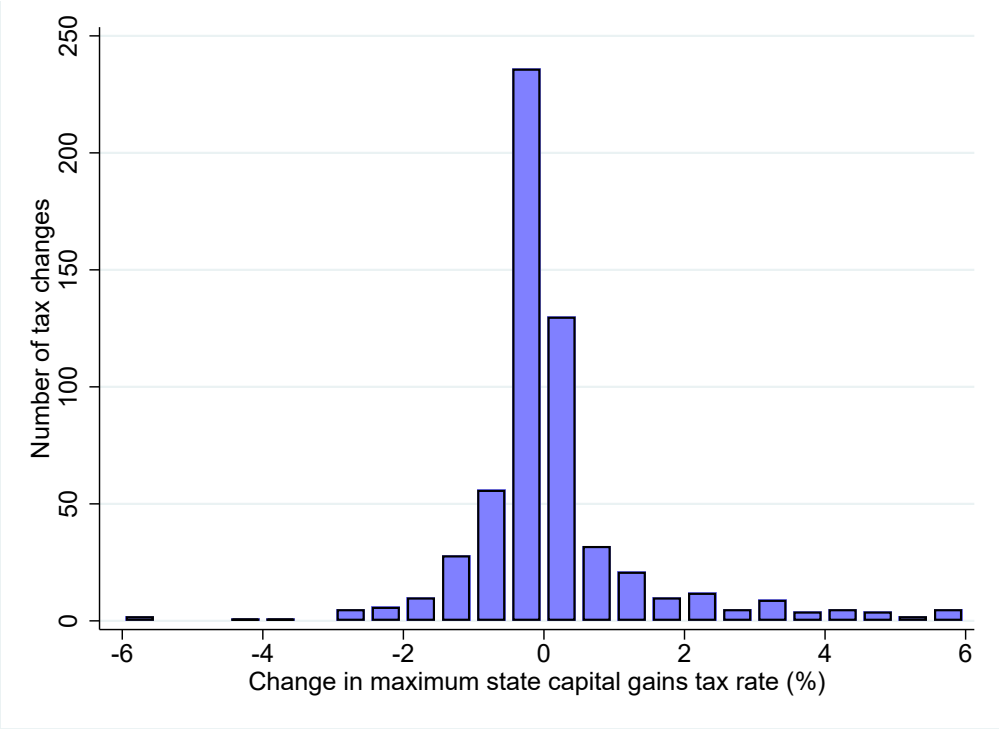


(c) Revenue-maximizing tax rate



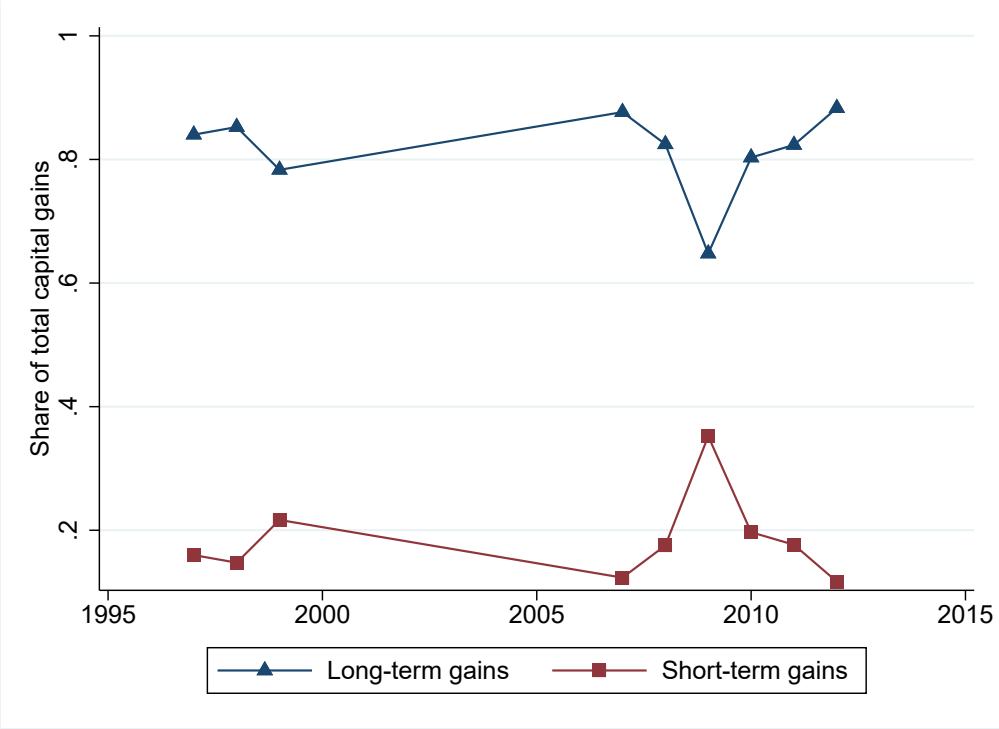
Notes: This figure compares our results with those from the previous literature on capital gains. With the exception of Dowd, McClelland and Muthitacharoen (2015), the elasticities reported in panel (a) are calculated with respect to a tax rate of 22% and are taken from Tables 1 and B-1 from Gravelle (2020). The relevant marginal tax rates in the Dowd et al. paper are those in place between 1999 and 2008, which average around 17 percent. The tax increase analyzed in panel (b) is a 5 percentage point increase, from 17.8% (the effective federal capital gains tax rate in 2018) to 22.8%, assuming that realizations and revenues begin at their 2018 levels ($891 B and $158 B respectively). We calculate revenue-maximizing rate using equation (2.6), using an average state tax rate of 6.4% (its 2016 population-weighted value).

Figure 2.5: Histogram of State Capital Gains Tax Rate Changes



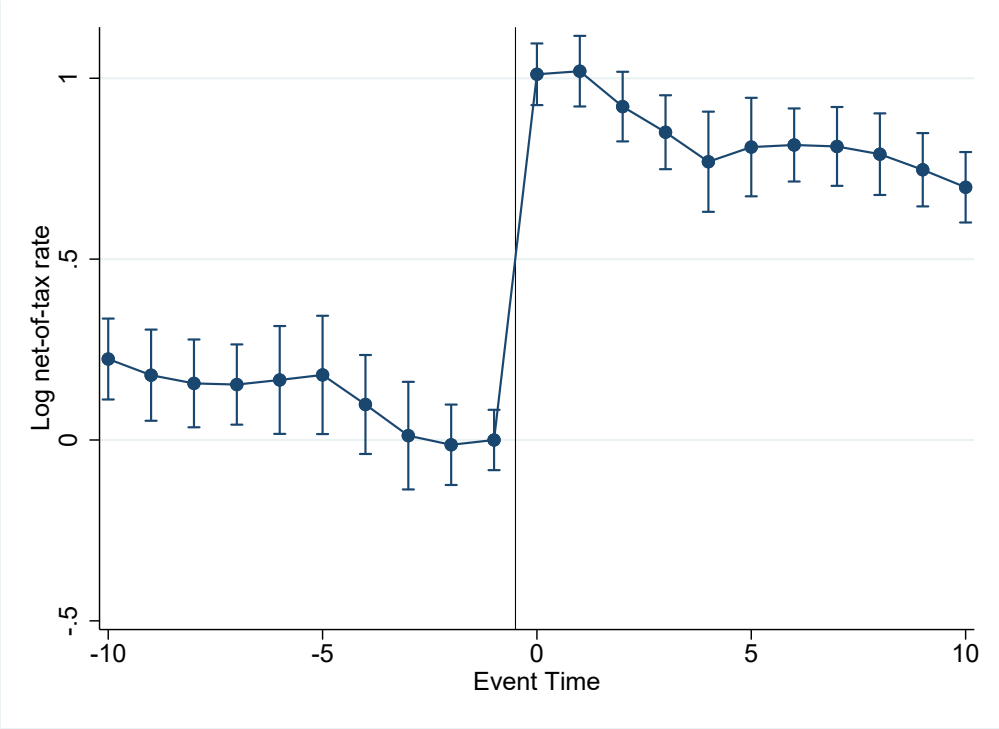
Notes: This figure shows the distribution of the size of changes in the top state capital gains tax rate throughout our panel, including changes to the statutory tax rate as well as deductibility and other minor provisions of the tax code. It has been censored at 6%, so tax changes of more than 6 percentage points in absolute value appear in the left- or rightmost bin in this figure. The figure does not include state-years where the tax rate stayed the same, i.e. changed by 0.

Figure 2.6: Long- and short-term share of capital gains



Notes: Data from the IRS SOI series "Short-term and Long-term Capital Gains and Losses" classified by size of adjusted gross income and selected asset type (accessible at `https://www.irs.gov/statistics/soi-tax-stats-sales-of-capital-assets-reported-on-individual-tax-returns`). More precisely, the long-term gain series corresponds to total gains reported in returns with long-term gain transactions. The short-term gain series corresponds to total gains reported in returns with short-term gain transactions. We calculate total capital gains as the sum of these two series. We use cross-sectional data, which are available for 1985, 1997-1999, and 2007-2012. Although panel data are available for 1999-2007, the IRS warns that extrapolations from the panel to the general population should be made with "extreme caution". Due to the IRS' warning, we have omitted data for these years.

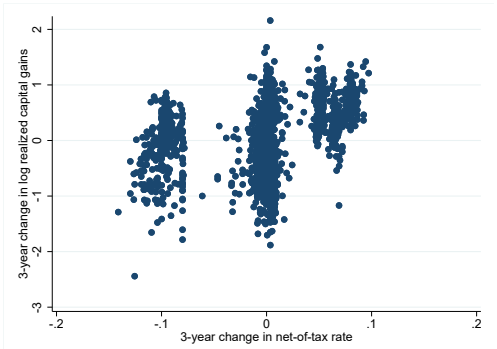
Figure 2.7: The Effect of Net-of-Tax Rate Changes on the Net-of-Tax Rate over Time



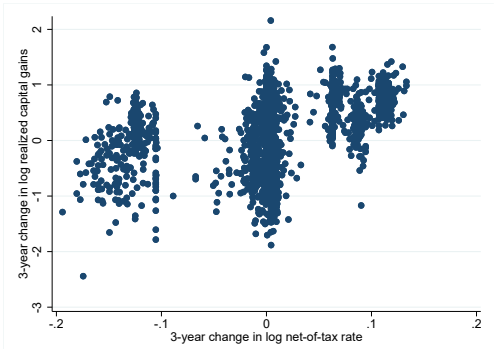
Notes: This figure uses our direct projections framework from equation (2.1), but with the outcome variable being the net-of-tax rate itself, and excluding the controls for capital gains tax reforms in other years. This creates a specification where the treatment effect from time -1 to time 0 is 1 by construction, but where the point estimates in other years show the degree of mean reversion in tax rates. A decline in point estimates following the initial jump indicates mean reversion, as states that lower tax rates tend to partially revert back towards higher tax rates over time and vice versa. The estimated coefficients are normalized to equal 0 at time -1. Standard errors are clustered at the state level.

Figure 2.8: Scatterplots, 3-year Changes in Log Capital Gains, Net-of-Tax Rate, and Log Net-of-Tax Rate

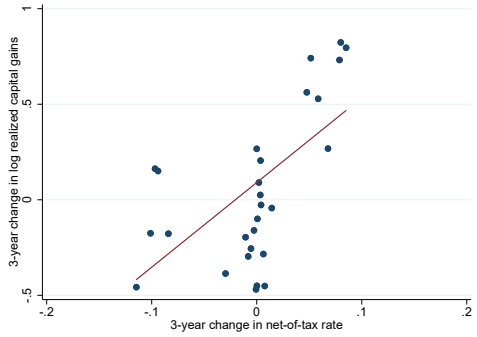
(a) Log capital gains and net-of-tax rate



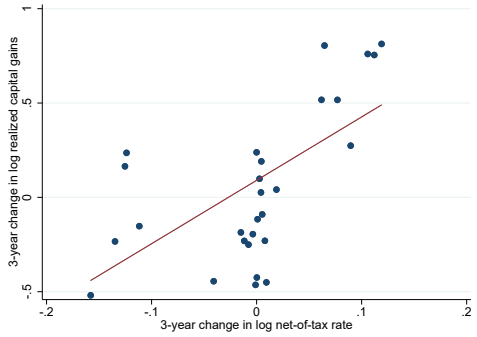
(b) Log capital gains and log net-of-tax rate



(c) Binscatter, log capital gains and net-of-tax rate



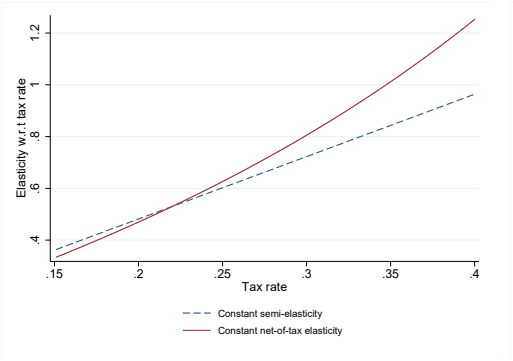
(d) Binscatter, log capital gains and log net-of-tax rate



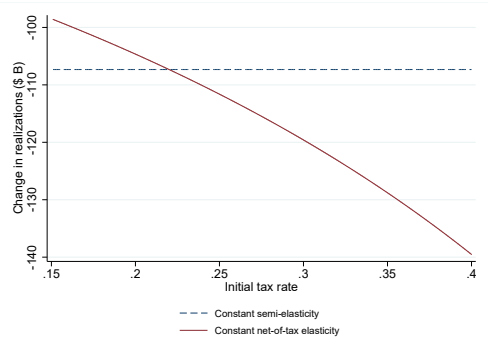
Notes: This figure shows two scatterplots of our state-year data on tax rates and realized capital gains, as well as corresponding binned scatterplots. In all plots, the variable on the vertical axis is the three-year change in log realized capital gains (2018 dollars). In the plots on the left, the variable on the horizontal axis is the three-year change in the net-of-tax rate by state (using the maximum combined federal and state capital gains tax rate, as we describe in section 2), while in the plots on the right, the horizontal axis shows the three-year change in the log net-of-tax rate. The plots in the top row are standard scatterplots, while the plots in the bottom row are the corresponding binned scatter plots, using 30 bins. The regression corresponding to the line of best fit in the scatterplot in panel (c) has an R^2 of 0.1304, and in panel (b) the R^2 is 0.1352.

Figure 2.9: Comparing Elasticity Assumptions

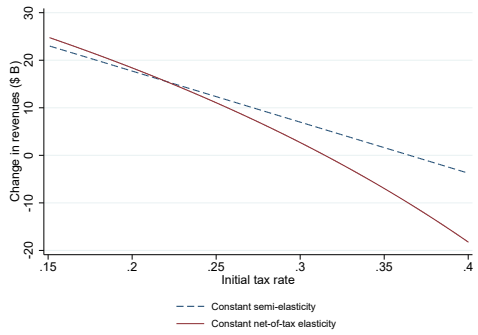
(a) Elasticities



(b) Realizations



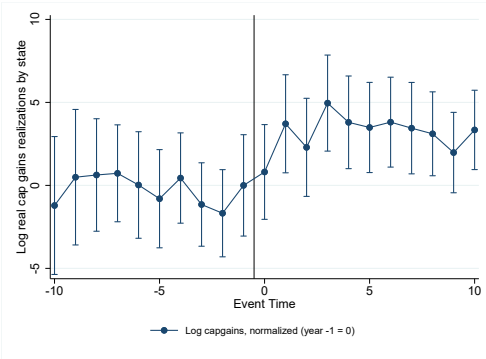
(c) Revenues



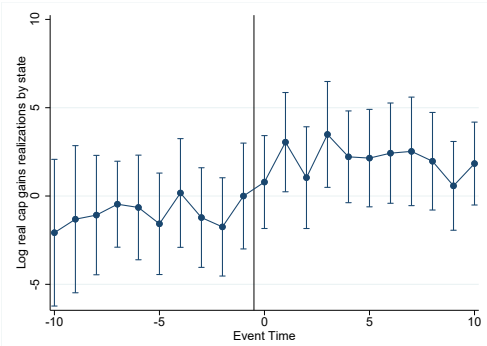
Notes: This figure supplements the discussion in Section 2.C of models of capital gains realizations that either assume a constant semi-elasticity (models of the form $\log CG_t = \gamma \cdot \tau_t$), or a constant net-of-tax elasticity (of the form $\log CG_t = \delta \cdot \log(1 - \tau_t)$). We here show some implications of these models at various tax rates when either the semi-elasticity or the net-of-tax elasticity is assumed to be constant for all tax rates. All calculations are based on a tax rate elasticity of 0.53 at 22%, corresponding to a semi-elasticity of 2.41 and a net-of-tax rate elasticity of 1.88. Panel (a) shows, at each given tax rate, what is obtained when converting the semi-elasticity or net-of-tax elasticity to an elasticity with respect to the tax rate. We also evaluate the implications of both models for realizations (panel b) and tax revenue (panel c) under a 5 percentage point tax increase relative to the tax rate on the horizontal axis. We assume initial realizations equal to $891 billion regardless of starting tax rate. This figure is equal to net capital gains realizations in 2018 according to data from Piketty, Saez and Zucman (2018). Total tax rates observed in our data range from 0.15 to 0.37, comparable to the range of rates presented here.

Figure 2.10: The Effect of Net-of-Tax Rate Changes on Capital Gains Realizations, by Specification

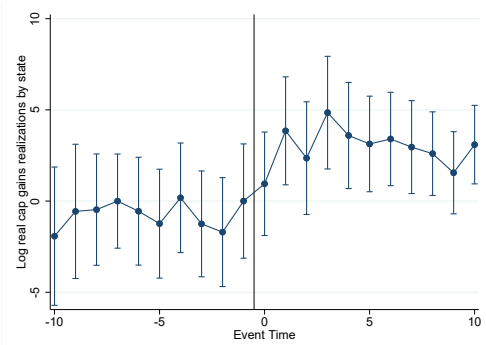
(a) Control for GDP



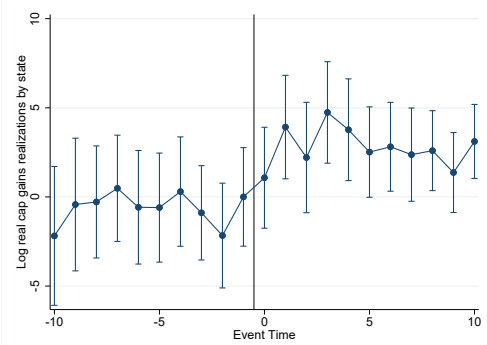
(b) Control for GDP growth*year



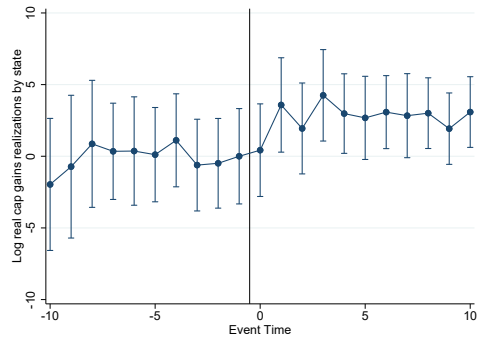
(c) Control for unemployment



(d) Large reforms only

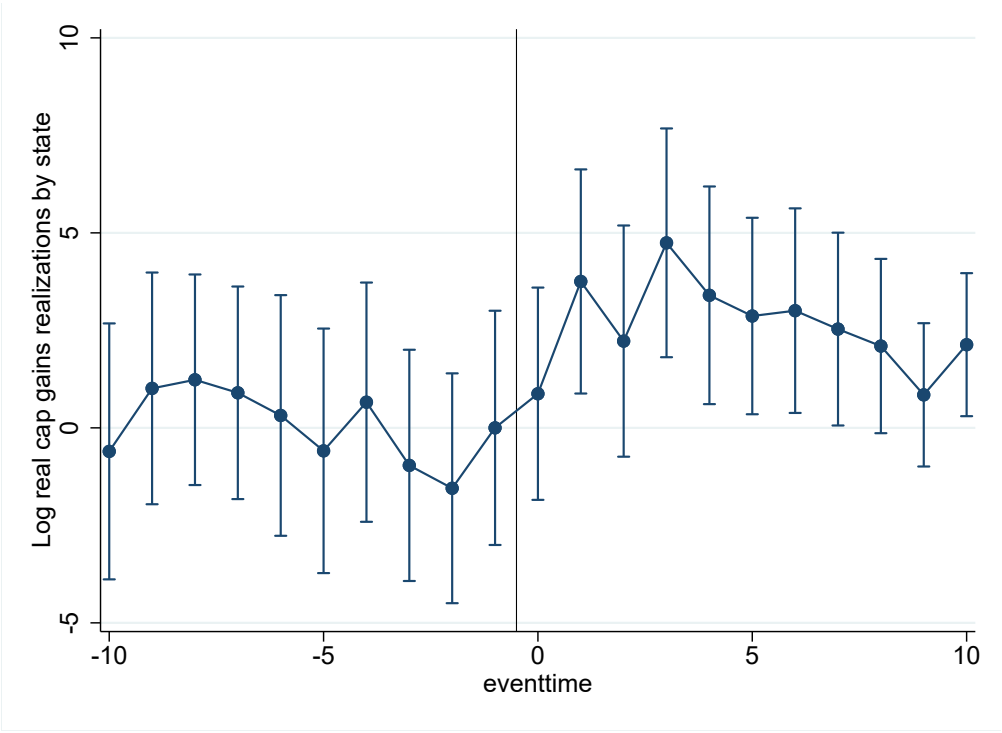


(e) Control for other taxes



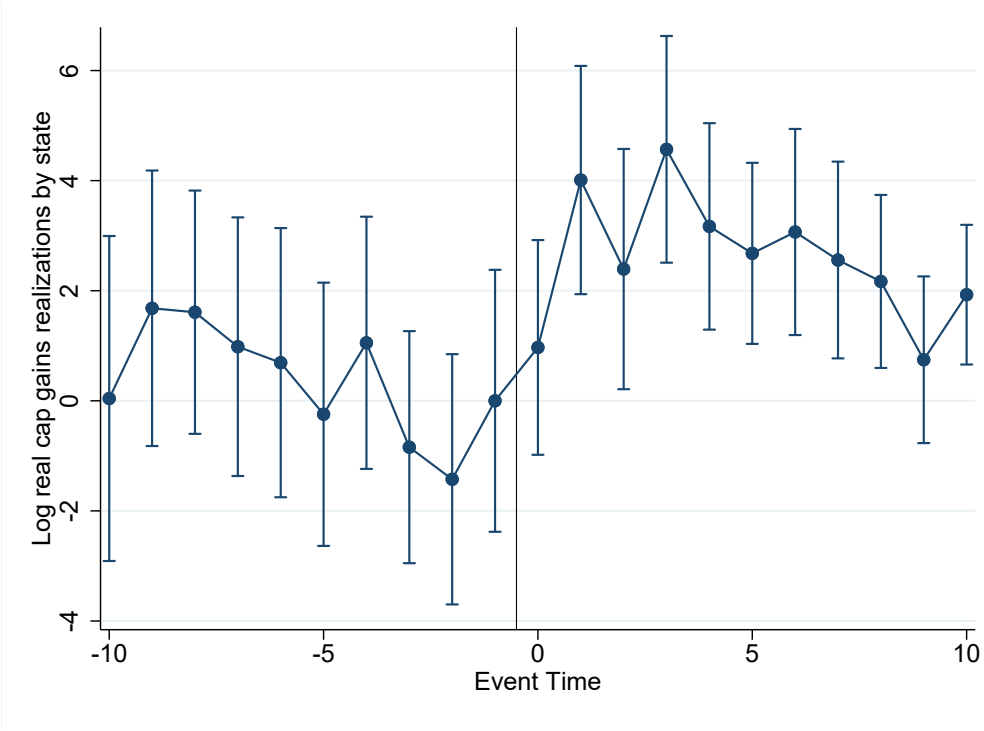
Notes: This figure shows the five series from panel (b) of Figure 2.2 individually. Panel (a) in this figure controls linearly for GDP in the year before the reform; panel (b) controls for a set of dummies interacting GDP growth tertiles and years, where the GDP growth tertile is determined separately in each year by grouping states according to GDP growth over the three years prior to the reform; panel (c) controls for the unemployment rate in the year prior to the capital gains tax reform. Panel (d) uses a specification that interacts the variable for changes in the net-of-tax rate with indicators for reforms greater or smaller than 1 percentage point in magnitude – the reported point estimates are for the interaction with the large reform dummy, such that the effects shown in this panel are idetified only from larger reforms. Panel (e) includes controls for changes in the state personal income and corporate tax rates in each year from 10 years before until 10 years after the capital gains reform. In all cases, we plot the estimated coefficients for the impact of a one-period change in the total (federal and state) log net-of-tax capital gains tax rate on log capital gains realizations. Capital gains are in real terms and standard errors are clustered at the state level. Estimates have been normalized to 0 in period -1. We include state and year fixed effects.

Figure 2.11: Baseline Specification as Event Study



Notes: This figure shows a variation of our baseline graph in Figure 2.2 that uses a classic event study specification in which all plotted coefficients are estimated within the same single regression rather than the direct projections approach. The specification used is $y_{s,t} = \sum_{h=-10}^{10} \beta_h \Delta \log(1 - \tau_{s,t-h}) + \gamma_s + \phi_t + \varepsilon_{s,t}$ (where Δ indicates a one-period change), and the figure plots the point estimates for $\beta_{-10}, \beta_{-9}, \ldots, \beta_{10}$. Capital gains are in real terms and the estimated coefficients are normalized to equal 0 at time -1. Standard errors are clustered at the state level.

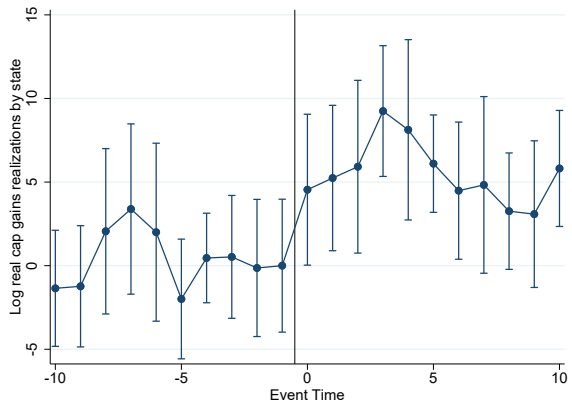
Figure 2.12: The Effect of Net-of-Tax Rate Changes on Capital Gains Realizations – No Controls for Other Capital Gains Tax Reforms



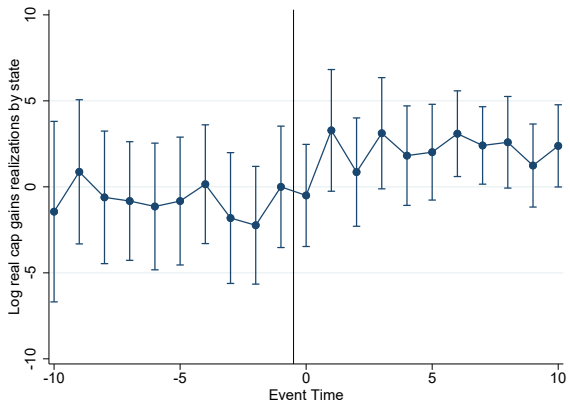
Notes: This figure is analogous to panel (a) of Figure 2.2, but without controls for other changes in the capital gains tax rate before and after the reform of interest. Realized capital gains are in real terms, and the estimated coefficients are normalized to equal 0 at time -1. Standard errors are clustered at the state level.

Figure 2.13: The Effect of Net-of-Tax Rate Changes on Capital Gains Realizations – Tax Cuts and Increases

(a) Tax cuts only

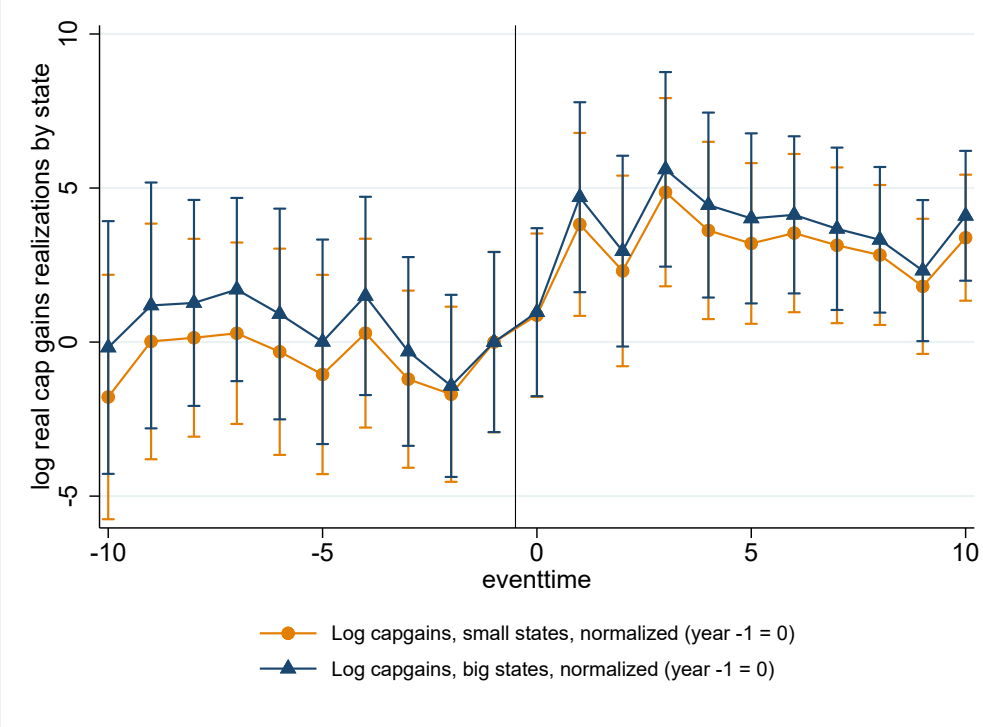


(b) Tax increases only



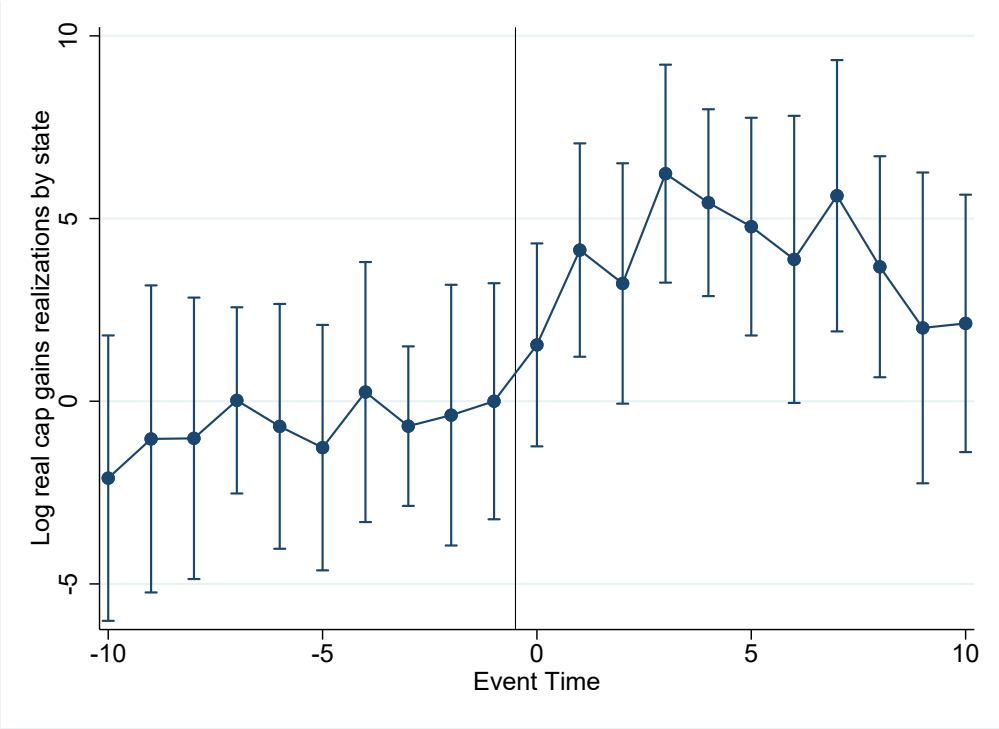
Notes: This figure is analogous to panel (a) of Figure 2.2, but shows the effects of tax cuts and tax increases separately. Specifically, we modify the baseline specification in equation (1) by estimating separate coefficients for tax increases and decreases, i.e., we use the specification $y_{s,t+h} = \beta_h^{cut} \times \mathbf{1}(\Delta\tau_{s,t} < 0) \times \Delta \log(1 - \tau_{s,t}) + \times \beta_h^{inc} \times \mathbf{1}(\Delta\tau_{s,t} > 0) \times \Delta \log(1 - \tau_{s,t}) + \mathbf{X}'_{s,t}\mathbf{\Lambda}_h + \gamma_{s,h} + \phi_{t,h} + \varepsilon_{s,t,h}$, where $\mathbf{1}(\Delta\tau_{s,t} < 0)$ is an indicator for tax cuts, $\mathbf{1}(\Delta\tau_{s,t} > 0)$ is an indicator for tax increases, and β_h^{cut} and β_h^{inc} are the corresponding tax-change-direction-specific coefficients. Panel (a) reports the coeffcients $\{\beta_h^{cut}\}_{h=-10}^{10}$, and panel (b) reports $\{\beta_h^{inc}\}_{h=-10}^{10}$ from this specification. Realized capital gains are in real terms, and the estimated coefficients are normalized to equal 0 at time -1. Standard errors are clustered at the state level.

Figure 2.14: The Effect of Net-of-Tax Rate Changes on Capital Gains Realizations in the 10 most populous states versus the rest



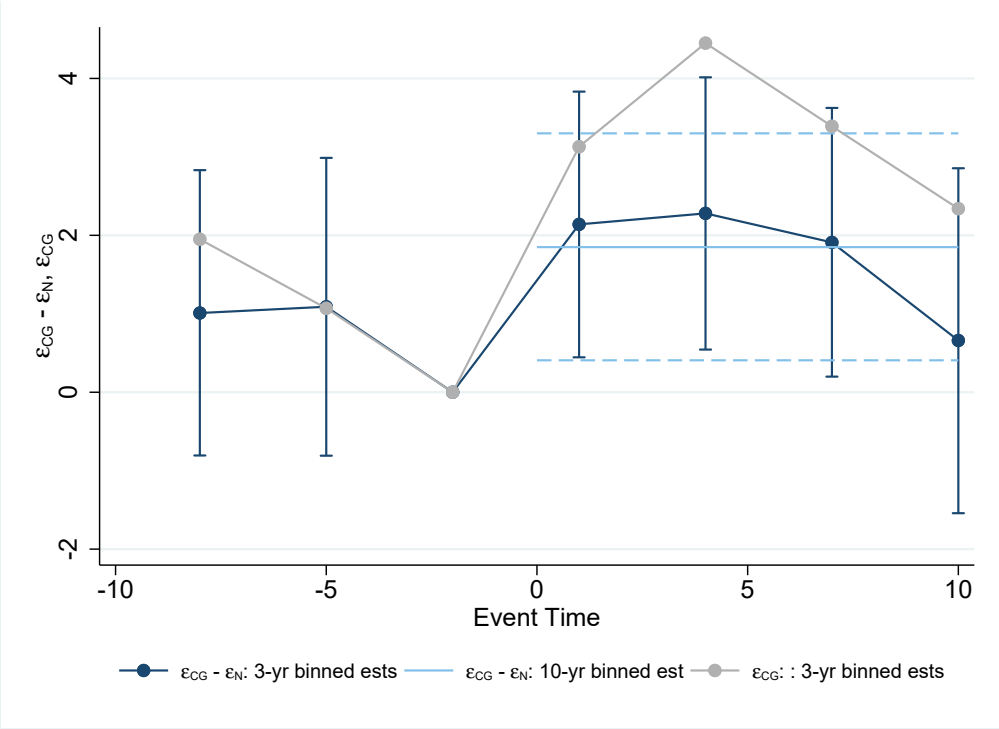
Notes: This figure is analogous to panel (a) of Figure 2.2, but interacts the main right-hand-side variable of interest with an indicator for belonging to the 10 most populous states, lagged by 10 years relative to the reform to avoid any potential confounding effects of recent population growth or decline. This thus creates two sets of treatment effects, one for larger and one for smaller states. Realized capital gains are in real terms, and the estimated coefficients are normalized to equal 0 at time -1. Standard errors are clustered at the state level.

Figure 2.15: The Effect of Net-of-Tax Rate Changes on Capital Gains Realizations After 1990



Notes: This figure is analogous to panel (a) of Figure 2.2, but restricts the specification to only estimate the effects of reforms after 1990. We do use realizations data from prior to 1990 to estimate the pre-trend, and we also include reforms from before 1990 when in our set of controls for capital gains tax reforms in the years surrounding the reform of interest. Realized capital gains are in real terms, and the estimated coefficients are normalized to equal 0 at time -1. Standard errors are clustered at the state level.

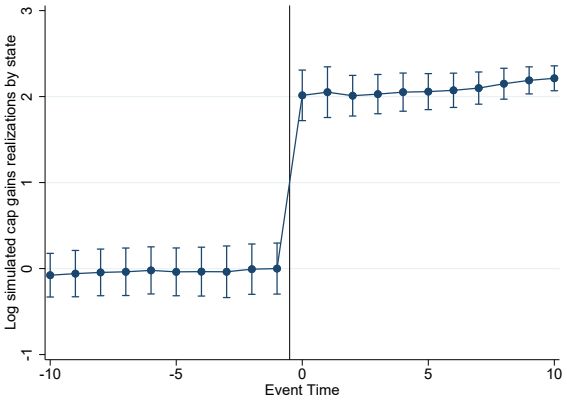
Figure 2.16: Binned Elasticity Estimates, Event Study Specification



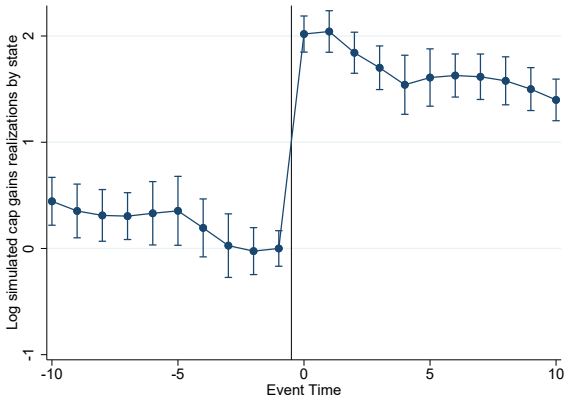
Notes: This figure is analogous to panel (c) of Figure 2.3, but uses an event study specification where elasticities at different horizons are estimated jointly within the same regression model rather than using separate regressions for each coefficient as in the direct projections case. Specifically, for the three-year bins, instead of equation 2, we run a single regression of the form $y_{s,t} = \beta_{-9}\Delta_3 \log(1-\tau_{s,t+9}) + \beta_{-6}\Delta_3 \log(1-\tau_{s,t+6}) + \cdots + \beta_9\Delta_3 \log(1-\tau_{s,t-9}) + \gamma_s + \phi_t + \varepsilon_{s,t}$. Using capital gains realizations as the outcome variable, the grey line then plots elasticity estimates $\hat{\varepsilon}^{CG}$ as the difference between the coefficients $\beta_t - \beta{-}3$ from this regression. Migration elasticities $\hat{\varepsilon}^{N}$ are calculated correspondingly, using our migration variable from the main text as the outcome variable, and the dark blue line then plots the estimated policy-relevant elasticity, $\hat{\varepsilon}^{R} = \hat{\varepsilon}^{CG} - \hat{\varepsilon}^{N}$. For the estimates in the 0-10 year bin, we run a separate regression $y_{s,t} = \tilde{\beta}_{-9}\Delta_3 \log(1-\tau_{s,t+9}) + \tilde{\beta}_{-6}\Delta_3 \log(1-\tau_{s,t+6}) + \tilde{\beta}_{-3}\Delta_3 \log(1-\tau_{s,t+3}) + \tilde{\beta}_0\Delta_{11} \log(1-\tau_{s,t}) + \tilde{\gamma}_s + \tilde{\phi}_t + \tilde{\varepsilon}_{s,t}$ and estimate the 0-10 year elasticity as the difference $\tilde{\beta}_0 - \tilde{\beta}_{-3}$. Standard errors are clustered at the state level.

Figure 2.17: Estimating Effects of Simulated Tax Changes Assuming a Constant Elasticity

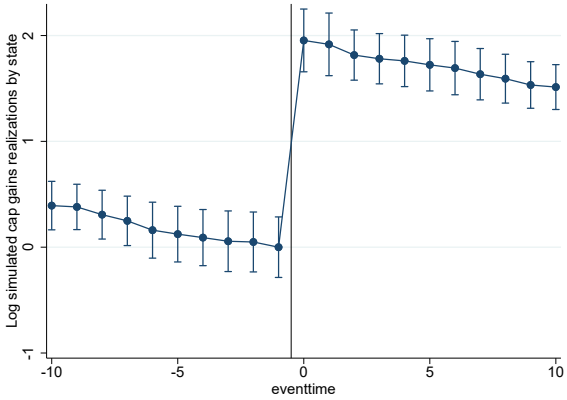
(a) Direct projection, control for reforms in surrounding years



(b) Direct projection, no controls for other reforms



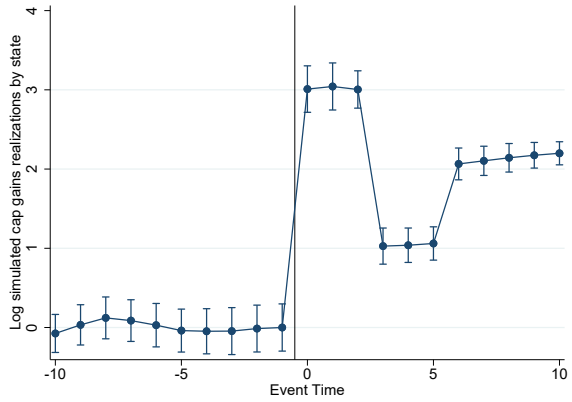
(c) Event study



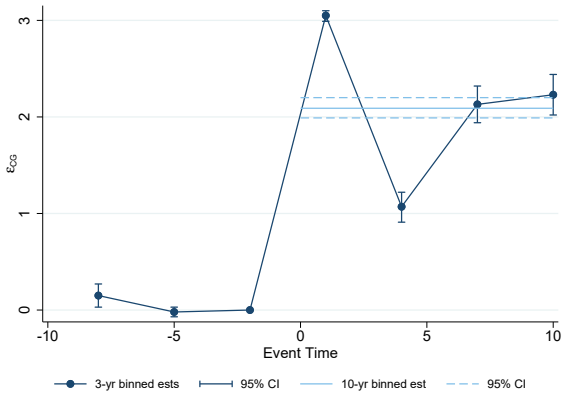
Notes: This figure shows the outcome of the simulation exercise detailed in Appendix Section 2.D. We simulate a log capital gains realizations variable which is generated in such a way that it has a true empirical elasticity of 2 with respect to the net-of-tax rate, constant across all horizons. The figures use three different methods to estimate the impact of changes in the net-of-tax rate on this simulated outcome variable. Panel (a) shows the outcome of a direct projections estimation exactly as specified in equation (1), panel (b) shows the outcome of a similar specification that excludes the controls for tax reforms in other years, and panel (c) shows the outcome of an event study specification. Estimates have been normalized to 0 in period -1. We include state and year fixed effects. See Appendix Section 2.D for further details.

Figure 2.18: Estimating Effects of Simulated Tax Changes Assuming a Time-Varying Elasticity

(a) Direct projection, control for reforms in surrounding years



(b) 3-year and 10-year binned estimates



Notes: This figure shows the outcome of the simulation exercise detailed in Appendix Section 2.D. Unlike Figure 2.17, we here generate an outcome variable for which the true elasticity with respect to the net-of-tax rate is 3 in years 0-2 following the reform, 1 in years 3-5, and 2 in the following years. Panel (a) shows the outcome of a direct projections estimation exactly as specified in equation (1). Panel (b) shows the outcome of the estimation using binned specifications as outlined in section 2.2. We include state and year fixed effects. See Appendix Section 2.D for further details.

Table 2.3: Largest Capital Gains Tax Changes by State-Year

	Year	Δ tax rate	Post-change tax rate	Δ net of tax rate	Post-change net of tax rate
Largest tax decreases					
Oklahoma	2005	-7.97	0.19	5.19	83.83
Alaska	1980	-5.80	0.00	1.79	72.79
Rhode Island	2003	-4.46	4.82	2.67	75.82
Montana	2005	-3.84	6.11	2.50	79.98
New York	1978	-3.00	7.50	9.90	81.75
Vermont	2002	-2.99	5.20	1.81	75.65
Vermont	1994	-2.94	9.28	1.76	65.21
Kansas	1988	-2.84	3.87	1.34	69.21
Wisconsin	1984	-2.60	4.00	1.30	78.00
North Carolina	2014	-2.18	5.80	1.27	71.73
Largest tax increases					
Maine	1989	8.50	8.50	-6.12	65.88
Hawaii	2016	7.30	14.55	-4.51	66.50
Utah	1987	6.14	7.75	-11.96	67.23
Minnesota	1987	5.99	9.00	-12.03	66.46
District of Columbia	1987	5.60	10.00	-11.95	65.85
New Mexico	1987	5.38	8.50	-11.67	66.77
Idaho	1987	5.20	8.20	-11.54	66.96
Delaware	1987	4.92	8.80	-11.47	66.59
Kansas	1987	4.83	6.71	-11.19	67.87
Montana	2003	4.71	9.38	-3.12	72.86

Notes: This table shows the largest year-by-year changes in top marginal state capital gains tax rates, measured using TAXSIM and presented as percentages. Tax changes are relative to previous year.

Table 2.4: Policy Regressions

	Tax Increases		**Tax Decreases**	
	Estimate	Standard Error	Estimate	Standard Error
Unemployment (t-1)	1.2372	0.4009	0.7959	0.5581
Unemployment (t-4)	0.2352	0.4537	-0.6874	0.5517
GDP per capita, USD 1000 (t-1)	-0.0000	0.0020	0.0080	0.0044
GDP per capita, USD 1000 (t-4)	-0.0014	0.0020	-0.0077	0.0044
Max. state income tax (t-1)	0.9204	0.4928	1.9205	0.7455
State corporate tax (t-1)	0.6303	0.2792	-0.0142	0.3590
State in middle CGT tertile (t-1)	0.0119	0.0268	0.0425	0.0378
State in highest CGT tertile (t-1)	0.0018	0.0339	0.0440	0.0484
Constant	0.0036	0.0474	-0.0132	0.0452

Notes: This table shows the results of two OLS regressions aiming to determine whether economic conditions, previous capital gains rates, and other tax rates predict changes in the capital gains tax rate. The outcome variables are dummies for increases and decreases in the state capital gains tax rate at time t, respectively. The variables "Unemployment", "Max state income tax" and "State corporate tax" are measured as decimal numbers, not percentages. We do not include fixed effects. Standard errors are clustered at the state level.

Table 2.5: Capital Gains and Income Tax Changes by State-Year

	Capital gains tax rate			
Income tax rate	Decreased	Unchanged	Increased	Total
Decreased	75.1% (259)	0.7% (10)	13.8% (33)	14.4% (302)
Unchanged	13.6% (47)	98.5% (1484)	18.8% (45)	75.4% (1576)
Increased	11.3% (39)	0.9% (13)	67.4% (161)	10.2% (213)
Total	100.0% (345)	100.0% (1507)	100.0% (239)	100.0% (2091)

Notes: This table relates the direction of year-by-year changes in state capital gains and income tax rates. Within each column, the percentages show the share of income tax changes that went in the given direction. The rates used are maximum state capital gains and wage tax rates, calculated for a hypothetical high-income taxpayer for years 1977-2017 by Daniel Feenberg for NBER. Tax changes are relative to previous year. Observation (state-year) counts are in parentheses.

Table 2.6: Capital Gains Semi-Elasticities and Revenue-Maximizing Tax Rates

Specification	Total elast. γ^{CG}	Policy elast. $\gamma^R = \gamma^{CG} - \gamma^N$	Laffer rate, $\tau^* = \frac{-1}{\gamma^R}$	Elasticity with respect to tax, $\varepsilon^{tax} = \gamma^R \cdot \frac{-0.22}{1-0.22}$	χ^2 test: $\varepsilon^{tax} = -1$
Baseline					
0-10 years	-5.53	-2.89	0.35	-0.64	1.81
	(1.35)	(1.23)	(0.15)	(0.27)	(0.18)
0-2 years	-5.23	-3.21	0.31	-0.71	1.12
	(1.32)	(1.26)	(0.12)	(0.28)	(0.29)
3-5 years	-7.36	-3.34	0.30	-0.74	0.94
	(1.49)	(1.24)	(0.11)	(0.27)	(0.33)
6-8 years	-6.59	-3.05	0.33	-0.67	1.25
	(1.70)	(1.33)	(0.14)	(0.29)	(0.26)
6-10 years	-6.04	-2.39	0.42	-0.53	2.38
	(1.81)	(1.40)	(0.24)	(0.31)	(0.12)
Big changes only					
0-10 years	-4.80	-2.37	0.42	-0.52	3.25
	(1.34)	(1.20)	(0.21)	(0.26)	(0.07)
0-2 years	-5.55	-3.74	0.27	-0.82	0.35
	(1.33)	(1.36)	(0.10)	(0.30)	(0.55)
3-5 years	-7.62	-3.49	0.29	-0.77	0.68
	(1.48)	(1.28)	(0.11)	(0.28)	(0.41)
6-8 years	-6.27	-2.67	0.37	-0.59	1.98
	(1.68)	(1.33)	(0.19)	(0.29)	(0.16)
6-10 years	-5.01	-1.78	0.56	-0.39	3.66
	(1.87)	(1.44)	(0.45)	(0.32)	(0.06)
Control for other taxes					
0-10 years	-4.02	-1.61	0.62	-0.35	3.18
	(1.81)	(1.65)	(0.64)	(0.36)	(0.07)
0-2 years	-3.88	-1.90	0.53	-0.42	2.62
	(1.63)	(1.63)	(0.45)	(0.36)	(0.11)
3-5 years	-5.68	-2.35	0.42	-0.52	2.24
	(1.72)	(1.46)	(0.26)	(0.32)	(0.13)
6-8 years	-5.66	-2.28	0.44	-0.50	1.94
	(2.02)	(1.63)	(0.31)	(0.36)	(0.16)
6-10 years	-5.24	-2.00	0.50	-0.44	2.28
	(2.20)	(1.69)	(0.42)	(0.37)	(0.13)

Notes: This table corresponds to 2.2 in the main text, but uses a specification in which the right-hand side variable of interest is the change in the tax rate, rather than in the log net-of-tax rate. Specifically, we use the exact estimation method described in Section 2.2.3, with the exception that in equation (2.2), we replace the variable $\Delta_3 \log(1-\tau_{s,t})$ with $\Delta_3 \tau_{s,t}$, and correspondingly, in equation (3), $\Delta_{11} \log(1-\tau_{s,t})$ is replaced with $\Delta_{11}\tau_{s,t}$. Similarly, the vector of controls for reforms in surrounding years, $\mathbf{X}_{s,t}$, is also specified using the tax rate in levels rather than the log net-of-tax rate. With this change of right-hand-side variables, columns 1 and 2 estimate semi-elasticities in a way completely analogously to how we estimate the corresponding elasticities in Table 2. In column 3, we calculate the revenue-maximizing tax rate given the policy-relevant semi-elasticity from column 2. In a model where the semi-elasticity γ is assumed constant, the revenue-maximizing tax rate can be found as $\tau^* = -\frac{1}{\gamma}$. In column 4, we convert our semi-elasticity to an elasticity with respect to the tax rate simply by multiplying by a tax rate of 0.22. More details on our calculations can be found in Appendix 2.C. Standard errors are clustered at the state level. Values in parentheses in columns 1-4 represent standard errors; values in parentheses in column 5 (χ^2 test) represent p-values.

Table 2.7: Capital Gains Elasticities and Revenue-Maximizing Tax Rates - Event Study

Specification	Total elast/ ε^{CG}	Policy elast. $\varepsilon^{R} = \varepsilon^{CG} - \varepsilon^{N}$	Laffer rate, $\tau^{*} = \frac{1-\bar{\tau}_S}{1+\varepsilon^{R}}$	Elasticity with respect to tax, $\varepsilon^{tax} = \varepsilon^{R} \cdot \frac{-0.22}{1-0.22}$	χ^2 test: $\varepsilon^{tax} = -1$
Baseline					
0-10 years	3.35	1.85	0.33	-0.52	5.26
	(0.90)	(0.74)	(0.08)	(0.21)	(0.02)
0-2 years	3.13	2.14	0.30	-0.60	2.65
	(0.97)	(0.86)	(0.08)	(0.24)	(0.10)
3-5 years	4.45	2.28	0.29	-0.64	2.05
	(1.09)	(0.89)	(0.08)	(0.25)	(0.15)
6-8 years	3.39	1.91	0.32	-0.54	3.50
	(1.04)	(0.87)	(0.10)	(0.25)	(0.06)
6-10 years	3.00	1.41	0.39	-0.40	6.34
	(1.01)	(0.85)	(0.14)	(0.24)	(0.01)
Big changes only					
0-10 years	2.68	1.57	0.37	-0.44	7.22
	(0.94)	(0.74)	(0.10)	(0.21)	(0.01)
0-2 years	3.48	2.72	0.25	-0.77	0.72
	(0.99)	(0.97)	(0.07)	(0.27)	(0.40)
3-5 years	4.84	2.75	0.25	-0.78	0.68
	(1.15)	(0.97)	(0.06)	(0.27)	(0.41)
6-8 years	3.38	2.00	0.31	-0.57	2.86
	(1.05)	(0.91)	(0.09)	(0.26)	(0.09)
6-10 years	2.44	1.30	0.41	-0.37	6.38
	(1.04)	(0.89)	(0.16)	(0.25)	(0.01)
Control for other taxes					
0-10 years	2.70	1.35	0.40	-0.38	6.27
	(0.96)	(0.88)	(0.15)	(0.25)	(0.01)
0-2 years	2.31	1.27	0.41	-0.36	4.57
	(1.08)	(1.07)	(0.19)	(0.30)	(0.03)
3-5 years	3.47	1.69	0.35	-0.48	4.03
	(1.10)	(0.92)	(0.12)	(0.26)	(0.04)
6-8 years	3.01	1.61	0.36	-0.45	3.67
	(1.12)	(1.01)	(0.14)	(0.29)	(0.06)
6-10 years	2.60	1.25	0.42	-0.35	5.23
	(1.10)	(1.01)	(0.19)	(0.28)	(0.02)

Notes: This table corresponds to Table 2.2, but uses an event study specification where elasticities at different horizons are estimated jointly within the same regression model. Specifically, for the three-year bins, instead of equation (2.2), we run a single regression of the form $y_{s,t} = \beta_{-9}\Delta_3 \log(1-\tau_{s,t+9}) + \beta_{-6}\Delta_3 \log(1-\tau_{s,t+6}) + \cdots + \beta_{9}\Delta_3 \log(1-\tau_{s,t-9}) + \gamma_s + \phi_t + \varepsilon_{s,t}$. Using capital gains realizations as the outcome variable, we then provide elasticity estimates as the difference between the coefficients $\beta_t - \beta_{-3}$ from this regression. Migration elasticities are calculated correspondingly, using our migration variable from the main text as the outcome. For the estimates in the 0-10 year and 6-10 year rows, we run separate regressions. For the 0-10 year bin, we run the regression $y_{s,t} = \tilde{\beta}_{-9}\Delta_3 \log(1-\tau_{s,t+9}) + \tilde{\beta}_{-6}\Delta_3 \log(1-\tau_{s,t+6}) + \tilde{\beta}_{-3}\Delta_3 \log(1-\tau_{s,t+3}) + \tilde{\beta}_{0}\Delta_{11} \log(1-\tau_{s,t}) + \tilde{\gamma}_s + \tilde{\phi}_t + \tilde{\varepsilon}_{s,t}$ and estimate the 0-10 year elasticity as the difference $\tilde{\beta}_0 - \tilde{\beta}_{-3}$. Similarly, for the 6-10 year bin, we run the regression $y_{s,t} = \check{\beta}_{-9}\Delta_3 \log(1-\tau_{s,t+9}) + \cdots + \check{\beta}_{0}\Delta_3 \log(1-\tau_{s,t}) + \check{\beta}_{3}\Delta_3 \log(1-\tau_{s,t-3}) + \check{\beta}_{6}\Delta_5 \log(1-\tau_{s,t-6}) + \check{\gamma}_s + \check{\phi}_t + \check{\varepsilon}_{s,t}$ and identify the 6-10 year elasticity as $\check{\beta}_6 - \check{\beta}_{-3}$. In the specification that controls for other taxes, we control for changes in state labor and corporate tax rates over the same year bins as for capital gains taxes. In the "big changes only" specification, we replace the $\Delta_q \log(1-\tau_{s,t-r})$-variables with two variables that each sum changes that are smaller and larger than 1 pp, respectively, over the relevant time horizon. See the notes to Table 2.2 for details. After estimating the total capital gains and migration elasticities in this way, all other calculations follow exactly as in Table 2.2. Standard errors are clustered at the state level. Values in parentheses in columns 1-4 represent standard errors; values in parentheses in column 5 (χ^2 test) represent p-values.

Table 2.8: Capital Gains Elasticities and Revenue-Maximizing Tax Rates with Controls for Unemployment

Specification	**Total elast.** ε^{CG}	**Policy elast.** $\varepsilon^{R} = \varepsilon^{CG} - \varepsilon^{N}$	**Laffer rate,** $\tau^{*} = \frac{1-\bar{\tau}_S}{1+\varepsilon^{R}}$	**Elasticity with respect to tax,** $\varepsilon^{tax} = \varepsilon^{R} \cdot \frac{-0.22}{1-0.22}$	χ^2 **test:** $\varepsilon^{tax} = -1$
Baseline					
0-10 years	3.39	1.87	0.33	-0.53	3.38
	(1.01)	(0.91)	(0.10)	(0.26)	(0.07)
0-2 years	3.32	2.09	0.30	-0.59	2.58
	(0.97)	(0.91)	(0.09)	(0.26)	(0.11)
3-5 years	4.78	2.28	0.29	-0.64	2.05
	(1.10)	(0.88)	(0.08)	(0.25)	(0.15)
6-8 years	4.07	1.94	0.32	-0.55	3.02
	(1.20)	(0.92)	(0.10)	(0.26)	(0.08)
6-10 years	3.66	1.47	0.38	-0.41	4.54
	(1.27)	(0.97)	(0.15)	(0.27)	(0.03)
Control for unemployment					
0-10 years	3.24	1.79	0.34	-0.51	3.87
	(0.98)	(0.89)	(0.11)	(0.25)	(0.05)
0-2 years	3.36	2.03	0.31	-0.57	2.74
	(0.98)	(0.92)	(0.09)	(0.26)	(0.10)
3-5 years	4.71	2.15	0.30	-0.61	2.80
	(1.14)	(0.84)	(0.08)	(0.24)	(0.09)
6-8 years	3.87	1.76	0.34	-0.50	4.24
	(1.26)	(0.87)	(0.11)	(0.24)	(0.04)
6-10 years	3.45	1.39	0.39	-0.39	5.10
	(1.22)	(0.96)	(0.16)	(0.27)	(0.02)
Control for unemployment growth					
0-10 years	3.04	1.94	0.32	-0.55	3.20
	(0.94)	(0.90)	(0.10)	(0.25)	(0.07)
0-2 years	3.02	2.17	0.30	-0.61	2.14
	(0.94)	(0.94)	(0.09)	(0.27)	(0.14)
3-5 years	4.56	2.37	0.28	-0.67	1.91
	(1.07)	(0.85)	(0.07)	(0.24)	(0.17)
6-8 years	3.92	1.93	0.32	-0.54	2.89
	(1.24)	(0.95)	(0.10)	(0.27)	(0.09)
6-10 years	3.59	1.63	0.36	-0.46	3.99
	(1.18)	(0.96)	(0.13)	(0.27)	(0.05)

Notes: This table provides additional alternative specifications to our main estimation, beyond those shown in Table 2.2. The top panel of this table, labelled "Baseline", corresponds exactly to the Baseline panel in Table 2.2. The next two panels add controls for unemployment in two different ways. The middle panel includes a linear control for the state's unemployment rate prior to the capital gains tax reform in question. The bottom panel instead controls for a set of dummies interacting unemployment growth tertiles and years, where the unemployment growth tertile is determined separately in each year by grouping states according to unemployment growth over three years prior to the reform. Standard errors are clustered at the state level. Values in parentheses in columns 1-4 represent standard errors; values in parentheses in column 5 (χ^2 test) represent p-values.

Table 2.9: Capital Gains Elasticities and Revenue-Maximizing Tax Rates with Controls for Other Taxes and Unemployment

Specification	Total elast. ε^{CG}	Policy elast. $\varepsilon^{R} = \varepsilon^{CG} - \varepsilon^{N}$	Laffer rate, $\tau^{*} = \frac{1-\bar{\tau}_S}{1+\varepsilon^{R}}$	Elasticity with respect to tax, $\varepsilon^{tax} = \varepsilon^{R} \cdot \frac{-0.22}{1-0.22}$	χ^2 test: $\varepsilon^{tax} = -1$
Control for other taxes					
0-10 years	2.28	1.01	0.47	-0.29	4.64
	(1.32)	(1.18)	(0.27)	(0.33)	(0.03)
0-2 years	2.38	1.25	0.42	-0.35	3.88
	(1.19)	(1.16)	(0.21)	(0.33)	(0.05)
3-5 years	3.58	1.64	0.36	-0.46	3.54
	(1.24)	(1.01)	(0.14)	(0.29)	(0.06)
6-8 years	3.32	1.40	0.39	-0.39	3.57
	(1.43)	(1.14)	(0.19)	(0.32)	(0.06)
6-10 years	2.98	1.18	0.43	-0.33	3.96
	(1.54)	(1.19)	(0.24)	(0.34)	(0.05)
Control for taxes, unemployment					
0-10 years	2.22	0.96	0.48	-0.27	4.83
	(1.30)	(1.18)	(0.29)	(0.33)	(0.03)
0-2 years	2.35	1.20	0.43	-0.34	3.92
	(1.23)	(1.18)	(0.23)	(0.33)	(0.05)
3-5 years	3.51	1.57	0.36	-0.44	3.84
	(1.33)	(1.01)	(0.14)	(0.28)	(0.05)
6-8 years	3.18	1.31	0.41	-0.37	4.11
	(1.47)	(1.10)	(0.19)	(0.31)	(0.04)
6-10 years	2.94	1.13	0.44	-0.32	4.23
	(1.48)	(1.18)	(0.24)	(0.33)	(0.04)
Control for taxes, unemployment growth					
0-10 years	2.21	1.06	0.45	-0.30	4.67
	(1.20)	(1.15)	(0.25)	(0.32)	(0.03)
0-2 years	2.31	1.31	0.41	-0.37	3.68
	(1.12)	(1.16)	(0.20)	(0.33)	(0.05)
3-5 years	3.55	1.70	0.35	-0.48	3.68
	(1.20)	(0.96)	(0.12)	(0.27)	(0.05)
6-8 years	3.38	1.38	0.39	-0.39	4.01
	(1.41)	(1.08)	(0.18)	(0.30)	(0.05)
6-10 years	3.08	1.25	0.42	-0.35	4.23
	(1.45)	(1.12)	(0.21)	(0.31)	(0.04)

Notes: This table provides additional alternative specifications to our main estimation, beyond those shown in Table 2.2. The top panel of this table, labelled "Control for other taxes", controls for changes in state labor and corporate taxes, and corresponds exactly to the bottom panel in Table 2.2 – see the notes to Table 2.2 for further details on how this is specified. The next two panels in this table include the same controls for other state tax changes, but also control for unemployment in two different ways. The middle panel includes a linear control for the state's unemployment rate prior to the capital gains tax reform in question. The bottom panel instead controls for a set of dummies interacting unemployment growth tertiles and years, where the unemployment growth tertile is determined separately in each year by grouping states according to unemployment growth over three years prior to the reform. Standard errors are clustered at the state level. Values in parentheses in columns 1-4 represent standard errors; values in parentheses in column 5 (χ^2 test) represent p-values.

2.B Data Appendix

This appendix covers our data sources in more detail and clarifies adjustments and assumptions that have been made in cleaning and processing the data.

Tax rates

- Maximum federal, state, and total tax rates on long-term capital gains for the years 1977 to 2017 are obtained from NBER TAXSIM data. These data are calculated by simulating the effect of a change in income for a hypothetical high earner in the TAXSIM model, and thus incorporate statutory state tax rates as well as phaseouts of exemptions and itemized deductions. In some specifications, we control for changes in the maximum state income tax and corporate tax rates, which we also obtain from this dataset.[15]

- Table 1 shows maximum federal long-term capital gains tax rates, which we obtain from the US Department of the Treasury.[16] As of the time of writing, the Treasury data starts in 1954, but unfortunately only goes to 2014. Knowing that the maximum statutory tax rate on capital gains did not change between 2014 and 2017, we manually extend the data forwards, assuming a tax range that is unchanged from its 2014 level. Since the data also includes the effects of other minor provisions on maximum tax rates, it is possible that we are leaving out smaller changes in the maximum tax rate through our extrapolation. However, we expect any such discrepancies to be small, as the maximum effective tax rate is very close to the maximum statutory rate for recent years. In the data, certain years contain two different tax rates. In these cases, the capital gains tax rate changed at some point during the year, and different tax rates applied for gains that were realized in different parts of the years. Since our data on realized capital gains are at the year level, we are unable to apply these different tax rates over the year. Therefore, for these few years, we assume the relevant tax rate to be the midpoint between the two tax rates that applied over the year.

[15] Further background information and the dataset itself are available at `https://users.nber.org/~taxsim/state-rates/`.

[16] These data are currently available in excel format at `https://www.treasury.gov/resource-center/tax-policy/tax-analysis/Documents/Taxes-Paid-on-Capital-Gains-for-Returns-with-Positive-Net-Capital-Gains.xlsx`.

- Data on UK tax rates comes from HM Revenue and Customs.[17] Note that for some of the years in question, the UK had different tax rates for individuals and trusts. For the purposes of our analysis, we ignore trusts, which account for a comparatively small and stable amount of capital gains across the years we study.

Realized capital gains

- Data on realized capital gains at the federal level are obtained from the appendix tables of Piketty, Saez and Zucman (2018).[18] Specifically, the numbers we use can be found in column 21 of Table C1. These numbers were calculated by Piketty, Saez and Zucman (2018) as aggregates of the NBER micro-files samples of tax returns and correspond to the *fikgi* variable.[19] Note that these are net long-term plus short-term capital gains, i.e., any capital losses are subtracted from the total in each year.[20]

- Data on realized capital gains at the state level are taken from Smith, Zidar and Zwick (2020). The series that we use comes from a state-year collapse of the *fikgi* variable in the IRS SOI sample files. We also use data from the same source on the number of residents by state within the top 1 and 10% of the national wealth distribution.

- UK data on realized net capital gains comes from HM Revenue and Customs.[21] These data are given separately for individuals and trusts, which are taxed at different rates over the period that we study. We only use the data for individuals, who are taxed at a 40 percent rate throughout all the years that we study. For specific types of trusts, the tax rate did change slightly over the period that we study. Our choice to only use realized capital gains by individuals is therefore based on an assumption that any spillover effects between the two tax bases resulting from changes in the trust tax rate are negligible. We view this as a reasonable assumption since capital gains by trusts is a small and relatively stable share

[17] Specifically, it can be found at `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/764247/Table_A1.pdf`. Note that we do not currently use these data directly in our analysis, except to verify that the UK did not change its capital gains tax rate at any time during the periods in which we use it as a control group to the US. Capital gains were treated as part of regular income throughout all of the years in our UK sample. The maximum income tax rate throughout this period was 40 percent.

[18] Currently, these can be found under `http://gabriel-zucman.eu/files/PSZ2017AppendixTablesI(Macro).xlsx`

[19] See the Piketty, Saez and Zucman (2018) codebook here `http://gabriel-zucman.eu/files/PSZCodebook.pdf`.

[20] In future work, it would be valuable to consider gains and losses separately and test where the inclusion of net short term gains and losses and netting of loss carryovers from past years have different effects.

[21] They can be found at `https://www.gov.uk/government/statistics/capital-gains-tax-statistical-tables`

of total capital gains, since only certain types of trusts (i.e, interest in possession trusts and personal representative) were subject to tax changes over the period we study, and since the tax changes were very small, changing by only 3 percentage points over the period.

Other variables

- In some specifications, we control for state GDP numbers obtained from the Bureau of Economic Analysis.[22]
- Other specifications control for unemployment, which we obtain from the Bureau of Labor Statistics.[23]
- To generate variables in per capita terms, we use annual estimates of US state populations from the US Census Bureau, obtained through the St. Louis Fed.[24]
- We convert dollar amounts into real values using the CPI for urban consumers (CPI-U), obtained from the US Bureau of Labor Statistics.[25]
- Nominal values from the UK are deflated using the Retail Price Index, avalable from the Office for National Statistics. [26]

[22]These can be found using the Regional Economic Accounts download tool at `https://apps.bea.gov/regional/downloadzip.cfm`.

[23]Available at `https://www.bls.gov/lau/#data`.

[24]These can be obtained as a single zip file by using the download tool at `https://research.stlouisfed.org/pdl/628/download`.

[25]Available at `https://www.bls.gov/cpi/data.htm`.

[26]`https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceindices`.

2.C On Elasticities and Semi-Elasticities

Our approach in equation (2.1) relates log capital gains realizations to the log net-of-tax rate, which is a standard approach in much of the recent literature on the elasticity of taxable income (e.g., Gruber and Saez, 2002; Saez, Slemrod and Giertz, 2012*a*; Kleven and Schultz, 2014; Doerrenberg, Peichl and Siegloch, 2017), where the log net-of-tax rate (or changes in the log net-of-tax rate) is often used as the primary explanatory variable. One reason for this formulation is that the net-of-tax price is often the relevant price governing behavior in standard economic models.[27] However, most of the previous literature on capital gains taxes (e.g., Bogart and Gentry, 1995; Bakija and Gentry, 2014; Dowd, McClelland and Muthitacharoen, 2015) has related log capital gains to the linear tax rate or net-of-tax rate, fitting specifications of the semi-log form:

$$\log CG_t = \gamma \cdot \tau_t. \tag{2.7}$$

In a model of this form, the coefficient γ corresponds to a semi-elasticity of capital gains realizations with respect to the tax rate. Empirically, given the variation we observe in tax rates, it is hard to determine whether our baseline model in equation (2.1) fits the data better than a model of this type.[28] However, the two models generate different predictions about how behavioral responses scale with tax rates. Whereas our baseline model would imply that a 1 percentage point change in the tax rate induces larger realizations responses when the tax rate is higher, the model in (2.7) implies that a 1 percentage point change in tax rates would always generate the same percentage change in realizations. As we illustrate in Appendix Figure 2.9, the difference between the predictions of these models grows larger when the tax rate is extrapolated further from the rate at which both semi-elasticity and net-of-tax rate elasticity are calculated. This difference can be important, for instance, when using these models to calculate the implied revenue-maximizing tax rate.

We use the model in equation (2.1) as our baseline model because it is standard in the broader taxable elasticity literature and may generate more realistic predictions when considering implied behavioral responses at large tax rates. For example, the model in (2.7) implies that the percentage

[27] For example, labor supply decisions depend on the net-of-tax wage, i.e., $(1-\tau)w$, in standard models of labor supply.

[28] Figure 2.8 shows that there is little difference in how strongly changes in log realized capital gains correlate with changes in net-of-tax rates and log net-of-tax rates, respectively.

change in realizations would be the same whether tax rates were increased from 20 to 21 percent or from 99 to 100 percent, which seems less accurate given standard predictions about distortions growing with the square of the tax rate. Our model, in contrast, would predict much larger responses when tax rates are high.

2.C.1 Empirical Estimates Using a Semi-Log Specification

We provide results estimation using semi-log specifications to facilitate comparisons to the prior literature and to assess the robustness of our estimates. We report these semi-log results in Table 2.6, which has the same structure as Table 2.2 with our binned multi-year estimates from the main text, but with various changes as outlined below.

In the first column of Table 2.6, we estimate the empirical semi-elasticity, which we label γ^{CG}. We do so using the same procedure described in Section 2.2.3 in the main text, but with the right hand-side variable of interest changed. For instance, instead of equation (2.2), we estimate:

$$y_{s,t+h} = \tilde{\theta}_h \Delta_3 \tau_{s,t} + \mathbf{X}'_{s,t} \tilde{\mathbf{\Lambda}}_h + \tilde{\mu}_{s,h} + \tilde{\phi}_{t,h} + \tilde{\varepsilon}_{s,t,h}. \tag{2.8}$$

Just as we do for the elasticities in the main text, we calculate the semi-elasticity based on these estimates as the difference between post- and pre-reform point estimates. For example, the 0-2 year semi-elasticity is calculated as $\tilde{\gamma}_0 \equiv \tilde{\theta}_0 - \tilde{\theta}_{-3}$, the 3-5 year semi-elasticity as $\tilde{\gamma}_3 \equiv \tilde{\theta}_3 - \tilde{\theta}_{-3}$, and so on. Also note that in this specification, the vector of controls for reforms in surrounding years, $\mathbf{X}_{s,t}$, is also specified in terms of the tax rate in levels, i.e., it contains the variables $\Delta_3 \tau_{s,t+r}$ for $r = -9, -6, -3, 3, 6, 9$.

In the second column of Table 2.6, we obtain a semi-elasticity that differences out the migration effect by subtracting the point estimates of γ^N from a specification like (2.8) with $\theta_s \ln N_{s,t}^{P99-P100} + (1-\theta_s) \ln N_{s,t}^{P90-P100}$ as the outcome variable. This step is completely analogous to what we do in Table 2 for our main estimates.[29]

[29] One small difference is that unlike in Table 2, we do not multiply by the minor adjustment factor that we describe in Appendix 2.E. This adjustment is not needed in a model where semi-elasticities are assumed constant for all tax rates. We can see this point by repeating the simplification from equation (2.15) in the context of a semi-elasticity model. Specifically, let $\gamma \equiv \frac{\mathrm{d}CG}{\mathrm{d}\tau} \cdot \frac{1}{CG}$ be the policy-relevant semi-elasticity at a national level (corresponding to ε in

We then go on to calculate the revenue-maximizing tax rate implied by the estimated policy-relevant semi-elasticities. As in Appendix 2.E, this is the tax rate that solves the federal government's revenue maximization problem,

$$\begin{aligned}\tau_F^* &= \arg\max_{\tau_F} \tau_F \cdot CG \\ &= \arg\max_{\tau_F} \log \tau_F + \log CG\end{aligned}$$

Taking the first order condition for maximization with respect to τ_F in the second line, we get:

$$0 = \frac{1}{\tau_F} + \frac{\mathrm{d} \log CG}{\mathrm{d}\tau_F},$$

which rearranges to

$$\tau_F^* = -\frac{1}{\gamma},$$

where $\gamma \equiv \frac{\mathrm{d}CG}{\mathrm{d}\tau_F} \cdot \frac{1}{CG} = \frac{\mathrm{d} \log CG}{\mathrm{d}\tau_F}$ is the policy-relevant semi-elasticity at a national level. Note that unlike in our main analysis, where we find the formula for the revenue-maximizing tax rate in equation (2.6), we do not adjust for the average state taxes here. That is because in a model with a constant semi-elasticity, state taxes are irrelevant for the federal tax; differences in the baseline tax rate do not affect the magnitude of behavioral responses. In other words, in this model, if the semi-elasticity is -4, the revenue-maximizing federal tax rate is 25%, regardless of whether average state taxes are 0 or 50%.

(2.15)), and let $\gamma^R \equiv \frac{\mathrm{d} \log R_s}{\mathrm{d}\tau} = \frac{\mathrm{d}R_s}{\mathrm{d}\tau} \cdot \frac{1}{R_s}$ be the realizations semi-elasticity at the state level. We then find:

$$\begin{aligned}\gamma &= \frac{\mathrm{d}CG}{\mathrm{d}\tau} \cdot \frac{1}{CG} \\ &= \left[\sum_{s \in S} N_s \frac{\mathrm{d}R_s}{\mathrm{d}\tau}\right] \cdot \frac{1}{\sum_{s \in S} N_s R_s} \\ &= \left[\sum_{s \in S} N_s \gamma^R \cdot R_s\right] \cdot \frac{1}{\sum_{s \in S} N_s R_s} \\ &= \gamma^R \left[\sum_{s \in S} N_s R_s\right] \cdot \frac{1}{\sum_{s \in S} N_s R_s} \\ &= \gamma^R.\end{aligned}$$

Finally, in the fourth column of Table 2.6, we convert the estimated policy-relevant semi-elasticity to an elasticity with respect to the tax rate, evaluated at 22% to facilitate comparison to Table 2.2 and the prior literature as referenced in Gravelle (2020). Since the semi-elasticity is given by $\gamma = \frac{\mathrm{d}CG}{\mathrm{d}\tau} \cdot \frac{1}{CG}$, it can be converted into a tax elasticity at any given tax rate τ by simply multiplying it by τ.

Our results in Table 2.6 are broadly consistent with what we find in Table 2.2. While the first two columns are not directly comparable, we find that the semi-log specification delivers tax elasticities that are slightly larger than those in our main specification. However, despite differences in how revenue-maximizing tax rates are calculated in the two models, we get similar estimates of the revenue-maximizing tax rate across our various specifications.

2.D Estimation Using Simulated Data

This section presents the results of two simulation exercises that aim to examine whether our estimation methods are biased when estimating the empirical elasticity of capital gains with respect to the net-of-tax rate. We use our actual empirical data on the maximum combined state and federal capital gains tax rate, but rather than use empirical data on realized capital gains as the outcome variable, we simulate this variable in such a way that we know the true empirical elasticity for this simulated variable with respect to the net-of-tax rate. We start with a simulation that features an elasticity that is constant across all time horizons, before moving on to the more complex case of a time-varying elasticity.

2.D.1 Simulation with constant elasticity

First, we generate the simulated log capital gains variable according to the following formula:

$$\log CG_{s,t}^{sim} = 10 + 1 \cdot fips_s + 2 \cdot t + 2 \cdot \log\left(1 - \tau_{s,t}\right) + \eta_{s,t}. \tag{2.9}$$

From this data-generating process, it is clear to see that the "true" empirical elasticity of capital gains with respect to the net-of-tax rate will be $\varepsilon^{CG} = \frac{\partial \log CG_{s,t}^{sim}}{\partial \log(1-\tau_{s,t})} = 2$. The constant 10 and the terms $1 \cdot fips_s + 2 \cdot t$ are chosen arbitrarily—they are simply there to generate a time trend and some variation across states which will be absorbed by state and year fixed effects in our estimation anyway—mimicking structural time trends and differences across states which are unrelated to the tax rate.[30] The term $\eta_{s,t}$ is a random variable with mean 0 and standard deviation 0.001. In reality, the short-term and long-term elasticities likely differ, which we explore in more detail in subsection 2.D.2. Here, we have here simplified matters and assumed a single permanent elasticity of 2, which holds for both the short and the long term. This is for expositional purposes, since it will make any bias in the different estimation methods easier to see visually.

Having generated this simulated capital gains realizations variable, we now estimate empirically the effect of capital gains tax reforms on it, using various methods. The results of these estimations are shown in Appendix Figure 2.17. First, in panel (a), we use our main specification—the baseline

[30] $fips_s$ is literally just the FIPS code corresponding to state s, which is included as a crude way to ensure that the mean of $\log CG^{sim}$ will vary across states.

direct projections specification from equation (1), which includes a set of controls for other capital gains reforms in the 10 years before and after each reform that enters our regressions. This figure is thus analogous to Figure 2(a) in the main text, but with simulated outcome data. If our estimation method is accurate, the figure should display a completely flat trend in the pre-and post-periods, with a single upward jump of size 2 from year -1 to 0. The results show we accurately identify the true elasticity of 2 in the simulated data. There is a clear jump from time -1 to time 0, the pre-trend is flat, and the trend in the post-period is almost entirely flat as well, with just a very slight upward trend.[31]

Panel (b) of Figure 2.17 uses the same estimation method, but omits our controls for capital gains reforms at different horizons. We see that not controlling for other capital gains tax changes performs worse. While the jump at time 0 is still exactly 2, we now have a clear downward trend both before and after the reform year. This pattern reflects the fact that tax reforms in our data tend to be somewhat mean-reverting, as we show in Appendix Figure 2.7 and discuss briefly in section 2. Because of this mean-reversion, the difference between tax rates at time 0 and -1 will tend to be an overestimate of the difference between rates at time h and -1 for $h > 0$, and because we are overestimating the size of the tax change, we will correspondingly underestimate the magnitude of the elasticity at these longer horizons.

Finally, panel (c) of Figure 2.17 uses an event study method. We use the following specification:

$$\log CG_{s,t}^{sim} = \sum_{h=-10}^{10} \beta_h \Delta \log (1 - \tau_{s,t-h}) + \gamma_s + \delta_t + \varepsilon_{s,t}, \tag{2.10}$$

and the figure plots the point estimates that are normalized relative to period -1, i.e. it shows $\beta_{-10} - \beta_{-1}, \beta_{-9} - \beta_{-1}, \ldots, \beta_{10} - \beta_{-1}$. This differs from our baseline direct projections specification from equation (1) in that all the point estimates are generated within the same regression. We see that this method performs better than the direct projections method without controls, but worse than the direct projections with controls, since there is still somewhat of a downward trend. This result is likely because given the specification that we use, we are unable to bin tax reforms that

[31] An alternative specification that is similar to equation (1), but using $y_{s,t+h} - y_{s,t-1}$ as the outcome variable, would in fact remove this very slight bias. However, we have opted not to use this specification for our empirical analysis, as it would not extend naturally to our long multi-year binned specifications, and therefore we do not present it here either.

occur beyond the endpoints of our window of 10 leads and lags around the realization year.[32]

Overall, these results show that our main direct projections method with controls is accurately identifying the true elasticity in the longer run. In the next subsection, we extend the simulation exercise to show that our estimation methods also accurately identify true elasticities when they differ across time horizons.

2.D.2 Simulation with time-varying elasticity

To simulate a log realized capital gains variable with a varying net-of-tax elasticity over time, we slightly alter the simulated variable we introduced in equation (2.9). It now becomes:

$$\begin{aligned}\log CG_{s,t}^{sim} =& 10 + 1 \cdot fips_s + 2 \cdot t + 2 \cdot \log(1 - \tau_{s,t}) \\ &+ 1 \cdot [\Delta \log(1 - \tau_{s,t}) + \Delta \log(1 - \tau_{s,t-1}) + \Delta \log(1 - \tau_{s,t-2})] \\ &- 1 \cdot [\Delta \log(1 - \tau_{s,t-3}) + \Delta \log(1 - \tau_{s,t-4}) + \Delta \log(1 - \tau_{s,t-5})] \\ &+ \eta_{s,t}.\end{aligned} \tag{2.11}$$

In this specification, the long-run elasticity of capital gains with respect to the net-of-tax rate is still 2, but the added terms modify this elasticity in the short and medium run. The first bracketed term increases the impact of the change in the net-of-tax rate on realizations by 1 for reforms that happened 0, 1 or 2 years ago, which means that the elasticity in these three years will now be 3. Similarly, the second bracketed term decreases the impact of a reform that happened 3, 4 or 5 years ago on capital gains realizations, lowering the elasticity in each of these three years to 1.

We now estimate our baseline specification from equation (1) with this new outcome variable. The result is shown in panel (a) of Figure 2.18. We clearly see that our estimation method accurately picks up on the time-varying elasticity across different years, which confirms that this specification also does a good job of capturing the dynamics of the empirical capital gains elasticity around a tax reform.

Panel (b) of Figure 2.18 estimates the impact of capital gains reforms on the same simulated

[32]Specifiying the right-hand side variables in the event study specification in levels rather than first differences could help deal with this problem, but would also make our empirical estimates harder to compare directly with those from the direct projections specification.

outcome variable using the binned specifications in equations (2) and (3). We see that the 3-year binned specification also captures the dynamics of the elasticity, and the long 0-10-year specification captures the average elasticity across time (the three short-run years with an elasticity of 3 and the three medium-run years with an elasticity of 1 average out to an elasticity of 2 across these six years). Overall, these results confirm that our empirical methods are well-suited to capture the magnitude and evolution of empirical capital gains elasticities over time.

2.E Full Model

This section contains further details and derivations for the model outlined in section 4.1 of the main text. We first show the derivations of the formula for the revenue-maximizing tax rate provided in the text. Consider the revenue maximization problem of the national government.[33] Let $\bar{\tau}_S$ denote the average state tax rate.[34] The government sets τ_F to maximize revenue:

$$\max_{\tau_F} \tau_F \cdot CG, \tag{2.12}$$

where $CG = \sum_{s \in S} CG_s = \sum_{s \in S} N_s \left(1 - \tau_F - \tau_s, \boldsymbol{\tau_{-s}}\right) R_s \left(1 - \tau_F - \tau_s\right)$. Taking logs in (2.12), we get the first order condition

$$\frac{1}{\tau_F} = -\frac{\mathrm{d} \log CG}{\mathrm{d}\tau_F},$$

which is equivalent to

$$\begin{aligned}
\frac{1}{\tau_F} &= \frac{\mathrm{d} \log CG}{\mathrm{d} \left(1 - \tau_F - \bar{\tau}_S\right)} \\
&= \frac{\mathrm{d} \log CG}{\mathrm{d} \log \left(1 - \tau_F - \bar{\tau}_S\right)} \cdot \frac{\mathrm{d} \log \left(1 - \tau_F - \bar{\tau}_S\right)}{\mathrm{d} \left(1 - \tau_F - \bar{\tau}_S\right)} \\
&= \frac{\mathrm{d} \log CG}{\mathrm{d} \log \left(1 - \tau_F - \bar{\tau}_S\right)} \cdot \frac{1}{1 - \tau_F - \bar{\tau}_S}.
\end{aligned}$$

We can rewrite this expression to find that the revenue-maximizing tax rate is

$$\tau_F^* = \frac{1 - \bar{\tau}_S}{1 + \varepsilon}, \tag{2.13}$$

where

$$\varepsilon = \frac{\mathrm{d} \log CG}{\mathrm{d} \log \left(1 - \tau_F - \bar{\tau}_S\right)} = \frac{\mathrm{d} CG}{\mathrm{d} \left(1 - \tau_F - \bar{\tau}_S\right)} \cdot \frac{\left(1 - \tau_F - \bar{\tau}_S\right)}{CG}. \tag{2.14}$$

[33] Changes in the capital gains rate may affect other types of income (e.g., dividends, wage and salary income, etc.). This formula does not account for spillovers to other parts of the tax base, so we may actually be underestimating the revenue maximizing capital gains rate. Estimating these fiscal spillovers precisely is difficult empirically. Another relevant issue is the negative vertical fiscal externality between federal and state governments—maximizing federal revenues is different from maximizing federal plus state revenues.

[34] This average is weighted by state populations N_s, so

$$\bar{\tau}_S = \frac{\sum_{s \in S} N_s \tau_s}{\sum_{s \in S} N_s}.$$

We call this elasticity the *policy-relevant elasticity* since it is a sufficient statistic for determining revenue-maximizing rates at the federal level. We would like to express this elasticity in terms of the realization elasticities ε^R that we described in equation (5), since they are what we can empirically estimate. To simplify the expression in (2.14), we impose two fairly reasonable assumptions.

Assumption 1: In every region, n_s is unaffected by the federal tax rate. Mathematically, this means that

$$\frac{\mathrm{d}n_s\left(1-\tau_F-\tau_1, 1-\tau_F-\tau_2, \ldots, 1-\tau_F-\tau_s, \ldots\right)}{\mathrm{d}\tau_F} = 0 \quad \text{for all } s \in S.$$

In reality, it might be hard to find a function for which this is indeed exactly true everywhere on its domain. However, within the range of usual tax rates that we consider, it seems like a reasonable first degree approximation that a change in national tax rates wouldn't directly cause internal migration between regions.

Assumption 2: Realization elasticities are homogenous across regions, $\varepsilon_s^R = \varepsilon^R$ for all $s \in S$.

Given these two assumptions, we can simplify the expression in (2.14):

$$\begin{aligned}
\varepsilon &= \frac{\mathrm{d}CG}{\mathrm{d}\left(1-\tau_F-\bar{\tau}_S\right)} \cdot \frac{\left(1-\tau_F-\bar{\tau}_S\right)}{CG} \\
&= \left[\sum_{s\in S} N_s \frac{\mathrm{d}R_s}{\mathrm{d}\left(1-\tau_F-\tau_s\right)}\right] \cdot \frac{\left(1-\tau_F-\bar{\tau}_S\right)}{\sum_{s\in S} N_s R_s} \\
&= \left[\sum_{s\in S} N_s \varepsilon^R \cdot \frac{R_s}{\left(1-\tau_F-\tau_s\right)}\right] \cdot \frac{\left(1-\tau_F-\bar{\tau}_S\right)}{\sum_{s\in S} N_s R_s} \\
&= \varepsilon^R \left[\sum_{s\in S} N_s \frac{R_s}{\left(1-\tau_F-\tau_s\right)}\right] \cdot \frac{\left(1-\tau_F-\bar{\tau}_S\right)}{\sum_{s\in S} N_s R_s} \\
&= \varepsilon^R \left[\sum_{s\in S} n_s \frac{R_s}{\left(1-\tau_F-\tau_s\right)}\right] \cdot \frac{\left[\sum_{s\in S} n_s \cdot \left(1-\tau_F-\tau_s\right)\right]}{\left[\sum_{s\in S} n_s R_s\right]},
\end{aligned} \tag{2.15}$$

where n_s is the population of the state expressed as a share of the national population.[35] This expression now gives us the policy-relevant national elasticity as a function of the local realizations elasticity, ε^R, times an adjustment term which is somewhat difficult to interpret but deals with

[35] Formally, this is defined as

$$n_s = \frac{N_s}{\sum_{q\in S} N_q}.$$

weighting among states. Essentially, it corrects for the fact that an elasticity measured at the aggregate level is not necessarily exactly equal to the mean of elasticities measured at a more granular level. Note that in our empirical setting, in every single year for which we have data, the adjustment factor is very close to 1 – its minimum value across all years is 0.993, and its maximum is 1.004. For this reason, ε and ε^R are in practice almost identical, which is why we present equation (6) for the revenue-maximzing tax rate in the main text with an ε^R, although it should be an ε to be completely precise, as in equation (2.13). We use the 2016 value, which is 1.0025, to adjust our empirical estimates when estimating ε^R and the revenue-maximizing tax rates in section 4.3.

Equations (2.13) and (2.15) together tell us directly how to get an estimate of the revenue-maximizing tax rate from local elasticities. Since our empirically estimated elasticities at the state level give us $\varepsilon^{CG} = \varepsilon^N + \varepsilon^S$, we need an estimate of ε^N to get the realizations elasticity. We do this in section 4 of the paper by running a version of our direct projections with the outcome variable being a weighted combination of the log number of state residents that belong to the top 10% and the top 1% of the national wealth distribution, respectively. This can be thought of as applying our outlined model only to the top of the wealth distribution. Capital gains tax reforms are likely to have a limited impact on migration of individuals outside the top 10%, and even if there is a migration effect for these individuals, it will not substantially affect overall capital gains realizations in the state, since individuals outside the top 10% only account for a very small share of capital gains realizations.

Two-region example

To make the above derivations a little more concrete, we present the following simple numerical example. Suppose the country consists of two regions: California (c) and everywhere else (e).

Assume the following values:[36]

$$N_c = 20$$

$$N_e = 80$$

$$\kappa_c = 25$$

$$\kappa_e = 22$$

$$\tau_c = 0.14$$

$$\tau_e = 0.05$$

$$\tau_F = 0.25$$

Suppose we estimate a total elasticity of capital gains $\hat{\varepsilon}^{CG} = 3$ and a migration elasticity $\hat{\varepsilon}^{N} = 1.5$.[37] This implies a realization elasticity of $\hat{\varepsilon}^{R} = 1.5$.

The average population-weighted state tax rate is $\bar{\tau} = 0.2 \cdot 0.14 + 0.8 \cdot 0.05 = 0.068$. Using equation (2.15) above, we thus find the policy-relevant elasticity ε above to be

$$\begin{aligned}
\varepsilon &= \varepsilon^{\kappa} \left[\sum_{s \in S} n_s \frac{R_s}{(1 - \tau_F - \tau_s)} \right] \cdot \frac{\left[\sum_{s \in S} n_s \left(1 - \tau_F - \tau_s\right) \right]}{\left[\sum_{s \in S} n_s R_s \right]} \\
&= 1.5 \cdot \left[\frac{0.2 \cdot 25}{0.61} + \frac{0.8 \cdot 22}{0.7} \right] \cdot \frac{0.2 \cdot 0.61 + 0.8 \cdot 0.7}{0.2 \cdot 25 + 0.8 \cdot 22} \\
&= 1.5 \cdot 1.006 \\
&= 1.509.
\end{aligned}$$

The revenue-maximizing national capital gains tax rate implied by this elasticity is thus

$$\tau_F^* = \frac{1 - \bar{\tau}_S}{1 + \varepsilon} = \frac{1 - 0.068}{1 + 1.509} = 0.371.$$

[36] These values are roughly consistent with actual numbers for people in the top 10% of the wealth distribution in the US in 2016, with the population rebased to 100 individuals and the realized capital gains per capita given in thousands of dollars.

[37] These values are close to what we find in our specification using big tax changes only.

Chapter 3

Tax Complexity as Price Discrimination: Theory and Simulation[1]

Why are most modern tax systems so complex? The benefits of simplifying the tax code would seem obvious at a first glance: Lower compliance costs, less filing requirements and administrative work both for taxpayers and tax authorities, fewer loopholes that can be used for tax avoidance. Many academic researchers have found significant costs associated with complex tax systems (Benzarti, 2020; Pitt and Slemrod, 1989) and have recommended simplifying the tax code (Slemrod, 1989; Fuest, Peichl and Schaefer, 2006). And importantly, simpler tax systems seem to be overwhelmingly popular with voters and the public – for instance, Blesse, Buhlmann and Dörrenberg (2019) find that over 90% of Germans support tax simplification. Nevertheless, complex tax systems continue to be the norm in much of the world.

In this chapter, I demonstrate one possible reason why governments might prefer a complex tax system: When taxpayers have heterogeneous elasticities of taxable income (ETI), tax complexity can serve as a form of price discrimination. A monopolist engaging in second-degree price discrimination can extract a higher total profit from its customers by offering bulk deals, limited-time sales or other

[1]I thank Annette Alstadsæter, Julie Brun Bjørkheim, Ron Davies, Henrik Kleven, John Sturm, Damian Vergara, Owen Zidar and seminar participants at Princeton and the DAEiNA annual congress for helpful comments. I am grateful for financial support for this research from the Skatteforsk Centre for Tax Research at the Norwegian University of Life Sciences

advantages that cater to more price-sensitive customers, while the less elastic shoppers often pay full price. In a similar way, governments can offer deductions, exemptions and loopholes that can reduce the tax burden for those willing to invest in knowledge about the tax code, while the less knowledgeable end up paying more. In the theoretical section of this paper, I first show that a hypothetical government with perfect information and the power to set individualized tax rates would optimally set a lower tax rate on more elastic taxpayers. Of course, this is not feasible as governments cannot observe individual elasticities or set individual tax rates, but I go on to demonstrate that a complex tax system can achieve a similar outcome through a self-selection mechanism. I develop a model of tax complexity in which taxpayers can decide to invest in costly knowledge, which will reduce their tax burden. I prove that the more elastic a taxpayer is, the more tax knowledge they will choose to acquire, since a higher elasticity means their individual welfare gains from reducing tax distortions are larger. This demonstrates that tax complexity can serve as a self-selection mechanism akin to price discrimination.

I also show that if elasticities are increasing with income, this self-selection effect can allow the government to raise a higher overall tax revenue than the maximum possible in a "simple" tax system, and without leaving anyone worse off. The reason is that when taxpayers have heterogeneous ETIs, the tax rate that maximizes overall revenues will be inefficiently high for the most elastic individuals. Allowing for complexity gives them a way to reduce their distortionary tax burden and increase their labor supply significantly, while at the same time, the government can continue to extract full tax revenue from those with lower elasticities who will not find it worth the cost to invest in tax knowledge.

While my theoretical model shows that tax complexity can be efficiency-improving under some circumstances, there are of course many negative aspects of tax complexity as well. Generally, complexity benefits the wealthiest and most elastic taxpayers, and therefore increases inequality, which will be at odds with many policymakers' objectives. In a simulation exercise, I demonstrate that when the initial tax rate is substantially below the revenue-maximizing rate, complexity can boost labor output but will generally not lead to increased tax revenues. A government interested in boosting growth might therefore use tax complexity as a tool, but this will come at the expense of increased inequality and somewhat lower tax revenues.

Overall, my findings illustrate a possible reason why tax complexity may be advantageous

to governments: It can allow them to raise more tax revenues by lowering costly distortions for elastic taxpayers in a more targeted way than through general tax cuts, but this comes at a cost of increased inequality. Although maximizing tax revenue is often not the overall goal for governments, I focus on this outcome as it is important in certain contexts. Tax researchers sometimes impose "Rawlsian" social preferences, which aim to simply increase the welfare of the worst-off member of society, assuming that this person is not a taxpayer (Hellwig, 1986). Under such a model, the optimal tax rate would indeed be the one that raises the most tax revenue. In a less extreme case, optimal tax models often assume a welfare weight of zero on the highest earners in an economy (Saez and Stantcheva, 2016; Piketty and Saez, 2013), meaning that the optimal tax rate for this group would be the revenue-maximizing one. If we consider my model in the context of very high earners, it might make sense why tax systems are often especially complex for them. Many of the highest earners can also be highly tax elastic, since they have more opportunities to move their income, assets, or residence into tax havens. A complex tax code can provide hidden benefits that encourage them to remain in their home country, while at the same time raising a higher tax revenue from those high earners who are less elastic. For the wealthiest, a complex tax system with exemptions and loopholes might therefore be a more efficient way of reducing the tax burden than simply lowering their headline tax rates. From the government's point of view, an additional benefit could be that these complex tax incentives are less apparent to the general public than more straightforward tax breaks at the top, which may be unpopular with voters. Therefore, although tax complexity leads to increased inequality, the fact that this is less salient may itself be a desirable feature to policymakers. My findings highlight the importance of considering possible effects on both tax revenue and inequality when carrying out reforms that significantly alter the complexity of the tax systems.

My work relates to several strands of the economics literature. Multiple previous studies have developed theoretical models of complex tax systems and on optimal taxation when agents lack information or are not fully rational (Goldin, 2015; Rees-Jones and Taubinsky, 2020; O'Donoghue and Rabin, 2003). Craig and Slemrod (2022) develop one of the most comprehensive models of tax complexity, and briefly touch upon the implications of private costly investments in tax knowledge, like this paper. However, their approach involves directly modelling the beliefs and misperceptions of taxpayers, and is therefore less suited to delivering empirical predictions without making explicit

assumptions about these beliefs. The model used by Slemrod (1989) is quite similar to mine, also featuring costly tax knowledge in a very simple framework, but his main focus is on the choice between the itemized and standard deduction in a U.S. context. Krause (2000) is also similar in spirit, but models imperfect information by both the taxpayer and the tax authority, and focuses on compliance and audit risk.

While not many researchers have explored possible benefits of tax complexity, some have studied a similar idea for welfare transfers. Nichols and Zeckhauser (1982) pointed out the possible use of ordeals as a screening mechanism to discourage the less needy from applying for welfare benefits, and this idea has since been explored further by Alatas et al. (2016) and Sunstein (2018), among others. Kleven and Kopczuk (2011) study the optimal complexity of benefit programs in a context where higher complexity improves the government's ability to detect who is eligible, but also imposes a cost on applicants that discourages some from applying.

In the remainder of this paper, I first provide a simple example to illustrate the key ideas at play. I then develop a model that illustrates how a complex tax system can act in a similar fashion to price discrimination. By formulating conditions under which a complex tax system can generate higher tax revenues than a simple one, I show that this increase in tax revenues comes at no efficiency cost. I prove that in this model, more elastic taxpayers will acquire more tax knowledge and thus get a lower effective tax rate, and finally, I demonstrate some further implications of the model through a numerical simulation.

3.1 First-best case: Optimal individualized tax rates

Consider an economy consisting of N workers with quasi-linear and isoelastic[2] utility functions of the following form:

$$U_i(C, L) = C - \frac{e^{-\alpha_i/\varepsilon_i}}{1 + 1/\varepsilon_i} L^{1+\frac{1}{\varepsilon_i}}$$

$$C = (1 - \tau)\, L,$$

[2]We see that preferences are isoelastic by substituting in the budget constraint and taking the first order condition. Doing so and rearranging yields:

$$L = \left(1 - T'(L)\right)^{\varepsilon_i} e^{\alpha_i}.$$

where wages are normalized to 1, L is labor supply, and τ is the tax rate, assumed to be a flat tax in this simple example. For convenience, I will use the notation $\nu = 1 - \tau$ to refer to the net-of-tax rate in the following. ε_i is individual i's elasticity of labor supply with respect to net-of-tax rate ν, and α_i is a preference parameter which shifts overall labor supply without affecting elasticity.

A classic optimal tax problem for this model might derive the flat net-of-tax rate ν which maximizes some social welfare function W subject to a government tax revenue requirement R. We will consider a slight variation of this problem. Suppose the government has perfect information and knows ε_i for every citizen, and that it has the ability to set a separate tax rate τ_i for every single person in the economy based on this information. What will the set of optimal tax rates be in this case? Letting $V_i(\nu)$ and $L_i(\nu)$ denote, respectively, the indirect utility function and the labor supply function that follows from the optimal solution to the individual's problem, we could write this variant problem as:

$$\max_{\nu_i, \lambda} W\left(V_1(\nu_1), \ldots, V_N(\nu_N)\right) + \lambda\left(\sum_{i=1}^{N}(1-\nu_i)L_i(\nu_i) - R\right). \tag{3.1}$$

Taking the first order condition with respect to ν_i for $i = 1, \ldots, N$, we get:

$$\frac{\partial W}{\partial V_i} \cdot \frac{dV_i}{d\nu_i} + \lambda\left(-L_i(\nu_i) + (1-\nu_i)\frac{dL_i}{d\nu_i}\right) = 0. \tag{3.2}$$

By the envelope theorem, we have

$$\begin{aligned}\frac{dV_i}{d\nu_i} &= \frac{\partial U_i\left(L_i(\nu_i), \nu_i\right)}{\partial \nu_i} \\ &= L_i(\nu_i).\end{aligned}$$

If we let $\psi_i = \frac{\partial W}{\partial V_i}$ denote person i's welfare weight, we can rewrite (3.2) as:

$$\psi_i L_i(\nu_i) = -\lambda\left(-L_i(\nu_i) + (1-\nu_i)\frac{dL_i}{d\nu_i}\right).$$

Inserting $L_i(\nu_i) = \nu_i^{\varepsilon_i} e^{\alpha_i}$ in this and rearranging, we can ultimately reduce this equation to:

$$\nu_i = \frac{\varepsilon_i}{1 - \frac{\psi_i}{\lambda} + \varepsilon_i}.$$

This is the optimal net-of-tax rate for person i. It is increasing in elasticity ε_i, and approaches 1 as the elasticity goes to infinity – in other words, the more elastic someone is, the less they should be taxed.[3] This result is probably not surprising – from classical optimal tax theory, we know that the higher the aggregate elasticity in an economy, the lower the optimal tax will be, and our example here simply shows that a similar result will hold on an individual level.

We can think of this example as analogous to first-degree price discrimination. When a monopolist is assumed to have perfect information and the ability to charge a separate price for each customer, the optimal price will be higher for less elastic customers. Although the government in this example isn't motivated by profit maximization, a similar result holds.

Of course, the assumptions made in the first-degree price discrimination model are extremely strong: The monopolist must have perfect information and the ability to set unique prices for everyone. The analogous assumptions in our tax example may be slightly less unrealistic: Many governments have vast information on taxpayers, and unlike private companies, governments usually do not face legal restrictions that would in principle prevent them from individualizing tax rates to some extent. Nevertheless, an individual's tax elasticity is not directly observable, and even if the government could estimate it, setting individualized tax rates would likely run counter to common notions of horizontal tax fairness.

3.2 Complexity as second-degree price discrimination

While first-degree price discrimination functions primarily as an illustrative device in theoretical models, second-degree price discrimination is very often seen in practice. The core idea of second-degree price discrimination is to offer different "menus" such that customers will self-select into different options based on their price elasticity and other characteristics, ultimately allowing the monopolist to extract a higher revenue. This can be done in multiple ways, for instance through bulk discounts, coupons, or limited time promotions. At a very high level, these pricing practices create a trade-off between price and convenience: If you are willing to do the requisite planning to buy in bulk, or to cut out coupons to bring to the store, you are rewarded with a lower overall price. Ngwe (2017) finds that the location choices of outlet stores works by a similar logic: They are often

[3] This is of course assuming that everything else is kept equal. If those with lower elasticities have higher welfare weights, it would be possible that more elastic taxpayers have a higher optimal tax rate overall.

placed in comparatively remote locations to avoid cannibalizing sales at the "flagship" stores too much, while simultaneously allowing for price-sensitive customers to travel the extra distance and save money. Essentially, these strategies work by creating an *ordeal* that requires some effort to overcome in order to access the lower price.

We can think of a complex tax system as functioning in a similar way. Those who are highly sensitive to taxes will spend time and effort on understanding the tax system – or pay someone else to do this work for them – and can thus lower their total tax rate, for example by taking advantage of obscure deductions, optimizing the balance between wages and dividends, optimally timing realizations of capital gains and losses, or setting up holding companies or other legal structures. A complex tax system allows the government to extract as much tax revenue as possible from those less sensitive to taxes, while creating an ordeal that requires some wasteful effort to overcome, but which lets the more elastic taxpayers reduce their tax burden. It may even be the case that this complexity can lead to efficiency gains, if the cost of learning about the tax system is outweighed by the reduced distortionary effects of taxation for the more elastic taxpayers. To answer this question, we formulate a mathematical model of tax complexity.

We now write the taxpayer's problem as:

$$\begin{aligned} \max_{C,L,S} U_i(C,L) &= C - D_i(L) \\ \text{s.t. } C &= w_i L \cdot N(S) - Q(S) \\ L, S &\geq 0, \end{aligned}$$

where we think of S as time spent studying the tax system. $D_i(L)$ can be interpreted as the disutility of labor, $Q(S)$ as the disutility or cost of acquiring tax knowledge, and $N(S)$ as the net-of-tax rate associated with knowledge level S.[4] We assume that $N(S)$ has a weakly negative first derivative with respect to S everywhere, as knowledge about the tax system may help the taxpayer find strategies to reduce the overall tax burden. The government's problem will then be to decide on the complexity of the tax system by choosing the shape of the function $N(S)$.

The solution to the problem is pinned down by the first order conditions. In the case where

[4]Since we are considering an economy consisting of multiple taxpayers with differing elasticities, a completely correct notation would include taxpayer-specific subscripts on the choice variables: L_i, C_i, S_i. However, we omit these for readability whenever it does not lead to ambiguities.

$S \geq 0$, these can easily be found to be:

$$wN(S) = D'(L)$$
$$wLN'(S) = Q'(S). \tag{3.3}$$

Without making further assumptions on functional forms, we cannot analytically derive a closed-form solution. However, we can characterize how introducing a marginal amount of complexity into the model will affect outcomes. To do this, we make the following definitions:

Definition 1. *We define a* simple problem *under this model as a consumer utility maximization problem with no complexity, i.e. where $N(S) = 1 - \tau_0$ is a simple net-of-tax rate that does not depend on S. Trivially in the simple problem, any optimal solution will have $S = 0$ since any positive S will incur a cost and yield no tax benefits.*

Definition 2. *Given a simple problem, we define a corresponding* marginally complex problem *as one which introduces a small degree of complexity, replacing $N(S) = 1 - \tau_0$ with $N_\delta(S)$ such that*

- $N_\delta(0) = 1 - \tau_0$
- $N'_\delta(S) > 0$ *for* $S < \delta$
- $N'_\delta(S) = 0$ *for* $S \geq \delta$

for small δ.

This definition helps us to analyze how complexity will affect decisions at the margin without having to worry about discrete responses, since only the initial marginal amount of tax knowledge will reduce the tax burden. We are now ready to find necessary and sufficient conditions for marginal complexity to generate increased tax revenue.

Proposition 1. *Consider a taxpayer i with wage w_i, elasticity of taxable income ε_i and labor output L_i^0 at the optimal solution to the simple problem. If we introduce marginal complexity to the model, i.e. replace the simple net-of-tax rate $1 - \tau_0$ with $N_\delta(S)$ that satisfies $N_\delta(0) = 1 - \tau_0$, $N'_\delta(S) > 0$ for $S < \delta$ and $N'_\delta(S) = 0$ for $S \geq \delta$ with $\delta > 0$ sufficiently small, then tax revenues generated by*

taxpayer i will increase relative to the simple problem if and only if

$$w_i L_i^0 N_\delta'(0) > Q'(0)$$

and

$$\varepsilon_i > \frac{1-\tau_0}{\tau_0}.$$

Proof. We first note that the first condition, $w_i L_i^0 N_\delta'(0) > Q'(0)$, is a necessary condition for i to optimally choose a positive level of tax knowledge. If this does not hold, then for a sufficiently small δ, i will simply prefer a corner solution with $S = 0$, and therefore tax revenue will be unchanged relative to the simple problem.

Now, to see how optimal labor supply will change as a function of tax knowledge, consider the first-order condition associated with the labor variable around $S = 0$:

$$D'(L) = w_i N_\delta(S)$$

Notice that the above expression will hold with equality both at $L = L_i^0$ and $S = 0$ for the optimal solution to the simple problem, and at any (L, S) that solves the complex problem. We can therefore find the derivative of labor as a function of tax knowledge around L_i^0 by taking the total derivative of this expression with respect to S around $S = 0$ and rearranging. This yields:

$$\left.\frac{\mathrm{d}L}{\mathrm{d}S}\right|_{S=0} = \frac{w_i N_\delta'(0)}{D''\left(L_i^0\right)},$$

which we notice is positive whenever D is strictly convex. Now we turn to tax revenue. This can be written as $R = w_i L_i \left(1 - N_\delta(S)\right)$, which we can again totally differentiate around $S = 0$ to get

$$\mathrm{d}R = w_i \left(1 - N_\delta\left(0\right)\right) \mathrm{d}L - N_\delta'\left(0\right) w_i L_i^0 \mathrm{d}S.$$

Rearranging yields

$$\begin{aligned}
\left.\frac{\mathrm{d}R}{\mathrm{d}S}\right|_{S=0} &= w_i\left(1-N_\delta\left(0\right)\right)\left.\frac{\mathrm{d}L}{\mathrm{d}S}\right|_{S=0} - N_\delta'\left(0\right)w_iL_i^0 \\
&= w_i\tau_0\frac{w_iN_\delta'(0)}{D''\left(L_i^0\right)} - N_\delta'\left(0\right)w_iL_i^0 \\
&= w_iL_i^0\tau_0\frac{w_iN_\delta'(0)}{D''\left(L_i^0\right)L_i^0} - N_\delta'\left(0\right)w_iL_i^0 \\
&= w_iL_i^0N_\delta'(0)\left[\frac{\tau_0w_i}{D''\left(L_i^0\right)L_i^0}-1\right]. \qquad (3.4)
\end{aligned}$$

Introducing a marginal amount of complexity to the simple model results in an increase in tax revenue from person i if and only if the expression above is greater than zero. To rewrite this in terms of the ETI, notice from the first-order condition to the simple problem, we have that

$$L_i^0 = \left[D'\right]^{-1}\left(w_i\left(1-\tau_0\right)\right).$$

Letting $g = D'$ and $n = 1-\tau$ for ease of notation, and taking the derivative of this with respect to the net-of-tax rate n, we get:

$$\begin{aligned}
\left.\frac{\mathrm{d}L_i}{\mathrm{d}n}\right|_{\tau=\tau_0} &= w_i\left[g^{-1}\right]'\left(w_i\left(1-\tau_0\right)\right) \\
&= \frac{w_i}{g'\left(g^{-1}\left(w_i\left(1-\tau_0\right)\right)\right)} \\
&= \frac{w_i}{D''\left(L_i^0\right)}.
\end{aligned}$$

We can therefore write the elasticity of taxable income at L_i^0 as

$$\begin{aligned}
\varepsilon_i &= \frac{\mathrm{d}(w_iL_i)}{\mathrm{d}n}\cdot\frac{1-\tau_0}{w_iL_i^0} \\
&= \frac{\mathrm{d}L_i}{\mathrm{d}n}\cdot\frac{1-\tau_0}{L_i^0} \\
&= \frac{w_i\left(1-\tau_0\right)}{D''\left(L_i^0\right)L_i^0}.
\end{aligned}$$

Using this, we can now write (3.4) as

$$\left.\frac{\mathrm{d}R}{\mathrm{d}S}\right|_{S=0} = w_i L_i^0 N_\delta'(0)\left[\frac{\tau_0}{1-\tau_0}\varepsilon_i - 1\right],$$

which is positive if and only if $\varepsilon_i > \frac{1-\tau_0}{\tau_0}$. □

In the above result, note that the first condition, $w_i L_i^0 N_\delta'(0) > Q'(0)$, is simply a participation constraint which states that the benefit of the first marginal unit of tax knowledge – income times the slope of the net-of-tax function – must exceed the cost. If this condition isn't satisfied, the individual will optimally choose no tax knowledge, and the tax revenue derived from them will trivially be the same as in the simple problem. The second condition, $\varepsilon_i > \frac{1-\tau_0}{\tau_0}$, can be rewritten as $\tau_0 > \frac{1}{1+\varepsilon_i}$. This simply states that the tax rate, τ_0, is above the revenue-maximizing rate for this particular individual. The result is therefore very intuitive: Introducing complexity allows the taxpayer to obtain a lower tax rate, and beyond the peak of the Laffer curve, this lower tax rate is associated with higher revenues. Note that even though tax knowledge is costly, the increased tax revenue does not come at an overall efficiency cost, as the taxpayer's opportunity set has only increased. Remaining at the optimum from the simple model is still an option, and the taxpayer must therefore be made better off by choosing to obtain tax knowledge.

Corollary 1. *Starting from an economy with a single flat tax rate, introducing tax complexity can increase the maximum tax revenue that the government can collect if there exists some income threshold $\tilde{I}$ such that those with incomes above $\tilde{I}$ have a larger aggregate ETI than those with incomes below $\tilde{I}$.*

Proof. Let $\bar{\varepsilon}$ be the aggregate ETI of the full population. It is a standard result from the optimal tax literature that the revenue-maximizing flat tax rate in this economy will be $\tau^* = \frac{1}{1+\bar{\varepsilon}}$. Let H denote the set of taxpayers with incomes above $\tilde{I}$. The government can design a complex tax system with functions N_δ and Q determined such that $\tilde{I} = \frac{Q(0)}{N_\delta'(0)}$. From the proof of the previous proposition, we know this ensures that for δ sufficiently small, the set of those who will invest in tax knowledge will approach H. Let ε_H be the aggregate ETI for the group of taxpayers in H. By the definition of the income threshold $\tilde{I}$, the aggregate ETI of those who invest in tax knowledge will be higher than for those who do not, and therefore also higher than the ETI for the total population,

so thus $\varepsilon_H > \bar{\varepsilon}$. We show that overall tax revenue is increased by extending the argument from the previous proof to an aggregate population.

Let R_H denote aggregate tax revenue collected for the group H, and let I_H denote their aggregate income. As before using n for the net-of-tax rate, notice that we can write the aggregate elasticity for the high-income group as

$$\begin{aligned}\varepsilon_H &= \frac{\mathrm{d}I_H}{\mathrm{d}n} \cdot \frac{n}{I_H} \\ &= \sum_{i \in H} w_i \frac{\mathrm{d}L_i}{\mathrm{d}n} \cdot \frac{n}{I_H} \\ &= \sum_{i \in H} \frac{w_i^2}{D''\left(L_i^0\right)} \cdot \frac{n}{I_H}\end{aligned}$$

Using our derivation from the previous proof, we can write the change in aggregate tax revenue when we allow for a marginal amount of tax complexity as:

$$\begin{aligned}\left.\frac{\mathrm{d}R_H}{\mathrm{d}S}\right|_{S=0} &= \sum_{i \in H} w_i \left(1 - N_\delta\left(0\right)\right) \left.\frac{\mathrm{d}L_i}{\mathrm{d}S}\right|_{S=0} - \sum_{i \in H} N_\delta'\left(0\right) w_i L_i^0 \\ &= \sum_{i \in H} w_i \tau^* \frac{w_i N_\delta'(0)}{D''\left(L_i^0\right)} - \sum_{i \in H} N_\delta'\left(0\right) w_i L_i^0 \\ &= N_\delta'\left(0\right) \left[\tau^* \sum_{i \in H} \frac{w_i^2}{D''\left(L_i^0\right)} - I_H\right] \\ &= N_\delta'\left(0\right) \left[\tau^* \varepsilon_H \cdot \frac{I_H}{1 - \tau^*} - I_H\right] \\ &= N_\delta'\left(0\right) I_H \left[\varepsilon_H \cdot \frac{\tau^*}{1 - \tau^*} - 1\right],\end{aligned}$$

which is positive whenever the term in brackets is greater than zero. But this follows from the definition of τ^* from the beginning of the proof, since

$$\varepsilon_H \cdot \frac{\tau^*}{1 - \tau^*} > \bar{\varepsilon} \cdot \frac{\tau^*}{1 - \tau^*} = \bar{\varepsilon} \cdot \frac{1}{\bar{\varepsilon}} = 1.$$

□

Just like in the previous result, increased tax revenue comes at no efficiency cost since introducing complexity makes the opportunity set larger for everyone, so no one can end up worse off. Thus, in

cases where the result applies, complexity actually leads to an efficiency improvement by allowing those who are on the downward-sloping side of the Laffer curve to gain tax knowledge and lower their effective tax rate while increasing their labor supply enough to offset the mechanical loss in tax revenue. If the surplus tax revenue is spent on e.g. a small lump-sum transfer to everyone in the economy, the introduction of complexity will make everyone strictly better off.

Note that the previous result assumed a single, flat tax rate. However, the same logic and arguments would apply in the case of a progressive tax rate: Complexity can generate increased tax revenue whenever there is an income threshold above which the aggregate ETI is large enough that this subpopulation are on the downward-sloping part of the Laffer curve.

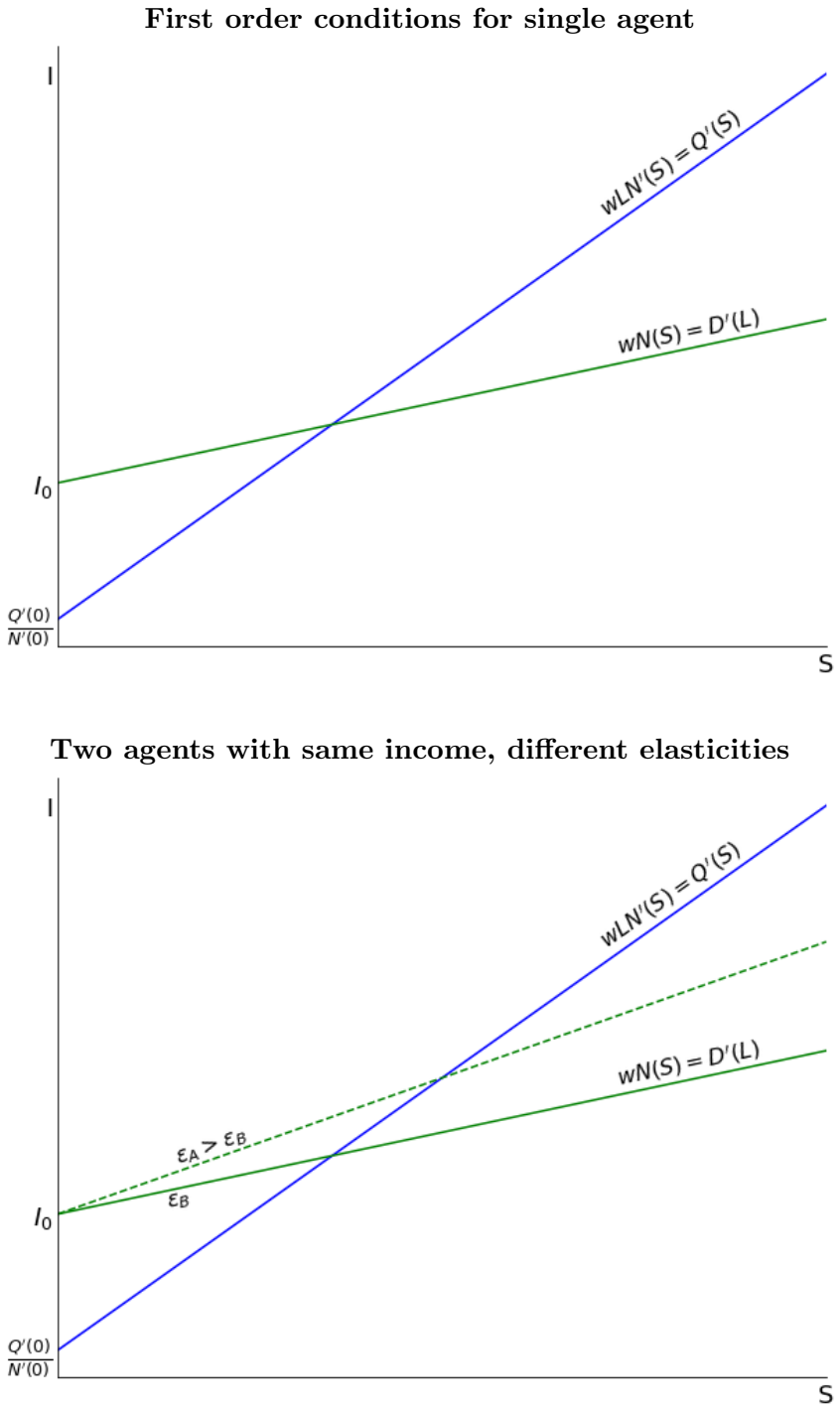
The results we have seen so far have relied on more elastic taxpayers investing in tax knowledge while less elastic taxpayers choose not to. However, this has been driven by our assumptions on differences in income: For the initial unit of tax knowledge, elasticity does not even factor into the decision to invest or not, as this is fully determined by the condition $w_i L_i^0 N_\delta'(0) > Q'(0)$. We might therefore worry about what happens outside of the context of marginal complexity, when we allow for larger investments in tax knowledge. The next result shows there is indeed a positive link between elasticity and investment in tax knowledge once we move beyond the marginal case.

Proposition 2. *Consider two taxpayers, A and B, with identical income $I_0 > \frac{Q'(0)}{N'(0)}$ in the simple equilibrium. If A has a higher ETI than B across the entire income spectrum, i.e. $\varepsilon_A > \varepsilon_B$, then under complexity, A will invest in more tax knowledge than B, have a higher overall income and a lower effective tax rate than B. If the two taxpayers have identical elasticities across the income spectrum, they will have the same income and tax knowledge under complexity.*

Proof. In short, this proposition states that if we keep income fixed in the simple equilibrium, investments in tax knowledge will be a monotonous function of a person's elasticity of taxable income. However, note that it's not enough for the ETI at the initial simple equilibrium to be the same, as elasticity may change over the income distribution moving towards the complex equilibrium.

The proof can most easily be understood through a graphical explanation. The first-order conditions for the complex problem in (3.3) can be represented as curves in (I, S)-space as shown in the first panel of Figure 3.1. The blue line represents the first-order constraint for tax knowledge,

Figure 3.1: Graphical representation of first-order conditions in (I, S)-space



Notes: This figure shows a graphical representation of the first-order conditions for the agent's optimization problem under tax complexity. The bottom panel illustrates that if two taxpayers have the same initial income but different tax elasticities, the more elstic taxpayer will have a steeper increase in labor earnings as a function of tax knowledge. See the proof of Proposition 2 for further details.

and the green one for labor. Note that while the choice variables in the problem as originally stated are L and S, since $I = wL$ and w is assumed fixed for any given taxpayer, we can also think of I as a choice variable directly. Income in the initial, simple equilibrium is given by I_0 – the condition that this is greater than $\frac{Q'(0)}{N'(0)}$ ensures that at least some investment in S is beneficial once complexity is introduced. The equilibrium under complexity will correspond to the point where the two lines meet.

Consider taxpayers A and B. The proof is done if we can show that the intersection of the two lines will occur at a higher (I, S)-pair for A than for B, as illustrated in the second panel of Figure 3.1. First, notice that the blue line (the constraint on S) will be identical for the two taxpayers, as it only depends on income and on the functions N and Q, which are assumed to be independent of the identity of the taxpayer. Also notice that the intercept of the green line is simply income in the simple equilibrium, which is I_0 for both taxpayers. Therefore, the proof is complete if we can show that the slope of the green line will be steeper for the more elastic taxpayer. We can see this by totally differentiating the constraint associated with the green line, which yields

$$\frac{\mathrm{d}L_i}{\mathrm{d}S_i} = \frac{w_i N'(S_i)}{D''(L_i)}, \tag{3.5}$$

as we showed for the special case of $S_i = 0$ in the proof of Proposition 1. Now, recall that the ETI of taxpayer i is defined as $\varepsilon = \frac{\mathrm{d}I_i}{\mathrm{d}n} \cdot \frac{n}{I_i}$ where n is the net-of-tax rate, and where we evaluate the response to an *exogenous* change in the net-of-tax rate – that is, at a given effective net-of-tax rate $N(S_i)$, the ETI is the percentage-wise increase in income that i would earn if $N(S_i)$ were exogenously lowered by 1 percent. Therefore, we can find an expression for the elasticity by differentiating the first-order condition for the *simple* problem at a labor supply level of L_i and net-of-tax rate $N(S_i)$. Recall that in the simple problem, labor supply is pinned down by the first order condition $L_i = [D']^{-1}(w_i n)$, so just as in the proof of Proposition 1, we can express elasticity

as

$$\begin{aligned}\varepsilon_i &= \frac{\mathrm{d}(w_i L_i)}{\mathrm{d}n} \cdot \frac{N(S_i)}{w_i L_i} \\ &= \frac{\mathrm{d}L_i}{\mathrm{d}n} \cdot \frac{N(S_i)}{L_i} \\ &= \frac{w_i N(S_i)}{D''(L_i) L_i},\end{aligned}$$

which corresponds to $\frac{w_i}{D''(L_i)} = \frac{\varepsilon_i L_i}{N(S_i)}$. Insert this in equation (3.5) and use the identity $I_i = w_i L_i$ to find:

$$\frac{\mathrm{d}I_i}{\mathrm{d}S_i} = w_i \frac{\mathrm{d}L_i}{\mathrm{d}S_i} = w_i \frac{\varepsilon_i L_i N'(S_i)}{N(S_i)} = \varepsilon_i I_i \frac{N'(S_i)}{N(S_i)}.$$

This completes the proof: Since both taxpayers are assumed to have the same income at $S_i = 0$, if $\varepsilon_A > \varepsilon_B$, we do indeed see that $\frac{\mathrm{d}I_A}{\mathrm{d}S} = \varepsilon_A I_A \frac{N'(S)}{N(S)} > \varepsilon_B I_B \frac{N'(S)}{N(S)} = \frac{\mathrm{d}I_B}{\mathrm{d}S}$ at every point along the curve, so the constraint for taxpayer A is indeed steeper than that of B as illustrated in the second panel of Figure 3.1, leading to an equilibrium further up along the blue curve, which involves a higher income, higher S, and a lower effective tax rate. Conversely, if $\varepsilon_A = \varepsilon_B$, the two curves will have the exact same slope at every point, and will hence coincide such that A and B end up with the same equilibrium income and effective tax rate under complexity. □

3.3 Simulation

The theoretical results in this section have proved that under some circumstances, tax complexity can lead to increased tax revenue without sacrificing efficiency. However, those propositions assume marginal complexity and only provide sufficient, but not necessary conditions for the results to hold. Therefore, there is scope to explore in further detail what the model implies under different circumstances – for example, when we consider tax complexity beyond the marginal unit, or when we apply complexity to a tax system at an initial tax rate which is lower than the revenue-maximizing one. This subsection investigates these questions by simulating particular specifications of the model in Python.

To parameterize the model, I assume the following class of taxpayer utility functions:

$$U_i(C, L, S) = C - \frac{e^{-\alpha_i/\varepsilon_i}}{1 + 1/\varepsilon_i} L^{1+\frac{1}{\varepsilon_i}} - 0.05S$$

I model an economy with two types, a *high type* with ETI $\varepsilon_h = 1.25$ and a *low type* with $\varepsilon_l = 0.75$. I assume each of these types makes up half of the population. In this version of the model, I have normalized wages to 1, such that incomes are equal to labor supply. The parameter α_i will shift labor output independently of the elasticity – we initially set $\alpha_h = 1.5$ and $\alpha_l = 1$, which ensures that the high type has higher labor supply and income than the low type, in line with the assumptions in our model. I analyze the difference between a simple model with a fixed tax rate, and a complex model that allows taxpayers to reduce their effective tax rate. In the first two cases I examine, I let the complexity function be exponential:

$$N(S) = 1 - \tau \left(0.2e^{-0.3S} + 0.8\right) \tag{3.6}$$

This function generates a net-of-tax rate of $1 - \tau$ if $S = 0$, and will decline asymptotically towards $1 - 0.8\tau$ as S grows large, so that the effective tax rate can be reduced by at most 20% of its initial value.

The first case I examine illustrates the general idea behind Corollary 1. I set a baseline tax rate of $\tau = 0.495$, which is the flat tax rate that maximizes government revenue in this case.[5] The results are shown in the first three columns of Table 3.1, labelled "Model A". The first column shows tax revenue, labor output and utility by type under the simple model, i.e. at a flat tax rate of $\tau = 0.495$ with no possibility to reduce this through tax knowledge. The second column shows the corresponding values after we allow for complexity as described by equation (3.6), and the third column shows the difference between these two cases. We see that allowing for complexity in this case will cause the high type to invest in $S_h = 0.5369$ units of tax knowledge, which causes their effective tax rate to fall by roughly 1.5 percentage points, while the low type will not invest in tax knowledge. Consequently, the high type ends up with higher labor output and higher utility overall.[6] Tax revenues also increase, confirming the result from Corollary 1 in a simulation that

[5]The aggregate ETI $\bar{\varepsilon}$ will be the income-weighted average of the type-specific ETIs, and the revenue-maximizing tax rate can then be calculated as $\tau^* = \frac{1}{1+\bar{\varepsilon}}$.

[6]The fact that utility increases is true by construction, as allowing for complexity simply expands everyone's

Table 3.1: Simulation results

	Model A			Model B			Model C		
	Simple	Complex	Diff.	Simple	Complex	Diff.	Simple	Complex	Diff.
Tax revenue	0.8752	0.8779	0.0027	0.8021	0.7989	-0.0032	0.8081	0.8091	0.0009
S_l		0			0			0.5	
S_h		0.5369			0.359			1	
τ_l	0.495	0.495	0	0.35	0.35	0	0.5	0.4838	0.0162
τ_h	0.495	0.4803	-0.0147	0.35	0.3429	-0.0071	0.5	0.4690	0.0310
L_l	1.6284	1.6284	0	1.9678	1.9678	0	1.6163	1.6555	0.0392
L_h	1.9079	1.9777	0.0698	2.6157	2.6517	0.0360	1.6163	1.7425	0.1262
Utility, low	0.4699	0.4699	0	0.7309	0.7309	0	0.4618	0.4634	0.0016
Utility, high	0.4282	0.4299	0.0018	0.7556	0.7565	0.0009	0.3592	0.3612	0.0021

Notes: This table shows results of simulations of my theoretical model done in Python. S_i represents tax knowledge, τ_i the effective tax rate and L_i labor output for group $i = h, l$. In all cases, the simulations compare the difference between a simple model with a fixed tax rate and a corresponding model that allows for complexity. The first three columns labelled "Model A" illustrate a case where complexity allows for increased tax revenues and increases efficiency starting from the revenue-maximizing flat tax rate. "Model B" shows the same model starting from a lower tax rate, where complexity no longer causes a tax revenue increase. "Model C" shows a case where two individuals with initially identical incomes but different elasticities will self-separate as a consequence of complexity, once again leading to increased tax revenue and welfare. See the main text for more details.

allows for tax knowledge to increase beyond just a marginal unit.

The next three columns, labelled "Model B", repeat the same calculations, but starting from a baseline tax rate of 0.35. This example illustrates the effects of introducing complexity in an environment where the baseline tax rate is below the revenue-maximizing rate. The low type still chooses not to invest in tax knowledge, whereas the high type will invest, but by less than in the previous example. Tax revenues now decrease once complexity is introduced – the reduced distortion on the high-type taxpayer is not enough to make up for the mechanical reduction in revenues. However, we do see that labor output increases for the high type. In fact, if the government has a goal of increasing labor output, complexity remains a more efficient way to achieve this. To see this, note that tax revenues in the complex version of Model B are 0.7989, and the average labor output between the two types is 2.3097. A simple model with a tax rate of 0.347 would achieve the same tax revenue, but average labor output in that case is only 2.3030. This illustrates another case for complexity beyond just revenue maximization: Total economic output is a metric that governments are often interested in increasing, and tax complexity may be a tool that helps achieve this aim.

The goal of Model C is to show that the assumption of a positive correlation between income and elasticity is not a necessary condition for tax complexity to lead to increased tax revenue. In

opportunity set. No one can be made worse off, as they can still choose not to invest in tax knowledge.

this example, I show that starting from the same initial income and with both types investing in positive amounts of tax knowledge, complexity can still generate higher tax revenues and be efficiency-improving. I set the initial tax rate to 0.5 and change the value of α_h to 1.346574, which will cause both types to have the same initial labor supply and income under the simple tax system.[7] I also change the complexity function slightly. Instead of (3.6), I now use the following piecewise linear function:

$$N(S) = \begin{cases} 1 - \tau\left(1 - 0.065S\right) & \text{if } S \leq \frac{1}{2} \\ 1 - \tau\left(1 - \frac{0.065}{2} - 0.059\left(S - \frac{1}{2}\right)\right) & \text{if } \frac{1}{2} < S \leq 1 \\ 1 - \tau\left(1 - \frac{0.065}{2} - \frac{0.059}{2}\right) & \text{if } S > 1 \end{cases} \cdot \tag{3.7}$$

The outcome of this model is shown in the last three columns of Table **??**. We see that low types will invest in tax knowledge up to the first kink point, where $S = \frac{1}{2}$, whereas high types will locate at the second kink point with $S = 1$. Both types now gain utility in the complex model over the simple one, while overall tax revenue and labor output will also increase. This result illustrates that complexity can also generate horizontal tax inequality: Two individuals starting at the same initial income may self-separate according to their elasticity and achieve different effective tax rates under complexity.

3.4 Discussion and Conclusion

In this article, I show that tax complexity can indeed serve a purpose, as it may allow the government to extract a higher tax revenue than with a simple flat tax rate, assuming that elasticities are increasing with income. This result would apply in the exact same way if we had instead assumed that the baseline tax is progressive, as is the case in most countries. These findings give us an indication of why a complex tax system might serve a useful function for a government: It allows for a progressive tax schedule at a surface level, giving the appearance of a tax system that redistributes income, while at the same time providing deductions and other loopholes such that top earners in practice pay less than the headline rate. It can thus to some extent bridge the gap between two

[7] In fact, the high-elasticity type will have a very slightly lower initial income, so that income and initial elasticity are actually negatively correlated in this example.

principles of taxation that are at odds with each other: General notions of fairness and public opinion usually favor progressive taxes (Tolbert, Witko and Wolbers, 2019; Barnes, 2022), and this is also often assumed for papers in the tax literature through a social welfare function that places less weight on high earners (Saez and Stantcheva, 2016; Piketty and Saez, 2013). However, several results in the optimal tax literature have shown that distortions from taxation are higher at the top, and that in the extreme, the marginal tax rate should be 0 for the highest individual earner (Mirrlees, 1971; Seade, 1977). Tax complexity may provide a more covert way to reduce distortions at the top, delivering a tax system which is less redistributional than it seems and therefore perhaps less likely to be unpopular with voters.

My last result shows that complexity does indeed cause taxpayers to self-separate according to elasticity, such that starting from identical incomes under the simple model, a more elastic taxpayer will obtain a lower effective tax rate. This highlights another possible purpose of tax complexity, namely to allow for horizontal inequity in taxation. Striving for horizontal equity in taxation is a widely accepted principle of fairness (Musgrave, 1990), but if there are heterogeneous tax elasticities at a given income level, horizontal tax inequality may improve efficiency and lead to increased tax revenues, as we demonstrated in the previous subsection. Again, this is a case where complexity can lead to unequal treatment that may be desirable to the government, without this being directly codified in the tax law.

As with any economic model, I have made several simplifying assumptions in order to derive meaningful results. I discuss some of these assumptions here, and how they relate to instances of tax complexity that we might find in the real world.

My most unrealistic assumption is perhaps that the model implies perfect knowledge, in some sense. While taxpayers do not start with full knowledge of the tax system, the model assumes that they "know what they don't know" – i.e., that they know precisely which effective tax rate they would end up with if they had invested in S units of tax knowledge. A more realistic model might assume that taxpayers search over some policy space to find tax provisions which may lead to a lower effective tax rate to a smaller or larger degree. Some taxpayers may end up making sub-optimal decisions in the tax knowledge space – they might not invest in knowledge even though the benefit from doing so would have outweighed the cost, or they might conversely overestimate how much they can reduce their tax burden, and invest too much. While a model incorporating

these nuances might be more realistic, it would also lead to many additional questions: If people don't know how much they can lower their effective tax rate by investing in tax knowledge, they need to have some belief about it. If this belief isn't accurate, then what causes the discrepancy, and how should this be modelled? Do taxpayers learn their true tax rate before they make their decision of how much labor to supply? Other papers on tax complexity, such as Craig and Slemrod (2022) do include these richer models of beliefs, but models of this sort are less useful in delivering predictions on e.g. tax revenue unless they impose other, more specific assumptions.

I believe that the assumption of "full knowledge", while not quite realistic, is useful for tractability of the model and in fact makes its predictions more conservative. Allowing for inaccurate beliefs would provide another channel through which complexity could lead to increased government revenue. It is likely that wealthier individuals would on average have a network that gives them access to better high-level information on what deductions are available, what strategies can be used to minimize capital gains taxes, and so on (Alstadsæter, Kopczuk and Telle, 2019). My results show that the government can increase tax revenues through complexity simply as a result of heterogeneous elasticities, and not because of inaccurate beliefs which would tend to skew the knowledge gap between the rich and poor even further.

Another simplifying assumption in the model is that tax knowledge uniformly lowers the tax rate independently of a taxpayer's identity or initial income. This is clearly not the case in practice: The complexities in the tax code that can be used by low-income wage earners to lower their tax burden are very different than the ones employed by wealthy business owners. This means that the government has much more flexibility to differentiate tax complexity according to individual circumstances than our model implies, so this is again a case where my assumption is relatively conservative.

Finally, I assumed that the cost of tax knowledge, $Q(S)$, does not vary by individual. This assumption is debatable: We might assume that more elastic individuals would also have a higher cost of acquiring tax knowledge. If we imagine a model where a fixed time endowment is split into leisure, productive labor, and time spent studying the tax system, we could think of the latter as simply another form of labor which is governed by the same elasticity parameter. In reality though, the time spent on learning about the tax system is minuscule for most people compared to the time spent on normal labor, and it isn't clear that adding these two quantities up into a

single "total labor" variable would be a good modelling choice in that case. And on the other hand, if we work under the assumption that income and elasticity are positively correlated – for instance, if more elastic people have higher productivity and wages – it might not be unreasonable to assume that this higher productivity also applies to the ability to understand the tax system. In that case, each "unit" of tax knowledge would be cheaper for the more productive taxpayers to acquire. Ultimately, I have chosen not to assume that one of these opposite effects is stronger than the other, but to keep the cost of tax knowledge independent of the identity of the taxpayer. This would be consistent with a model where tax knowledge is a good that can be bought for a particular price, for instance by hiring an accountant or tax advisor.

This article has shown that within a simple theoretical model, allowing for tax complexity can lead to increased tax revenue and welfare by facilitating self-selection according to type in a very similar fashion to second-degree price discrimination. The main results focus on the consequences of introducing marginal complexity starting from a "simple" tax system, since this is a case where it is possible to derive general results with few assumption about the form of the taxpayer's utility function. However, the simulation shows that complexity can also lead to efficiency and tax revenue improvements when discrete responses are involved. Future work might explore under which conditions this is true. Another avenue for future research may be to define ways to enumerate the complexity of a given tax system, and to empirically measure whether simplifications or complexifications of the tax system lead to changes in the distribution of effective tax rates.

Chapter 4

Complexity and Tax Overpayment: Evidence from Norway[1]

How do complex tax systems affect individual decision-making? Does complexity lead people to make mistakes that leave them worse off? These questions can be difficult to answer, because tax systems that are complex to the taxpayer will also often be similarly complex to the economists studying them. While it may be possible to quantify how large the tax liability will be in any given situation, this does not tell us much about any possible mistakes that are made if the taxpayer's choice also involves a "real" response such as changing one's labor supply, purchasing a different product, shopping in another jurisdiction with a different tax rate, or selling a certain asset rather than holding on to it. These decisions all have consequences for tax liability, but also for real outcomes experienced by the individual in one way or another. If we observe a person driving to another US state with a higher sales tax rate and making a purchase there, when a lower tax would have been available closer to their home, we might assume the person is making a mistake due to an incomplete understanding of the tax system, but we can never be sure – they might enjoy the drive and/or the shopping experience in the other state more than at home.

Several past empirical studies of tax complexity have found at a population level that some

[1]This chapter was co-authored with Julie Brun Bjørkheim. We thank Annette Alstadsæter, Ron Davies, Henrik Kleven, John Sturm, Damian Vergara, Owen Zidar and seminar participants at Princeton and the DAEiNA annual congress for helpful comments. We are grateful for financial support for this research from the Skatteforsk Centre for Tax Research at the Norwegian University of Life Sciences. Financial support from The Research Council of Norway (Grant number 341289) is gratefully acknowledged.

people must be behaving suboptimally, but cannot determine which specific taxpayers are making mistakes. This is true for e.g. Benzarti (2020), who shows that there is "missing mass" in the distribution of itemized deductions just above the standard deduction amount in the US, indicating that some filers who would have benefited from itemizing choose not to, but we cannot observe which individuals are making this choice, and whether it is a mistake or simply a conscious decision to avoid the administrative burden of itemizing. Chetty, Looney and Kroft (2009) show that customers on average underreact to sales taxes when they are not included in the posted price, but it is impossible to tell if any individual consumer who doesn't see the tax-inclusive price would have made the purchase if they had seen it.

In this paper, we improve on those past studies by identifying a particular instance of tax complexity in Norway under which we can identify suboptimal behavior at the individual level with very little ambiguity. Our study utilizes de-identified register data from Statistics Norway on full owners of Norwegian stock-based companies (*aksjeselskap*), i.e. individuals who own 100% of shares in their company. Stock-based companies are separate tax entities, which means that the owners encur a tax liability at the time when money is transferred from the firm to the owner. This can be done in two different forms, which only differ in their tax treatment: As wages, or as dividends. Taking as given the full pre-tax amount that owners transfer, we can therefore determine if the owners split the amount between wages and dividends in the way that minimizes their tax liability, and if not, how large their tax overpayment ends up being.

Our results reveal that many taxpayers behave suboptimally, and that as a consequence, some pay significantly higher effective tax rates than they would have with optimal behavior. The size of tax overpayment differs along demographic dimensions, with women overpaying slightly more than men and immigrants more than native-born Norwegians on average. We also find that complexity is very regressive: Wealthy taxpayers overpay significantly less than others. The strongest predictor of overpayment that we can see in the data is the amount spent within firms on accountants and financial advisors: Those with high spending make much fewer and smaller mistakes. While we cannot know if these services specifically include tax advice, we think of this spending measure as a proxy for acquisition of tax knowledge, so it is sensible that it would be highly predictive.

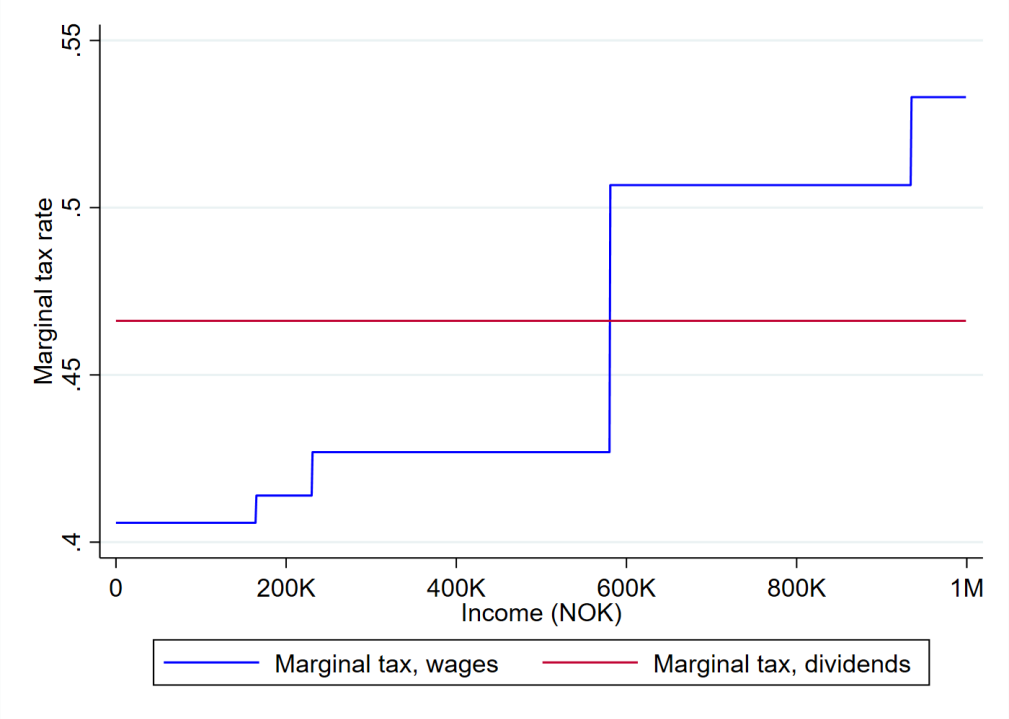
Finally, to test implications of the theoretical model covered in the previous chapter, we investigate if higher tax overpayment is indeed associated with lower elasticity as the theoretical

results predict. We study the behavioral effect of an earlier tax reform on our study population, and estimate separate elasticities of taxable income for groups based on the magnitude of their later tax overpayment. Since our measure of overpayment relies on income shifting between tax bases for business owners, this could introduce a problem when attempting to estimate elasticity based on this measure: High tax knowledge can lead to more precise optimization, which might in itself cause estimated tax elasticity to be higher, and so any test of a correlation between overpayment and elasticity could be tautological. To alleviate this concern, we limit our sample for this test to those taxpayers who were not yet self-employed at the time of the earlier tax reform. We do find indicative results that higher tax overpayment is associated with lower elasticity, in line with theoretical predictions. Further, we find that elasticity estimates are smaller for women, immigrants, people with lower incomes and those below the threshold for the wealth tax. Given that these groups tend to overpay more, this aligns with our theoretical predictions, though we note that many other factors in addition to elasticity differences could be at play.

Our paper relates to previous empirical studies of tax complexity, misinformation, and tax overpayment. Aghion et al. (2017) estimate how much taxpayers value simplicity by studying a discrete choice between different tax regimes for small businesses in France, which differ in both tax incentives and simplicity. This is one of the few studies besides ours to also quantify tax misoptimization at an individual level – some taxpayers choose a tax regime which is both more complex and leads to a higher tax liability. Abeler and Jäger (2015) demonstrate misoptimization in an experimental study mimicking a complex tax system. This study also highlights the role of cognitive ability in tax optimization. The findings of Bastani and Waldenström (2021) imply that tax elasticity itself may be increasing with cognitive ability. Relatedly, Alstadsæter, Kopczuk and Telle (2019) find that networks play an important role in tax avoidance: People are more likely to take advantage of a particular tax avoidance strategy if others in their close family use it too. While our findings focus on the link between overpayment and elasticity, it is clear that there may be other reasons why tax complexity tends to lead to higher inequality, as our findings show.

The rest of this article proceeds as follows: Section 4.1 provides further detail on the institutional background and our study design. Section 4.2 describes the data and details our methodology for calculating tax overpayment. Section 4.3 provides evidence on the size and distribution of tax overpayments overall and for particular demographic, income and wealth groups. Section 4.4

Figure 4.1: Tax scheme for labor and dividend income in Norway



Notes: This figure illustrates the tax scheme in Norway, featuring a progressive tax rate for labor income and a flat tax rate on dividends. The marginal tax rate for dividends is calculated as the combined tax rate on corporate profits and dividend payments. The marginal tax rate for labor income reflects personal income taxes, employer social security contributions, and payroll taxes. Payroll taxes are differentiated by geographic location of the employee – the wage tax scheme shown here reflects the payroll taxes for Zone 1, which encompasses the majority of Norway's population and has the highest tax rate. This figure is for the year 2017; tax rates vary slightly over the years, but the overall pattern is similar.

extends our panel of business owners backwards in time and analyzes their income response to a previous tax reform, examining if there is a link between estimated elasticity and subsequent tax overpayment. Finally, section 4.5 concludes.

4.1 Institutional Background

In this paper, we explore empirical implications of tax complexity using rich register data from Norway. We are among the first researchers able to quantify the extent to which complexity leads to overpayment or mistakes at the individual level. In our setting, we define a tax mistake as a situation where, for a given gross income, a taxpayer makes sub-optimal decisions at a pure accounting level, leaving them with a higher effective tax rate than would have been otherwise

possible. This section gives a brief overview of our strategy, which we expand on in the following sections.

Norway has a dual income tax system where labor and capital income are taxed separately. Labor income taxes follow a progressive schedule with multiple brackets, while capital income is taxed at a flat rate. Stock-based companies (*aksjeselskap*) are subject to a corporate tax rate on profits in the year they accrue, but are considered separate tax entities, such that individual income taxes are only applied when money is paid out from the firm to an individual.

The specific setting we study involves entrepreneurs who own all shares in their company, and who therefore face a choice for any money that they transfer out of the company to pay themselves: They can pay themselves as an employee in the form of wages, as an owner in the form of dividends, or split the payment between these two types in any way they like. These two payment forms are functionally equivalent except for their tax treatment: Wages are taxed progressively, and dividends are subject to a flat tax. For most taxpayers, the marginal tax rates for the two income types cross at a certain point, meaning that there is one particular way to minimize tax liability within a given year: Paying out income as wages up until the point where the marginal tax rates cross, and any remaining income as dividends. Given the total gross amount that the owner transfers from the firm to themselves, the choice between these two income types is purely a question of accounting involving no real costs, and when owners fail to allocate funds optimally, we can therefore unambiguously say that this represents a tax overpayment.

We interpret these cases of overpayment as being the result of tax complexity: If taxpayers had complete information about the tax system and were fully rational, everyone should theoretically be able to optimize their tax payment. It is also worth noting that the tax treatment of this income is indeed quite complex: In the case of both wages and dividends, determining the effective tax rate involves calculating the combination of individual taxes on wages and dividends with taxes that are levied at the firm level – payroll tax and corporate profit tax. Furthermore, payroll taxes vary across different geographical regions of the country, wages generate public pension savings while dividends do not, and dividends are themselves subject to a complicated risk-free return deduction that can cause some or all paid-out dividends to be free of taxes. Our unique dataset allows us to account for all of these complications at the individual level when we determine the combined tax rate on either form of compensation.

We note that our setting is a very specific and quite narrow example of the much broader concept of tax complexity. Not every instance of tax complexity and resulting overpayment comes down to a pure question of costless accounting – in many cases, to take advantage of a complex tax system, people have to pay a real cost, but one which may be outweighed by the tax savings. For example, consider the complexities generated by geographical differences in tax rates across jurisdictions in the U.S. and elsewhere. People can and do take advantage of this complexity by living in areas with lower property tax rates, shopping in states with lower sales taxes, and so on. These decisions can generate tax savings, but also involve a real cost which can be difficult for researchers to quantify. If someone chooses to live in a jurisdiction with high property taxes, we generally do not know if they are unaware that they could pay a lower tax rate by moving elsewhere, or if they simply value their chosen location enough to offset the higher cost of living. Economists tend to assume the latter by imposing rationality in economic models, but in reality, such complexities may cause people to make less-than-optimal choices due to lack of information. The strength of our setting is that there are no real costs of reclassifying income as one type or the other, so we generally avoid ambiguities that could arise when real costs differ across individuals. There is one exception to this, namely a potential ambiguity generated by the public pension system, which we deal with by making conservative assumptions. We explain this further in the following section.

4.2 Data

We use Norwegian de-identified administrative data obtained through Statistics Norway. Our main data used to calculate tax mistakes cover the period 2011-2017, but for additional analysis in Section 4.4, we will use income data going as far back as 1996. The data link firms with their owners and employees, and include detailed information on wages and dividends paid, as well as other accounting data. We have information on all owners of shares in Norwegian stock-based companies, which allows us to identify firms that are wholly owned by a single person. Our main dataset thus consists of annual linked firm-owner data, but in order to accurately calculate tax overpayment, we make further sample restrictions as we explain in the following section.

We have rich demographic data on individuals including age, gender, education, immigration status, and geographic location. We also have detailed data on income and wealth, which allows

us to calculate the marginal tax rates that apply for each business owner.

4.2.1 Quantifying Tax Mistakes

This section explains in detail how we calculate the size of tax mistakes, including any assumptions and sample restrictions made to ensure our calculations are as accurate as possible.

Our calculation of tax mistakes is based on the marginal tax rates that apply when full business owners – defined as those who own all shares of a company – transfer money from their firm to their personal accounts. Whether the money is transferred in the form of dividends or wages, taxes are incurred both at the business and the individual level. We assume that owners would prefer to minimize the total tax liability, i.e. the sum of taxes applied on the business and on themselves as individuals.

Dividends are paid out from firm profits, and are thus taxed first at the corporate tax rate τ_{corp} when profits accrue, and subsequently at the dividend tax rate τ_{div} when paid out to the owner. We can therefore calculate the total marginal tax rate on dividends as:

$$1 - (1 - \tau_{corp})(1 - \tau_{div}).$$

However, not all dividends are subject to taxes. The Norwegian tax code features a deduction equivalent to the risk-free return on investments (*skjermningsfradrag*), calculated based on yields on Norwegian government bonds. Only dividends that are in excess of the theoretical return on such a bond are taxed at rate τ_{div}. Our data include detailed information on this deduction, including how much of it is applied to individual dividend payments. In our calculations of tax mistakes, we only consider the taxed portion of dividend payments, leaving out any dividends that are untaxed due to this deduction.

We now turn to the calculation of taxes on wage income. When a business owner receives wages from their own company, they are treated like any other employee from a tax code perspective. The wage is considered a business expense, so the amount spent on wages is not subject to corporate taxes. However, the company does owe payroll taxes (*arbeidsgiveravgift*) on the wage amount, at a rate τ_{pr}. This tax is not taken out of the gross wage paid to workers, but is instead paid directly

by the firm in adddition to gross wages, so the marginal tax rate associated with this tax is $\frac{\tau_{pr}}{1+\tau_{pr}}$.[2]

The payroll tax is geographically differentiated into seven zones based on the physical location of the firm, with taxes being lower in more rural areas. Zone 1, which encompasses nearly all major urban areas and contained approximately 78% of the population in 2017, has a payroll tax rate of 14.1%, while the tax rate is 0% in the zone that covers Norway's extreme north. Our data contains information on the location of each firm, so we apply the appropriate payroll tax rate for all firms in our calculations. One zone has a multi-bracket system for the payroll tax, where wage expenses are taxed at a lower rate up to a certain threshold. For firms in this zone, we use data on the firm's total wage expenses to apply the appropriate tax rate, taking wage expenses for all other employees as given.

Received wages are subject to personal income tax τ_{inc} at the individual level.[3] Income tax rates are constant across the country, with the exception of the sparsely populated extreme northern zone, whose residents face slightly lower income tax rates. Using data on the municipality of owner-employees, we apply these lower tax rates for residents of this zone.

Combining payroll and individual income taxes, the effective marginal tax rate on wage income for a business owner can be calculated as:

$$\frac{\tau_{inc} + \tau_{pr}}{1 + \tau_{pr}}$$

However, there is one additional complication when calculating the effective tax rate for wage income, namely the Norwegian pension system. Government pensions in Norway are calculated based on lifetime earnings in a quite similar manner to Social Security in the United States. Importantly, only income from wages and certain other sources counts towards pensions, but not dividend income. Prior to 2011, the system was very intricate, involving a complex calculation with multiple brackets and only considering the highest earning years within a person's career.

[2]For a simple example demonstrating the logic behind this, consider a hypothetical tax rate of 100%. For a regular income tax, this would mean that the worker would be left with nothing. However, in the case of the payroll tax, it would simply mean that for every Norwegian krone paid out to the worker, the government would receive one krone from the firm as well. The effective marginal tax rate (ignoring subsequent income taxes) would thus be $\frac{\tau_{pr}}{1+\tau_{pr}} = \frac{1}{1+1} = \frac{1}{2}$.

[3]Norway differentiates between personal income taxes and social security taxes (*trygdeavgift*), which are used to finance welfare benefits for e.g. unemployment, maternity leave, and disability, but not old-age pensions. Since social security taxes are levied at the individual level, we use the term personal income tax to refer to the combined rates of these and "actual" income taxes.

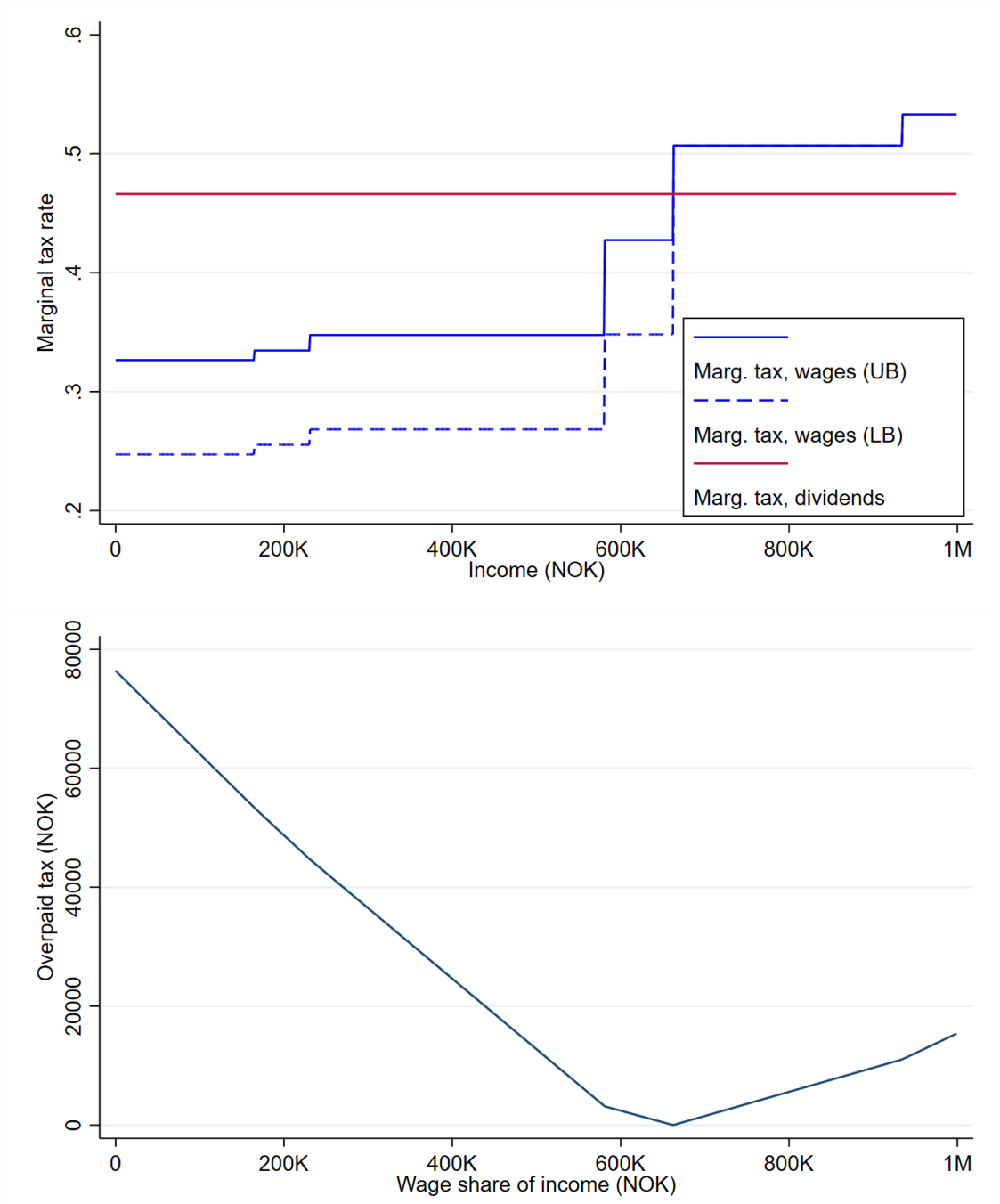
However, in 2011, the system was simplified for individuals born after 1962. For these people, any wage income in a given year up to a certain threshold will generate pension savings equal to 18.1% of earnings, while any wage income beyond the threshold will not contribute to pensions. Accumulated pension savings for an individual are adjusted upwards every year at a rate equal to average wage growth. Upon retirement, average life expectancy for a cohort is used to convert pension savings to an annuity, which is paid out until death.

Because pension savings are generated from wage earnings but not dividends, we can think of them as a negative tax on wage income. The pre-2011 system is too complex for us to reliably calculate an effective tax rate associated with pension savings, which is why we limit our analysis to the years from 2011 onwards and to business owners born in 1963 or later. For these people, we could in principle think of the pension savings as a negative tax of 18.1% up to the threshold. However, we want to be conservative in our assumptions and rationalize as much as possible of what we observe through potential individual preferences rather than labelling it a mistake. There are several reasons why a hypothetical perfectly rational individual might not value these forced savings at their "nominal" value of 18.1%. For instance, they might be able to obtain a higher average return on their other investments than the growth rate of wages. They might anticipate that they will live fewer years than the average member of their cohort, in which case the expected value of the annuity they receive after retirement will be lower.[4] Or they might simply have intertemporal preferences under which they would prefer saving less than 18.1% of their income. For these reasons, we cannot simply consider pension savings to be a negative tax of 18.1% up to the threshold in our calculation of tax mistakes. If we were to do so, we might overestimate the tax mistake for someone who pays themselves "too much" in dividends relative to wage income. Therefore, we assume that the lower bound of the tax value of pension savings corresponds to half the nominal rate – 9.05% – when calculating tax mistakes for those who overpay themselves in dividends. We view this as a reasonable and conservative assumption, since for most taxpayers, these pension payments are a significant part of old-age income and must be assumed to have at least some value.

The effect of including pensions in our marginal tax rate calculations is shown in the top panel of Figure 4.2. This graph corresponds to the tax scheme shown in Figure 4.1, but with marginal tax

[4]Of course, there is the converse possibility that some taxpayers think they will live longer than average, which would raise their valuation of the pension savings.

Figure 4.2: Calculating tax overpayment from marginal tax rates



Notes: The top panel shows tax schedules in 2017 for wages and dividends. Wage incomes up to NOK 662,295 generate pension savings, which reduces the effective marginal tax rate below this point. We assume taxpayers would value these savings at most at their nominal value of 18.1% of income, and at least at half this amount. These assumptions give rise to the two bounds on wage tax rates.
The bottom panel shows the total tax overpayment for a hypothetical business owner transferring money from their business to their personal account. Wage amounts below the optimum (the point where marginal tax rates cross) only lead to tax overpayment when the taxpayer has made corresponding dividend payments, which this graph assumes. The maximum overpayment shown at a wage amount of 0 thus only applies if the taxpayer has made at least NOK 662,295 in dividend payments – lower dividend amounts lead to less tax overpayment. In cases below the optimum, the calculated overpayment is based on the upper bound of the wage tax range, which leads to the smallest overpayment.

rates on wage income adjusted downwards below the pension threshold. This creates an additional bracket at the threshold, and two bounds on marginal tax rates below this point. The dotted line corresponds to the assumption that pension savings are valued at their full nominal value of 18.1%, while the solid line corresponds to a valuation of 9.05%, which leads to a higher effective tax rate on wage income.

The top and bottom panels of Figure 4.2 together illustrate how we use marginal tax rates to calculate tax overpayment. We can think of the bottom panel as illustrating the tax overpayment associated with possible allocations between wages and dividends for a hypothetical taxpayer transfering a total of NOK 1 million out of their firm. The horizontal axis shows the share of this amount which is paid in the form of wages. Recall that the optimal strategy for a fully rational taxpayer with perfect information would be to pay the transfer in the form of wages up to the point where marginal tax rates cross – at NOK 662,295 in this case – and the rest as dividends. This strategy corresponds to the minimum point of the graph and would generate a tax overpayment of 0. If instead the taxpayer were to pay themselves the entire NOK 1 million in the form of wages, this would correspond to the rightmost point on the graph and generate a tax overpayment of NOK 15,370 - equivalent to the area between the marginal tax rate graphs above the intersection point in the top panel.

If we instead suppose the firm owner pays themselves the NOK 1 million entirely in the form of dividends, this would correspond to the leftmost point on the graph and lead to a tax overpayment of NOK 76,390. This amount is calculated by using the *upper* bound of marginal taxes on labor – i.e. as the area between the solid blue line and the red line below their intersection point – as this is the assumption that gives rise to the lowest estimated overpayment in this case. Note that wage payments below the intersection point only cause tax overpayments to the extent that the owner also pays themselves in the form of dividends. A business owner paying herself NOK 400,000 in the form of wages and nothing in dividends would incur no tax overpayment. If she were to pay herself 400,000 as wages and 100,000 as dividends, her overpayment would be $100,000 \cdot 0.118 = 11,800$, where 0.118 is the difference in marginal tax rates on wages (upper bound) and dividends over the interval from 400K to 500K in the top panel.

The above illustrates the general principles behind our calculations of tax overpayment. However, certain complications apply in specific cases. Some individuals in our sample are full owners of more

than one firm, in which case we aggregate up all wage and dividend payments from each firm, and apply the same calculations as if the payments derived from one large firm. Some taxpayers have wage income from other sources than their own firm. In our calculations of tax overpayment, we take this wage income as given – for instance, if a taxpayer has NOK 300,000 of wage income from other sources, our method applies the same calculations as in Figure 4.2, but starting from an income of 300,000 rather than 0. Finally, we note that in some cases in the extreme northern zone of Norway, the reduced rates for individual income and payroll taxes mean that the marginal tax rate on wage income will never exceed the tax rate on dividend income. In these cases, taxes are optimized when everything is paid as wages. A business owner paying any taxable dividend to themselves will always be making a tax mistake, and we calculate overpayment correspondingly.

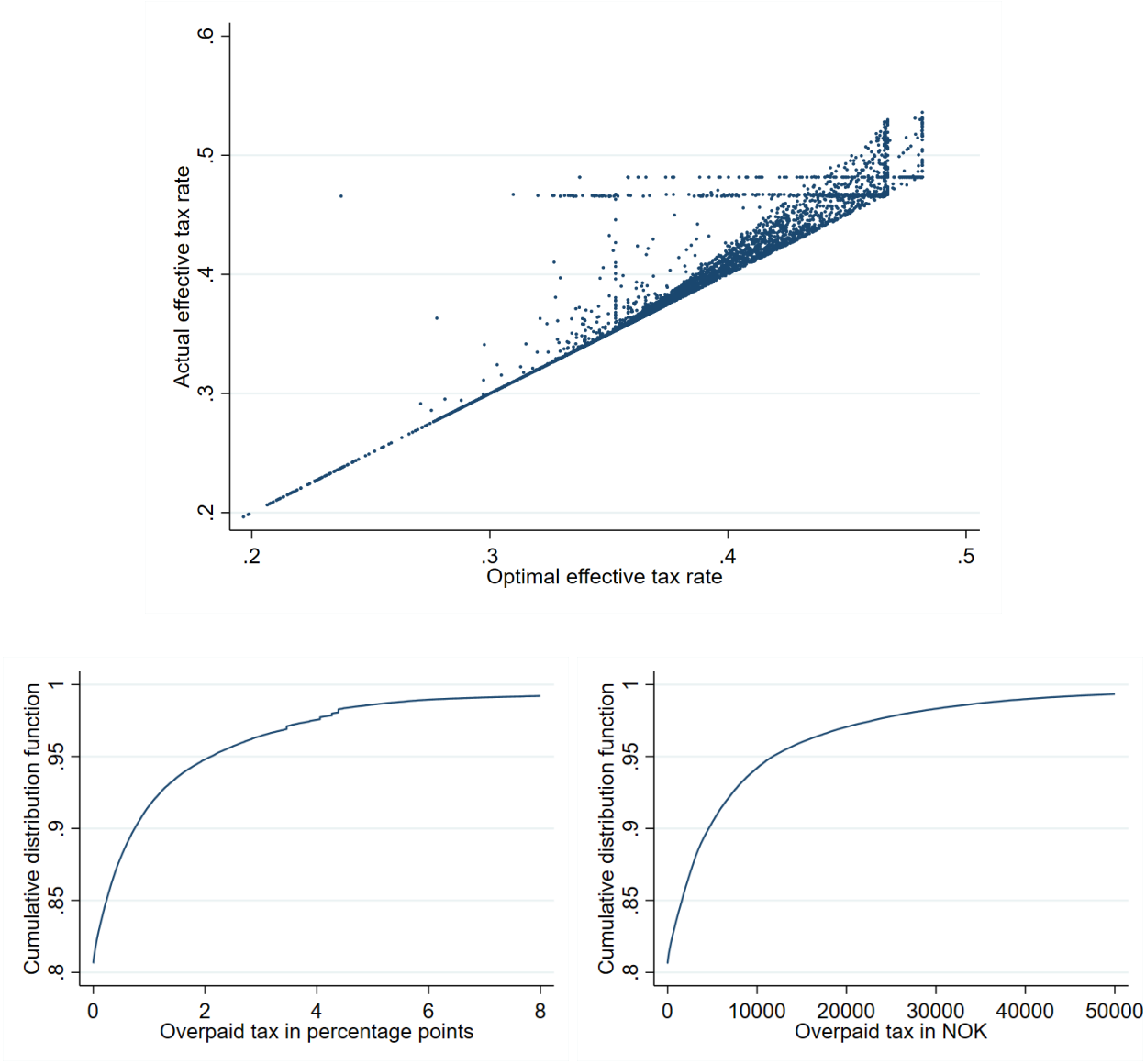
4.3 Tax overpayment: Empirical evidence

We now present evidence on tax overpayment, its distribution within the study population, and how this varies for different population groups. We emphasize that our study population consists of people who are full owners of a business, and who will therefore in many ways be unrepresentative of the general population. The setting we study here deals with one particular form of tax complexity and the dispersion in effective tax rates generated from this, but tax complexity in other areas may lead to very different outcomes. Nevertheless, we believe that some of the evidence presented here can at least be indicative of more general trends.

Figure 4.3 provides an overview of how tax overpayment is distributed in the general population. The top panel plots actual effective tax rates against the theoretical minimized effective tax rates that each individual could have achieved if they had behaved completely optimally. The actual effective tax rate, shown on the vertical axis, is calculated by adding up the total tax amount paid on dividends and wages (including both personal and business taxes) and dividing it by the sum of wages and taxable dividends. The optimal effective tax rate is a result of redoing the same calculation, but using the total tax liability that would have applied if the taxpayer had optimized. In order to more clearly see individual points and patterns in the data, the plot shows a randomly selected sample of 10,000 individual-year observations from our full study population.

We see a clear pattern here: For those with low wage incomes, who thus also have a low optimal

Figure 4.3: Distribution of tax overpayment



Notes: The top panel shows a scatterplot that relates actual effective tax rates on business income to the hypothetical effective tax rate that could be attained under optimal behavior, for a random sample of 15,000 taxpayers from our study population. The actual effective tax rate is calculated as the total tax liability on business income divided by the gross amount of this income. The "optimal effective tax rate" assumes the same total gross income from the business, but instead uses the hypothetical tax liability if this income had been optimally split between wages and dividends for the given taxpayer. A small number of taxpayers on the bottom left of this panel have very low effective tax rates due to favorable tax treatment of residents in the extreme northern part of Norway.
The bottom two panels show cumulative distribution functions for tax overpayments. The bottom left panel shows the distribution of surplus tax in percentage points, equivalent to the vertical distances between points in the scatterplot and the 45-degree line. The bottom right panel shows the distribution of tax overpayments in absolute amounts. In the years 2011-2017 which we study, the exchange rate fluctuated roughly between 6 and 8 Norwegian kroner per US dollar. The CDFs start at a value of roughly 0.81 since this is the fraction of taxpayers who do not overpay at all in a given year. Both figures omit some extreme outliers on the top end.

effective tax rate, mistakes are very rare – in the bottom left corner of the scatterplot, nearly everyone is clustered along the 45-degree line. Further up the income distribution, we see much more dispersion in actual effective tax rates. The reason for this pattern is quite simple: Most business owners in most years pay themselves entirely in the form of wages. For taxpayers whose annual wage incomes are low enough, this happens to be the optimal strategy, so there is no mistake made. In the plot, we can clearly see the point where marginal tax rates for wage income start to exceed the tax rate on dividends, since the distribution starts fanning out beyond this point, as many taxpayers pay themselves too much in the form of wages. The two horizontal lines of points at the top of the graph consist of those taxpayers who pay themselves entirely in the form of dividends.[5]

The bottom two panels of Figure 4.3 show two versions of cumulative distribution functions for tax overpayment. In the bottom left panel, overpayments are expressed as a percentage of total business income – corresponding to the vertical distance between the 45-degree line and each point in the scatterplot. This is essentially a measure of how much higher each taxpayer's effective tax rate becomes as a result of misoptimization. The bottom right panel shows the overpayment distribution in pure money terms. For this panel, note that across the study period, the USD-NOK exchange rate has fluctuated between 6 and 8. For instance, the 95th percentile of this distribution is NOK 11,747, which would roughly correspond to between 1500 and 2000 US dollars.

In both of the lower panels, the CDFs start at a value of approximately 0.81, indicating that 81% of taxpayers do not make mistakes. The vast majority of these are people whose total business income does not reach the point in the tax scheme where marginal taxes are higher for wages than for dividends, and who are therefore behaving optimally when they take everything out as wages. Essentially, this means that in this particular case, tax complexity increases with income. Low-income business owners have a very simple optimal strategy, whereas at higher incomes, tax optimization requires splitting income between dividends and wages, which is a more complicated process. Because we are mostly interested in taxpayers who face a substantially complex tax system, much of our subsequent analysis will drop those taxpayers whose total business income is below the

[5]There are two separate lines since the marginal tax rate on dividends was increased slightly in 2016 and 2017 compared to previous years. The two vertical lines seen near the top of the distribution represents business owners who earn enough income from other sources that they should optimally pay out all business income as dividends, but who nevertheless pay themselves fully or partly in wages.

crossing point of marginal tax rates.[6] We will refer to the remaining taxpayers as the "high-income" sample. These make up 51,343 of our total 335,704 individual-year observations.[7]

Figure 4.4 is equivalent to Figure 4.3, but narrowed down to our high-income sample. The scatterplot now excludes the long "tail" of taxpayers in the lower part of the tax distribution, but otherwise shows all of the same patterns as the scatterplot in Figure 4.3. Looking at the two panels with CDFs, we see that unsurprisingly, tax overpayment is much more common within this subgroup, as they face a much more complex tax situation. Only 22% of this group do not overpay any tax at all, and the majority of those are individuals who have so much income from other sources that paying all their business income as dividends is optimal.

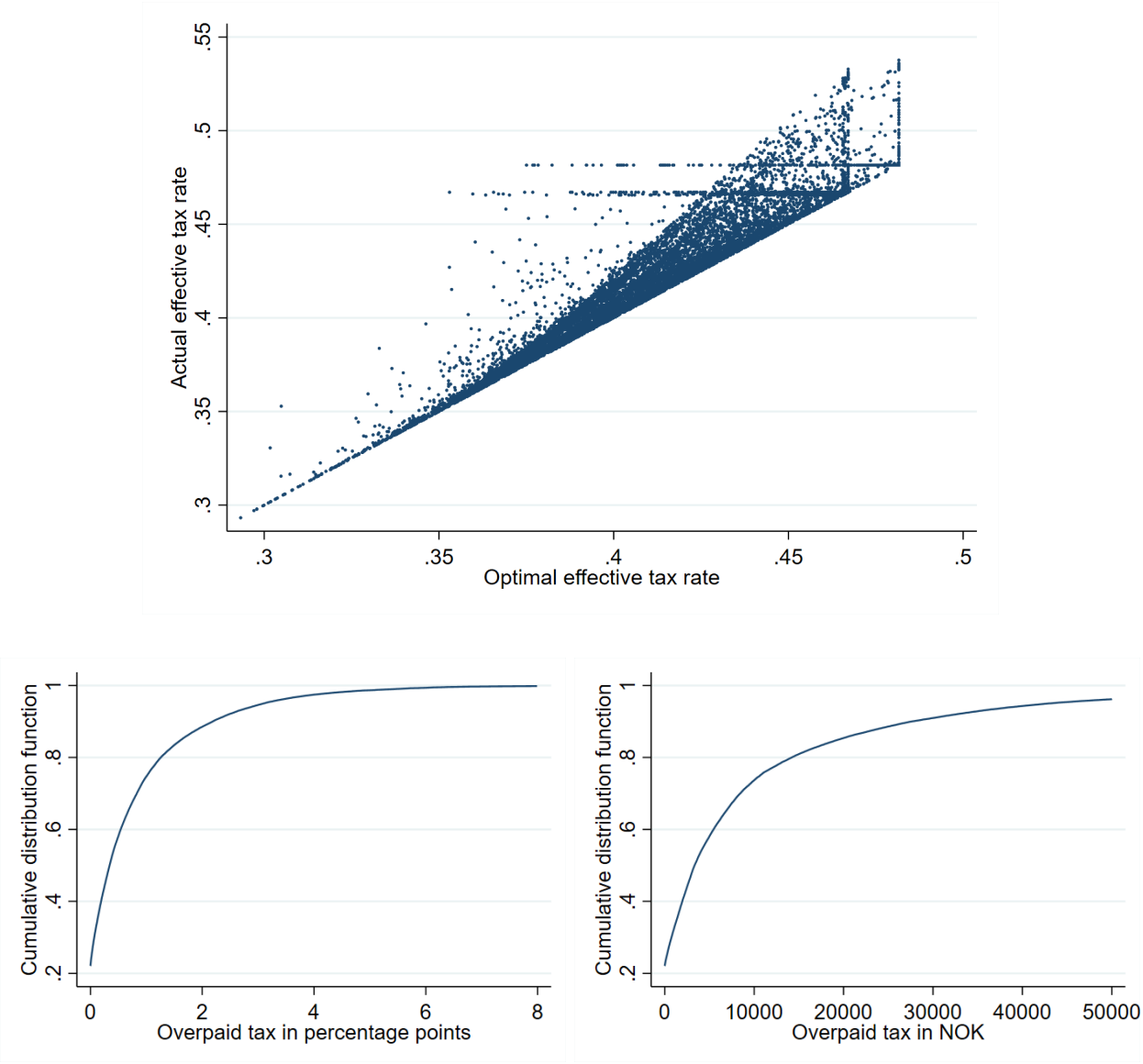
Next, we examine how tax overpayment varies for certain demographic subgroups that we can observe in our data. We may have reason to believe that tax overpayments would differ by demography – either through the mechanism demonstrated in our model, if tax elasticities vary by demographic group, or for other reasons which our model cannot capture. For instance, Alstadsæter, Kopczuk and Telle (2019) have demonstrated that network effects play an important role in tax knowledge acquisition. If some demographic groups have access to more tax knowledge than others through their network, they may require less effort to learn about certain tax minimization strategies.

Figure 4.5 shows how tax elasticities are distributed within the high-income sample based on gender and immigration status. Immigrants are defined as those born abroad to non-Norwegian parents; native-born Norwegians are defined in this context as those born in Norway to two Norwegian parents. The Norwegian system for classification of immigration background includes other categories for more complex situations, which are dropped from the figure. We clearly see that tax overpayments are larger for women relative to men and immigrants – both when measured in

[6] Note that while most of these low-income taxpayers behave optimally by paying only wages, a few do pay some or all of their business income in the form of taxable dividends. Some of these taxpayers can end up making extremely large mistakes – for example, in the top panel of Figure 4.3, the sample includes one outlier at the top left who overpays their taxes by over 20 percentage points, and ends up paying more than twice as much in taxes as they optimally could have.

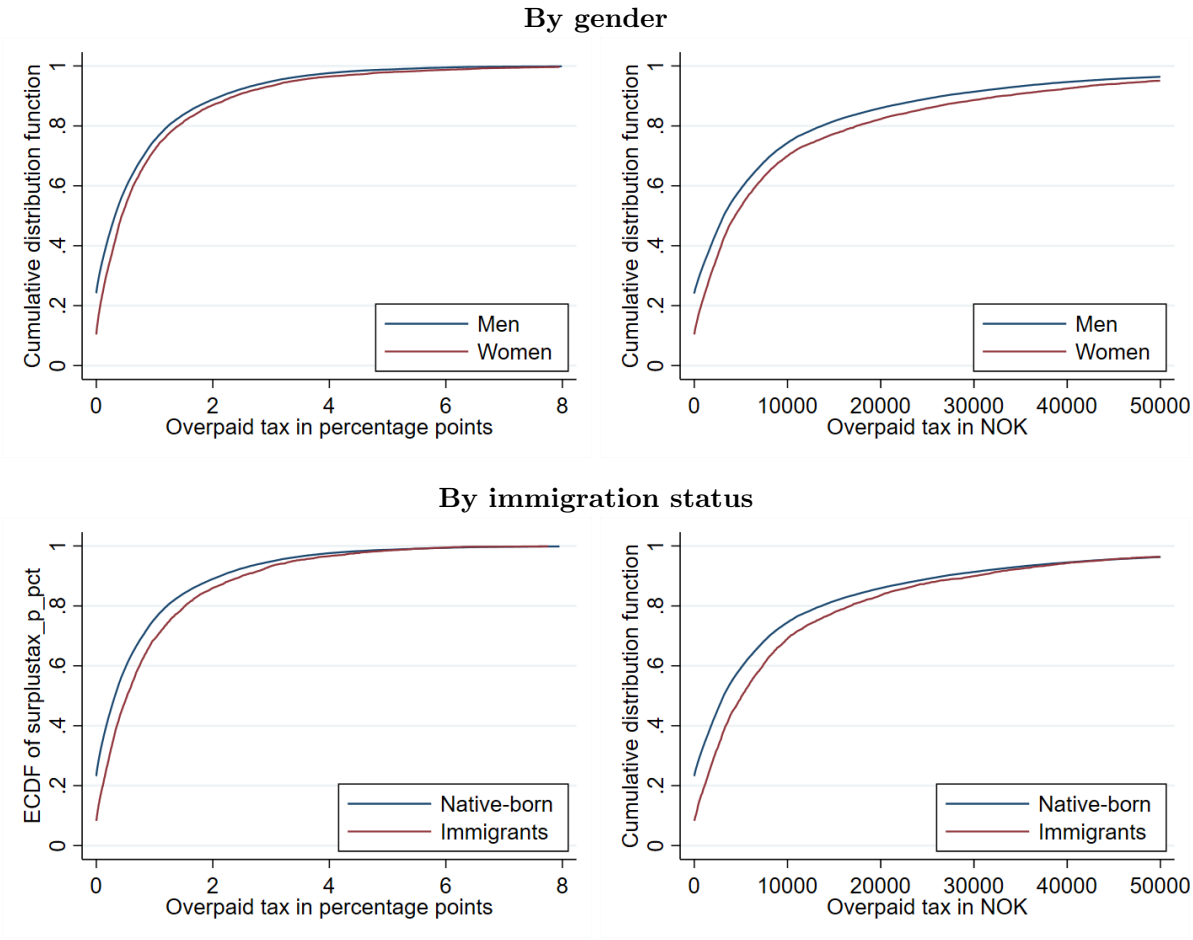
[7] The term "high-income" may be misleading in some cases, as business owners can have a high economic income in their business, but only transfer a small amount to their personal accounts. In fact, this can in itself be the outcome of another tax optimization strategy: Keeping retained earnings in the firm and reinvesting them from there will generally incur less capital taxes than paying them out immediately as they are earned. This illustrates once more that tax complexity is much broader than the specific example we study. We note that this additional dimension of tax complexity is not a threat to our strategy: Conditioning on the amount of money that owners transfer out of the firm, we do measure overpayment accurately. It just means that they may be leaving even more money on the table than we can accurately measure.

Figure 4.4: Distribution of tax overpayment: High-income sample



Notes: This figure corresponds to Figure 4.3, but reduced to our "high-income" sample of those business owners whose combined wage and dividend income from their business in a given year exceeds the point where marginal tax rate schedules for wages and dividends cross. In other words, this is the subpopulation who, given their total transfer, face a more complex tax situation since they cannot optimally pay the entire amount out as wages. The high-income sample comprises 51,343 individual-year observations out of the total 335,704 in our full study population. To improve readability, the scatterplot is based on 15,000 randomly drawn observations from this sample. See the notes to Figure 4.3 for further details.

Figure 4.5: Tax overpayment by demographic groups



Notes: This figure shows distributions of tax overpayments for demographic subgroups within the high-income sample. In the bottom two panels, immigrants are defined as those born abroad to two non-Norwegian parents, whereas native-born refers to people born in Norway to Norwegian parents. Norwegian immigration classifications includes separate categories for people born abroad with one or two Norwegian parents, and those born in Norway with one or two non-Norwegian parents. These categories are dropped from this figure.

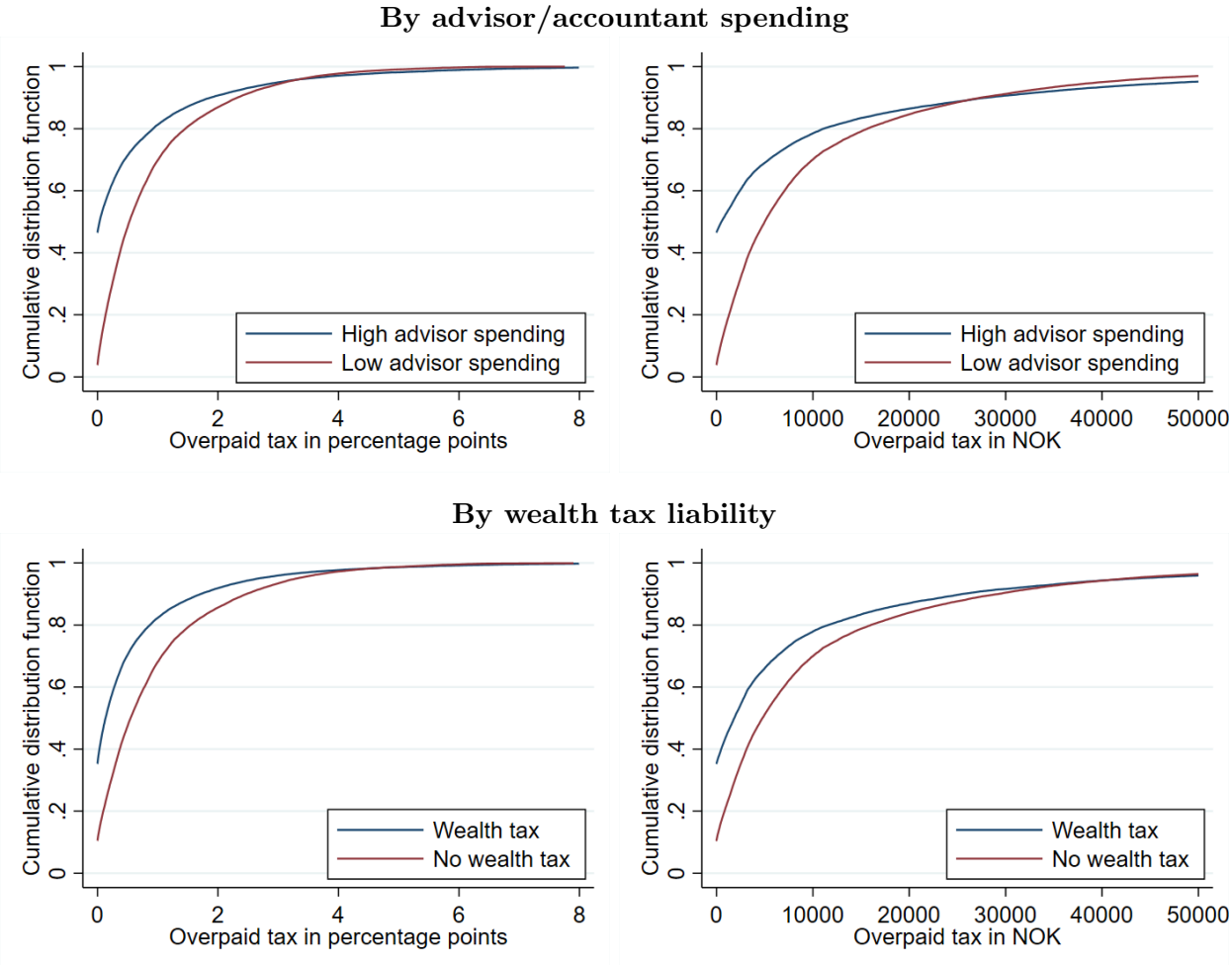
percentage points and in absolute terms, and that this is true across nearly the entire distribution. The mean overpayment for women in this sample is NOK 11092 (0.9 percentage points), versus NOK 9322 (0.75 pp) for men. For immigrants, the mean overpayment is NOK 10998 (0.95 pp), while it is NOK 9307 (0.75 pp) for native Norwegians.

Next, we look at tax mistakes by within-firm advisor spending and by wealth tax liability in Figure 4.6. We have fairly detailed financial data for all firms, including a line item capturing expenses on outside services such as accounting and financial advisors. We calculate the ratio of advisor spending to total firm assets and assign as "high-spending" those for whom this value is above the median. We can think of this as a proxy for tax knowledge, although we emphasize that it is a crude measure – this spending covers all accounting and financial advice, not necessarily just tax advice. And even if the firm has hired a tax advisor, they might only take the firm's financial situation into account and not the owner's personal taxes. Nevertheless, we see an extremely strong effect: Almost half of those with high advisor spending do not overpay taxes at all, relative to a negligible number for those with low spending. The difference in overpayment is quite marked over most of the distribution, though it does reverse near the top. The average overpayment for those with high advisor spending in this sample is 8310 (0.6 pp), versus 10553 (0.9 pp) for those with low spending.

Norway applies a wealth tax to individuals with net wealth above a certain threshold, which ranges over our study period from NOK 700,000 in 2011 to 1,480,000 in 2017. All assets, including those owned indirectly through a business, are included in this calculation.[8] We think of wealth tax liability as simply a proxy for being wealthy – 28.2% of our full sample and 46.7% of our high-income sample pay wealth tax. Importantly, being liable for the wealth tax does not change any incentives to minimize personal and business taxes when transferring money out of the business. We see that the wealthy overpay significantly less than the rest of the sample, and are less likely to overpay at all. Average overpayment for high-wealth tax payers is 8327 (0.58 pp), while it is 10704 (0.94 pp) for the less wealthy. This makes sense given our model's prediction that tax knowledge is increasing in income, if we use wealth as a proxy for this: People who earn enough to pay wealth

[8]Some directly owned assets are applied a discount before being included in the calculation. For example, the tax value of a primary residence is only counted as 25% of its assessed value. Married couples are taxed jointly on their wealth with thresholds being twice as large as those that apply for individuals. The wealth tax rate was 1.1% in 2011-2013, 1% in 2014 and 0.85% from 2015 onwards.

Figure 4.6: Tax overpayment by wealth and advisor expenditure



Notes: This figure shows distributions of tax overpayments for different subgroups within the high-income sample. In the top two panels, taxpayers are grouped by the size of their within-firm spending on accountants divided by total assets, relative to the median. Wealth tax liability in the bottom two panels is based on total assets owned, including those held indirectly through the business.

taxes typically have more to gain from saving on their taxes, so they are more likely to find it worth the cost to invest in tax knowledge. Also, the wealth tax itself might generate additional complexity which can make these taxpayers more likely to hire a tax advisor for their individual taxes, who may also help them optimize their tax situation in this specific case. However, since we cannot observe any privately hired tax advisors, this is difficult to verify.

We note that our restriction to the high-income sample is important in these figures, because as evidenced by the top panel of Figure 4.3, tax complexity rises with income in our empirical setting. If we simply look at the unrestricted study population, overpayments are larger and more frequent for those who pay wealth tax and for native-born Norwegians, which is due to the fact that these groups on average earn more than their counterparts, and thus are more likely to be in the income region where complexity is high. Restricting to high-income taxpayers is one way to get around this issue, but as a further robustness check, we estimate the difference for our full sample. We regress total tax overpayment at the person-year level on a group indicator and control flexibly for the total of wage and taxed dividend income using a 10-part spline, as well as for the optimal effective tax rate, shown on the horizontal axis in the scatterplot in Figure 4.3. Because the outcome variable is 0 for over 80% of our full sample, we use a "Tobit" censored regression model, which accounts for the outcome variable being restricted to values weakly greater than zero (Tobin, 1958). Results are shown in Table 4.1 for each of our indicator variables separately, as well as all together, and roughly confirm what we have seen in the figures. We do note, however, that the estimate for wealth taxes is small and insignificant compared to the relatively large apparent effect in the corresponding graph. This is likely due to some large outliers, as we do see a comparatively fatter tail for the wealth tax payers in Figure 4.6.

Overall, we have seen that tax complexity leads to dispersion in tax rates and overpayment relative to the optimal tax rate. This overpayment can be very large in some cases and is on average larger for women and immigrants relative to men and native Norwegians, respectively. Wealthier individuals make fewer mistakes, highlighting the adverse effect of complexity on inequality, and those with high in-firm advisor spending make fewer mistakes, which indicates that differences in tax knowledge does explain some of the dispersion.

Table 4.1: Group differences in tax overpayment

	(1)	(2)	(3)	(4)	(5)
Female	3315.2				3378.4
	(257.3)				(265.4)
Immigrant		1010.0			599.6
		(285.5)			(281.5)
Wealth tax			-255.4		148.4
			(246.7)		(240.9)
High advisor spend				-3941.2	-4030.2
				(247.9)	(246.2)

Notes: This table shows Tobit regressions of calculated tax overpayment in NOK on group indicators. The outcome variable is left-censored at 0. Each regression controls for the optimal effective tax rate and a spline over total income deciles, where total income is the sum of taxed dividends and all wage income from the owned business and other sources. Standard errors are clustered at the individual level.

4.4 Linking overpayment and elasticity

The previous section demonstrated that in a setting with tax complexity, there is considerable variation in how optimally people behave, and subsequently what effective tax rates are achieved. This section investigates whether heterogeneous elasticities can explain some of the variation we observe in effective tax rates.

For this purpose, we use our existing panel of business owners and estimate their behavioral response to an earlier tax reform which was phased in from 2005 onwards. An overview of this reform is shown in Table 4.2. The reform kept the base tax rate unchanged at 28% while lowering tax rates for higher brackets over the years 2005 and 2006. The middle bracket rate was lowered from 41.5% to 37%, and the highest bracket rate was lowered from 47.5% to 40%. At the same time, the cutoff for the highest bracket was gradually adjusted downwards.[9] Using income data for the taxpayers in our panel for the years around the reform, we can estimate the elasticity of taxable income (ETI). We group taxpayers by the magnitude of the overpayment we observe them making in later years, and test if taxpayers who overpay more are also less elastic when reacting to a tax reform, as our model predicts.

[9]The cutoff for the middle bracket increased slightly over these years, but this happens routinely to account for inflation and avoid bracket creep, even outside of reform years.

Table 4.2: 2005-2006 tax reform

	2004	2005	2006
Base tax rate	28%	28%	28%
Bracket 1 cutoff (NOK)	354,300	381,000	394,000
Bracket 1 tax rate	41.5%	40%	37%
Bracket 2 cutoff (NOK)	906,900	800,000	750,000
Bracket 1 tax rate	47.5%	43.5%	40%

We estimate the ETI using the following panel regression for the years $t = 2003$-2006:

$$\Delta \log(Y_{it}) = \alpha_0 + \alpha_1 \mathbf{1}_g + \beta_0 \Delta \log(NTR_{it}) + \beta_1 \mathbf{1}_g \Delta \log(NTR_{it}) + \gamma X_{it} + \mu_i + \theta_t + u_{it}, \quad (4.1)$$

where Δ represents the two-year difference in values of a variable, Y_{it} represents taxable income, and NTR_{it} is the marginal net-of-tax rate faced by person i in year t. $\mathbf{1}_g$ is an indicator for person i belonging to a particular group, which we will use to test if the ETI differs for taxpayers based on various criteria, e.g. magnitude of tax overpayment in our later study years. The coefficient β_0 provides an estimate of the ETI for those not in group g, whereas β_1 estimates the difference in the ETI for group g relative to non-group members. μ_i and θ_t are individual and year fixed effects, which we use to control for income trends by person and generally over time. In order to consistently estimate the person fixed effects and avoid multicollinearity, we include a few years prior to the 2005 tax reform, which helps establish a baseline income growth trend for each taxpayer.

It is well known that the main regressor of interest, $\Delta \log(NTR_{it})$, will be endogenous, since the net-of-tax rate depends on a taxpayer's bracket, which is a direct function of Y_{it} and hence affected by the error term u_{it} (Saez, Slemrod and Giertz, 2012*b*). To address this issue, we follow Weber (2014), who recommends instrumenting $\Delta \log(NTR_{it})$ with a variable that takes the hypothetical difference in log net-of-tax rates under the tax systems in two different years, but calculated using income from some year prior to the base year. The instrument will thus measure the effect of the change in tax system, without being affected by concurrent random fluctuations in income. In our case, we instrument with the hypothetical change in marginal net-of-tax rates using income two years prior to the base year – e.g. we calculate the difference in net-of-tax rates an individual would face under the 2006 and 2004 tax systems, respectively, if their income were kept fixed at

its level in 2002.[10] To instrument for the interaction $\mathbf{1}_g \Delta \log(NTR_{it})$, we use the interaction of the Weber instrument with $\mathbf{1}_g$. Weber (2014) notes that endogeneity will persist if income growth rates are different along the income distribution, and following her, we include as a remedy a income splines for the same year used in calculating the instrument. We drop individuals earning under NOK 100,000 from our sample, and following Saez, Slemrod and Giertz (2012*b*), we trim values of $\Delta \log(Y_{it})$ at 2% on each tail to limit the influence of outliers.

One possible objection to this study design might be that tax knowledge and elasticities are difficult to completely disentangle empirically. An individual with more tax knowledge will be better at tax optimization, and this might empirically manifest as a larger estimated tax elasticity in response to a reform like this one, even if it were just due to income shifting as described in the previous section, rather than a "real" response. To alleviate this concern, we estimate (4.1) only for those taxpayers whose firms were founded in 2007 or later, such that they are not self-employed during the tax reform.

We look at several groups of taxpayers in our sample to study how their elasticity estimates differ from the rest of the population. First, to investigate our model's prediction that higher tax knowledge is correlated with higher elasticity, we group our sample by the average magnitude of the tax overpayments that we observe them making in later years. We look at two measures: Whether an individual's average tax overpayment over the years 2011-2017 is greater than the median, and whether the average overpayment in percent exceeds 1 percentage point. As we have described in previous sections, there is a strong correlation between income and tax complexity in the full sample, so to avoid accidentally selecting on average income when constructing these samples, we only include in our calculations those observation years where an individual belongs to the high-income sample as defined in the previous section, i.e. the years where total business income is large enough that a simple strategy of paying everything as wages would not be optimal. This limits the sample size quite a lot for these regressions, as anyone who is never part of the high-income sample in 2011-2017 is dropped.

We also study elasticities by income and wealth. Several results from our theoretical section assume a positive correlation between income and elasticity, so we are interested in finding out

[10] Formally, the exclusion restriction requires using an income year with a sufficiently long lag that random shocks to the income process will be serially uncorrelated with those in the base year. Using data from Denmark, Weber finds that a two-year lag relative to the base year is sufficient to eliminate this serial correlation.

Table 4.3: Elasticity estimates for 2005-2006 tax reform

$\Delta \log I$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ log NTR	0.085	0.739	0.267	-0.064	-0.043	0.203	0.107
	(0.160)	(0.904)	(0.651)	(0.172)	(0.164)	(0.171)	(0.167)
Δ log NTR * group		-0.988	-0.485	0.808	0.931	-1.365	-0.059
		(0.815)	(0.627)	(0.233)	(0.599)	(0.619)	(0.838)
Group		> median	>1pp	High inc	WT	Women	Immigr.
Observations	130,858	17,078	16,968	130,858	130,858	130,858	122,558

Notes: This table shows instrumental variable estimates of ETI for various subgroups within our sample. Δ represents a two-year difference. All regressions cover the years 2003-2006 and include individual and year fixed effects. Standard errors are clustered at the individual level. See the main text for further details on the regression specification and group definitions.

if this holds empirically. We look at ETI estimates for high-income taxpayers, defined as those earning an average of at least 500,000 annually, and for taxpayers who pay the wealth tax. Basing this on concurrent income or wealth would introduce endogeneity since these are directly influenced by the error term, so we instead group individuals by wealth tax liability in the year 2000 – before our study period begins – and average income in the years 1997-2000. Finally, we look at elasticities for women relative to men, and for immigrants relative to native Norwegians.

Our results are shown in Table 4.3. The first column just estimates the elasticity for the full sample, omitting any group interaction. The next two columns show results based on the magnitude of later tax overpayment. Column 2 estimates elasticities for those with overpayment above vs. below the median in money terms, and column 3 for those with overpayment of at least 1 percentage point compared to the rest. In both columns, we can see that the estimated effects are numerically quite large in the direction that our model would suggest – taxpayers with large overpayments are on average less elastic – but the small sample size also leads to large standard errors, so neither of these estimates is statistically significant. Although we cannot conclusively prove a link between low elasticity and high tax overpayment based on these regressions, we do consider this to be at least an indication of a possible connection.

Columns 4 and 5 show our results for high-income and high-wealth taxpayers. We see that the estimated elasticity is significantly larger for higher earners, in line with previous estimates by e.g. Kumar and Liang (2020), and that the estimated difference by wealth tax liability is

also positive, though not statistically significant. Finally, we see that women have a surprisingly strongly negative elasticity estimate, while immigrants have an estimated very small difference to the native population. The large negative difference estimate for women is statistically significant, and note that the hypothesis that the elasticity for women is 0, is just barely rejected at a 5% level ($F = 3.97, p = 0.046$). The strongly negative estimate is implausible, and could be due to selection issues – our sample of business owners is over 80% male, and if women in our sample are concentrated in certain business areas or fields, results could be influenced by variations in income trends for these fields that happen to covary with the tax reform. However, we also note that a slightly negative elasticity estimate, as we see for some other groups, is not necessarily at odds with standard consumer theory, since the income effect may well dominate the substitution effect around the tax threshold for certain demographics.

Overall, these results lend some credence to the hypothesis that elasticity differences lead to differences in overpayment, though we cannot say for sure that they are the main driver. Many other factors, including differences in cognitive ability or networks, may also influence people's ability to optimize their taxes. Even if elasticity differences are not the main driver of overpayment, the fact that they are correlated with overpayment means that complexity may still boost tax revenue and labor output along the lines of what is described in the model from the previous chapter.

4.5 Conclusion

In this paper, we study the topic of tax complexity using government microdata from Norway. We look at business owners who face the choice of how to pay themselves when transferring funds out of their business. As full shareholders, they are free to pay themselves any amount in either taxes or dividends, but these two types of compensation have different tax implications. We identify deviations from optimal behavior, which are common for taxpayers with sufficiently high business income and can be very large in some cases. We are able to calculate overpaid tax and show that controlling for income, women and immigrants overpay more, as do those with less wealth. We also see that owners of firms that spend more on advisory services have substantially less tax overpayment, indicating that knowledge is a likely driver of the differences we see. Using a prior tax reform, we estimate tax elasticity by group and find that those who go on to overpay more do

appear to have lower elasticities.

Further research on this topic could help shed light on questions that we have been unable to address. For example, what is the role of heterogeneous elasticities in driving differences in tax knowledge, relative to other possible explanatory factors, such as networks or cognitive ability? How would a tax reform that substantially simplifies or complexifies the tax system impact overpayment overall, and by different demographic and income groups? Can we observe that people learn about the tax system and become more knowledgeable over time? Answering these questions would help us better understand the tradeoffs associated with a simpler tax system.

Bibliography

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, 105(490): 493–505.

Abeler, Johannes, and Simon Jäger. 2015. "Complex Tax Incentives." *American Economic Journal: Economic Policy*, 7(3): 1–28.

Agersnap, Ole, Amalie Jensen, and Henrik Kleven. 2020. "The Welfare Magnet Hypothesis: Evidence from an Immigrant Welfare Scheme in Denmark." *American Economic Review: Insights*, 2(4): 527–542.

Agersnap, Ole, and Owen Zidar. 2020. "The Tax Elasticity of Capital Gains and Revenue-Maximizing Rates." *American Economic Review: Insights*, 3(4): 399–416.

Aghion, Philippe, Ufuk Akcigit, Matthieu Lequien, and Stefanie Stantcheva. 2017. "Tax Simplicity and Heterogeneous Learning."

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2016. "Self-Targeting: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy*, 124(2): 371–427. Publisher: The University of Chicago Press.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. 2018. "Immigration and Redistribution." NBER Working Paper 24733.

Alesina, Alberto, Elie Murard, and Hillel Rapoport. 2019. "Immigration and Preferences for Redistribution in Europe." NBER Working Paper 25562.

Alstadsæter, Annette, Wojciech Kopczuk, and Kjetil Telle. 2019. "Social networks and tax avoidance: evidence from a well-defined Norwegian tax shelter." *International Tax and Public Finance*, 26(6): 1291–1328.

Andersen, Lars Højsgaard, Christian Dustmann, and Rasmus Landersø. 2019. "Lowering Welfare Benefits: Intended and Unintended Consequences for Migrants and their Families." The Rockwool Foundation Research Unit Study Paper 4.

Auerbach, Alan J, and Jonathan M Siegel. 2000. "Capital-gains realizations of the rich and sophisticated." *American Economic Review*, 90(2): 276–282.

Auten, Gerald, and David Joulfaian. 2004. "Taxes and Capital Gains Realizations: Evidence from a Long Panel." Vol. 5.

Auten, Gerald E, and Joseph J Cordes. 1991. "Policy watch: cutting capital gains taxes." *Journal of Economic Perspectives*, 5(1): 181–192.

Bakija, Jon M, and William M Gentry. 2014. "Capital gains taxes and realizations: Evidence from a long panel of state-level data." *Unpublished Manuscript, Williams College*.

Barnes, Lucy. 2022. "Taxing the rich: public preferences and public understanding." *Journal of European Public Policy*, 29(5): 787–804. Publisher: Routledge _eprint: https://doi.org/10.1080/13501763.2021.1992485.

Bastani, Spencer, and Daniel Waldenström. 2021. "The ability gradient in tax responsiveness." *Journal of Public Economics Plus*, 2: 100007.

Benzarti, Youssef. 2020. "How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs." *American Economic Journal: Economic Policy*, 12(4): 38–57.

Blank, Rebecca M. 1988. "The Effect of Welfare and Wage Levels on the Location Decisions of Female-Headed Households." *Journal of Urban Economics*, 24(2): 186–211.

Blesse, Sebastian, Florian Buhlmann, and Philipp Dörrenberg. 2019. "Do People Really Want a Simple Tax System? Evidence on Preferences Towards Income Tax Simplification."

Boeri, Tito. 2010. "Immigration to the Land of Redistribution." *Economica*, 77(308): 651–687.

Bogart, William T, and William M Gentry. 1995. "Capital gains taxes and realizations: Evidence from interstate comparisons." *The Review of Economics and Statistics*, 267–282.

Borjas, George J. 1999. "Immigration and Welfare Magnets." *Journal of Labor Economics*, 17(4): 607–637.

Buhlmann, Florian, Philipp Doerrenberg, Benjamin Loos, and Johannes Voget. 2020. "How do taxes affect the trading behavior of private investors? Evidence from individual portfolio data." *Working Paper*.

Burman, Leonard E. 2010. *The labyrinth of capital gains tax policy: A guide for the perplexed.* Brookings Institution Press.

Burman, Leonard E, and William C Randolph. 1994. "Measuring permanent responses to capital-gains tax changes in panel data." *The American Economic Review*, 794–809.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145–1177.

Craig, Ashley C., and Joel Slemrod. 2022. "Tax Knowledge and Tax Manipulation: A Unifying Model."

De Giorgi, Giacomo, and Michele Pellizzari. 2009. "Welfare Migration in Europe." *Labour Economics*, 16(4): 353–363.

Diamond, Peter, and Emmanuel Saez. 2011. "The case for a progressive tax: from basic research to policy recommendations." *Journal of Economic Perspectives*, 25(4): 165–90.

Dodson, III, Marvin E. 2001. "Welfare Generosity and Location Choices among New United States Immigrants." *International Review of Law and Economics*, 21(1): 47–67.

Doerrenberg, Phillipp, Andreas Peichl, and Sebastian Siegloch. 2017. "The elasticity of taxable income in the presence of deduction possibilities." *Journal of Public Economics*, 151: 41–55.

Dowd, Tim, and Robert McClelland. 2019. "The bunching of capital gains realizations." *National Tax Journal*, 72(2): 323–358.

Dowd, Tim, Robert McClelland, and Athiphat Muthitacharoen. 2015. "New evidence on the tax elasticity of capital gains." *National Tax Journal*, 68(3): 511.

Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm. 2019. "Refugee Migration and Electoral Outcomes." *The Review of Economic Studies*, 86(5): 2035–2091.

Feldstein, Martin, Joel Slemrod, and Shlomo Yitzhaki. 1980. "The effects of taxation on the selling of corporate stock and the realization of capital gains." *The Quarterly Journal of Economics*, 94(4): 777–791.

Fiva, Jon H. 2009. "Does Welfare Policy Affect Residential Choices? An Empirical Investigation Accounting for Policy Endogeneity." *Journal of Public Economics*, 93: 529–540.

Fuest, Clemens, Andreas Peichl, and Thilo Schaefer. 2006. "Does Tax Simplification Yield More Equity and Efficiency? An Empirical Analysis for Germany."

Gelbach, Jonah B. 2004. "Migration, the Life Cycle, and State Benefits: How Low Is the Bottom?" *Journal of Political Economy*, 112(5): 1091–1130.

Goldin, Jacob. 2015. "Optimal tax salience." *Journal of Public Economics*, 131: 115–123.

Gravelle, Jane. 1991. "Limits to Capital Gains Feedback Effects." Congressional Research Service, Library of Congress.

Gravelle, Jane G. 2020. "Capital Gains Tax Options: Behavioral Responses and Revenues." Vol. 41364.

Gruber, Jon, and Emmanuel Saez. 2002. "The elasticity of taxable income: evidence and implications." *Journal of Public Economics*, 84: 1–32.

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica: Journal of the econometric society*, 153–161.

Hellwig, Martin F. 1986. "The optimal linear income tax revisited." *Journal of Public Economics*, 31(2): 163–179.

Joint Committee on Taxation, Staff. 1990. "Explanation of Methodology Used to Estimate Proposals Affecting the Taxation of Income from Capital Gains." Vol. JCS-12-90.

Kaushal, Neeraj. 2005. "New Immigrants' Location Choices: Magnets without Welfare." *Journal of Labor Economics*, 23(1): 59–80.

Kleven, Henrik, and Esben Schultz. 2014. "Estimating Taxable Income Responses Using Danish Tax Reforms." *American Economic Journal: Economic Policy*, 6: 271–301.

Kleven, Henrik, Camille Landais, Mathilde Muñoz, and Stefanie Stantcheva. 2019. "Taxation and Migration: Evidence and Policy Implications." *Journal of Economic Perspectives*. Forthcoming.

Kleven, Henrik Jacobsen, and Wojciech Kopczuk. 2011. "Transfer Program Complexity and the Take-Up of Social Benefits." *American Economic Journal: Economic Policy*, 3(1): 54–90.

Krause, Kate. 2000. "Tax Complexity: Problem or Opportunity?" *Public Finance Review*, 28(5): 395–414. Publisher: SAGE Publications Inc.

Kumar, Anil, and Che-Yuan Liang. 2020. "Estimating taxable income responses with elasticity heterogeneity." *Journal of Public Economics*, 188: 104209.

Mirrlees, J. A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *The Review of Economic Studies*, 38(2): 175–208.

Musgrave, Richard A. 1990. "Horizontal equity, once more." *National Tax Journal*, 43(2): 113–122. Publisher: The University of Chicago Press.

Ngwe, Donald. 2017. "Why Outlet Stores Exist: Averting Cannibalization in Product Line Extensions." *Marketing Science*, 36(4): 523–541. Publisher: INFORMS.

Nichols, Albert L., and Richard J. Zeckhauser. 1982. "Targeting Transfers through Restrictions on Recipients." *The American Economic Review*, 72(2): 372–377. Publisher: American Economic Association.

O'Donoghue, Ted, and Matthew Rabin. 2003. "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes." *American Economic Review*, 93(2): 186–191.

OECD. 1990–2017. "International Migration Database." `https://stats.oecd.org/Index.aspx?DataSetCode=MIG`.

Piketty, Thomas, and Emmanuel Saez. 2013. "Chapter 7 - Optimal Labor Income Taxation." In *Handbook of Public Economics*. Vol. 5 of *handbook of public economics, vol. 5*, , ed. Alan J. Auerbach, Raj Chetty, Martin Feldstein and Emmanuel Saez, 391–474. Elsevier.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. "Distributional national accounts: methods and estimates for the United States." *Quarterly Journal of Economics*, 133(2): 553–609.

Pitt, Mark M., and Joel Slemrod. 1989. "The Compliance Cost of Itemizing Deductions: Evidence from Individual Tax Returns." *The American Economic Review*, 79(5): 1224–1232. Publisher: American Economic Association.

Poterba, James M. 2002. "Taxation, risk-taking, and household portfolio behavior." In *Handbook of Public Economics*. Vol. 3, 1109–1171. Elsevier.

Razin, Assaf, and Jackline Wahba. 2015. "Welfare Magnet Hypothesis, Fiscal Burden, and Immigration Skill Selectivity." *The Scandinavian Journal of Economics*, 117(2): 369–402.

Rees-Jones, Alex, and Dmitry Taubinsky. 2020. "Measuring "Schmeduling"." *The Review of Economic Studies*, 87(5): 2399–2438.

Rosholm, Michael, and Rune Vejlin. 2010. "Reducing Income Transfers to Refugee Immigrants: Does Start-Help Help you Start?" *Labour Economics*, 17(1): 258–275.

Saez, Emmanuel. 2001. "Using elasticities to derive optimal income tax rates." *The Review of Economic Studies*, 68(1): 205–229.

Saez, Emmanuel, and Stefanie Stantcheva. 2016. "Generalized Social Marginal Welfare Weights for Optimal Tax Theory." *American Economic Review*, 106(1): 24–45.

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

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

Seade, J. K. 1977. "On the shape of optimal tax schedules." *Journal of Public Economics*, 7(2): 203–235.

Slemrod, Joel. 1989. "The Return To Tax Simplification: an Econometric Analysis." *Public Finance Quarterly*, 17(1): 3–27. Publisher: SAGE Publications.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the Twenty-first Century." *Quarterly Journal of Economics*, 134(4): 1675–1745.

Smith, Matthew, Owen M Zidar, and Eric Zwick. 2020. "Top Wealth in America: New Estimates and Implications." Working Paper.

Statistics Denmark. 1980–2017. "Central person registry and migration registers, 1980-2017." `https://www.dst.dk/en/TilSalg/Forskningsservice`.

Statistics Finland. 1990–2019. "Country of birth according to age and sex by region, 1990-2019." `http://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__vrm__vaerak/statfin_vaerak_pxt_11rp.px/`.

Statistics Norway. 1970–2020. "Immigrants, by sex and country background 1970-2020." `https://www.ssb.no/en/statbank/table/05184/`.

Statistics Sweden. 2000–2019*a*. "Immigrations and emigrations by country of birth and sex. Year 2000-2019." `http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101__BE0101J/ImmiEmiFod/`.

Statistics Sweden. 2000–2019*b*. "Population by country of birth, age and sex. Year 2000-2019." `http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101__BE0101E/FodelselandArK/`.

Sunstein, Cass R. 2018. "Sludge and ordeals." *Duke Law Journal*, 68: 1843.

Tobin, James. 1958. "Estimation of Relationships for Limited Dependent Variables." *Econometrica*, 26(1): 24–36. Publisher: [Wiley, Econometric Society].

Tolbert, Caroline J., Christopher Witko, and Cary Wolbers. 2019. "Public Support for Higher Taxes on the Wealthy: California's Proposition 30." *Politics and Governance*, 7(2): 351–364.

Weber, Caroline E. 2014. "Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences." *Journal of Public Economics*, 117: 90–103.